Some properties of the operator algebra generated by Hodge's star and the exterior derivative

C. S. Sharma and U. Egele
Department of Mathematics, Birkbeck College, London WC1E 7HX, England

(Received 7 November 1980; accepted for publication 3 April 1981)

Some properties of the operator algebra generated by Hodge's star and the exterior derivative are established and in particular it is shown that viewed as an algebra over its center, this algebra is four-dimensional.

PACS numbers: 02.10.Pk, 11.10.Np

One of the most promising recent developments in mathematical physics has been the discovery of the nice vector bundle structure associated with gauge fields such as the electromagnetic field and the Yang–Mills field. In further developments of the theory both the exterior derivative d and Hodge's star $*$ are bound to play a very important part. With a suitable normalization of the star, they have the following properties:

$$d^2 = 0, \tag{1}$$

$$*^2 = \mathbb{1}, \tag{2}$$

and

$$*d \neq d*, \tag{3}$$

where $\mathbb{1}$ is the identity operator. It can, therefore, be useful to study the purely algebraic structure of the algebra over $\mathbb{C}$ with unity generated by these two operators, or more precisely the algebra $\mathbb{A}$ over $\mathbb{C}$ which is spanned by $\mathbb{1}$, d, $*$, $(d*)^n$, $(*d)^n$, $(d*)^n d$, and $*(d*)^n$ with $n \in \mathbb{Z}^+$ and d and $*$ satisfying the above relations. Such an algebra has recently been studied by Plebański.[1] However, as was pointed out by one of us,[2] this study contains a mistaken identification of the center of the algebra. It is easy to verify that the center of the algebra is the polynomial ring $\mathbb{C}[k]$ over $\mathbb{C}$ generated by k (k for Kendra meaning center in Sanskrit) defined by

$$k = d* + *d. \tag{4}$$

$\mathbb{C}[k]$ has many nice properties: it is a Euclidean ring which is an integral domain, that is, in other words, it is a Euclidean domain. In particular, $\mathbb{C}[k]$ does not have any zero divisors.

The main result of this work is a proof that viewed as an algebra over its center, $\mathbb{A}$ is a four-dimensional. For this purpose we define c(c for commutator) by

$$c = *d - d*. \tag{5}$$

We can now easily establish the following identities:

$$\begin{aligned} *d &= \tfrac{1}{2}(*d + d*) + \tfrac{1}{2}(*d - d*) \\ &= \tfrac{1}{2}(k\mathbb{1} + c), \end{aligned} \tag{6}$$

and similarly

$$d* = \tfrac{1}{2}(k\mathbb{1} - c), \tag{7}$$

$$dc = kd = -cd, \tag{8}$$

and

$$\begin{aligned} *c &= d - *d* = 2d - d - *d* \\ &= 2d - (d* + *d)* = 2d - k* \\ &= -c*. \end{aligned} \tag{9}$$

We can, therefore, construct the following multiplication table for the elements of $\mathbb{A}$ viewed as an algebra over its center:

$\times$	$\mathbb{1}$	d	$*$	c
$\mathbb{1}$	$\mathbb{1}$	d	$*$	c
d	d	0	$\tfrac{1}{2}(k\mathbb{1} - c)$	kd
$*$	$*$	$\tfrac{1}{2}(k\mathbb{1} + c)$	$\mathbb{1}$	$2d - k*$
c	c	$-kd$	$k* - 2d$	$k^2\mathbb{1}$

This shows that the dimension of $\mathbb{A}$ as an algebra over its center is not greater than four. We next show that this dimension cannot be less than four. To do this, suppose that for

$$\begin{aligned}
& \alpha_1, \alpha_2, \alpha_3, \alpha_4 \in \mathbb{C}[k], && \\
& \alpha_1\mathbb{1} + \alpha_2 d + \alpha_3 * + \alpha_4 c = 0 && \\
\Rightarrow\quad & \alpha_1 d + (1/2)\alpha_3 && \\
& \times(k\mathbb{1} - c) + \alpha_4 kd = 0 && \text{(multiplication on the left by } d\text{)} \\
\Rightarrow\quad & (1/2)\alpha_3(kd - cd) = 0 && \text{(multiplication on the right by } d\text{)} \\
\Rightarrow\quad & \alpha_3 kd = 0 && \text{[identity (8)]} \\
\Rightarrow\quad & \alpha_3 = 0 && \\
\Rightarrow\quad & \alpha_1\mathbb{1} + \alpha_2 d + \alpha_4 c = 0 && \\
\Rightarrow\quad & \alpha_1 d + \alpha_4 kd = 0 && \text{(multiplication on the left by } d\text{)}
\end{aligned}$$

and

$$\begin{aligned}
& \alpha_1 d - \alpha_4 kd = 0 && \text{(multiplication on the right by } d\text{)} \\
\Rightarrow\quad & \alpha_1 d = 0 &&
\end{aligned}$$

and

 0022-2488/81/081519-02$01.00 © 1981 American Institute of Physics

$$\alpha_4 kd = 0$$
$$\Rightarrow \quad \alpha_1 = 0$$
and
$$\alpha_4 = 0,$$
which with
$$\alpha_3 = 0$$
$$\Rightarrow \quad \alpha_2 * = 0$$
$$\Rightarrow \quad \alpha_2 = 0.$$
This completes the proof of our result.

A more detailed study of the algebra is under progress. However, we would like to make two points. First, that since d is a differential operator, it is not the algebra over $\mathbb{C}$ which is going to be useful but the algebra over the ring $C^{\omega}(M, \mathbb{C})$ of analytic functions from a finite dimensional differential manifold M to $\mathbb{C}$. Second, that though $*$ can be normalized (cf. Ref. 1) so that it can have the nice property of being an involution, the normalizing factor for $*$ as an operator is different for $*$ in $*$ and $*d$ and therefore $*$ in actual applications will not have the purely algebraic character postulated above.

The description of the algebra given by Plebański[1] is, of course, perfectly self-consistent, but the description of the algebra as an algebra over the center as in this work is clearly simpler than Plebański's description as an algebra over a subcenter: for one thing, our description has a much lower dimension. To indicate some applications of our work, we wish to say that every object (operator) in Plebański's[1] work, of course, belongs also to our algebra, where it has a simpler description. Further, Plebański[1] in defining his extended algebra finds it necessary to postulate the existence of $\Delta^{-1/2}$ and says something about the difficulty in giving a rigorous meaning to his postulate. If our description is used $\Delta^{-1/2}$ is simply the inverse of k and k is invertible if we take the quotient space of the domain of k by its kernel. We believe that every application given by Plebański[1] becomes simpler in our description of the algebra. As already stated, further work on both the theory and its applications is in progress and will be reported in due course.

[1]J. F. Plebański, J. Math. Phys. 20, 1415 (1979).
[2]C. S. Sharma, Math. Rev. 80, 2549 (1980).

Fermionic coherent states in a fock superspace

J. W. F. Valle
Physics Department, Syracuse University, Syracuse, New York 13210

(Received 13 January 1981; accepted for publication 6 February 1981)

Fermionic coherent states (FCS) are constructed via the Weyl supergroup of isometries of a Fock superspace. Their properties are derived and the connection with the pseudo-mechanics formalism is pointed out.

PACS numbers: 02.10.Sp

1. INTRODUCTION

Coherent states[1] are usually introduced for *bosonic* oscillators via the Weyl group arising from the canonical commutation relations. Less is known about coherent states for systems with finite spectrum as arise in dealing with internal degrees of freedom. An approach in this direction was given by Perelomov[2] in generalizing the concept of coherent state for an arbitrary Lie group. Here we present another logical possibility. The idea is that observables with finite spectrum do not have a strict classical analog, i.e., do not admit a c number classical description.[3] They should then be described "classically" by anticommuting dynamical Grassmann variables, as in the pseudomechanics[4] formalism. Upon quantization one obtains the usual Clifford algebra of fermionic oscillators. According to this view then all degrees of freedom of a finite spectrum are fermionic.[5] This argument has been the basis of recent attempts at a unified description of leptons and quarks as fermionic oscillator excitations.[6]

In this paper we extend the notion of coherent states to the fermionic case. The arena of the formalism is a Fock superspace[7] where the Weyl supergroup[8] acts as a unitary group generating the coherent states. The underlying number system is a Grassmann algebra (GA) with involution used in substitution for the complex number field.[9] The Weyl supergroup entails a correspondence between pseudoclassical anticommuting dynamical variables and quantum mechanical Fermi-Dirac operators. Such correspondence was first suggested by Schwinger[10] and is the basis of the pseudomechanics formalism. The mathematical foundations were given by Berezin.

The organization of the material is as follows. In Sec. 2 we review basic properties of GA's and settle notation. In Sec. 3 we sketch the construction of the Fock superspace. In Sec. 4 we construct the Weyl supergroup and study some of its properties. The families of FCS[11,12] are studied in Sec. 5. Section 6 briefly gives the Bargmann–Segal representation. Additional discussion is given in Sec. 7.

2. PRELIMINARIES

A GA $\mathscr{A}$ over the complex field $\mathbb{C}$ is a set of elements x which is a vector space (under complex linear combinations) and has an extra $\mathbb{Z}_2$-graded product operation; any element $x \in \mathscr{A}$ can be split into an even part x_0 and an odd part x_1

$$x = x_0 + x_1 \tag{2.1}$$

such that the product of any two elements x_α of parity α, x'_β of parity β, obeys

$$x_\alpha x'_\beta = (-1)^{\alpha\beta} x'_\beta x_\alpha . \tag{2.2}$$

(We use the first greek letters for $\mathbb{Z}_2$-valued indices) To denote that the parity of x_α is α we write

$$\sigma(x_\alpha) = \alpha. \tag{2.3}$$

Thus $\mathscr{A}$ has a $\mathbb{Z}_2$-decomposition

$$\mathscr{A} = \bigoplus_{\alpha \in \mathbb{Z}_2} \mathscr{A}_\alpha , \tag{2.4}$$

where the set $\mathscr{A}_0$ is a subalgebra.

We assume $\mathscr{A}$ to have a bar involution obeying

$$\begin{aligned} \overline{px} &= p^* \bar{x}, \\ \overline{x+y} &= \bar{x} + \bar{y}, \\ \overline{xy} &= \bar{y}\bar{x} \quad \forall x,y \in \mathscr{A}, \quad p \in \mathbb{C}. \end{aligned} \tag{2.5}$$

We realize the elements of $\mathscr{A}$ as formal polynomials $x(\theta,\bar{\theta})$ on generators θ_a and $\bar{\theta}_a$ obeying

$$\{\theta_a,\theta_b\} = \{\bar{\theta}_a,\theta_b\} = \{\bar{\theta}_a,\bar{\theta}_b\} = 0. \tag{2.6}$$

The involution (2.5) is an inner automorphism of $\mathscr{A}$. We leave the dimensionality of $\mathscr{A}$ unspecified.

Integration[13] is a linear operator defined by

$$\int d\theta_a \, \theta_b = \delta_{ab} = \int d\bar{\theta}_a \, \bar{\theta}_b \tag{2.7}$$

and extended to all products in a straightforward way. Rules of a graded calculus are obeyed. From (2.7) two convenient measures can be defined

$$d\theta d\bar{\theta} = \prod_a d\theta_a d\bar{\theta}_a$$

and

$$d\mu(\bar{\theta}\theta) = \prod_a d\mu_a \quad d\mu_a = \exp \bar{\theta}_a\theta_a d\theta_a d\bar{\theta}_a .$$

For example,

$$\int d\theta_a d\bar{\theta}_a \bar{\theta}_a \theta_a = 1 \quad \forall a, \tag{2.8}$$

$$\int d\mu(\bar{\theta}\theta)\bar{\theta}_a\theta_a = 1 \quad \forall a. \tag{2.9a}$$

Clearly the μ-measure is normalized to one,

$$\int d\mu\,(\bar{\theta}\theta) = 1. \tag{2.9b}$$

We define a complex-valued inner product

$$\langle x,y \rangle = \int d\mu\,(\bar{\theta}\theta)\ \overline{x(\theta,\bar{\theta})}\,y(\theta,\bar{\theta}) \tag{2.10}$$

 0022-2488/81/081521-05$01.00 © 1981 American Institute of Physics

obeying the Hermiticity property

$$\langle x,y\rangle = \langle y,x\rangle^* \quad x,y\in\mathscr{A}.$$

The corresponding squared norm is positive definite when restricted to the "analytic" subalgebra $\mathscr{A}_+$ of elements of form $x = x(\theta)$ and indefinite otherwise, in particular on "antianalytic" elements $x = x(\bar\theta)\in\mathscr{A}$. In either case $x_\pm \in\mathscr{A}_\pm$, its real part $R(x_+)$ defined as the first coefficient in the polynomial expansion, can be obtained by integration with respect to the Gaussian measure.

$$R(x_+) = \int d\mu\,(\bar\theta\theta)x_+, \quad x_+\in\mathscr{A}_+, \quad R(x_\pm)\in\mathbb{C}.$$

An element x with $R(x)\neq 0$ is invertible.[11] The set of such elements is closed under multiplication and contains the unit element. Hence it is a multiplicative group. Clearly it has an abelian normal subgroup containing only the even elements.[14]

3. FOCK SUPERSPACE

Consider an assembly of N fermion oscillators, described by the algebra

$$\{c_i,c_j\} = 0 = \{c_i^\dagger,c_j^\dagger\}, \qquad (3.1)$$
$$\{c_i,c_j^\dagger\} = \delta_{ij}.$$

From the vacuum state defined by

$$c_i|0\rangle = 0 \qquad (3.2)$$

we define the Fock basis $\mathscr{B}_i$, two-dimensional for each degree of freedom i,

$$|\alpha\rangle_i = (c_i^\dagger)^{\alpha_i}|0\rangle_i\in\mathscr{B}_i \qquad (3.3)$$

which is clearly orthonormal

$$\langle\alpha_i|\beta_i\rangle = \delta_{\alpha_i\beta_i}. \qquad (3.4)$$

The usual Fock space $\mathscr{V}$ for the system is a complex vector space whose basis $\mathscr{B}$ is the tensor product $\mathscr{B} = \otimes_i\mathscr{B}_i$. We now upgrade $\mathscr{V}$ from a vector space over $\mathbb{C}$ to a vector space $\mathscr{H}$ over a GA $\mathscr{A}$. For simplicity, we consider one degree of freedom and delete the index i. The canonical basis vectors $|\alpha\rangle$ are said to have parity α. They can only be multiplied by Grassmann coefficients of definite parity. The two possibilities give rise to two classes $\mathscr{H}_\alpha$ of vectors

$$|h_\alpha\rangle = \sum_{\beta\in Z_2} h_{|\alpha+\beta|}|\beta\rangle\in\mathscr{H}_\alpha, \qquad (3.5)$$

where $h_{|\alpha+\beta|}\in\mathscr{A}_{|\alpha+\beta|}$, $|\alpha+\beta|\equiv(\alpha+\beta) \bmod 2$. $|h_\alpha\rangle$ is said to have parity α. Addition is defined only for vectors of the same parity. The Grassmann coefficients can be written on either side of the *canonical* basis vectors $|\beta\rangle\in\mathscr{B}$. (In general, care has to be taken with ordering). Clearly, $\mathscr{H}_\alpha$ is closed under addition and multiplication by even Grassmanns. The dual (bra) vector associated with (3.5) is

$$\langle h_\alpha| = \sum_{\beta\in Z_2}\langle\beta|\bar h_{|\alpha+\beta|}. \qquad (3.6)$$

We define a Grassman-valued inner product

$$\langle h'_\alpha|h_\beta\rangle = \sum_{\gamma\in Z_2}\bar h'_{|\alpha+\gamma|}h_{|\beta+\gamma|}, \qquad (3.7)$$

clearly linear in the second factor and antilinear [in the sense of the involution (2.5)] in the first factor

$$\langle h'_\alpha|h_\beta x_\gamma\rangle = \langle h'_\alpha|h_\beta\rangle x_\gamma, \qquad (3.8)$$
$$\langle h'_\alpha x_\gamma|h_\beta\rangle = \bar x_\gamma\langle h'_\alpha|h_\beta\rangle, \qquad (3.9)$$

where $|h_\alpha x_\beta\rangle\equiv|h_\alpha\rangle x_\beta$, $x_\beta\in\mathscr{A}_\beta$, $|h_\alpha\rangle\in\mathscr{H}_\alpha$. The parity of the inner product (3.7) is $|\alpha+\beta|$ and its Hermiticity property is

$$\langle h'_\alpha|h_\beta\rangle = \overline{\langle h_\beta|h'_\alpha\rangle}. \qquad (3.10)$$

In (3.8) and (3.9) we choose to write the Grassman coefficient to the right of the vectors. It can be transferred to the left by resolving the vectors in the canonical basis and using (2.2).

Linear operators L_γ in $\mathscr{H}$ also belong to two categories: grade-preserving (or even: $\gamma = 0$) and grade-flipping (or odd: $\gamma = 1$). They map $\mathscr{H}_\alpha$ to $\mathscr{H}_{|\alpha+\gamma|}$. Examples of even operators are the identity and the number operators. Creation and annihilation operators are odd. The canonical basis for the algebra of operators is

$$L_{\alpha\beta} = |\alpha\rangle\langle\beta|, \qquad (3.11)$$

with parity $|\alpha+\beta|$. A linear operator L_γ is a linear combination

$$L_\gamma = \sum_{\alpha,\beta\in Z_2} l^\gamma_{\alpha\beta}L_{\alpha\beta}, \qquad (3.12)$$

with Grassmann-valued matrix elements

$$\langle\alpha|L_\gamma|\beta\rangle = l^\gamma_{\alpha\beta} \qquad (3.13)$$

having parity

$$\sigma(l^\gamma_{\alpha\beta}) = |\alpha+\beta+\gamma|. \qquad (3.14)$$

To the product of linear operators, there corresponds a supermatrix product. Grade-preserving operators are closed under multiplication. The adjoint $L^\dagger_\gamma$ of L_γ is defined as

$$\langle h'_\beta|L_\gamma h_\alpha\rangle = \langle L^\dagger_\gamma h'_\beta|h_\alpha\rangle. \qquad (3.15)$$

To any super-matrix l^γ we associate an adjoint matrix $l^{\gamma\dagger}$ defined by

$$l^{\gamma\dagger}_{\alpha\beta} = \bar l^\gamma_{\beta\alpha} \qquad (3.16)$$

in analogy with the usual definition, algebraic involution replacing complex conjugation. Hermiticity is defined with respect to (3.7). Hermitian supermatrices obey

$$h^\gamma_{\alpha\beta} = \bar h^\gamma_{\beta\alpha}$$

and give rise to Hermitian operators. Similarly we define antiHermiticity. Unitary operators are exponentials of antiHermitean ones.

4. THE HEISENBERG SUPERALGEBRA AND WEYL SUPERGROUP

Complete the algebra of canonical anticommutation brackets (3.1) with the identity I,

$$[I,c_i] = 0 = [I,c_i^\dagger] = [I,I], \qquad (4.1)$$

so as to form a graded Lie algebra $\mathscr{W}_N$ over $\mathscr{A}$. The identity generates a one-dimensional center. An element $\omega\in\mathscr{W}_N$ is written as

$$\omega = \omega_0 + \omega_1\cdot c^\dagger + \omega_1'\cdot c. \qquad (4.2)$$

The $2N$ parameters (ω_1,ω_1') are odd and ω_0 is even. With the help of the involution (2.5) we define a Hermitean conjugation in $\mathscr{W}'_N$ which takes Grassmann coefficients into their involutes. It obeys

$$\begin{aligned}
&(\omega^\dagger)^\dagger = \omega,\\
&(\omega + \omega')^\dagger = \omega^\dagger + \omega'^\dagger,\\
&(x\omega)^\dagger = \omega^\dagger \bar{x},\\
&[\omega,\omega']^\dagger = [\omega'^\dagger,\omega^\dagger].
\end{aligned} \tag{4.3}$$

Therefore we can extract from $\mathscr{W}'_N$ an isometry subalgebra $u(1;N)$ defined by

$$a^\dagger = -a. \tag{4.4}$$

An element $a \in u(1;N)$ is written as

$$a = a_0 + \theta \cdot c^\dagger - \bar{\theta} \cdot c, \tag{4.5}$$

where a_0 is an imaginary even parameter ($\bar{a}_0 = -a_0$) and $\{\theta_i,\theta_j\} = \{\bar{\theta}_i,\bar{\theta}_j\} = \{\theta_i,\bar{\theta}_j\} = 0$. Notice that

$$[a_i,a_j] = 0, \tag{4.6}$$

where $a_i = \theta_i c_i^\dagger - \bar{\theta}_i c_i$.

We call $u(1;N)$ the Heisenberg superalgebra. It is clearly not simple since the identity generates an abelian ideal. All the generalized Jacobi identities are trivial for $u(1;N)$.

A convenient matrix representation is

$$\begin{aligned}
&|0\rangle_i \rightarrow \begin{pmatrix}1\\0\end{pmatrix}_i \qquad |1\rangle_i \rightarrow \begin{pmatrix}0\\1\end{pmatrix}_i\\
&c_i = \begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}_i \quad c_i^\dagger = \begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix}_i
\end{aligned} \tag{4.7}$$

Denote the $2N+1$ generators generically by X_α^A, $A \in \{1,\cdots,2N+1\}$. They are all traceless

$$\mathrm{str} X_\alpha^A = 0, \tag{4.8}$$

where the supertrace of a matrix M is defined by

$$\mathrm{str} M_{\alpha\beta} = \sum_{\alpha \in Z_2} (-1)^\alpha M_{\alpha\alpha}.$$

For any pair of elements $a,a' \in u(1;N)$ the formula

$$\langle a,a'\rangle = \mathrm{str}\, a^\dagger a' = \langle a',a\rangle \tag{4.9}$$

defines an invariant $\mathscr{A}_0$-valued inner product.

The corresponding supergroup $\mathrm{U}(1;N)$ is obtained via exponentation,

$$g = \exp a = \exp a_0 \prod_i g_i, \tag{4.10}$$

where $g_i = \exp a_i$. The supergroup manifold[15] is parametrized by

$$z = (a_0;\theta,\bar{\theta}).$$

Clearly, $\mathrm{U}(1;N)$ is a unitary supergroup

$$g^\dagger(z)g(z) = I = g(z)g^\dagger(z), \tag{4.11}$$

or

$$g^\dagger(z) = g^{-1}(z) = g(-z).$$

Therefore it is an automorphism group of the algebra of canonical anticommutation relations. Its commutator

$$g(z)g(z')g^{-1}(z)g^{-1}(z') = \exp(\bar{\theta}'\cdot\theta - \bar{\theta}\cdot\theta')$$

belongs to the normal ungraded $\mathrm{U}(1)$ subgroup. The coset space $\mathrm{U}(1;N)/\mathrm{U}(1)$ is parametrized by $(\theta,\bar{\theta})$ with a generic element denoted $\hat{g}$,

$$\hat{g} = \exp(\theta\cdot c^\dagger - \bar{\theta}\cdot c). \tag{4.12}$$

In the matrix representation (4.7) $\hat{g}$ is represented by a supermatrix $\langle\alpha|\hat{g}|\beta\rangle = U_{\alpha\beta}$ given by

$$\mathrm{U}_{\alpha\beta} = \bigotimes_i \begin{pmatrix} \exp\frac{1}{2}\theta_i\bar{\theta}_i & -\bar{\theta}_i\\ \theta_i & \exp\frac{1}{2}\bar{\theta}_i\theta_i \end{pmatrix}, \tag{4.13}$$

which is clearly not only unitary but *unimodular* as well. This follows from (4.8) and the general relation $s\det U = \exp \mathrm{str} \ln U$.[16]

5. FERMIONIC OSCILLATOR COHERENT STATES

The use of the Weyl supergroup gives a natural formulation for the coherent states considered in Ref. 11. Fermionic coherent states were also discussed in Ref. 12. The use of anticommuting numbers is crucial since the nilpotence of the Fermi–Dirac operators precludes their having c-number eigenvalues. The underlying number system is then taken to be a GA which is a product of factors pertaining to each degree of freedom:

Theorem: There exist in the Fock superspace $\mathscr{H}$ families $\mathscr{F}_\alpha \subset \mathscr{H}_\alpha$ of vectors $|\psi_\alpha\rangle$ obtained from the canonical Fock basis via $\hat{g} \in \mathrm{U}(1;N)/\mathrm{U}(1)$,

$$|\psi_\alpha\rangle = \hat{g}(\theta,\bar{\theta})|\alpha\rangle \tag{5.1}$$

obeying the properties (a),...,(f)

(a) $c_i|\psi_0\rangle = \theta_i|\psi_0\rangle$, (5.2)

(b) $c_i^\dagger|\psi_1\rangle = -\bar{\theta}_i|\psi_1\rangle$, (5.3)

(c) $\langle\psi_\alpha|\psi_\beta\rangle = \delta_{\alpha\beta}$, (5.4)

(d) The identity I is resolved as

$$I = \int d\theta\, d\bar{\theta}\, |\psi_\alpha\rangle(-1)^{|\alpha+1|}\langle\psi_\alpha|, \tag{5.5}$$

corresponding to which there are reproducing kernels $K_{\alpha\beta}\,[\sigma(K_{\alpha\beta}) = |\alpha+\beta|]$ defined by

$$K_{\alpha\beta}(\theta,\bar{\theta};\theta',\bar{\theta}') \equiv \langle\alpha|\hat{g}^\dagger\hat{g}'|\beta\rangle \equiv \langle\psi_\alpha|\psi_\beta'\rangle, \tag{5.6}$$

where $\hat{g}'$ is evaluated at $(\theta',\bar{\theta}')$, etc. They obey

$$(-1)^{|\alpha+1|}|\psi_\beta'\rangle = \int d\theta\, d\bar{\theta}\, |\psi_\alpha\rangle\,\langle\psi_\alpha|\psi_\beta'\rangle. \tag{5.7}$$

For example, we give the explicit expressions for $K_{\alpha\alpha}$,

$$K_{00}(\theta,\bar{\theta};\theta',\bar{\theta}') = \exp[-\tfrac{1}{2}(\bar{\theta}\cdot\theta + \bar{\theta}'\cdot\theta') + \bar{\theta}\cdot\theta'], \tag{5.8}$$

$$K_{11}(\theta,\bar{\theta};\theta',\bar{\theta}') = \exp[\tfrac{1}{2}(\bar{\theta}\cdot\theta + \bar{\theta}'\cdot\theta') + \theta\cdot\bar{\theta}'], \tag{5.9}$$

which are "bi-analytic" on (θ,θ') and their involutes.

(e) The operators

$$\xi_i = (\hbar/2)^{1/2}(c_i^\dagger + c_i), \qquad \pi_i = i(\hbar/2)^{1/2}(c_i^\dagger - c_i) \tag{5.10}$$

have their uncertainty products minimized in $\mathscr{F}_\alpha$,

$$\Delta\xi_i\Delta\pi_i = \hbar/2. \tag{5.11}$$

[$\Delta A = \sqrt{\langle A^2\rangle - \langle A\rangle^2}$ is the r.m.s deviation computed for a coherent state].

(f) The resolution property (d) is not restricted to the identity operator; the operators $L_{\alpha\beta} = |\alpha\rangle\langle\beta|$ can be resolved with respect to the μ-measure (2.9) as

$$L_{\alpha\beta} = \int d\mu\,(\bar\theta\theta)\,d\mu\,(\bar\theta'\theta')\langle\gamma|C_\gamma(\hat g)L_{\alpha\beta}\hat g|\gamma\rangle \exp(\bar\theta\cdot\theta' - \bar\theta'\cdot\theta)P_\gamma(\theta',\bar\theta'), \tag{5.12}$$

where

$$C_0(\hat g) = \hat g,\ C_1(\hat g) = \hat g^\dagger,\ \hat g \equiv \hat g(\theta,\bar\theta) \text{ and } \gamma = |\alpha+\beta|.$$

$P_\gamma(\theta',\bar\theta')$ are the projectors

$$P_\gamma(\theta',\bar\theta') \equiv \hat g'|\gamma\rangle\langle\gamma|\hat g'^\dagger \equiv |\psi'_\gamma\rangle\langle\psi'_\gamma|. \tag{5.13}$$

The corresponding supermatrices are

$$P_0 = \otimes_i \begin{pmatrix} 1-\bar\theta_i\theta_i & \bar\theta_i \\ \theta_i & -\bar\theta_i\theta_i \end{pmatrix},$$

$$P_1 = \otimes_i \begin{pmatrix} \bar\theta_i\theta_i & -\bar\theta_i \\ -\theta_i & 1+\bar\theta_i\theta_i \end{pmatrix}, \tag{5.14}$$

and clearly obey

$$P_\alpha P_\beta = \delta_{\alpha\beta}P_\beta \quad P_\alpha^\dagger = P_\alpha. \tag{5.15}$$

We now sketch the proof of the above.

Eigenproperties (a) and (b) follow from $\theta_i^2 = 0 = \bar\theta_i^2$. The unitarity of $\hat g$ assures that the vectors $|\psi_\alpha\rangle$ are orthonormal, property (c). To prove (d), Eq. (5.5), notice that from (5.1), (4.12), and (4.6) it is sufficient to consider the case of one degree of freedom. Expand $|\psi_\alpha\rangle$ in the canonical basis,

$$|\psi_\alpha\rangle = \sum_\gamma U_{\gamma\alpha}|\gamma\rangle,$$

where $U_{\alpha\beta} \equiv \langle\alpha|\hat g|\beta\rangle$. Then we find

$$\int d\theta\,d\bar\theta\,|\psi_\alpha\rangle(-1)^{|\alpha+1|}\langle\psi_\alpha|$$

$$= \sum_{\beta,\gamma}\int d\theta\,d\bar\theta\,U_{\gamma\alpha}\bar U_{\beta\alpha}|\gamma\rangle\,\langle\beta|(-1)^{|\alpha+1|} = \sum_\beta|\beta\rangle\,\langle\beta| = I$$

In the above we used

$$\int d\theta\,d\bar\theta\,U_{\gamma\alpha}\bar U_{\beta\alpha} \equiv \int d\theta\,d\bar\theta\,[P_\alpha]_{\gamma\beta} = (-1)^{|\alpha+1|}\delta_{\beta\gamma}, \forall\alpha\in\mathbb{Z}_2,$$

which follows from (4.13) and (2.8).

Formula (5.7) is a corollary of (5.5). Property (e) follows from (a) and (b):

$$\langle\psi_\alpha|\xi_i|\psi_\alpha\rangle^2 = 0 = \langle\psi_\alpha|\pi_i|\psi_\alpha\rangle^2$$

and

$$\langle\psi_\alpha|\xi_i^2|\psi_\alpha\rangle = \hbar/2 = \langle\psi_\alpha|\pi_i^2|\psi_\alpha\rangle.$$

The Fock states $|\alpha\rangle\in\mathscr{B}$ are particular solutions of (5.11) corresponding to zero eigenvalue in (a) and (b). We now verify formula (5.12) for the case of the annihilation operator (one degree of freedom). The rhs of (5.12) becomes

$$\int d\mu\,(\bar\theta\theta)\,d\mu\,(\bar\theta'\theta')\langle 1|\hat g^\dagger|0\rangle \times\langle 1|\hat g|1\rangle\exp(\bar\theta\theta' - \bar\theta'\theta)P_1(\theta',\bar\theta').$$

Using the representation (4.7), we substitute the projection matrix P_1 form (5.14) in the above, together with

$$\langle 1|\hat g^\dagger|0\rangle = \bar U_{01} = -\theta$$
$$\langle 1|\hat g|1\rangle = U_{11} = 1+\bar\theta\theta/2.$$

The final result after μ and μ' integration is

$$\begin{pmatrix} 0 & 1 \\ 0 & 0\end{pmatrix},$$

which is the appropriate matrix for c.

Similarly, we can complete the proof. Thus (5.12) shows that any operator can be reconstructed from its "diagonal" matrix elements. As an example, the identity $I = \sum_\alpha L_{\alpha\alpha}$ admits a μ-measure resolution

$$I = \int d\mu\,(\bar\theta\theta)\,d\mu\,(\bar\theta',\theta')\sum_\alpha\langle 0|\hat g|\alpha\rangle\langle\alpha|\hat g|0\rangle \times\exp(\bar\theta\theta' - \bar\theta'\theta)\mathrm{P}_0(\theta',\bar\theta') \tag{5.16}$$

$$\equiv\sum_\alpha\int d\mu\,(\bar\theta\theta)\,d\mu\,(\bar\theta'\theta')\langle -\theta,-\bar\theta;0|\alpha\rangle \times\langle\alpha|\theta,\bar\theta;0\rangle\exp(\bar\theta\theta' - \bar\theta'\theta)|\theta',\bar\theta';0\rangle\langle\theta',\bar\theta';0|,$$

where $|\theta,\bar\theta;0\rangle \equiv \hat g(\theta,\bar\theta)|0\rangle \equiv |\psi_0\rangle$.

6. THE BARGMANN–SEGAL REALIZATION FOR THE FERMIONIC CASE

We now briefly sketch how the fermionic Fock superspace is realized in terms of wavefunctions in a way entirely analogous to the bosonic case. The wave functions are analytic on a set of complex [in a sense analogous to (2.5)] *anticommuting* Grassmann variables u_i. The creation and annihilation operators are u_i and $\partial/\partial u_i$ obeying

$$\{u_i,u_j\} = 0 = \{\partial/\partial u_i, \partial/\partial u_j\}, \tag{6.1}$$
$$\{u_i,\partial/\partial u_j\} = \delta_{ij}.$$

For simplicity we take the case of one degree of freedom. The vectors $|\psi_\alpha\rangle\in\mathscr{F}_\alpha$ are realized by

$$\langle u|\psi_\alpha\rangle \equiv \psi_\alpha(u) = \sum_{\beta\in\mathbb{Z}_2} u^\beta U_{\beta\alpha}, \tag{6.2}$$

so, from (4.13),

$$\psi_0(u) = (1-\bar\theta\theta/2)(1+u\theta),$$
$$\psi_1(u) = (1+\bar\theta\theta/2)(u-\bar\theta). \tag{6.3}$$

Clearly both the *argument* (u) as well as the *label* $(\theta,\bar\theta)$ of the coherent wavefunctions are anticommuting, the corresponding GA's being independent. We list below some of the properties of the functions $\psi_\alpha(u)$,

$$\frac{\partial}{\partial u}\psi_0(u) = \theta\psi_0(u), \tag{6.4}$$

$$u\psi_1(u) = -\bar\theta\psi_1(u), \tag{6.5}$$

$$\int d\mu\,(\bar u u)\,\overline{\psi_\alpha(u)}\psi_\beta(u) = \delta_{\alpha\beta}. \tag{6.6}$$

The identity, i.e., the reproducing kernel for analytic wavefunctions $f_\beta(u)$,

$$\langle u|u'\rangle = \exp u\bar u' \tag{6.7}$$

obeying

$$f_\beta(u) = \int d\mu\,(\bar u'u')\exp u\bar u' f_\beta(u') \tag{6.8}$$

is resolved as

$$e^{u\bar u'} = \int d\theta\,d\bar\theta\,\psi_\alpha(u)(-1)^{|\alpha+1|}\,\overline{\psi_\alpha(u')}\,. \tag{6.9}$$

We can easily compute the reproducing kernels $K_{\alpha\beta}$, Eq. (5.6), in this realization. For example,

$$K_{00}(\theta,\bar{\theta};\theta',\bar{\theta}') = \int d\mu(\bar{u}u)\ \overline{\psi_0(u)}\,\psi_0'(u). \tag{6.10}$$

Therefore

$$K_{00}(\theta,\bar{\theta};\theta',\bar{\theta}') = e^{-\bar{\theta}\theta/2} e^{\bar{\theta}'\theta'/2} \int d\mu\ (\bar{u}u) e^{\bar{\theta}u} e^{u\theta'}$$
$$= \exp\,[\,-\tfrac{1}{2}(\bar{\theta}\theta + \bar{\theta}'\theta') + \bar{\theta}\theta'],$$

and so on. The kernel (6.9) also admits a μ-resolution corresponding to (5.16)

$$\exp u\bar{u}' = \int d\mu\ (\bar{\theta}\theta)\ d\mu(\bar{\theta}'\theta') \sum_{\alpha} U_{o\alpha}\, U_{\alpha o} \times \exp(\bar{\theta}\cdot\theta' - \bar{\theta}'\cdot\theta)\psi_0'(u)\ \overline{\psi_0'(u')}. \tag{6.11}$$

7. DISCUSSION

We have shown how the coherent state formalism for fermionic oscillators emerges naturally from the Weyl supergroup by upgrading the fermionic Fock space into a superspace. The treatment is analogous to the bosonic case but is purely algebraic; the underlying number system is a GA instead of the field of complex numbers. This enlargement permits the extra solutions (5.1) to (5.11). Built into the formalism is the idea that "classical" analogs of fermions are anticommuting. For example, the expectation value of a fermionic operator in a coherent state is an odd Grassmann.

We now compare the present formalism with the bosonic case. In both cases, the whole structure is multiplicative. The distinguishing feature is the use of nilpotents. This makes this formalism applicable even to the case of an infinite number of degrees of freedom, whereas in the bosonic situation one usually truncates to assure convergence. The algebraic nature of the present approach completely evades convergence questions.

ACKNOWLEDGMENTS

I thank Dr. L. Lusanna, Dr. F. Rohrlich, Dr. L. Horwitz, and Dr. J. Schechter for helpful comments. This work was sponsored by the National Research Council (CNPq), Brazil.

[1]J. R. Klauder and E. C. G. Sudarsham, *Fundamentals of Quantum Optics* (Benjamin, New York, 1968).
[2]A. M. Perelomov, Commun. Math. Phys. **26**, 222 (1972); **44**, 197 (1975).
[3]A. Barducci, R. Casalbuoni, and L. Lusanna, CERN report TH 2874 (1980).
[4]R. Casalbuoni, Nuovo Cimento A **33**, 389 (1976); F. A. Berezin and M. S. Marinov, Ann. Phys. **104**, 336 (1977); A. Barducci, R. Casalbuoni, and L. Lusanna, Nuovo Cimento Lett. **19**, 581 (1977).
[5]The importance of the FCS goes, however, beyond this universality hypothesis. For example, as first noted by Klauder [Ann. Phys. **11**, 123 (1960)] they are useful in the evaluation of the path integral and the partition function in statistical mechanics. For recent references see Ref. 11.
[6]R. Casalbuoni and R. Gatto, Phys. Lett. B **90**, 81 (1980).
[7]The present FCS construction develops the superspace point of view sketched in the first paper of Ref. 11. Related FCS constructions, of which we did not known until after the completion of the present work, are also listed in Ref. 11.
[8]The prefix "super" to an algebraic structure implies Z_2-grading, Z_2 being the set $\{0,1\}$ with modulo two addition.
[9]This contrasts with usual assumptions in quantum mechanics [see D. Finkelstein, J. M. Jauch, S. Schiminovich, and D. Speiser, J. Math. Phys. **3**, 207 (1962)]. The number system here is *not* a division algebra.
[10]J. Schwinger, Phil. Mag. **44**, 1171 (1953).
[11]J. W. F. Valle, Int. J. Theor. Phys. **18**, 923 (1979); Y. Onuki and T. Kashiwa, Prog. Theor. Phys. **60**, 548 (1978); Y. Ohnuki and S. Kamefuchi, J. Math. Phys. **21**, 601, 609 (1980).
[12]J. R. Klauder, Phys. Rev. D **19**, 2349 (1979); C. L. Hammer, J. E. Shrauner and B. de Facio, Phys. Rev. D **18**, 373 (1978); B. de Facio and C. L. Hammer, J. Math. Phys. **17**, 267 (1976); H. Kuratsuji and T. Suzuki, J. Math. Phys. **21**, 472 (1980); M. M. Nieto, Phys. Rev. D **22**, 391 (1980).
[13]F. A. Berezin, *The Method of Second Quantization* (Academic, New York, 1966).
[14]In connection with vector space construction over a GA, notice that it should be possible, from this discussion, despite having nilpotents in the basic number system, to perform meaningful vector space manipulations. The possibility of constructing supergroups of linear transformations is a partial realization of this conjecture.
[15]F. A. Berezin and G. Kac, Mat. Sb. **82**, 124 (1970).
[16]R. Arnowitt, P. Nath and B. Zumino, Phys. Lett. B **56**, 81 (1975).

Canonical transformations relating the oscillator and Coulomb problems and their relevance for collective motions

M. Moshinsky[a] and T. H. Seligman
Instituto de Física, UNAM, Apdo. Postal 20-364, México 20, D. F., Mexico

(Received 22 January 1981; accepted for publication 27 February 1981)

The present paper can be viewed from two standpoints. The first is that it derives the canonical transformation that takes the Hamiltonian of the Coulomb problem (in the Fock–Bargmann formulation) into that of the harmonic oscillator, while transforming the angular momenta of both problems into each other. The second is the one in which the solution of the previous problem is required if we wish to find the canonical transformation relating microscopic and macroscopic collective models, where the former is derived from a system of A particles moving in two dimensions and interacting through harmonic oscillator forces. The canonical transformation shows the existence of a U(3) symmetry group in the microscopic collective model corresponding to that of the three-dimensional oscillator which is the Hamiltonian of the macroscopic collective model. The importance of this result rests on the fact that had the motion of the particles taken place in the physical three-dimensional space, rather than the hypothetical two-dimensional one discussed here, the symmetry group would have been U(6) rather than U(3). Thus, the group theoretical structure of an *s-d* boson picture or, equivalently, of a generalized Bohr–Mottelson approach, is present implicitly in an A-body system interacting through harmonic oscillator forces.

PACS numbers: 02.20. – a, 21.60.Ev, 03.65.Fd, 03.65.Ca

I. INTRODUCTION AND SUMMARY

In a recent article Chacón, Moshinsky, and Vanagas[1] discussed the relations between macroscopic and microscopic nuclear collective models. The former have had a long history starting from the original liquid drop model of Bohr,[2] through the unified model of Bohr and Mottelson,[3] and continuing up to the present time in the interacting boson approximation (IBA).[4] The latter have been viewed from many angles,[5] but one that was stressed in Ref. 1 concerned the transformation of the single particle variables to a set of coordinates that included in an explicit fashion six that could be identified with collective degrees of freedom.[6,7] It is possible then to project out a collective Hamiltonian from an A-nucleon system by restricting the Hamiltonian of the latter to a single representation[8,9] of the orthogonal group $O(A-1)$ associated with the $A-1$ Jacobi coordinates. This could be the lowest weight irreducible representation of the $O(A-1)$ group consistent with the Pauli principle, as suggested by Filippov and his collaborators.[8] Alternatively, as suggested by Vanagas,[9] it could be the scalar representation of the $O(A-1)$ group. It is the latter viewpoint that was considered in Ref. 1, where an explicit procedure was implemented to go from A particles interacting through harmonic oscillator forces to what was called the *microscopic* collective (MC) model, arriving finally at an oscillator boson approximation (OBA) which can be viewed as a *macroscopic* collective model. The last step, illustrated in Fig. 1, required a canonical transformation in the classical picture, for which only the explicit representation in quantum mechanics was available.[1] *The purpose of the present paper is to find this canonical transformation explicitly when the A particles move in a two-dimensional space*, in which case all the steps of the quantum mechanical analysis where implemented in Ref. 1.

To achieve our objective we start in Sec. 2 by reviewing the transformation of coordinates that brings out the collective degrees of freedom.[6,7] If the motion takes place in two-dimensional space the scalar part with respect to the $O(A-1)$ group of the A-particle Hamiltonian[1] reduces to the Coulomb problem as discussed by Fock and Bargmann.[10,11] The states of this Coulomb problem are characterized not by the standard angular momentum $\mathbf{L}=(L_1, L_2, L_3)$, but by the SU(2) group whose generators are $\mathbf{J}=(A_1, A_2, L_3)$ with A_1, A_2 being the first two components of the Runge Lenz vector.[1]

We also indicate in Sec. 2 that the macroscopic collective Hamiltonian is that of the three-dimensional oscillator whose states, in the rotational limit,[1] are characterized by the standard angular momentum vector $\mathscr{L}=(\mathscr{L}_1, \mathscr{L}_2, \mathscr{L}_3)$.

The canonical transformation we are interested in is the one that takes the Hamiltonian of the Coulomb problem into that of the harmonic oscillator with the added condition that $\mathbf{J}=\mathscr{L}$.

In Sec. 3 we discuss the canonical transformation that leaves the Hamiltonian of the Coulomb problem invariant but takes the vector $\mathbf{J}$ into $\mathbf{L}$. We then require only the ca-

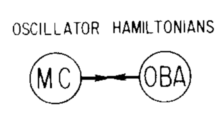


FIG. 1. The present figure symbolizes the canonical transformation relating the microscopic collective (MC) Hamiltonian with the oscillator boson approximation (OBA). This figure should be seen in the context of Fig. 1 of Ref. 1 where the full relation between macroscopic and microscopic collective models, both for oscillator and arbitrary interactions, is presented.

[a] Member of the Instituto Nacional de Investigaciones Nucleares and El Colegio Nacional.

 0022-2488/81/081526-10$01.00 © 1981 American Institute of Physics

nonical transformation that takes the Coulomb into the oscillator problem with $\mathbf{L} = \mathscr{L}$, which is much easier to derive than the one mentioned in the previous paragraphs.

In Sec. 4, with the help of the dynamical group O(4,2) of the Coulomb problem, we find operators functions of the generators of O(4,2) whose matrix elements with respect to the eigenstates of the Coulomb problem are the same as those of the standard creation and annihilation operators with respect to the eigenstates of the oscillator problem.

In Sec. 5 we pass to the classical limit of the operator functions mentioned in the previous paragraph and show that they lead to the canonical transformations relating the Coulomb and oscillator Hamiltonian with $\mathbf{L} = \mathscr{L}$. This canonical transformation is non-bijective (i.e., not one to one onto) and in Sec. 6 we find explicitly the ambiguity group[12] that relates the points in the Coulomb phase space that map on a single point in the oscillator phase space. This ambiguity group will be obtained with the help of the fact that the energy levels of the Coulomb problem have a two to one correspondence with those of the oscillator.

Finally in the concluding Sec. 7 we discuss the implications of the canonical transformations we derived and their possible generalizations to the physical case when the A particles move in three dimensions.

Note that Secs. 4,5, and 6 may be read independently of 2,3, and 7 by a person interested in the relations between Coulomb and oscillator Hamiltonians rather than in problems associated with collective motion in nuclei.

II.THE MICROSCOPIC AND MACROSCOPIC COLLECTIVE HAMILTONIANS

In this section we briefly review the derivation of the microscopic collective Hamiltonians and its integrals of motion to establish its connection with the macroscopic collective Hamiltonian, from which we will obtain the canonical transformations that relate them.

As mentioned in the introduction, we restrict ourselves to motion in a two-dimensional space, and for the A-particle system we have $2A-2$ Jacobi coordinates x_α^s; $\alpha = 1,2$; $s = 1,2,\cdots,A-1$. The translationally invariant Hamiltonian for particles interacting through harmonic oscillator forces takes then the form[1]

$$H_0 = \tfrac{1}{2}\sum_{s=1}^{A-1}\sum_{\alpha=1}^{2}\left[(x_\alpha^s)^2 + (p_\alpha^s)^2\right], \tag{2.1}$$

where p_α^s is th momentum canonically conjugate to x_α^s, and we use units in which $\hbar$, the mass of the particle, and frequency of the oscillator are 1.

The coordinate transformation that brings out the collective degrees of freedom was introduced by Dzublik *et al.*[6] and Zickendraht,[7] and for this problem it has the form[1]

$$x_\alpha^s = \sum_{\beta=1}^{2}\rho_\beta D^1_{\beta\alpha}(\vartheta)D^1_{A-3+\beta,s}(\chi), \tag{2.2}$$

where ρ_1^2,ρ_2^2 are connected with the principal moments of inertia of the A-body system, ϑ is the Euler angle taking us from the frame of reference fixed in the body to the one fixed in space, and we have $2A-5$ coordinates more, denoted by χ's, that parametrize the orthogonal group $O(A-1)$ mentioned previously. In (2.2)

$$\mathbf{D}^1(\vartheta) = \|D^1_{\beta\alpha}(\vartheta)\| = \begin{bmatrix} \cos\vartheta & \sin\vartheta \\ -\sin\vartheta & \cos\vartheta \end{bmatrix} \tag{2.3}$$

is a 2×2 matrix for the defining irreducible representation charcterized by 1 (which is the reason for the upper index of D^1) of the O(2) group. We have similar interpretation for $\|D^1_{ts}(\chi)\|$, only that now the group is $O(A-1)$, and as we do not need the full matrix of the representaion but just the rows $t = A-3+\beta$, $\beta = 1,2$, we have only[1] $2A-5$ of the χ's rather than the full complement[1] of $(\frac{1}{2})(A-1)(A-2)$.

Carrying out the transformation (2.2) for the Hamiltonian (2.1) where $p_\alpha^s = -i\partial/\partial x_\alpha^s$, we can express it in terms of the ρ_1,ρ_2,ϑ, and $(2A-5)$ χ's as well as their derivatives. If we restrict ourselves to a scalar representation of the $O(A-1)$ orthogonal group, the Hamiltonian[1] (which we designate by H_c rather than the $1/2\,H_c$ or Ref. 1) will depend only on $\rho_1,\rho_2,\vartheta,\partial/\partial\rho_1,\partial/\partial\rho_2,\partial/\partial\vartheta$, and if we carry out the point transformation

$$\rho_1 = \rho\cos\gamma, \tag{2.4a}$$
$$\rho_2 = \rho\sin\gamma, \tag{2.4b}$$
$$r = \rho^2/2, \tag{2.4c}$$
$$\Theta = 2\gamma + \pi/2, \tag{2.4d}$$
$$\varphi = 2\vartheta, \tag{2.4e}$$

it becomes[1]

$$H_c = \tfrac{1}{2}r(-\nabla^2 + 1), \tag{2.5}$$

in which the Laplacian ∇^2 is given in terms of the spherical coordinates (r,Θ,φ). This Hamiltonian is identical to the one of the Coulomb problem[10,11] when we replace the radial coordinate r by $r/(n+1)$, where n is the total quantum number[1] starting at $n=0$.

The Hamiltonian (2.1) commutes with the generators of its U(2) symmetry group given by[1]

$$\mathscr{C}_{\alpha\beta} = \tfrac{1}{2}\sum_{s=1}^{A-1}(x_\alpha^s x_\beta^s + p_\alpha^s p_\beta^s) + \frac{i}{2}\sum_{s=1}^{A-1}(x_\alpha^s p_\beta^s - p_\alpha^s x_\beta^s);\ \alpha,\beta = 1,2, \tag{2.6}$$

and thus also with those of it SU(2) subgroup given by

$$J_1 = \tfrac{1}{2}(\mathscr{C}_{11} - \mathscr{C}_{22}),$$
$$J_2 = -\tfrac{1}{2}(\mathscr{C}_{12} + \mathscr{C}_{21}),\quad J_3 = \tfrac{1}{2}i(\mathscr{C}_{12} - \mathscr{C}_{21}). \tag{2.7}$$

As shown in Ref. 1, $J_\alpha = A_\alpha$, $\alpha = 1,2$, where A_α are the first two components of the Runge–Lenz vector, while $J_3 = L_3$, where L_3 is the third component of the angular momentum. These operators satisfy the standard commutations relations of the generators of the SU(2) group. The microscopic collective states we are interested in will then be eigenstates of H_C, J^2, J_3, where all those operators are expresed[1] in terms of the position vector $\mathbf{r} = (x_1,x_2,x_3)$ whose spherical coordinates are given by the r,Θ,φ of (2.4c)–(2.4e) and the corresponding momenta $\mathbf{p} = (p_1,p_2,p_3) = -i\nabla$.

As discussed in Ref. 1, the macroscopic collective model starts from the idea of quadrupole vibrations of a two-dimensional liquid drop that leads to two component δ bosons. To this we add a scalar σ boson to get a σ-δ three-component boson analogous to the six-dimensional s-d boson system[4] in three-dimensional physical space. The macroscopic collective Hamiltonian $\mathscr{H}_C$ is then the one of the three-dimensional oscillator.

$$\mathscr{H}_C = \tfrac{1}{2}(P^2 + R^2), \tag{2.8}$$

where here we designate the coordinates $\mathbf{R} = [X_1, X_2, X_3]$ and momenta $\mathbf{P} = [P_1, P_2, P_3]$ by capital letters. The integrals of motion we are interested in for $\mathscr{H}_C$, when we consider the rotational limit,[1] are those associated with the standard angular momentum

$$\mathscr{L} = \mathbf{R}\times\mathbf{P}. \tag{2.9}$$

The corresponding eigenstates will then be characterized by the eigenvalues of $\mathscr{H}_C$, $\mathscr{L}^2$, $\mathscr{L}_3$.

The main objective of this paper will be to find the canonical transformation

$$X_i = X_i(\mathbf{r},\mathbf{p}), P_i = P_i(\mathbf{r},\mathbf{p}), \{X_i,P_j\} = \delta_{ij} \tag{2.10}$$

(where the braces $\{\ ,\ \}$ stand for the Poisson bracket) such that

$$H_C = \mathscr{H}_C, \quad \mathbf{J} = \mathscr{L}. \tag{2.11}$$

To achieve this purpose, we shall first indicate in the next section how can we find a canonical transformation taking $\mathbf{J} = (A_1, A_2, L_3)$ into $\mathbf{L} = (L_1, L_2, L_3)$. Thus, the remaining problem will be to find the canonical transformation (2.10) that relates

$$H_C = \mathscr{H}_C, \quad \mathbf{L} = \mathscr{L}, \tag{2.12}$$

which will be discussed in Secs. 4 and 5.

III. CANONICAL TRANSFORMATIONS RELATING DIFFERENT SETS OF INTEGRALS OF MOTION OF THE COULOMB PROBLEM

The Coulomb Hamiltonian H_C of (2.5) admits as integrals of motion the angular momentum L_i and Runge-Lenz vectors A_i,[10,11] whose components $i = 1,2,3$ satisfy the Poisson bracket relations

$$\{L_i,H_C\} = 0 \tag{3.1a}$$

$$\{A_i,H_C\} = 0, \tag{3.1b}$$

$$\{L_i,L_j\} = L_k, \tag{3.1c}$$

$$\{A_i,A_j\} = L_k, \tag{3.1d}$$

$$\{L_i,A_j\} = A_k, \tag{3.1e}$$

where i,j,k are cyclic permutations of 1,2,3.

In this section we shall deal with H_C,L_i,A_j; $i,j = 1,2,3$, as classical concepts and denote by T some linear combination of them, i.e.,

$$T = aH_C + \sum_{i=1}^{3} b_i L_i + \sum_{i=1}^{3} c_i A_i, \tag{3.2}$$

where a,b_i,c_i are arbitrary constants. The T can act on an observable $F(x_i,p_j)$ through the Poisson bracket relation

$$\{T,F\} = \sum_{k=1}^{3}\left(\frac{\partial T}{\partial x_k}\frac{\partial F}{\partial p_k} - \frac{\partial T}{\partial p_k}\frac{\partial F}{\partial x_k}\right). \tag{3.3}$$

If we want to obtain the one parameter group of canonical transformations associated with T, where we designate the parameter by s in the interval $0 \leqslant s < \infty$, we need to determine $F(x_i,p_j,s)$ that satisfies the first order linear differential equation[13]

$$\frac{dF(x_i,p_j,s)}{ds} = \{T,F(x_i,p_j,s)\}, \tag{3.4}$$

with the initial condition

$$F(x_i,p_j,0) = F(x_i,p_j). \tag{3.5}$$

Defining the operator[13]

$$(T)_{\mathrm{op}} = \sum_{k=1}^{3}\left(\frac{\partial T}{\partial x_k}\frac{\partial}{\partial p_k} - \frac{\partial T}{\partial p_k}\frac{\partial}{\partial x_k}\right), \tag{3,6}$$

it is clear from (3.2)-(3.4) that

$$F(x_i,p_j,s) = \exp[s(T)_{\mathrm{op}}]F(x_i,p_j), \tag{3.7}$$

where

$$\exp[s(T)_{\mathrm{op}}] = \sum_{n=0}^{\infty}(n!)^{-1}s^n[(T)_{\mathrm{op}}]^n. \tag{3.8}$$

If F is the coordinate x_i or momentum p_j, we obtain for a given observable T and value s of the parameter, the new coordinate $x_i(s)$ or momentum $p_j(s)$ which are related with $x_i(0) = x_i$ or $p_j(0) = p_j$ through the canonical transformation (3.7).

Let us consider

$$T = A_3, \tag{3.9a}$$

$$s = \tfrac{1}{2}\pi, \tag{3.9b}$$

and see how it affects, through (3.7), the H_C,L_i,A_i themselves. Clearly, from (3.1), it leaves H_C, L_3, A_3 invariant, but from

$$(A_3)_{\mathrm{op}}L_1 = \{A_3,L_1\} = A_2, \tag{3.10a}$$

$$(A_3)_{\mathrm{op}}L_2 = \{A_3,L_2\} = -A_1, \tag{3.10b}$$

$$(A_3)^2_{\mathrm{op}}L_1 = \{A_3,A_2\} = -L_1, \tag{3.10c}$$

$$(A_3)^2_{\mathrm{op}}L_2 = -\{A_3,A_1\} = L_2, \tag{3.10d}$$

we get

$$\begin{aligned}\exp[(\pi/2)(A_3)_{\mathrm{op}}]\begin{pmatrix}L_1\\L_2\end{pmatrix} &= \cosh[(\pi/2)(A_3)_{\mathrm{op}}]\begin{pmatrix}L_1\\L_2\end{pmatrix} + \sinh[(\pi/2)(A_3)_{\mathrm{op}}]\begin{pmatrix}L_1\\L_2\end{pmatrix}\\ &= \cos(\pi/2)\begin{pmatrix}L_1\\L_2\end{pmatrix} + \sin(\pi/2)\begin{pmatrix}A_2\\-A_1\end{pmatrix} = \begin{pmatrix}A_2\\-A_1\end{pmatrix}.\end{aligned} \tag{3.11}$$

Thus $\exp[(\pi/2)(A_3)_{\mathrm{op}}]$ when applied to the vector $\mathbf{L} = (L_1,L_2,L_3)$ transforms it into the vector $(A_2, -A_1, L_3)$.

In a similar fashion from the fact that

$$(L_3)_{\mathrm{op}}A_1 = \{L_3,A_1\} = A_2, \tag{3.12a}$$

$$(L_3)_{\mathrm{op}}A_2 = \{L_3, A_2\} = -A_1, \tag{3.12b}$$

$$(L_3)^2_{\mathrm{op}}A_1 = \{L_3,A_2\} = -A_1, \tag{3.12c}$$

$$(L_3)^2_{\mathrm{op}}A_2 = -\{L_3A_1\} = -A_2, \tag{3.12d}$$

we obtain

$$\exp[(\pi/2)(L_3)_{\mathrm{op}}]\begin{pmatrix}A_1\\A_2\end{pmatrix} = \begin{pmatrix}A_2\\-A_1\end{pmatrix}. \tag{3.13}$$

Thus we see that

$$\exp[-(\pi/2)(L_3)_{\rm op}]\exp[(\pi/2)(A_3)_{\rm op}]\begin{pmatrix}L_1\\L_2\\L_3\end{pmatrix}=\begin{pmatrix}A_1\\A_2\\L_3\end{pmatrix}, \tag{3.14}$$

providing us with the cannonical transformation that takes the vector $\mathbf{L}=(L_1,L_2,L_3)$ into $\mathbf{J}=(A_1,A_2,L_3)$. We note that as $[L_3,A_3]=0$ we can also write the operator in (3.14) as $\exp[(\pi/2)(A_3-L_3)_{\rm op}]$.

From the above discussion we see that we can obtain the canonical transformation we are looking for by restricting ourselves to the one that maps the Coulomb on the oscillator problem taking $\mathbf{L}$ into the $\mathscr{L}$ of (2.9). To achieve this purpose we shall derive first in the next section a quantum mechanical operator relation between the generators of the dynamical groups associated with the oscillator and Coulomb problems.

IV. MATRIX ELEMENTS OF THE GENERATORS OF THE DYNAMICAL GROUPS OF THE OSCILLATOR AND COULOMB PROBLEMS IN THEIR RESPECTIVE BASIS

We shall start our discussion with the oscillator problem. Normally one speaks of Sp (6) as its dynamical group,[13] but another possibility is to consider the group whose generators are the creation and annihilation operators

$$\boldsymbol{\eta}=1/\sqrt{2}(\mathbf{R}-i\mathbf{P}), \tag{4.1a}$$

$$\boldsymbol{\xi}=1/\sqrt{2}(\mathbf{R}+i\mathbf{P}), \tag{4.1b}$$

plus the number and angular momentum operators

$$\mathscr{N}=\boldsymbol{\eta}\cdot\boldsymbol{\xi}=\mathscr{H}_C-\tfrac{3}{2}, \tag{4.1c}$$

$$\mathscr{L}=-i(\boldsymbol{\eta}\times\boldsymbol{\xi})=\mathbf{R}\times\mathbf{P}. \tag{4.1d}$$

Clearly these operators together with 1 form a Lie algebra as their Poisson brackets (related to the commutators through $\{A,B\}=-i[A,B]$ are also in the set as shown in Table I.

Furthermore, any eigenstate $|NLM\rangle$ of the operators $\mathscr{N},\mathscr{L}^2,\mathscr{L}_3$ can be expressed as a polynomial function of the creation operators η_i applied to the ground state $|0\rangle$, i.e.,[14]

$$|NLM\rangle=A_{NL}(\boldsymbol{\eta}\cdot\boldsymbol{\eta})^{(N-L)/2}\mathscr{Y}_{LM}(\boldsymbol{\eta})|0\rangle, \tag{4.2a}$$

where $\mathscr{Y}_{LM}(\boldsymbol{\eta})$ is a solid spherical harmonic of the variable and the normalization constant is given by[14]

$$A_{NL}=(-1)^{(N-L)/2}[4\pi/(N+L+1)!!(N-L)!!]^{\frac{1}{2}}, \tag{4.2b}$$

with $N-L$ even.

TABLE I. Poisson brackets of the generators of the dynamical group for the oscillator problem. The square associated with a given row (e.g., $\mathscr{L}_i$) and column (e.g., η_j) is the Poisson bracket (e.g., $\{\mathscr{L}_i,\eta_j\}=\eta_k$), where i,j,k are cyclic permutations of 1,2,3. As $\{A,B\}=-\{B,A\}$ the expressions below the antidiagonal are suppressed.

	ξ_j	η_j	$\mathscr{L}_j$	$\mathscr{N}$
$\mathscr{N}$	$i\xi_j$	$-i\eta_j$	0	0
$\mathscr{L}_i$	ξ_k	η_k	$\mathscr{L}_k$	
η_i	$i\delta_{ij}$	0		
ξ_i	0			

Clearly then, all the states of the harmonic oscillator belong to a single representation of the Lie algebra whose generators are (4.1), and thus we can consider that they define a dynamical group for the oscillator.

As a last point related with the oscillator we consider the matrix elements of the generators (4.1) with respect to the states $|NLM\rangle$. Those of $\mathscr{N}$, $\mathscr{L}_i$ are obvious, while that of ξ_i is the Hermitian conjugate of that of η_i. The latter is a vector and taking it in spherical components η_τ, $\tau=1,0,-1$ instead of the cartesian ones η_i, $i=1,2,3$, we have

$$\langle N'L'M'|\eta_\tau|NL'M\rangle=\langle N'L'\|\eta\|NL\rangle\langle LM,1\tau|L'M'\rangle, \tag{4.3}$$

where the last bracket $\langle\ |\ \rangle$ is a Clebsch–Gordan coefficient. The reduced matrix elements are given in Ref. 14, and the only ones different from zero are

$$\langle N+1,L+1\|\eta\|NL\rangle=[(N+L+3)(L+1)/(2L+3 \tag{4.4a}$$

$$\langle N+1,L-1\|\eta\|NL\rangle=[(N-L+2)L/(2L-1)]^{1/2} \tag{4.4b}$$

We now turn our attention to the Coulomb problem. The dynamical group is then O(4,2), and its generators expressed in terms of the lower case coordinates $\mathbf{r}=(x_1,x_2,x_3)$ and momenta $\mathbf{p}=(p_1,p_2,p_2)$ discussed in Sec. 2 have the form,[15,16]

$$L_i=(\mathbf{r}\times\mathbf{p})_i\quad(i=1,2,3\text{ everywhere}), \tag{4.5a}$$

$$A_i=-\tfrac{1}{2}x_i(p_2-1)+p_i(\mathbf{r}\cdot\mathbf{p}), \tag{4.5b}$$

$$N_i=-A_i+x_i=\tfrac{1}{2}x_i(p^2+1)-p_i(\mathbf{r}\cdot\mathbf{p}), \tag{4.5c}$$

$$N_4=(2/3)(\mathbf{r}\cdot\mathbf{p})+(1/3)(\mathbf{p}\cdot\mathbf{r}), \tag{4.5d}$$

$$K_i=rp_i, \tag{4.5e}$$

$$K_4=\tfrac{1}{2}r(p^2-1)=H_C-r, \tag{4.5f}$$

$$\mathfrak{N}=\tfrac{1}{2}r(p^2+1)=H_C. \tag{4.5g}$$

The L_i and A_i, $i=1,2,3$ are respectively the angular momentum and Runge–Lenz vectors which are the six generators of the O(4) symmetry group of the Coulomb problem. The generators (4.5) form a Lie algebra as their Poisson brackets, given in Table II, are also in the set. The order in which the components x_i,p_j appear in the generators (4.5) was chosen in such a way so that Table II applies not only to the classical Poisson brackets but also to the quantum ones related to the commutators by $\{A,B\}=i[A,B]$. Note from Table II that the ten operators $\mathfrak{N},L_i,N_i,K_i$, $i=1,2,3$ close under the Poisson bracket relation, and thus form a subgroup of O(4,2) which can be identified[15] with O(3,2).

The eigenstates $|nlm\rangle$ of the operators, $\mathfrak{N}$, $L^2=L_1^2+L_2^2+L_3^2$, L_3 in (4.5) have the form[1,17]

$$|nlm\rangle=2^{l+\frac{1}{2}}[2(n-l)!/\Gamma(n+l+2)]^{1/2}\,r^le^{-r}L_{n-l}^{2l+1}(2r)Y_{lm}(\theta,\phi), \tag{4.6}$$

where $n=0,1,2\cdots$ is the total quantum number (i.e., eigenvalue[1] of $\mathfrak{N}-1$), L_{n-l}^{2l+1} is a Laguerre polynomial, and $Y_{lm}(\theta,\phi)$ a spherical harmonic. All these states belong to a single irreducible representation[15] of O(4,2) and in fact we show explicitly below that they can be related by the generators of the subgroup 0(3,2) mentioned in the previous paragraph.

TABLE II. Poisson brackets of the generators (4.5) of the dynamical group 0(4,2) for the Coulomb problem. The square associated with a given row (e.g., A_i) and a given column (e.g., N_j) is the Poisson bracket (e.g, $\{A_i, N_j\} = \delta_{ij} N_4$), where i, j, k are cyclic permutations of 1,2,3. As $\{A,B\} = -\{B,A\}$ the expressions below the antidigiagonal are suppressed.

	K_4	K_j	N_4	N_j	$\mathfrak{N}$	A_j	L_j
L_i	0	K_k	0	N_k	0	A_k	L_k
A_i	K_i	$-\delta_{ij} K_4$	N_i	$-\delta_{ij} N_4$	0	L_k	
$\mathfrak{N}$	N_4	N_j	$-K_4$	$-K_j$	0		
N_i	0	$\delta_{ij}\mathfrak{N}$	A_i	$-L_k$			
N_4	$\mathfrak{N}$	0	0				
K_i	A_i	$-L_k$					
K_4	0						

group 0(3,2) mentioned in the previous paragraph.

Our next step is to find the matrix elements of the generators (4.5) in the basis (4.6). Those of $\mathfrak{N}$, L_i, are obvious while those of A_i can be obtained by expanding $|nlm\rangle$ in terms of eigenfunctions in parabolic coordinates as shown in the appendix of Ref. 1. The matrix elements that will be of particular interest to us are those of N_i, K_i and specially of a linear combination of them denoted by $B_i^{\pm}$ which we define as

$$B_\tau^{\pm} = (N_\tau \mp i K_\tau), \quad \tau = 1,0,-1, \tag{4.7}$$

and write in terms of spherical components $\tau = 1,0,-1$ rather than the Cartesian ones $i = 1,2,3$. As the matrix element of $\mathbf{B}^-$ is related to the Hermitian conjugate of that of $\mathbf{B}^+$, we restrict ourselves to the latter which, being a vector, takes the form

$$\langle n'l'm'|B_\tau^+|nlm\rangle = \langle n'l'\|B^+\|nl\rangle\langle lm;1\tau|l'm'\rangle. \tag{4.8}$$

The reduced matrix element can be calculated by applying the operator

$$B^+{}_1 = \frac{x_-}{r}\mathfrak{N} - \left(r\frac{\partial}{\partial r} + 1 - r\right)\frac{\partial}{\partial x_+},$$

$$x_\pm = \mp\frac{1}{\sqrt{2}}(x_1 \pm i x_2), \tag{4.9}$$

to the state $|nll\rangle$ as discussed in Appendix A. The only reduced matrix elements that are different from zero then take the values

$$\langle n+1, l+1\|B^+\|n,l\rangle$$
$$= [(n+l+3)(n+l+2)(l+1)/(2l+3)]^{1/2}, \tag{4.10a}$$

$$\langle n+1, l-1\|B^+\|n,l\rangle$$
$$= -[(n-l+1)(n-l+2)\,l/(2l-1)]^{1/2} \tag{4.10b}$$

We would like now to find some function of $\mathbf{B}^+$ whose reduced matrix elements with respect to the states $|nlm\rangle$ would have an identical form as those of $\boldsymbol{\eta}$ with respect to $|NLM\rangle$ given in (4.4). This will allow us to establish an operator correspondence which in the classical limit provides the canonical transformation we want to obtain.

To achieve our objective we first note that from Table II and the definition (4.7), we have with i,j,k, being cyclic permutations of 1,2,3 that

$$\{L_i, B_j^+\} = B_k^+, \; i,j,k \tag{4.11}$$

and thus we conclude that

$$i\{L^2, \mathbf{B}^+\} = [L^2, \mathbf{B}^+] = -i[(\mathbf{L}\times\mathbf{B}^+) - (\mathbf{B}^+\times\mathbf{L})].$$

The reduced matrix element of $[L^2, B_\tau^+]$ with respect to the states $|nlm\rangle$ becomes then

$$\langle n+1, l+1\|[L^2, B^+]\|nl\rangle$$
$$= 2(l+1)[(n+l+3)(n+l+2)(l+1)/(2l+3)]^{\frac12}, \tag{4.13a}$$

$$\langle n+1, l-1\|[L^2, B^+]\|nl\rangle$$
$$= 2l\,[(n-l+1)(n-l+2)l/(2l-1)]^{\frac12}, \tag{4.13b}$$

as follows immediately from (4.10).

From (4.10) and (4.13) we get the following values for the reduced matrix elements of operators that are linear combinations of $[L^2, \mathbf{B}^+]$ and $\mathbf{B}^+$:

$$\langle n+1, l'\|[L^2, B^+] + 2l\,B^+\|nl\rangle$$
$$= 2(2l+1)(n+l+2)^{\frac12}[(n+l+3)(l+1)/(2l+3)]^{\frac12}\delta_{l'l+1}$$

$$\langle n+1, l'\|[L^2, B^+] - 2(l+1)B^+\|nl\rangle$$
$$= 2(2l+1)(n-l+1)^{\frac12}[(n-l+2)l/(2l-1)]^{\frac12}\delta_{l'l-1}. \tag{4.14b}$$

It is clear therefore that the linear combination

$$\frac{[L^2, B_\tau^+] + 2lB_\tau^+}{2(2l+1)(n+l+2)^{\frac12}} + \frac{[L^2, B_\tau^+] - 2(l+1)B_\tau^+}{2(2l+1)(n-l+1)^{\frac12}}, \tag{4.15}$$

has reduced matrix elements with respect to $|nlm\rangle$ of exactly the same form as those of η_τ in (4.4), if we replace in the latter the capital by lower case letters.

The expression (4.15) is not yet an operator as it contains the eigenvalues n,l of a specific ket. We can transform it though into an operator by recalling that $n+1$, $l(l+1)$ are eigenvalues of $\mathfrak{N}$, L^2, i.e.,[1]

$$\mathfrak{N}|nlm\rangle = (n+1)|nlm\rangle, \; L^2|nlm\rangle = l(l+1)|nlm\rangle. \tag{4.16}$$

Thus, for example, we can establish the correspondence

$$[2(2l+1)]^{-1} \leftrightarrow 1/4\,(L^2 + 1/4)^{\;\frac12}, \tag{4.17a}$$

when the operator acts on the ket $|nlm\rangle$. Similarly, we have the correspondence

$$[(n+1) \pm (l \pm 1 + \tfrac12) - \tfrac12]^{-\frac12}$$
$$\leftrightarrow [(\mathfrak{N} - 1) \pm (L^2 + 1/4)^{\frac12} - (1/2)]^{-\frac12} \tag{4.17b}$$

when the operator acts on the bra $\langle n+1, l\pm 1, m|$. Using

(4.12) we see then that the vector operator

$$\mathbf{I}^+ \equiv [(\mathfrak{R}-1)+(L^2+\tfrac{1}{4})^{\frac{1}{2}}-\tfrac{1}{2}]^{-\frac{1}{2}}$$
$$\times\{i(\mathbf{B}^+\times\mathbf{L})-i(\mathbf{L}\times\mathbf{B}^+)+\mathbf{B}^+[2(L^2+\tfrac{1}{4})^{\frac{1}{2}}-1]\}$$
$$\times(\tfrac{1}{4})^{-1/2}(L^2+\tfrac{1}{4})^{-\frac{1}{2}}+[(\mathfrak{R}-1)-(L^2+\tfrac{1}{4})^{\frac{1}{2}}-\tfrac{1}{2}]^{-\frac{1}{2}}$$
$$\times\{i(\mathbf{B}^+\times\mathbf{L})-i(\mathbf{L}\times\mathbf{B}^+)+\mathbf{B}^+[-2(\mathbf{L}^2+\tfrac{1}{4})^{\frac{1}{2}}-1]\}$$
$$\times\tfrac{1}{4}(L^2+\tfrac{1}{4})^{-\frac{1}{2}}, \tag{4.18}$$

has exactly the smae matrix elements with respect to the states $|nlm>$ as the vector operator $\boldsymbol{\eta}$ of (4.1a) has with respect to the states $|NLM>$.

The operator (4.18) is the main result of the present paper, as in the next section we will show that in the classical limit it will provide the canonical transformation we are looking for.

V. CANONICAL TRANSFORMATIONS RELATING THE COULOMB AND OSCILLATOR PROBLEMS

We wish now to express the vector operator $\mathbf{I}^+$ of (4.18) as a classical observable. We note first that if instead of dimensionless units we had used normal ones, the $\frac{1}{4}$ following L^2 would contain a factor $\hbar^2$, and the $\frac{1}{2}$, 1 appearing in (4.18) a factor $\hbar$. Thus, in the classical limit when $\hbar \to 0$ these factors would disappear and, in particular, $(L^2+\frac{1}{4})^{\frac{1}{2}}$ can be replaced by $L=(L_1^2+L_2^2+L_3^2)^{\frac{1}{2}}$. Furthermore, classically our observables commute and thus $(\mathbf{L}\times\mathbf{B}^+)=-(\mathbf{B}^+\times\mathbf{L})$. If we now equate the classical limit of $\mathbf{I}^+$ with $\boldsymbol{\eta}=(1/\sqrt{2})(\mathbf{R}-i\mathbf{P})$, replace in the former $\mathbf{B}^+$ by $\mathbf{N}-i\mathbf{K}$, where $\mathbf{N}$, $\mathbf{K}$ given by (4.5c) and (4.5e) are real vectors, and equate real and imaginary parts, we obtain

$$\mathbf{R}=W_+L^{-1}(\mathbf{K}\times\mathbf{L})+W_-\mathbf{N}, \tag{5.1a}$$
$$\mathbf{P}=-W_+L^{-1}(\mathbf{N}\times\mathbf{L})+W_-\mathbf{K}, \tag{5.1b}$$

where

$$W_\pm \equiv 1/\sqrt{2}[(\mathfrak{R}+L)^{-\frac{1}{2}}\pm(\mathfrak{R}-L)^{-\frac{1}{2}}]. \tag{5.1c}$$

Another form of the canonical transformation can be obtained by noting from (4.5c) and (4.5e) that

$$\mathbf{N}\times\mathbf{K}=\mathfrak{R}\mathbf{L}. \tag{5.2}$$

Replacing the vector $\mathbf{L}$ appearing in (5.1) by its value (5.2), the canonical transformation expressed in components takes the form

$$X_i=(\mathfrak{R}^{-1}L^{-1}W_+K^2+W_-)N_i-\mathfrak{R}^{-1}L^{-1}W_+(\mathbf{N}\cdot\mathbf{K})K_i, \tag{5.3a}$$

$$P_i=-\mathfrak{R}^{-1}L^{-1}W_+(\mathbf{N}\cdot\mathbf{K})N_i+(\mathfrak{R}^{-1}L^{-1}W_+N^2+W_-)K_i. \tag{5.3b}$$

While the matrix elements of $\mathbf{I}^+$ with respect to the states $|nlm\rangle$ suggest that in the classical limit $\boldsymbol{\eta}=\mathbf{I}^+$ will give the canonical transformation, it does not prove it. To achieve this we must show that the transformation (5.3) guarantees that, when considered as classical observables, we have

$$\{X_i,P_j\}=\delta_{ij} \tag{5.4}$$
$$\tfrac{1}{2}(P^2+R^2)=\tfrac{1}{2}r(p^2+1), \tag{5.5a}$$
$$\mathbf{R}\times\mathbf{P}=\mathbf{r}\times\mathbf{p}. \tag{5.5b}$$

This can be obtained, in a laborious but straightforward fashion, by using Table II for the Poisson brackets and relations such as (5.2) and its square, i.e.,

$$(\mathbf{N}\times\mathbf{K})^2=N^2K^2-(\mathbf{N}\cdot\mathbf{K})^2=\mathfrak{R}^2L^2, \tag{5.6}$$

as well as

$$N^2+K^2=\mathfrak{R}^2+L^2, \tag{5.7}$$

where the latter follows from the definitions (4.5).

We have then the canonical transformation that relates the observables (5.5a) and (5.5b) but as mentioned in the previous sections, we want to find the one that relates

$$\tfrac{1}{2}(P^2+R^2)=\tfrac{1}{2}r(p^2+1), \tag{5.8a}$$
$$\mathbf{R}\times\mathbf{P}=\mathbf{J}, \tag{5.8b}$$

where $\mathbf{J}=(A_1,A_2,L_3)$. We found though in Sec. 3, that the operator $\exp[(\pi/2)(A_3-L_3)_{\rm op}]$ acting on $\mathbf{L}$ transforms it into $\mathbf{J}$, while it leaves $\mathfrak{R}$ invariant, as $\mathbf{L},\mathbf{A}$ are integrals of motion of the Coulomb problem. Furthermore, from the Poisson bracket relations of Table II we see that for $j=1,2,3$,

$$(A_3)_{\rm op}N_j=\{A_3,N_j\}=-\delta_{j3}N_4, \tag{5.9a}$$
$$(A_3)^2_{\rm op}N_j=-\delta_{j3}\{A_3,N_4\}=-\delta_{j3}N_3, \tag{5.9b}$$
$$(A_3)_{\rm op}K_j=\{A_3,K_j\}=-\delta_{j3}K_4, \tag{5.9c}$$
$$(A_3)^2_{\rm op}K_j=-\delta_{j3}\{A_3,K_4\}=-\delta_{j3}K_3, \tag{5.9d}$$

EL4while with respect to $(L_3)_{\rm op}$, the observables N_j,K_j behave as ordinary vectors. Thus by a reasoning similar to the one given in Sec. 3 we conclude that

$$\exp[(\pi/2)(A_3-L_3)_{\rm op}]\begin{pmatrix}N_1\\N_2\\N_3\end{pmatrix}=\begin{pmatrix}-N_2\\N_1\\-N_4\end{pmatrix}, \tag{5.10}$$

$$\exp[(\pi/2)(A_3-L_3)_{\rm op}]\begin{pmatrix}K_1\\K_2\\K_3\end{pmatrix}=\begin{pmatrix}-K_2\\K_1\\-K_4\end{pmatrix}. \tag{5.11}$$

The canonical transformation associated with the relations (5.8) can then be obtained from (5.3) if in the right-hand side of the latter we carry out the following transformations

$$\mathfrak{R}\to\mathfrak{R},$$
$$\mathbf{L}=(L_1,L_2,L_3)\to\mathbf{J}=(A_1,A_2,L_3),$$
$$\mathbf{K}=(K_1,K_2,K_3)\to\mathbf{K}'\equiv(-K_2,K_1,-K_4),$$
$$\mathbf{N}=(N_1,N_2,N_3)\to\mathbf{N}'\equiv(-N_2,N_1,-N_4). \tag{5.12}$$

Thus we have achieved the purpose outlined in Sec. 2 of finding the canonical transformation that relates the macroscopic and microscopic nuclear collective Hamiltonians.

We now proceed to study the properties of the canonical transformation (5.3) starting from the fact that, as there is not a one to one correspondence between the energy levels of oscillator and Coulomb problems, the transformation is nonbijective.[12] In the next section we derive the ambiguity group associated with this nonbijectiveness.

VI. THE AMBIGUITY GROUP OF THE CANONICAL TRANSFORMATION

In Fig. 2 we draw the energy levels of the oscillator Hamiltonians $\mathscr{H}_C$ of (2.8) and the Coulomb one H_C of (2.5). Clearly the energy levels of the Coulomb case cover twice the energy levels of the oscillator as indicated in the figure. The

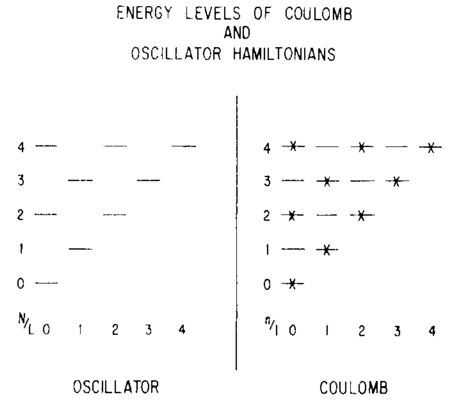


FIG. 2. The energy levels of the oscillator and Coulomb Hamiltonians [the latter in the Fock-Bargmann form $H_c = \frac{1}{2}r(p^2+1)$] as functions of the total and angular momentum quantum numbers. We note that there are twice as many levels in the Coulomb as in the oscillator case. We can establish a one to one relation $n = N$, $l = L$ between the levels marked by an x in the Coulomb case with those of the oscillator. Those unmarked can also be put into a one-to-one relation if we consider $n = N + 1$, $l = L$.

work of the authors[12] on the representation of canonical transformations immediately indicates that transformation (5.3) must then be nonbijective and that, in fact, there must be two points in the phase space $(\mathbf{r},\mathbf{p})$ of the Coulomb problem that map on a single one in the phase space $(\mathbf{R},\mathbf{P})$ of the oscillator problem, requiring two sheets for the latter.[12] Our first question will then be what is the canonical transformation that connects these two points, leading to the concept we denoted as the ambiuity group, associated in this case with (5.3).

To achieve this objective we start again with the quantum picture. We notice that the set of levels in the Coulomb case marked by x have the property that $n + l$; is even while for those unmarked $n + l$ is odd. There is one to one correspondence between the x levels on one hand and those unmarked on the other, with those of the oscillator, as indicated in Fig. 2. From the eigenvalues of $\mathfrak{N},L^2$ given in (4.16), it is clear that the operator written symbolically as

$$\mathcal{O} \equiv \exp\{i\pi[\mathfrak{N} + (L^2+\tfrac{1}{4})^{\frac{1}{2}} - (3/2)]\}, \tag{6.1}$$

when acting on the states $|nlm\rangle$ gives

$$\exp[i\pi(n+l)] = (-1)^{n+l}. \tag{6.2}$$

Thus it distinguishes between the states marked by x for which $n + l$ is even and those unmarked for which $n + l$ is odd.

A corresponding operator to $\mathcal{O}$ for the oscillator problem will be

$$\bar{\mathcal{O}} = \exp\{i\pi[\mathcal{N} + (\mathcal{L}^2+\tfrac{1}{4})^{1/2} - (\tfrac{1}{2})]\}, \tag{6.3}$$

which, when it acts on $|NLM\rangle$, gives

$$\exp[i\pi(N+L)] = (-1)^{N+L}. \tag{6.4}$$

We note though that as in the oscillator problem $N + L$ is even, $(-1)^{N+L}$ is always 1, and thus $\bar{\mathcal{O}}$ is the unit operator

Clearly then, when we apply the unitary transformation associated with $\mathcal{O}$ to $(\mathbf{r},\mathbf{P})$, i.e.,

$$\mathcal{O}^{\dagger}\begin{pmatrix}\mathbf{r}\\ \mathbf{p}\end{pmatrix}\mathcal{O} = \begin{pmatrix}\mathbf{r}'\\ \mathbf{p}'\end{pmatrix}, \tag{6.5}$$

we get other vector operators $(\mathbf{r}',\mathbf{p}')$, while the application of $\bar{\mathcal{O}}$ to $(\mathbf{R},\mathbf{P})$ leaves them invariant. We wish to implement the step (6.5), but in the classical picture. Again disregarding $\frac{1}{4}$, 3/2 as compared respectively with $L^2,\mathfrak{N}$ we can replace[13] the transformation (6.5) by

$$\exp[\pi(\mathfrak{N}+L)_{\rm op}]\begin{pmatrix}\mathbf{r}\\ \mathbf{p}\end{pmatrix}, \tag{6.6}$$

where $L = (L_1^2 + L_2^2 + L_3^2)^{\frac{1}{2}}$.

Rather than calculate this directly we first apply $\exp[\pi(\mathfrak{N}+L)_{\rm op}]$ to the vectors $\mathbf{K}$, $\mathbf{N}$, and $\mathbf{A}$. An analysis similar to that in Sec. 3, using Table II for the Poisson brackets, which is implemented in detail in Appendix B, shows that

$$\exp[\pi(\mathfrak{N}+L)_{\rm op}]\begin{bmatrix}\mathbf{K}\\ \mathbf{N}\\ \mathbf{A}\end{bmatrix} = \begin{bmatrix}\mathbf{K}\\ \mathbf{N}\\ -\mathbf{A}\end{bmatrix}. \tag{6.7}$$

We see then from (4.5) that

$$\mathbf{r} = \mathbf{N} + \mathbf{A}, \tag{6.8}$$

and thus

$$\begin{aligned}\exp[\pi(\mathfrak{N}+L)_{\rm op}]\mathbf{r} &= \mathbf{N} - \mathbf{A} = 2\mathbf{N} - \mathbf{r}\\ &= \mathbf{r}p^2 - 2\mathbf{p}(\mathbf{r}\cdot\mathbf{p}) = \mathbf{r}'.\end{aligned} \tag{6.9a}$$

Furthermore, again from (4.5), $\mathbf{p} = r^{-1}\mathbf{K}$, and thus from (6.7) and (6.9a),

$$\exp[\pi(\mathfrak{N}+L)_{\rm op}\mathbf{p} = |\mathbf{N}-\mathbf{A}|^{-1}\mathbf{K} = p^{-2}\mathbf{p} = \mathbf{p}']. \tag{6.9b}$$

As $\mathbf{K}$, $\mathbf{N}$ and also obviously $\mathfrak{N}$, L, remain invariant under the operation $\exp[\pi(N+L)_{\rm op}]$, it is clear that $(\mathbf{r},\mathbf{p})$, $(\mathbf{r}',\mathbf{p}')$ [where the latter are given by (6.9)], are two points in the phase space of the Coulomb problem that are mapped on a single point $(\mathbf{R},\mathbf{P})$ in the oscillator phase space through the transformation (5.3).

The ambiguity group of the canonical transformation (5.3) has then only the unit element e and the operation (6.9) that relates $(\mathbf{r},\mathbf{p})$ with $(\mathbf{r}',\mathbf{p}')$. One sees immediately that carrying the last operation twice we return to e.

It is of interest to note that in Secs. 4,5, and 6 we obtain explicitly the canonical transformation that maps the Coulomb Hamiltonian (in the Fock and Bargmann formulation) into the oscillator one, while taking the angular momenta of both problems into each other. Furthermore, if we write the total quantum numbers in the Coulomb and oscillator problems as

$$\begin{aligned}&n = 2s + l + \sigma,\ \ s,l = 0,1,2\cdots,\sigma = 0,1,\\ &N = 2S + L,\ \ S,L = 0,1,2\cdots,\end{aligned} \tag{6.10}$$

and denote the corresponding kets by

$$|nlm\rangle \equiv |slm,\sigma),\ \ |NLM\rangle \equiv |SLM\}, \tag{6.11}$$

we get for the representation of the canonical transformation mentioned above the expression[1]

$$\langle \mathbf{R}\sigma|U|\mathbf{r}\rangle = \sum_{s=0}^{\infty}\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\langle \mathbf{R}|slm\}(slm,\sigma|\mathbf{r}\rangle,$$

where $\sigma = 0,1$ is the ambiguity spin associated with the two irreducible representations of the ambiguity group mentioned in the previous paragraph.

We have here an example in three dimensions of the type of canonical transformations and their representations that were discussed in Ref. 12 only for one dimension.

In the concluding section we return to the problem of collective motions and discuss the implications of our results as well as their possible generalization when the A particles move in the physical three-dimensional space.

VII. CONCLUSION

The main conclusion we want to draw from this paper is that the microscopic collective Hamiltonian projected out of the A-body system through the scalar representation of $O(A-1)$ has a symmetry group U(3). This comes from the fact that U(3) is the symmetry group of the three-dimensional oscillator Hamiltonian associated with the σ-δ boson picture, and the canonical transformation (5.3) combined with (5.12) maps the microscopic collective Hamiltonian on it. Thus the symmetries of a σ-δ boson picture appear in a microscopic collective Hamiltonian in two-dimensional space and we could expect that the symmetries of the s-d boson picture[4] would appear in the physically relevant situation when the A particles move in three-dimensional space.

Considering first the problem classically the three-dimensional oscillator (2.8) has as generators of its U(3) symmetry group[18]

$$C_{ij} \equiv \eta_i \xi_j = \tfrac{1}{2}(X_i X_j + P_i P_j) + \tfrac{1}{2} i (X_i P_j - P_i X_j), \quad (7.1)$$

whose Poisson brackets take the form

$$\{C_{ij}, C_{i'j'}\} = -i(C_{ij'}\delta_{i'j} - C_{i'j}\delta_{ij'}). \quad (7.2)$$

If we substitute then X_i, P_j in (7.1) by their values (5.3) in which we make the replacements (5.12), we have a rather complex realization of the generators of U(3) in terms of $\mathbf{r} = (x_1, x_2, x_3)$, $\mathbf{p} = (p_1, p_2, p_3)$, the coordinates and momenta of the microscopic collective model. The appearace of U(3) was by no means expected. The Hamiltonian (2.1) for A particles interacting through harmonic oscillator forces in two-dimensional space, is equivalent to an oscillator in $(2A-2)$-dimensional space. Thus its symmetry group is $\mathbf{U}(2A-2)$ which, among others has the following subgroups

$$\begin{array}{cccc} U(2A-2) \supset & \mathscr{U}(2) \times & U(A-1) & \\ & \cup & \cup & \\ & S\mathscr{U}(2) & O(A-1) & \quad (7.3) \\ & \cup & \cup & \\ & \mathscr{O}(2) & S_A & \end{array}$$

In (7.3) the $\mathscr{O}$, O stand for orthogonal groups of the dimensions indicated and S_A is the symmetric group of permutations of A particles. The generators of $\mathscr{U}(2)$ are given by (2.6), those of $S\mathscr{U}(2)$ by $\mathbf{J} = (A_1, A_2, L_3)$ of (2.7), and the single generator of $\mathscr{O}(2)$ is $J_3 = L_3$. Nowhere in this picture is there a U(3), though from (2.11) the subgroup O(3) of U(3), whose generators are $\mathscr{L} = ((\mathscr{L}_1, \mathscr{L}_2, \mathscr{L}_3)$, coincides with $S\mathscr{U}(2)$, i.e, $\mathscr{L} = \mathbf{J}$. Thus the canonical transformation (5.3) with the replacements (5.12), was fundamental for detecting the presence or U(3) in the microscopic collective model.

If we want to pass to the quantum picture the generators of U(3) can be constructed if in the definition (7.1) we replace η_i by I_i^+ of (4.18) and the ξ_j by its Hermitian conjugate $(I_j^+)^\dagger \equiv I_j^-$, thus obtaining the operator

$$I_i^+ I_j^- \quad (7.4)$$

A note of warning should be given here. In Ref. 1 the Hamiltonian H_c of (2.5) was obtained not from projecting the scalar part of $O(A-1)$ from the A-body system, but by considering only the case $A = 3$. Vanagas[9] has shown though that in the collective coordinates $\rho_1, \rho_2, \vartheta$ the projection from the A-body system will add only the term

$$-\tfrac{1}{2}(A-3)\left(\frac{1}{\rho_1}\frac{\partial}{\partial \rho_1} + \frac{1}{\rho_2}\frac{\partial}{\partial \rho_2}\right) = -\frac{i}{2}(A-3)\left(\frac{1}{\rho_1}\pi_1 + \frac{1}{\rho_2}\pi_2\right)$$

(where $\pi_\alpha = -i\partial/\partial\rho_\alpha$, $\alpha = 1,2$) to the collective Haniltonian. This term is linear in the momenta π_α, and comparing it with the quadratic terms in π_α appearing in the rest of the kinetic energy, we conclude that, when we use normal rather than dimensionless units, it would contain an $\hbar$ and thus disappear in the classical limit. Therefore all the classical analysis presented here still holds when we project from the A rather than the $A = 3$ particle problem.

In the quantum case the Hamiltonian (2.5) is modified by a term depending on A. It is necessary then to find the dynamical group, which turns out to be another realization of O(3,2), for this Hamiltonian, and also to find the new kets $|nlm\rangle$. This has been done recently by Chacón and Moshinsky and in a future publication they will discuss the operators $I_i^\pm$ required in the quantum generators (7.4) of U(3).

As mentioned at the end of the first paragraph of this section, the interesting physical problem is related with the derivation of the canonical transformation that connects the microscopic collective Hamiltonian in three-dimensional space with an oscillator of appropriate dimension. In this case the collective variables are ρ_1, ρ_2, ρ_3 whose squares are related to the three principal moments of inertia and the three Euler angles $\vartheta_1, \vartheta_2, \vartheta_3$. Thus the microscopic collective Hamiltonian is in a configuration space of six and a phase space of twelve dimensions. We want to map it on a six-dimensional oscillator characterized by a single s and five $(m = 2,1,0,-1,-2)$ d bosons.

This problem will be considerably more difficult than the one discussed in this paper. Among other requirements, we will have to determine the generators of the dynamical group in terms of these six collective coordinates and their canonically conjugate momenta. Furthermore, we will need explicitly the eigenkets of the microscopic collective Hamiltonian associated with the irreducible representations (irreps) of the SU(3) symmetry group of this problem, which will be much more complicated to derive than $|nlm\rangle$ of (4.6). Once these eigenkets are available there is the problem of getting reduced matrix elements with respect to them of suitable operators such as the $\mathbf{B}^+$ in (4.10).

One may try to implement this ambitious program, though on the other hand, as the two-dimensional case already illustrates all the conceptual points, one may also work directly with the collective part i.e., the scalar representation of the $O(A-1)$ group] of Hamiltonian for A particles interacting through appropriate two-body forces, to see what are the predictions that have relevance for real nuclei.[19]

ACKNOWLEDGMENTS

The authors are indebted to Professor V. Vanagas and Professor K. B. Wolf for many helpful discussions in relation with the present problem.

APPENDIX A: REDUCED MATRIX ELEMENTS OF $\mathbf{B}^+$

In this Appendix we derive the reduced matrix elements (4.10) of $\mathbf{B}^+$ by applying the operator (4.9) for $B^{\pm}_{\pm 1}$ to the states $|nlm\rangle$ of (4.6). From the Wigner–Eckart theorem indicated in (4.8) we can restrict ourselves to applying B^{+}_{-1} to $|nll\rangle$, and from the commutation properties of $\mathbf{B}^+$ with $\mathfrak{N}$ and its polar vector character we obtain

$$B^{+}_{-1}|nll\rangle = |n+1,l+1,l-1\rangle \langle n+1,l+1\|B^+\|nl\rangle[(2l+1)(l+1)]^{-1/2} + |n+1,l-1,l-1\rangle\langle n+1,l-1\|B^+\|nl\rangle \times[(2l-1)/(2l+1)]^{1/2} \tag{A1}$$

where we already introduced in the square brakets on the right-hand side the relevant values of the Clebsch–Gordan coefficients.[20]

The ket $|nll\rangle$ can be written as

$$|nll\rangle = A_{\nu l} f_{\nu l}(r) x^l_+ , \tag{A2}$$

where we denote by ν the radial quantum number

$$\nu = n - l, \tag{A3}$$

and have

$$A_{\nu l} = [2^l(2l+1)!\nu!/\pi\Gamma(\nu+2l+2)]^{1/2}(l!)^{-1}, \tag{A4}$$

and

$$f_{\nu l}(r) = L^{2l+1}_{\nu}(2r)e^{-r}. \tag{A5}$$

A similar expression holds for $|n+1,l-1,l-1\rangle$ if we replace the corresponding quantum numbers. For $|n+1,l+1,l-1\rangle$ we have[20]

$$|n+1,l+1,l-1\rangle = A_{\nu l+1} f_{\nu l+1}(r)[(l+1)(2l+1)]^{\frac{1}{2}} x^{l-1}_+ \times[x_+x_- + l(2l+1)^{-1}r^2], \tag{A6}$$

where $x_{\pm} = \mp(1/\sqrt{2})(x_1 \pm ix_2)$, $x_0 = x_3$, $r^2 = -2x_+x_- + x_0^2$.

Applying the B^{+}_{-1} of (4.9) to $|nll\rangle$, we get

$$\begin{aligned} B^{+}_{-1}|nll\rangle &= [(x_-/r)\mathfrak{N} - (r\partial/\partial r + 1 - r)\partial/\partial x_+]A_{\nu l}f_{\nu l}(r)x^l_+ \\ &= A_{\nu l}x^{l-1}_+\{(x_-x_+/r)[(\nu+l+1)f_{\nu l} + (r\partial/\partial r + l + 1 - r)\partial f_{\nu l}/\partial r] - l(r\partial/\partial r + l - r)f_{\nu l}\} \\ &= A_{\nu l}x^{l-1}_+\{[x_-x_+ + l(2l+1)^{-1}r^2]Q - S\}, \end{aligned} \tag{A7}$$

where

$$Q \equiv 1/r[(\nu+l+1)f_{\nu l} + (r\partial/\partial r + l + 1 - r)\partial f_{\nu l}/\partial r], \tag{A8}$$

$$S \equiv l(r\partial/\partial r + l - r)f_{\nu l} + lr^2(2l+1)^{-1}Q. \tag{A9}$$

We now proceed to evaluate Q and S. To begin with let us denote

$$L^{2l+1}_{\nu}(2r) \equiv L(z), \quad z = 2r. \tag{A10}$$

We have then that

$$\begin{aligned} re^rQ &= (\nu+l+1)L + e^r(r\partial/\partial r + l + 1 - r)e^{-r}(\partial L/\partial r - L) \\ &= (\nu+l+1)L + (zd/dz + l + 1 - z)(2dL/dz - L) \\ &= 2zd^2L/dz^2 + (2l+2-3z)dL/dz + (\nu+z)L \\ &= -(2l+2+z)dL/dz + (z-\nu)L, \end{aligned} \tag{A11}$$

where the last line was obtained from the use of the equation for Laguerre polynominals[21]

$$zd^2L/dz^2 + (2l+2-z)dL/dz + \nu L = 0. \tag{A12}$$

Finally, with the help of Eqs. 2,4,5 of 8.971, p. 1037 of Ref. 21 we have

$$-(2l+2+z)dL^{2l+1}_{\nu}(z)/dz + (z-\nu)L^{2l+1}_{\nu}(z) = zL^{2l+3}_{\nu}(z), \tag{A13}$$

and thus

$$Q = 2e^{-r}L^{2l+3}_{\nu}(2r). \tag{A14}$$

Turning now our attention to S we see from (A9) and (A11) that it can be written as

$$\begin{aligned} e^rS &= l(zd/dz + l - z)L + [lz/(4l+2)][-(2l+2+2)dL/dz + (z-\nu)L] \\ &= [l/(4l+2)]\{(2l-z)zdL^{2l+1}_{\nu}(z)/dz + [2(2l+1)l - (4l+2+\nu)z + z^2]L^{2l+1}_{\nu}(z)\} \\ &\quad + l(4l+2)^{-1}(\nu+1)(\nu+2)L^{2l-1}_{\nu+2}(2r), \end{aligned} \tag{A15}$$

where we made use again of formulas 3–5 of 8.971, p. 1037 of Ref. 21.

We then get

$$\begin{aligned} B^{+}_{-1}|nll\rangle &= A_{\nu l}x^{l-1}_+\{[x_-x_+ + l(2l+1)^{-1}r^2]2e^{-r}L^{2l+3}_{\nu}(2r)\} \\ &\quad - A_{\nu l}x^{l-1}_+ l(4l+2)^{-1}(\nu+1)(\nu+2)L^{2l-1}_{\nu+2}(2r)e^{-r}. \end{aligned} \tag{A16}$$

Comparing it with (A1) and using (A2)-(A6), we immediately conclude that the reduced matrix elements $\langle n+1,l\pm 1\|B^+\|nl\rangle$ have the form (4.10).

APPENDIX B: DERIVATION OF THE EXPRESSION (6.7)

In this Appendix we apply $\exp[\pi(\mathfrak{N}+L)_{\rm op}]$ discussed in Sec. 6 to the vectors K_i, N_i, A_i, $i = 1,2,3$ so as to derive the expression (6.7). Note that we are dealing with a classical problem and thus from the definition (3.6) of $(\mathfrak{N}+L)_{\rm op}$ and Table II we get the Poisson brackets

$$\{\mathfrak{N}+L,K_i\} = \{\mathfrak{N},K_i\} + (2L)^{-1}\{L^2,K_i\} = N_i - L^{-1}(\mathbf{L}\times\mathbf{K})_i , \tag{B1}$$

$$\{\mathfrak{N}+L,N_i\} = \{\mathfrak{N},N_i\} + (2L)^{-1}\{L^2,N_i\} = -K_i - L^{-1}(\mathbf{L}\times\mathbf{N})_i, \tag{B2}$$

$$\{\mathfrak{N}+L,A_i\} = \{\mathfrak{N},N_i\} + (2L)^{-1}\{L^2,A_i\} = -L^{-1}(\mathbf{L}\times\mathbf{A})_i, \tag{B3}$$

as $L^2 = L_1^2 + L_2^2 + L_3^2$. Furthermore, as the Poisson brackets of L_i with $\mathfrak{R}$ and L vanish, we have also

$$\{\mathfrak{R} + L, L^{-1}(\mathbf{L}\times\mathbf{K})_i\} = L^{-1}(\mathbf{L}\times\{\mathfrak{R} + L,\mathbf{K}\})_i$$
$$= L^{-1}(\mathbf{L}\times[\mathbf{N} - L^{-1}(\mathbf{L}\times\mathbf{K})])_i$$
$$= L^{-1}(\mathbf{L}\times\mathbf{N})_i + \mathbf{K}_i$$
$$= -\{\mathfrak{R} + L, N_i\}, \quad \text{(B4)}$$

and similarly, we obtain

$$\{\mathfrak{R} + L, L^{-1}(\mathbf{L}\times\mathbf{N})_i\} = N_i + L^{-1}(\mathbf{L}\times\mathbf{K})_i = \{\mathfrak{R} + L, K_i\}, \quad \text{(B5)}$$

$$\{\mathfrak{R} + L, L^{-1}(L\times A)_i\} = A_i, \quad \text{(B6)}$$

where in all these relations we make use of the fact that from (4.5b), (4.5c), and (4.5e) we have $\mathbf{L}\cdot\mathbf{A} = \mathbf{L}\cdot\mathbf{N} = \mathbf{L}\cdot\mathbf{K} = 0$.

On the basis of the previous results we immediately conclude that for any integer ν we have the relations

$$(\mathfrak{R} + L)_{\text{op}}^{2\nu} K_i = (-1)^{\nu+1} 2^{2\nu-1}\{\mathfrak{R} + L, N_i\}, \quad \text{(B7)}$$

$$(\mathfrak{R} + L)_{\text{op}}^{2\nu+1} K_i = (-1)^{\nu} 2^{\nu}\{\mathfrak{R} + L, K_i\}, \quad \text{(B8)}$$

$$(\mathfrak{R} + L)_{\text{op}}^{2\nu} N_i = (-1)^{\nu} 2^{2\nu-1}\{\mathfrak{R} + L, K_i\}, \quad \text{(B9)}$$

$$(\mathfrak{R} + L)_{\text{op}}^{2\nu+1} N_i = (-1)^{\nu} 2^{2\nu}\{\mathfrak{R} + L, N_i\}, \quad \text{(B10)}$$

$$(\mathfrak{R} + L)_{\text{op}}^{2\nu} A_i = (-1)^{\nu} A_i, \quad \text{(B11)}$$

$$(\mathfrak{R} + L)_{\text{op}}^{2\nu+1} A_i = (-1)^{\nu+1} L^{-1}(\mathbf{L}\times\mathbf{A})_i. \quad \text{(B12)}$$

Writing now

$$\exp[\pi(\mathfrak{R} + L)_{\text{op}}] = \cosh[\pi(\mathfrak{R} + L)_{\text{op}}] + \sinh[\pi(\mathfrak{R} + L_{\text{op}})], \quad \text{(B13)}$$

and developing the cosh and sinh in power series we see that

$$\exp[\pi(\mathfrak{R} + L)_{\text{op}}]K_i = K_i - (\tfrac{1}{2})(\cos 2\pi - 1)\{\mathfrak{R} + L, N_i\} + (\tfrac{1}{2})\sin 2\pi \times \{\mathfrak{R} + L, K_i\} = K_i, \quad \text{(B14)}$$

$$\exp[\pi(\mathfrak{R} + L)_{\text{op}}]N_i = N_i + (\tfrac{1}{2})(\cos 2\pi - 1)\{\mathfrak{R} + L, K_i\} + (\tfrac{1}{2})\sin 2\pi \times \{\mathfrak{R} + L, N_i\} = N_i, \quad \text{(B15)}$$

$$\exp[\pi(\mathfrak{R} + L)_{\text{op}}]A_i = (\cos\pi)A_i - \sin\pi\, L^{-1}(\mathbf{L}\times\mathbf{A})_i = -A_i, \quad \text{(B16)}$$

and thus arrive at the expression (6.7).

[1]E. Chacón, M. Moshinsky, and B. Vanagas, J. Math. Phys. **22**, 605 (1981).
[2]N. Bohr. Nature **137**, 344 (1936).
[3]A. Bohr and B. Mottelson, Kgl. Danske Videnskab. Selskab. Mat. Fys. Medd. **27**, 16 (1953).
[4]A. Arima and F. Iachello, Ann. Phys. (N.Y.) **99**, 253 (1976); **111**, 201 (1978); A. Arima, F. Iachello, and O. Schotten, Ann. Phys. (N.Y.) **115**, 325 (1978).
[5]T. Otsuka, A. Arima, F. Iachello, and I. Talmi, Phys. Lett. B **66**, 205 (1977); B **76**, 139 (1978).
[6]A. Ya. Dzublik, V. I. Ovcharenko, A. I. Steshenko, and G. Filippov, Yad. Fiz. **15**, 869 (1972) [Sov. J. Nucl. Phys. **15**, 487 (1972)].
[7]W. Zickendraht, J. Math. Phys. **12**, 1663 (1971).
[8]G. F. Filippov, Fiz. Elementar. Chastico, Atoma yarda **4**, 992 (1973) [Sov. J. Particles Nuclei **4**, 405 (1974)].
[9]V. Vanagas, "The Microscopic Nuclear Theory," Lecture Notes, Dept. of Physics, University of Toronto, 1977.
[10]V. Fock, Z. Phys. **98**, 145 (1935); V. Bargmann, Z. Phys. **99**, 576 (1936).
[11]M. J. Engelfield, *Group Theory and the Coulomb Problem* (Wiley-Interscience, New York, 1972).
[12]M. Moshinsky and T. H. Seligman, Ann. Phys. (N.Y.) **114**, 243 (1978); **120**, 402 (1979); **127**, 458 (1980).
[13]M. Moshinsky, SIAM J. Appl. Math. **25**, 193 (1973).
[14]M. Moshinsky, *The Harmonic Oscillator in Modern Physics:From Atoms to Quarks* (Gordon and Breach, New York, 1969), pp. 21-22,29.
[15]K. B. Wolf, Suppl. Nuovo Cimento **5**, 1041 (1967).
[16]C. P. Boyer and K. B. Wolf, Lett. Nuovo Cimento **8**, 458 (1973).
[17]M. Moshinsky, T. H. Seligman, and K. B. Wolf, J. Math. Phys. **13**, 901 (1972).
[18]M. Moshinsky, *Group Theory and Many Body Problem* (Gordon and Breach, New York, 1968), p. 19.
[19]V. Vanagas, "The Microscopic Theory of the Collective Motion in Nuclei," Lecture Notes of the 1980 Latin American School of Physics edited by T. H. Seligman, AIP Conference Proceedings #71 (1981), p. 220.
[20]M. E. Rose, *Elementary Theory of Angular Momentum* (Wiley, New York, 1957), pp. 85-88, 225, 237.
[21]I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1965), pp. 1037–1039.

Representations and Clebsch–Gordan coefficients of Z-metacyclic groups

A. Bovier[a], M. Lüling[a], and D. Wyler
Physikalisches Instiut, Rheinische Friedrich-Wilhelms-Universität Bonn, 5300 Bonn 1, West Germany

(Received 15 July 1980; accepted for publication 16 January 1981)

We investigate finite groups of the form $\mathbb{Z}_m s\mathbb{Z}_n$ and give all irreducible representations and Clebsch–Gordan coefficients in analytic form. Two subclasses are considered which seem to be inportant for applications: the M-metacyclic groups which are important for spin systems, and the K-metacyclic groups which are the smallest finite groups that have an irreducible representation of dimension $p-1$, where p is prime.

PACS numbers: 02.20.Df

I. INTRODUCTION AND SUMMARY

Finite groups have in general found less applications in physics than continuous (Lie) groups, and their properties are not as well known, with the possible exception of the crystallographic and permutation groups.

However, in recent years finite groups have found new areas of application: They can be used as a symmetry of the flavor sector of unified gauge theories of the weak and electromagnetic interactions.[1] In studying the behavior of four-dimensional lattice gauge theories defined on Lie groups by using Monte Carlo calculations,[2] the results can be approximated with gauge theories defined on discrete subgroups.[3] In statistical mechanics discrete groups are considered in systems with finite symmetries (generalized Ising models).[4,5]

These developments motivated us to investigate more closely finite groups. In this paper we study groups G that are semidirect products[6] of two cyclic groups, $G=\mathbb{Z}_m s\mathbb{Z}_n$. Such groups of order mn are equivalent to the Z-metacyclic groups.[6,7] They are the simplest cases of semidirect products with an abelian normal subgroup. For such groups the respresentation theory is known (see Refs. 5 and 8, and especially the book by Mackey[9]), and the groups G are best suited to illustrate the methods involved.

As we will show, the special form of the groups not only allows one to find the representations but also the Clebsch–Gordan coefficients (and thus also 6-j symbols etc.) in closed analytic form. Especially, it provides for a solution of the "multiplicity problem." That is, if in the decomposition of the tensor product of two irreducible representations a given representation occurs several times, then we found a group theoretical way to label them.

The same methods can also be applied for semidirect products with several factors; for example we have obtained similar results[10] for the groups $(\mathbb{Z}_n \otimes \mathbb{Z}_n)s\mathbb{Z}_3$ and $(\mathbb{Z}_n \otimes \mathbb{Z}_n)sS_3$.[11]

Beside their mathematical simplicity, the groups $\mathbb{Z}_m s\mathbb{Z}_n$ are also attractive for physical applications. As shown by Marcu, Regev and Rittenberg,[5] they appear as global symmetries of spin systems. Then the analytic form of the coupling coefficients would allow one to obtain, for example, the finite temperature expansion in closed form. Also, $\mathbb{Z}_m s\mathbb{Z}_3$ groups appear as (finite) subgroups of SU(3).[10]

[a] Supported by Studienstiftuns des Deutchen Volkes.

The $\mathbb{Z}_m s\mathbb{Z}_n$ groups have also another nice property. If we ask the question, which is the smallest finite group which has an irreducible representation of a given dimension d? then in the cases where a unique answer is possible (in a sense to be made precise below) the group is a $\mathbb{Z}_m s\mathbb{Z}_n$ group [except for $d=3$, where the group is the tetrahedral group T, a product of the form $(\mathbb{Z}_2 \otimes \mathbb{Z}_2)s\mathbb{Z}_3$].

Let us sketch some properties of the groups $\mathbb{Z}_m s\mathbb{Z}_n$. They are defined by m and n *and* by a homomorphism H which maps $\mathbb{Z}_n$ into the automorphisms of $\mathbb{Z}_m$.[6] H is given by $H(\alpha):\beta \rightarrow a^\alpha \beta$, where $\alpha \in \mathbb{Z}_n$, $\beta \in \mathbb{Z}_m$; a is an integer such that $a^n = 1 \bmod m$ and a and m have no common divisor except 1. To given m, n there are generally several possible a. a and a' give equivalent groups if the least solutions l and l' to $a^x = 1 \bmod m$ and $a'^x = 1 \bmod m$ are equal. l is the dimension of the largest irreducible representations. The groups often factorize unless $n=l$, which is indeed the physically most interesting case.[5]

If $m=p=$ prime considerable simplifications take place (ZS-metacyclic groups[7]). For example the irreducible representations have only dimensions l and 1. If, furthermore, $n=-1$ (and $n=l$) we come to the K-metacyclic groups[7] which are the smallest groups with a $(p-1)$-dimensional representation.[12] It remains an interesting problem to look for the smallest group with an irreducible representation of arbitrary dimensions $d \neq p-1$.

Before closing this section we would like to discuss qualitatively the way how the multiplicity problem gets solved. It is closely related to the structure of the groups as a product of two abelian factors.

When one constructs the irreducible representations, one may first choose the representations R of the elements β of the normal subgroup $\mathbb{Z}_m$ to be diagonal, e.g.,

$$R(\beta) \sim \begin{bmatrix} \omega^{J_1\beta} & & & \\ & \omega^{J_2\beta} & & \\ & & \ddots & \\ & & & \omega^{J_l\beta} \end{bmatrix}, \quad \omega = e^{2\pi i/m},$$

where the J_i are elements of $\mathbb{Z}_m$; that is we can assign to each state a "$\mathbb{Z}_m$-quantum number" Q. The elements α of the factor group $\mathbb{Z}_n$ now act on the states e^{iJ_i} according to the homomorphism H used to define the group: If $H(\alpha)$: $J_i \rightarrow J_j$, then the representation matrix has for the ith column and jth row only a nonvanishing element at the intersection (j,i). The representation matrices of α can thus be thought of as chang-

 0022-2488/81/081536-07$01.00 © 1981 American Institute of Physics

ing the quantum numbers of states by a fixed amount $\Delta(\alpha)$.

Now we take the decomposition of the tensor product of two representations (by representation we always mean irreducible ones) with basis states $e_i, \tilde{e}_j$, and consider the product states $e_i \otimes \tilde{e}_j$. In general there are t ($t \geqslant 1$) such states which tranform in the same way, say as the k th component of a representation R. Then R occurs in t copies and the problem is to find a label η which characterizes those linear combinations of the $e_i \otimes \tilde{e}_j$ which go into one copy of R, denoted as R_η.

The action of the group elements of $\mathbb{Z}_n$ now consists in lifting the "quantum charge" by the amount Δ. Therefore, the difference in charges $\Delta Q_{ij} = Q_j - Q_i$ of the states $e_j$and e_i is constant and can be taken as η. In short, we may say that the group $\mathbb{Z}_n$ induces a grading of the vectors $e_i \otimes \tilde{e}_j$, ΔQ_{ij}, and by this grading we solve the multiplicity problem.

The actual realization of this mechanism is generally quite involved. For the groups described in this paper it is most clearly demonstrated by the the K-metacyclic groups. The mechanism can be observed quite clearly in the case of the $\Delta(3n^2)$ groups[10] which are of the form $(\mathbb{Z}_n \otimes \mathbb{Z}_n)s\mathbb{Z}_3$.

If the factor group ($\mathbb{Z}_n$ in this case) is not abelian it does not have a complete grading; in this case a complete labelling of the irreducible representations might become impossible. It is an interesting problem to investigate such a situation.

The remainder of this paper is organized as follows: In the next section we give some properties of the groups $G(m,n,a)$. Then, in Sec. III, the irreducible representation and Clebsch–Gordan coefficients are given. The representations have previously been presented by P. Tucker.[8] We rederive them in our notation for completeness and coherence, using the method of induced representations.[9] Then, in Sec. IV we consider the special cases that are of interest to physics; with an occasional glimpse to some earlier definition it can be read by itself. An appendix shows that the smallest group with a d-dimensional representation has $d(d+1)$ elements.

II. THE Z-METACYCLIC GROUPS

In this section we consider some properties which are relevant for our purposes of groups of the form

$$G = \mathbb{Z}_m s \mathbb{Z}_n, \quad m,n \in \mathbb{N}, \tag{2.1}$$

where s denotes the semidirect product.[6] We will see that the groups (2.1) are identical with the well-known Z-metacyclic groups.[7] But we will introduce a notation for the group elements, inspired by (2.1), which will enable us to give explicit (analytic) formulae for the irreducible representations and Clebsch–Gordan coefficients of the above groups.

Let us denote the groups G in (2.1) by $G(m,n,a)$. The index a distinguishes between the various automorphisms used to define the semidirect product in (2.1). The elements of $G(m,n,a)$ are denoted by

$$g^\alpha_\beta, \quad \alpha \in \mathbb{Z}_n, \quad \beta \in \mathbb{Z}_m, \tag{2.2}$$

with the product rule for two elements given by

$$g^{\alpha_1}_{\beta_1} g^{\alpha_2}_{\beta_2} = g^{\alpha_1 + \alpha_2}_{\beta_1 + a^{\alpha_1}\beta_2}. \tag{2.3}$$

The possible values of the indices m,n,a are restricted by the conditions

$$a^n = 1 \bmod m, \quad (a,m) = 1, \tag{2.4}$$

where (a,m) denotes the largest common divisor of a and m and mod stands for modulo.

First we demonstrate the connection to the mathematical nomenclature.[7] If we put

$$T = g^1_0 \quad S = g^0_1, \tag{2.5}$$

we see that S,T are generators of $G(m,n,a)$ and that they satisfy

$$T^n = g^0_0 = \mathbb{1}, \quad S^m = g^0_0 = \mathbb{1}, \quad TST^{-1} = g^0_a = S^a. \tag{2.6}$$

It is well known[7] that all finite Z-metacyclic groups (groups whose commutator subgroups and commutator quotient groups are cyclic) can be expressed by (2.6), and thus by (2.3).

Let us examine conditions (2.4) closer. We notice that for any m,a with $(a,m) = 1$ Euler's theorem[13] allows one to choose $n = \phi(m)$ [$\phi(m)$ is the number of integers $g_i < m$, with $(g_i,m) = 1$ and is usually called Eulers ϕ-function]. Thus, there exist Z-metacyclic groups for any value of m. Given m and a, all possible values of n are multiples of a number l, given as the smallest solution to

$$a^x = 1 \bmod m. \tag{2.7}$$

l will be seen to play an important role: the largest irreducible representation of $G(m,n,a)$ has dimension l. Not every choice of n leads to an interesting group, as may be seen by the following theorem

Theorem (Factorization): Let $m = k \cdot p$ with $(k,p) = 1$, $n = r \cdot s$ with $(r,s) = 1$, and furthermore $a^r = 1 \bmod k$ and $a^s = 1 \bmod p$. Then

$$G(m,n,a) = G(k,r,a) \otimes G(p,s,a). \tag{2.8}$$

To show this, consider the mappings

$$\phi: G(k,r,a) \to G(m,n,a)$$

$$\tilde{g}^\alpha_\beta \to \phi(\tilde{g}^\alpha_\beta) \equiv g^{s\alpha}_{p\beta}, \quad g^\alpha_\beta \in G(k,r,a),$$

$$\psi: G(p,s,a) \to G(m,n,a),$$

$$\tilde{\tilde{g}}^\alpha_\beta \to \psi(\tilde{\tilde{g}}^\alpha_\beta) \equiv g^{r\alpha}_{k\beta}, \quad \tilde{g}^\alpha_\beta \in G(p,s,a).$$

Now, any element of $G(m,n,a)$ may be written as follows:

$$\phi(\tilde{g}^\alpha_\beta)\psi(\tilde{\tilde{g}}^{\alpha'}_{\beta'}) = g^{s\alpha + r\alpha'}_{p\beta + a^{s\alpha}k\beta'} = g^{r\alpha' + s\alpha}_{p\beta + k\beta'} = g^{r\alpha' + s\alpha}_{k\beta' + a^{r\alpha}p\beta} = \psi(\tilde{\tilde{g}}^{\alpha'}_{\beta'})\phi(\tilde{g}^\alpha_\beta),$$

which proves (2.8). Notice that all the assumptions of the theorem enter the last equations.

Consider some special cases of this theorem:

(1) Let be $k = m, p = 1$. If $n = rs$, $(r,s) = 1$, and $a^r = 1 \bmod m$, then

$$G(m,n,a) = G(m,r,a) \otimes G(1,s,a) = G(m,r,a) \otimes \mathbb{Z}_s.$$

(2) Let be $r = n$, $s = 1$. Then factorization demands $a = 1 \bmod p$. We now put $n = 2$, $a = m - 1$ which implies $l = 2$. Thus,

$$G(2k,2,2k-1) = G(k,2,k-1) \otimes \mathbb{Z}_2 \quad (k \text{ odd}).$$

Observing that $G(2k,2,2k-1) \cong D_{2k}$; (dihedral group) we have recovered the well-known factorization theorem[3] for dihedral groups. We note, in passing, that the $G(k,4,2k-1)$

are the "double" dihedral groups.

(3) As an exercise we leave to the reader to show that

$$G(35,6,4) = G(7,3,4) \otimes G(5,2,4).$$

Thus, in most cases only the choice $n = l$, where l is the smallest solution of $a^x = 1 \bmod m$, gives groups that cannot be expressed as direct products of smaller groups. (We will see later that even if $(l,n/l) \neq 1$ and no immediate factorization occurs, l is still the maximal dimension of the irreducible representations, and only the case $n = l$ should be interesting for applications.)

Can different choices of the number a lead to equivalent groups? An answer is the following theorem.

Theorem: $G(m,n,a)$ and $G(m,n,a')$ are isomorphic if and only if the smallest solution of the equations

$$a^x = 1 \bmod m, \quad a'^x = 1 \bmod m$$

coincide and are given by l. Consider first the case where $n = l$.

Observe now that the two sets $\{a^\alpha | a \in \mathbb{Z}_l\}$ and $\{a'_\alpha | a \in \mathbb{Z}_l\}$ coincide. Thus $a' = a^k$ for some $k \in \mathbb{Z}_l$ and $(k,l) = 1$. The mapping $\alpha \to \alpha' = k\alpha$ satisfies $a^{k\alpha} = a'^\alpha$ The desired isomorphism between the groups is then given by $g^\alpha_\beta \to g^{k\alpha}_\beta$. Using the division properties of l,n,k, this isomorphism can be extended to *all possible* values of n.

The theorem tells us that there are as many inequivalent metacyclic groups to given m,n as there are divisors of n.

We finally come to the class structure which we need for the classification of the representations.

The element $g^{\alpha_1}_{\beta_1}$ is conjugate to

$$g^{-\alpha}_{a^{-\alpha}\beta} g^{\alpha_1}_{\beta_1} g^\alpha_\beta = g^{\alpha_1}_{a^{-\alpha}\beta_1 + (a^{\alpha_1} - 1)a^{-\alpha}\beta} \tag{2.9}$$

for all $\alpha \in \mathbb{Z}_n, \quad \beta \in \mathbb{Z}_m$.

(2.9) implies:

(i) Elements with different α_1 lie in different classes.

(ii) If $a^{\alpha_1} \neq 1 \bmod m$ all elements with the same α_1 are in the same class.

(iii) If $a^{\alpha_1} = 1 \bmod m$, that is, $\alpha_1 = k \cdot l$ for some $k \in \mathbb{Z}\nu$, $(\nu = n/l)$, $g^{\alpha_1}_{\beta_1}$ and $g^{\alpha_1}_{\beta'_1}$ are in the same class if and only if

$$\beta'_1 \in \{\beta_1\} \equiv \{\beta \,|\, \beta = \beta_1 a^\alpha \quad \text{for some } \alpha \in \mathbb{Z}_n\}. \tag{2.10}$$

We will call $\{\beta_1\}$ the orbit of β_1.

The orbits of two numbers β,β' are either identical or have no common element. We therefore must calculate numbers $j_1,\dots, j_q$ such that each element of $\mathbb{Z}_m$ is contained in the orbit of just one of the j's. We can now label the classes as follows:

$$\begin{aligned} &K_i = \{g^\alpha_\beta \,|\, \alpha = i \neq k \cdot l\}, \\ &K_{[k,j]} = \{g^\alpha_\beta \,|\, \alpha = k \cdot l, \beta \in \{j\}\}, \\ &\quad j \in \{j_1, \dots, j_q\}. \end{aligned} \tag{2.11}$$

In order to get a complete set of j's we may proceed as follows: put

$$j_1 = 0, \; j_2 = 1,$$

j_s = smallest element of $\mathbb{Z}_m$ which is not yet contained in one of the orbits $\{j_r\}$, $r < s$.[14] We will denote the number of elements in the orbit $\{j\}$ by l_j (length of the orbit). All these numbers will be extensively used later. Observe that for m = prime, the lengths of all orbits except l_0 are all equal. (An example may help the reader to become familiar with these constructions: Take $m = 7$, $a = 2$; then $\{j_1 j_2 \cdots\} = \{0,1,3\}$ and the lengths of $\{1\}$, $\{3\}$ are 3, while the length of $\{0\}$ is 1.)

III. REPRESENTATIONS AND CLEBSCH–GORDAN COEFFICIENTS

We now want to investigate the unitary irreducible representations of the groups $G(m,n,a)$ of the previous section. We will give a simple analytic expression for the matrix elements of the representations and the Clebsch–Gordan coefficients which simplifies all calculations done with these groups enormously and is thus especially important for applications. The representations are obtained using the method of induced representations.[9]

We denote the irreducible representations of the $G(m,n,a)$ by a pair of indices

$$[k,j], \tag{3.1}$$

where $j \in \{j_1,\dots,j_q\}$ and $k \in \mathbb{Z}_{\nu_j}, \nu_j = n/l_j$. (the j_i and l_i are defined in the previous section).

The dimension of the representation $[k,j]$ is l_j (see below). We introduce the following notation for the basis vectors in the representation space of $[k,j]$: Let $e'_0, e'_1, \dots, e'_{l_i - 1}$ be a standard (orthogonal) basis in $\mathbb{R}^{l_j}$. Arrange now the elements of the orbit $\{j\}$ in ascending order:

$$ja^{\alpha_0} < ja^{\alpha_1} < \dots < ja^{\alpha_{l_j - 1}}, \tag{3.2a}$$

and put

$$e'_i \equiv e_{ja^{\alpha_i}}. \tag{3.2b}$$

As an example take $m = 7, n = 3, a = 2$. The values of j are 0,1,3 (see previous section). For $j = 3$ we have $\{3\} = (3,6,5)$, and thus

$$e'_0 = e_3, \; e'_1 = e_5 \; e'_2 = e_6.$$

[Notice that the orderings of the ja^{α_i}, (3.2a) and of the α_i need not correspond.]

Using this notation, the group acts as follows on the basis vectors:

$$T^{[k,j]}(g^\alpha_\beta) e_{ja^s} = \omega^{ja^{(s-\alpha)}\beta} \epsilon^{kl_j\alpha} e_{ja^{(s-\alpha)}}, \tag{3.3}$$

$$\text{with } \omega = e^{2\pi i/m}; \quad \epsilon = e^{2\pi i/n}. \tag{3.4}$$

The matrix elements of the T's are thus given by

$$[T^{[k,j]}(g^\alpha_\beta)]_{ja', ja^s} = \omega^{ja^{(s-\alpha)}\beta} \epsilon^{kl_j\alpha} \delta^m_{ja', ja^{(s-\alpha)}}, \tag{3.5}$$

the superscript m at the Kronecker delta indicating that the lower indices are to be taken mod m. Equation (3.5) gives for the characters χ,

$$\chi^{[k,j]}(g^\alpha_\beta) = \sum_{s=0}^{l_j - 1} \omega^{ja^s\beta} \epsilon^{kl_j\alpha} \delta^m_{j,ja^{\alpha}}. \tag{3.6}$$

By means of (3.6) we can easily show that (3.3) are indeed all unitary irreducible representations of $G(m,n,a)$:

We have

$$\langle \chi^{[k,j]}, \chi^{[k',j']} \rangle$$

$$\equiv \sum_{g^\alpha_\beta \in G(m,n,a)} \chi^{[k,j]}(g^\alpha_\beta) \chi^{[k',j']}(g^\alpha_\beta)^*$$

$$= \sum_{\alpha=0}^{n-1}\sum_{\beta=0}^{m-1} \delta^m_{j,ja^{-\alpha}} \delta^m_{j',j'a^{-\alpha}} \epsilon^{\alpha(kl_j - k'l_{j'})} \cdot \sum_{s,s'} \omega^{\beta(ja^s - j'a^{s'})}$$

$$= \sum_{\alpha=0}^{n-1} \delta^m_{j,ja^{-\alpha}} \delta^m_{j',j'a^{-\alpha}} \epsilon^{\alpha(kl_j - k'l_{j'})} m \sum_{s,s'} \delta^m_{ja^s j'a^{s'}} = \delta^m_{jj'} \delta^{\nu_j}_{kk'} \cdot m \cdot n.$$

This proves inequivalence and irreducibility. The completeness follows by applying the theorem of Burnside[15]:

$$\sum_{\substack{j\in\{j_1 \cdots j_q\}\\ k\in Z_{\nu_j}}} (\dim)[k,j]^2 = \sum_j \nu_j l_j^2 = \sum_j n \cdot l_j = n \cdot m,$$

by definition of $\{j_1 \cdots j_r\}$.

We will give examples of the formulae (3.3–3.5) in the next section. But first we conclude this part with the expressions of the Clebsch–Gordan coefficients. Take the product of two representation $[k'j']$ and $[k''j'']$. The CG coefficients for the representation $[k,j]$ (basis vector e_{ja^s}) in the irreducible decomposition of the product $[k'j'] \otimes [k''j'']$ (basis vector $e_{j'a^{s'}} \otimes e_{j''a^{s''}}$) is

$$C^{[k,j]_\eta ja^s}_{[k'j']j'a^{s'};[k''j'']j''a^{s''}} \tag{3.7}$$

$$= \sqrt{\frac{l_j}{\bar{l}}} \sum_{r=0}^{\lambda-1} \delta^m_{j'a^r + j''a^\eta, j}\, \delta^n_{(k'l_{j'} + k''l_{j''} - kl_j)\bar{l},0}$$

$$\times \sum_{\gamma=0}^{\bar{l}/l_j - 1} \epsilon^{(k'l_{j'} + k''l_{j''} - kl_j)(l_j\gamma - s)} \times \delta^m_{j'a^{s'}, j'a^{(r+s-l_j\gamma)}}$$

$$\times \delta^m_{j''a^{s''}, j''a^{(\eta+s-l_j\gamma)}},$$

with $\bar{l} = \psi(\psi(l_j,l_{j'}),\psi(l_j,l_{j''}))$

$$\lambda = \psi(l_j,l_{j''}) \cdot l_{j'}/\bar{l}, \tag{3.8}$$

and where $\psi(a,b)$ denotes the least common multiple of a and b. The index η[16] distinguishes between various identical representations which may occur in the decomposition of the product of two representations. Its range is restricted to $0 \leqslant \eta < (l_j, l_{j''})$, but not all numbers in this range need give nonzero coefficients in (3.7). Its explicit appearance in the formula (3.7) is a noteworthy feature; it allows a group theoretical labelling of the several identical representations in the Kronecker product and thus a correspondence between the basis vectore $e_{j'a^{s'}} \otimes e_{j''a^{s''}}$ and e_{ja^s}. As is well known from, say SU (3), this is not always possible. From (3.7) we derive the expression for the CG series:

$$[k'j'] \otimes [k''j'']$$

$$= \bigoplus_{j\in\{j_1,\dots,j_q\}} \bigoplus_{k=0}^{\nu_j - 1} \bigoplus_{\eta=0}^{(l_j,l_{j''})-1} \sum_{r=0}^{\lambda-1} \delta^m_{j'a^r + j''a^\eta j}\, \delta^n_{(k'l_{j'} + k''l_{j''} - kl_j)\bar{l},0}\, [k,j]_\eta.$$

The appearance of the Kronecker-delta $\delta^m_{j'a^r + j''a^\eta j}$ in both (3.7) and (3.9) makes it necessary to analyze the solutions of

$$j'a^r + j''a^\eta = j \bmod m \tag{3.10}$$

for given j', j'' and a. As far as we know there does not exist a closed form for the solution of the diophantine Eq. (3.10). However, it is always easy to find by trial and error the solutions (r,η,j) to (3.10) for each special case. Thus by means of (3.7–3.9) the determination of all the CG coefficients of all Z-metacylic groups is reduced to the simple number-theoretical problem of solving (3.10). This result is similar to the one obtained for the groups $\mathscr{D}(n)$[17] which are semidirect products of three $\mathbb{Z}_n$ groups, and for the groups considered in Ref. 10). We conjecture that also other groups which are semidirect products of $\mathbf{Z}$ groups have this property.

We will see in the following chapter that for an important class of the Z-metacylic group, the K-metacylic groups (see below or Ref. 3 for a definition), the problem of solving (3.10) does not occur.

We close this paragraph by sketching the derivation of (3.9). The method employed is standard and described in detail in any textbook on representation theory (see Ref. 9). One first constructs the vectors

$$p_s^{[k,j]}(e_{j'a^{s'}} \otimes e_{j''a^{s''}}$$

$$\equiv \frac{l_j}{m \cdot n} \sum_{g^\alpha_\beta \in (G(m,n,a)} [T^{[k,j]}(g^\alpha_\beta)]^*_{ja^s,j}$$

$$\times (T^{[k'j']} \otimes (T^{[k''j'']})(g^\alpha_\beta) e_{j'a^{s'}} \otimes e_{j''a^{s''}}$$

$$= \frac{l_j}{mn} \sum_{\alpha=0}^{n-1} \sum_{\beta=0}^{m-1} [T^{[k,j]}(g^\alpha_\beta)]^*_{ja^s,j}$$

$$\times (T^{[k'j']} \otimes T^{[k''j'']})(g^\alpha_\beta) e_{j'a^{s'}} \otimes e_{j''a^{s''}}$$

$$= \frac{l_j}{mn} \sum_{\alpha=0}^{n-1}\sum_{\beta=0}^{m-1} \delta^m_{ja^s, ja^{-\alpha}} \omega^{(j'a^{(s'-\alpha)} + j''a^{(s''-\alpha)} - ja^s)\beta)}$$

$$\times \epsilon^{(k'l_{j'} + k''l_{j''} - kl_j)\alpha} \times (e_{j'a^{(s'-\alpha)}} \otimes e_{j''a^{(s''-\alpha)}})$$

$$= l_j/\bar{l}\, \delta^m_{j'a^{(s'+s)} + j''a^{(s''+s)} ja^s}\, \delta^n_{(k'l_{j'} + k''l_{j''} - kl_j)\bar{l},0}$$

$$\times \epsilon^{-s(k'l_{j'} + k''l_{j''} - kl_j)}$$

$$\times \sum_{\gamma=0}^{\bar{l}/l_j - 1} \epsilon^{l_j\gamma(k'l_{j'} + k''l_{j''} - kl_j)}$$

$$\times (e_{ja^{(s+s'-l_j\gamma)}} \otimes e_{j''a^{(s+s''-l_j\eta)}}).$$

Here we made use of the fact that

$$e_{j'a^{\bar{l}}} \otimes e_{j''a^{\bar{l}}} = e_{j'} \otimes e_{j''}.$$

Observe further, that

$$\sum_{\gamma=0}^{\bar{l}/l_j - 1} \epsilon^{l_j\gamma(k'l_{j'} + k''l_{j''} - kl_j)} (e_{j'a^{(s'+\eta l_j + s - l_j\gamma)}} \otimes e_{j''a^{(s''+\eta l_j + s - l_j\gamma)}})$$

$$= \epsilon^{l_j\lambda(k'l_{j'} + k''l_{j''} - kl_j)}$$

$$\times \sum_{\gamma=0}^{\bar{l}/l_j - 1} \epsilon^{(k'l_{j'} + k''l_{j''} - kl_j)l_j\gamma} (e_{j'a^{(s'+s-l_j\gamma)}} \otimes e_{j''a^{(s''+s-l_j\gamma)}}).$$

Thus, if the tupels $(\bar{s}', \bar{s}'')$ and (s',s'') differ only by $(\lambda l_j, \lambda l_j)$, $\lambda \in \mathbb{Z}_m$, the resulting vectors are proportional. We therefore select one of them to be our new basis state by restricting the range of s', s'' by

$$s' < \psi(l_j,l_{j'}) l_{j'}/l_j$$

$$s'' < (l_j,l_{j''}).$$

Now, we set $s=0$ in order to obtain the first basis vector in the representation space of $[k,j]$:

$$p_0^{[k,j]}(e_{j'a^{s'}} \otimes e_{j''a^{s''}})$$

$$= (l_j/\bar{l})^{1/2} \delta^m_{j'a^{s'} + j''a^{s''} j}\, \delta^n_{\bar{l}(k'l_{j'} + k''l_{j''} - kl_j),0}$$

$$\times \sum_{\gamma=0}^{\bar{l}/l_j - 1} \epsilon^{l_j\gamma(kl_{j'} + k''l_{j''} - kl_j)} (e_{ja^{(s'-l_j\gamma)}} \otimes e_{j''a^{(s''-l_j\gamma)}}).$$

The square root above appears for normalization of the new basis vectors. Applying $p_s^{[k,j]}$ to the above first basis vector, we get the other basis vectors:

$$p_s^{[k,j]} p_0^{[k,j]}$$

$$= \left(\frac{l_j}{\bar{l}}\right)^{1/2} \delta^m_{j'a^{s'} + j''a^{s''}, j} \delta^n_{\bar{l}(k'l_{j'} + k''l_{j''} - kl_j)}$$
$$\times \sum_{\gamma=0}^{\bar{l}/l_j - 1} \epsilon^{(k'l_{j'} + k''l_{j''} - kl_j)(l_j\gamma - s)} \left(e_{j'a^{(s'+s-l_j\gamma)}} \otimes e_{j''a^{(s''+s-l_j\gamma)}}\right).$$

If $\delta_{j'a^{s'} + j''a^{s''}, j} \neq 0$ for different tupels (s_1', s_1''), (s_2', s_2''), we obtain for each of them a distinct vector e_{ja^s} of the same representation. This signals, that the same representation occurs several times. It is easy to see that if $s_1' \neq s_2'$ then also $s_1'' \neq s_2''$ etc. Thus, we may choose any of the two indices s' and s'' to distinguish between the various identical representations. Calling this index η, we arrive at (3.7). The reason, why η appears in such an unsymmetrical way in (3.7) is thus a pure convention.

IV. SPECIAL CASES

We now want to apply the results of the previous section to those $G(m,n,a)$ Z-metacyclic groups that are important for applications. In the mathematical literature[7] two special cases of Z-metacyclic groups are generally introduced: the ZS-metacyclic groups and the K-metacyclic groups; the latter is of special interest to us. We would like to introduce still another special case which we shall call M-metacyclic; these groups have been shown by Marcu and Rittenberg[5] to appear as global symmetries in spin systems. The above groups are described as follows:

ZS-metacyclic: $(n,m) = 1 \bmod m$.

K-metacyclic: $m = p =$ prime, $n = p - 1$, and a is primitive root mod p, that is $p - 1$ is the smallest solution of $a^x = 1 \bmod p$. We consider the K-metacyclic groups in detail, for they are the smallest groups with a $(p-1)$-dimensional representation.

M-metacyclic: $m = p =$ prime; $n = l$, where l is the smallest solution of $a^x = 1 \bmod p$. [n is a divisor of $(p-1)$]. These groups are subgroups of affine transformations of $\mathbb{Z}_p$, that is of transformations of the type

$$g^\alpha_\beta : x \to x' = \alpha x + \beta \quad x,\beta \in \mathbb{Z}_p,\ \alpha \in \mathbb{Z}_p \setminus 0. \tag{4.1}$$

We observe that the product rule for these transformations is given by

$$g^{\alpha_1}_{\beta_1} g^{\alpha_2}_{\beta_2} = g^{\alpha_1\alpha_2}_{\beta_1 + \alpha_1\beta_2}, \tag{4.2}$$

which is isomorphic to the rule (2.3) if a is a primitive root. On the other hand, all subgroups of affine transformations (4.2) are obtained by choosing different a's and restricting the range of α to $\mathbb{Z}_l$. See Ref. 5 for applications of (4.1).

We begin with the M-metacyclic groups. Recall from the previous section that there exists a set of numbers, $j_1 = 0$, $j_2 = 1, j_3, \ldots, j_r$, $[r = (p-1)/l]$ constructed as follows: define the orbits $\{j'\}$ of j_i as $\{j_i\} = \{j_i, j_i a, j_i a^2, \cdots, \bmod m\}$ and j_s is the smallest number not in $\{j_r\}$, $r < s$. Each orbit, save $\{0\}$ has l elements. There are only l-dimensional and one-dimensional representations; since $l = n$ the l-dimensional representations can be labelled by $[0,j]$ [see Eqs. (2.11) and (3.1)] and the one-dimensional by $[k,0]$, $k \in \mathbb{Z}_l$. We have for their matrix elements

$$[T^{[0,j]}(g^\alpha_\beta)]_{ja^t, ja^s} = \omega^{ja^s\beta} \delta_{ja^t, ja^s\alpha},$$
$$j = 1,2,\ldots,(p-1)/l \tag{4.3}$$

(1-dimensional representations),

$$T^{[k,0]}(g^\alpha_\beta) = \epsilon^{k\alpha}, \quad k \in \mathbb{Z}_l \tag{4.4}$$

(one-dimensional representations). In (4.3) the indices ja^t, ja^s label the matrix elements between the basis vectors e_{ja^t} and e_{ja^s} (notice that the e_p are not necessarily labelled in the usual order $e_1, e_2, \ldots$; they are labelled in ascending order of ja^t mod m, *not* of t!). [For a more detailed description of the basis see (3.2) and the discussion there.] The characters of the representations (4.3) are

$$\chi^j(g^\alpha_\beta) = \sum_{s=0}^{l-1} \omega^{ja^s\beta} \delta_{j, ja^s\alpha} = \sum_{s=0}^{l-1} \omega^{ja^s\beta} \delta_{\alpha,0}. \tag{4.5}$$

The Clebsch–Gordan coefficients $C^R_{R',R''}$ for the representation R in the product $R' \otimes R''$ are

$$C^{[0,j]_\eta ja^s}_{[0,j'] j'a^{s'};[0,j''] j''a^{s''}} = \sum_{r=0}^{l-1} \delta^p_{j'a^r + j''a^\eta, j} \delta^p_{a^{(r+s)}, a^{s'}} \delta^p_{a^{(\eta+s)}, a^{s''}}, \tag{4.6}$$

$$C^{[k,0]}_{[0,j'] j'a^{s'};[0,j''] j''a^{s''}} = \frac{1}{\sqrt{l}} \sum_{r=0}^{l-1} \delta^p_{j'a^r + j'',0} \sum_{l=0}^{l-1} \epsilon^{-k\gamma} \delta^p_{a^{s'}, a^{(r+\gamma)}} \delta^p_{a^\gamma, a^{s''}}, \tag{4.7}$$

$$C^{[0,j] ja^s}_{[0,j'] j'a^{s'};[k'',0]} = \delta^p_{j,j'} \epsilon^{-k''s} \delta^p_{a^{s'}, a^s}, \tag{4.8}$$

$$C^{[k,0]}_{[k',0];[k'',0]} = \delta^l_{k'+k'',k}, \tag{4.9}$$

where η is an index to label identical representations which occur several times in $[0,j'] \otimes [0,j'']$.

As an example consider the group $G(7,3,2)$: we have $j = 0,1,3$. For $j = 1,3$ we have the bases

$j = 1: \{e_1, e_2, e_4\}$,

$j = 3: \{e_3, e_5, e_6\}$.

The one-dimensional representations are [(4.4)]

$$T^{[k,0]}(g^1_0) = \epsilon^k, \quad T^{[k,0]}(g^0_1) = 1, \quad \epsilon = e^{2\pi i/3},$$

while the three-dimensional one's are [(4.3)]

$$T^{[0,1]}(g^1_0) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix},$$
$$T^{[0,1]}(g^0_1) = \begin{pmatrix} \omega & 0 & 0 \\ 0 & \omega^2 & 0 \\ 0 & 0 & \omega^4 \end{pmatrix}, \tag{4.10.1}$$
$$T^{[0,3]}(g^1_0) = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix},$$
$$T^{[0,3]}(g^0_1) = \begin{pmatrix} \omega^3 & 0 & 0 \\ 0 & \omega^5 & 0 \\ 0 & 0 & \omega^6 \end{pmatrix}, \tag{4.10.2}$$
$$\omega = e^{2\pi i/7}.$$

In order to arrive at the expressions (4.10.2) we must remember that the order of basis vectors is e_3, e_5, e_6, which corresponds to the order 0,2,1 of the respective exponents of a (i.e, $3 = 3\cdot 2^0$, $5 = 3\cdot 2^2$, $6 = 3\cdot 2^1$).

We might wish to decompose the product $T^{[0,1]} \otimes T^{[0,1]}$. Equation (4.6) then yields

$$C^{[0,j]_\eta j2^s}_{[0,1] 2^{s'};[0,1] 2^{s''}} = \sum_{r=0}^{2} \delta^7_{2^r + 2^\eta, j} \delta^7_{2^{r+s}, 2^{s'}} \delta^7_{2^{\eta+s}, 2^{s''}}. \tag{4.11}$$

If $j = 1$, $r = \eta = 2$ and we have

$$C^{[0,1]_{(\eta=2)},2^s}_{[0,1],2^{s'};[0,1],2^{s''}} = \delta^7_{2^{2+s},2^{s'}}\delta^7_{2^{2+s},2^{s''}} = \delta^7_{2+s,s'}\delta^7_{2+s,s''}.$$

If $j = 3$, on the other hand, we have either $\eta = 0$, $r = 1$, or $\eta = 1$, $r = 0$. Thus

$$C^{[0,3]_0,3\cdot 2^s}_{[0,1],3\cdot 2^{s'};[0,1],3\cdot 2^{s''}} = \delta^7_{1+s,s'}\delta^7_{s,s''}, \tag{4.12.1}$$

$$C^{[0,3]_1,3\cdot 2^s}_{[0,1],3\cdot 2^{s'};[0,1],3\cdot 2^{s''}} = \delta^7_{s,s'}\delta^7_{1+s,s''}. \tag{4.12.2}$$

No one-dimensional representation occurs in the product, for $2^r + 1 = 0 \bmod 7$ has no solution. We have thus found

$$[0,1] \otimes [0,1] = [0,1] \oplus [0,3] \oplus [0,3] \tag{4.13}$$

We leave it to the reader to verify the following decompositions (3.9):

$$[0,3] \otimes [0,3] = [0,3] \oplus [0,1] \oplus [0,1]. \tag{4.14}$$

$$[0,1] \otimes [0,3] = [0,1] \oplus [0,3] \oplus [0,0] \oplus [1,0] \oplus [2,0]. \tag{4.15}$$

We now turn to the K-metacyclic groups, namely the $G(p,p-1,a)$ where a is a primitive root mod p. These groups have just one $(p-1)$-dimensional representation and $(p-1)$ one-dimensional ones. Their order being $p(p-1)$, they are the smallest groups with a $(p-1)$-dimensional representation. Observe, that there is just one K-metacyclic group of a given order.

The expressions above for the representations and CG coefficients simplify now a lot since there is no need to evaluate a set of j's as before [see Eq. (3.10)] or solve an equation for r and η. We have

$$T^{[k,0]}(g^\alpha_\beta) = \epsilon^{\alpha k}, \quad k \in \mathbb{Z}_{p-1} \tag{4.16}$$

(one-dimensional),

$$[T^{[0,1]}(g^\alpha_\beta)]_{t,s} = \omega^{sa^{\alpha}\beta}\delta^p_{t,sa^{\alpha}}, \quad t,s = 1,\dots,(p-1) \tag{4.17}$$

[$(p$-$1)$-dimensional].

Notice the simplified labelling for the basis in (4.17). The characters are

$$\chi^{[k,0]}(g^\alpha_\beta) = \epsilon^{k\alpha}, \tag{4.18}$$

$$\chi^{[0,1]}(g^\alpha_\beta) = \delta^{p-1}_{\alpha,0}(p\delta^p_{\beta,0} - 1).$$

The C G coefficients are simply

$$C^{[0,1]_\eta,s}_{[0,1],s';[0,1],s''} = \delta^p_{(1-\eta)s,s'}\delta^p_{\eta s,s''}, \quad \eta = 2,3,\dots,(p\text{-}1), \tag{4.19}$$

$$C^{[k,0]}_{[0,1],s';[0,1],s''} = \frac{1}{\sqrt{p-1}}\sum_{\gamma=0}^{p-2}\epsilon^{-k\gamma}\delta^p_{a^{\gamma},s'}\delta^p_{-a^{\gamma},s''}, \tag{4.20}$$

$$C^{[0,1],s}_{[0,1],s';[k'',0]} = \epsilon^{-k''\rho}\delta^p_{s,s'}, \quad \text{where } \rho{:}a^\rho = s \bmod p \tag{4.21}$$

$$C^{[k,0]}_{[k',0];[k'',0]} = \delta_{k'+k'',k}. \tag{4.22}$$

The C G series can be given explicitly in this case:

$$[k,0] \otimes [k',0] = [k+k',0], \tag{4.23}$$

$$[k,0] \otimes [0,1] = [0,1], \tag{4.24}$$

$$[0,1] \otimes [0,1] = \underbrace{[0,1] \oplus \dots \oplus [0,1]}_{p-1 \text{ times}} \oplus \bigoplus_{k=0}^{p-2}[k,0]. \tag{4.25}$$

As an example, consider the group $G(5,4,2)$. This group of order 20 has one four-dimensional representation

$$T^{[0,1]}(g^1_0) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix},$$

$$T^{[0,1]}(g^0_1) = \begin{pmatrix} \omega & 0 & 0 & 0 \\ 0 & \omega^2 & 0 & 0 \\ 0 & 0 & \omega^3 & 0 \\ 0 & 0 & 0 & \omega^4 \end{pmatrix},$$

$$\omega = e^{2\pi/5i}$$

The C G coefficients are for $[0,1] \otimes [0,1]$, given by

$C_{[0,1];[0,1]} =$

		$[k,0]$				$[0,1]^s_1$				$[0,1]^s_2$				$[0,1]^s_3$			
s'	s''	0	1	2	3	1	2	3	4	1	2	3	4	1	2	3	4
1	1										1						
1	2															1	
1	3								1								
1	4	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$												
2	1							1									
2	2												1				
2	3	$\frac{1}{2}$	$\frac{i}{2}$	$-\frac{1}{2}$	$\frac{-i}{2}$												
2	4													1			
3	1																1
3	2	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$												
3	3									1							
3	4						1										
4	1	$\frac{1}{2}$	$\frac{-i}{2}$	$-\frac{1}{2}$	$\frac{i}{2}$												
4	2					1											
4	3														1		
4	4											1					

(blank entries are zero)

Notice again how the multiplicity problem is solved; there is an index η, that labels in a group theoretical way the irreducible representations of $[0,1] \otimes [0,1]$ and allows a unique connection between the states $s' \cdot s''$ of $[0,1] \otimes [0,1]$ and the $[0,1]$ states s.

ACKNOWLEDGMENTS

The authors wish to express their gratitude to U. Abresch, M. Marcu, and V. Rittenberg for many helpful and inspiring discussions.

APPENDIX

Assuming a group G is the smallest with an n-dimensional representation, it has only one such representation R which is therefore real. Considering now the Clebsch–Gordan series of R with itself, the R cannot occur more than $(n-1)$ times: The number of thimes R occurs in the expansion of $R \otimes R$ is $n_R = \Sigma\chi_i^3/\Sigma\chi_i^2$ where χ_i is the character for a group element g_i in the representation R, and the sum extends over all $g_i \in G$. Since $\chi_i \leqslant n$ (because the representation is unitary, with $\chi_i = n$) only for the identity or for $n = 1$ (remember: all χ_i are real) we have

$$n_R < (=) n\Sigma\chi_i^2/\Sigma\chi_i^2 = n.$$

Thus, only $(n-1)\,n$ states of the n^2 states of $R \otimes R$ can be absorbed by R. The remaining n states must go into new representations. This is achieved most economically with one-dimensional representations, since higher dimensional representations, say of dimension t, swallow only t states but contribute a factor t^2 to the order of the group (or, more mathematically, one seeks to minimize $\Sigma_{i=1}^m n_i^2$ with fixed $\Sigma_{i=1}^m n_i$, m free). Thus the minimal group must have $n(n+1)$ elements.

[1]See, e.g., D. Wyler, Phys. Rev D **19**, 3369 (1979) and references therein.
[2]M. Creutz, L. Jacobs, and C. Rebbi, Phys. Rev. Lett. **42**, 1390 (1979); M. Creutz, Phys. Rev. Lett. **43**, 553 (1979).
[3]C. Rebbi, Phys. Rev. D **21**, 3350 (1980); D. Petcher and D. H. Weingarten, Indiana University Preprint IUHET-53, 1980.
[4]See, e.g., J. Kogut, Rev. Mod. Phys. **51**, 659 (1979).
[5]M. Marcu, A. Regev, and V. Rittenberg, Max Planck Institute Preprint MPI-PAE/PTh 35/80; M. Marcu and V. Rittenberg, *ibid.* 42/80.
[6]C. Curtis and I. Reiner, *Representation Theory of Finite Groups and Associative Algebras* (Interscience, New York, 1962).
[7]H. S. M. Coxeter and W.O.J. Moser, *Generators and Relations for Discrete Groups* (Springer, Berlin, 1957).
[8]P. A. Tucker, "On the reduction of induced representation of finite groups", Ph. D. thesis, Wisconsin, 1961.
[9]G. W. Mackey, *Induced Representations of Groups and Quantum Mechanics* (Benjamin, New York, 1968).
[10]A. Bovier, M. Lüling, and D. Wyler, J. Math. Phys. **22**, 1543 (1981).
[11]The method used here can in principle be extended to a group G with abelian normal subgroup A and an abelian factor group $A' = G/A$. See, e.g., Ref. 6.
[12]This follows because a group with a d-dimensional representation has at least $d(d+1)$ elements (see Appendix).
[13]See, for example, B. M. Stewart, *Theory of Numbers* (Macmillan, New York, 1964).
[14]In general, it is not possible to give a closed expression for the number q of j's.
[15]See for example G. Ljubarskii, *The Theory of Groups and Its Applications to Physics* (Pergamon, Oxford, 1960).
[16]The reader might wonder why η appears in an unsymmetric fashion with respect to $[k, j]$ and $[k' j']$. It is a pure convention as will be explained at the end of this section.
[17]U. Abresch, A. Bovier, O. Lechtenfeld, M. Lüling, V. Rittenberg, and G. Weymans, J. Math. Phys. **21**, 2481 (1980).

Finite subgroups of SU(3)

A. Bovier,[a] M. Lüling,[a] and D. Wyler
Physikalisches Institut, Universität Bonn, D-5300 Bonn 1, West Germany

(Received 23 December 1980; accepted for publication 30 January 1981)

We present a new class of finite subgroups of SU(3) of the form $\mathbb{Z}_m$ s $\mathbb{Z}_n$ (semidirect product). We also apply the methods used to investigate semidirect products to the known SU(3) subgroups $\Delta(3n^2)$ and $\Delta(6n^2)$ and give analytic formulae for representations (characters) and Clebsch–Gordan coefficients.

PACS numbers: 02.20.Df, 02.20.Rt

I. INTRODUCTION

In the last years lattice gauge theories with continuous gauge groups have become the subject of many studies, mainly using Monte-Carlo techniques.[1] At the same time the interest in such theories with discrete gauge groups increased. They are interesting in their own right[2] but also in connection with the continuous gauge theories. Rebbi[3] and Petcher and Weingarten[4] have shown that the SU(2) gauge theories can be approximated by using gauge theories defined on discrete subgroups of SU(2); and one hopes to attack the important SU(3) of quantum chromodynamics in the same way.

The finite subgroups of SU(3) are not well known. Miller, Blichfeldt, and Dickson[5] and Fairbairn, Fulton and Klink[6] gave six groups, corresponding to regular polyhedrons of SU(3) and two infinite sequences, the $\Delta(3n^2)$ and the $\Delta(6n^2)$ groups of order $3n^2$, $6n^2$, respectively. In a recent paper[7] we investigated finite groups of the form $\mathbb{Z}_m$ s $\mathbb{Z}_n$ (s denotes the semidirect product) of order $n{\cdot}m$. Their simple structure (semidirect product) enabled us to give not only all representations and characters, but also the Clebsch–Gordan coefficients in analytic form.[8] (For example, we can write a simple analytic formula for the characters of products of elements which can be used in lattice gauge theories.) These groups include, for example, the two sequences C_n and D_n of finite SU(2) subgroups. As we will see, they also generate new sequences of SU(3) subgroups. These groups, denoted here by T_n, are of the form $\mathbb{Z}_m$ s $\mathbb{Z}_3$ where m must contain at least one prime factor of the form $p = 3n + 1$ ($p = 7, 13,\cdots$ but not $5, 11,\cdots$). They are the generalizations of the dihedral groups D_n ("Trihedral" groups).

There is no closed formula for all possible m (essentially because there is no closed formula for prime numbers), but a possible choice is $m_n = 1 + n + n^2$, n integer. [The construction also holds for SU(N) if 3 is replaced by N.] The smallest T_m is T_7 of order 21. This group is the smallest SU(3) subgroup which is not also SU(2) subgroup.

The method of representing finite groups as semidirect products is not only useful for two factors. We will show that also the groups $\Delta(3n^2)$ and $\Delta(6n^2)$ can be cast into this form; namely $\Delta(3n^2) = (\mathbb{Z}_n \otimes \mathbb{Z}_n)$ s $\mathbb{Z}_3$ and $\Delta(6n^2) = (\mathbb{Z}_n \otimes \mathbb{Z}_n)$ s S_3 (note that for $n = 2$, $\Delta(3n^2)$ is the tetrahedral group T and $\Delta(6n^2)$ the octahedral group O). This will enable us to give again the irreducible representations and coupling coefficients in analytic form, a property which is most useful in numerical calculations.

From the physical point of view we would like to add the following note:

The $\Delta(3n^2)$ and the $\Delta(6n^2)$ groups have, in contrast to the D_n groups, a nontrivial $n\to\infty$ limit. For example, the $\Delta(6n^2)$ give $(\mathrm{U}(1) \otimes \mathrm{U}(1))$ s S_3 where S_3 connects the two U(1)'s. $\Delta(6\,\infty^2)$ consists of the elements corresponding to the Cartan subalgebra of SU(3) and combinations of the other SU(3) elements.

We also note that all the groups T_n, $\Delta(3n^2)$, and $\Delta(6n^2)$ do not contain the center of SU(3), $\mathbb{Z}_3$. This is analogous to the situation in SU(2). In this case, the center can be incorporated by making a "central extension" with multiplicator $\mathbb{Z}_2$ (double groups). One might consider such extensions with $\mathbb{Z}_3$ for SU(3) subgroups ("triple groups").

The remainder of this paper is organized as follows. In the next section we review the necessary aspects of Ref. 7 and identify the SU(3) subgroups of the form $\mathbb{Z}_m$ s $\mathbb{Z}_n$. We give the analytic formulae for the representations (the Clebsch–Gordon coefficients may be found in Ref. 7). In Sec. III we show how the $\Delta(3n^2)$ and $\Delta(6n^2)$ can be written as semidirect products and give representations and coupling coefficients. The Appendix contains a proof of a theorem in the next section.

II. "DIHEDRAL-LIKE" SUBGROUPS OF SU(3)

We begin by recalling the most important aspects of the groups $\mathbb{Z}_n$ s $\mathbb{Z}_m$. These groups are characterized by three integers m, n, a [we will denote them by $G(m,n,a)$] which satisfy $a^n = 1 \bmod m$ and $(a,m) = 1$ (the greatest common divisor of a and m is 1) with $a < m$. The elements of these groups are denoted by

$$g^\alpha_\beta, \quad \alpha\in\mathbb{Z}_n,\ \beta\in\mathbb{Z}_m, \tag{2.1a}$$

and satisfy the multiplication rule

$$g^\alpha_\beta \cdot g^{\alpha'}_{\beta'} = g^{\alpha+\alpha'}_{\beta+a^\alpha\beta'}. \tag{2.1b}$$

g^1_0 and g^0_1 generate the whole group. The order of the group is $n{\cdot}m$. An important role is played by l, the least nonzero solution to $a^x = 1 \bmod m$. It is the dimension of the largest irreducible representation of $G(m,n,a)$. n must always be a multiple of l.

The irreducible representations of the groups $G(m,n,a)$ are found as follows. First, one has to calculate a set of numbers $j_1, j_2,\cdots, j_g$ as follows: $j_1 = 0, j_2 = 1$, and j_s = smallest

[a]Supported by Studienstiftung des Deutschen Volkes.

 0022-2488/81/081543-05$01.00 © 1981 American Institute of Physics

number not contained in the set $\{j_r\}$, $r<s$, with

$$\{j_r\} = \{j_r, j_r a, j_r a^2, \cdots\} \tag{2.2}$$

and $j_r a^k \neq j_r a^i$ mod m if $k \neq i$. The number of elements (or length) of $\{j_r\}$ is obviously at most l; we will denote it by l_j.

We must continue this procedure until all numbers $\mathbb{Z}_m$ are used up in the various $\{j\}$. Defining furthermore, $v_j = n/l_j$ (n is always a multiple of l_j) and letting $k_j \in \mathbb{Z}_{vj}$, we can label the irreducible representations by $[k_j, j]$. Their dimension is l_j and their matrix form is

$$[T^{[k_j,j]}(g^\alpha_\beta)]_{ja^t,ja^s} = \omega^{ja^{(s-\alpha)}\beta}\epsilon^{k_j l_j \alpha}\delta^m_{ja^t,ja^{(s-\alpha)}}. \tag{2.3}$$

Here, $\omega = e^{2\pi i/m}$, $\epsilon = e^{2\pi i/n}$; the basis vectors are labeled by $j\cdot a^r$, $r = 0,1,\cdots,(l_j - 1)$, in ascending order of the $j\cdot a^r$, not of the r. Finally, the superscript m on the Kronecker-delta indicates that ja^t and $ja^{(s-\alpha)}$ have to be equal mod m. As an illustration let us take $m = 7$, $n = 3$, $a = 2$. Then $l = 3$, and we have $\{j = 0\} = 0$, $\{j = 1\} = 1,2,4$, and $\{j = 3\} = 3,6,5$. $j = 1,3$ correspond to three-dimensional representations whose basis vectors are (e_1,e_2,e_4) and (e_3,e_5,e_6), respectively. For $j = 0$ we have three one-dimensional representations, labelled by $k = 0,1,2$. As another example take arbitrary m; if we put $n = 2$, $a = m - 1$, we obtain the dihedral groups D_m; the well-known factorization theorem for the D_n is easily recovered using a general factorization theorem of Ref. 7.

We are now ready to investigate which of our groups $G(m,n,a)$ are SU(3) subgroups. To do this we require the following sufficient conditions:

(1) The group should possess a faithful three-dimensional unitary representation with determinant one.

(2) To ensure faithfulness we require this representation to be irreducible, and the group should not possess irreducible representations of higher dimension. Of course, the second condition is by no means necessary. But it is convenient to impose it, and a complete answer to the problem posed this way can be found. On the other hand, in the SU(2) case the dihedral groups are just those which fulfill the analogous conditions.

From the second condition we see that n has to contain a factor 3 and $a^3 = 1$ mod m.

A factorization theorem in Ref. 7 implies that for $n = 3^r q$ we have $G(m,n,a) = G(m,3^r,a) \otimes \mathbb{Z}_q$. We can therefore restrict our attention to the case $n = 3^r$. Now, $G(m,3^r,p)$ has representations $[k,j]$, where $k \in \mathbb{Z}_{3^r}/l_j$ for $l_j = 3$, $k \in \mathbb{Z}_{3^r}$. For $r > 1$ the representations with $k = 0$ are not faithful; on the other hand, for the three-dimensional representations

$$\det(T^{[k,j]}(g^1_0)) = \epsilon^{3k}, \quad \epsilon = e^{2\pi i/3^r}.$$

Thus only for the case $r = 1$ both our conditions can be fulfilled.

It remains to consider the other generator g^0_1. We must ask for which m there exists a number a such that for some representation $[0,j]$,

$$\det(T^{[0,j]}(g^0_1)) = \omega^{j(1+a+a^2)} = 1.$$

This condition can be rewritten in the following form:

$$a \neq 1 \bmod m, \quad j(1 + a + a^2) = 0 \bmod m,$$
$$a^3 = 1 \bmod m, \quad ja \neq j \bmod m.$$

One finds that this set of conditions can be fulfilled whenever m contains at least one prime factor p of the form $p = 3n + 1$, $n \in \mathbb{N}$. The complete answer to our problem is contained in the following theorem.

Theorem: Let

$$m = 3^i P\cdot Q, \ Q = q_1^{\alpha_1}\cdots q_s^{\alpha_s}, \ q_k \neq 3z + 1, \ q_K \neq 3$$
$$P = p_1^{\beta_1}\cdots p_r^{\beta_r}, \ p_k = 3z + 1; \ q,p \text{ prime},$$

and not all $\beta_k = 0$. Then there exists a number a such that the group $G(m,3,a)$ fulfills conditions (1) and (2) and is a subgroup of SU(3). Furthermore,

(i) if $i \neq 1$ then $G(3^i QP,3,a) = G(3^i P,3,a) \otimes \mathbb{Z}_Q$,

(ii) if $i = 1$ then $G(3QP,3,a) = G(P,3,a) \otimes \mathbb{Z}_{3Q}$.

The proof of this theorem involves some number-theoretical considerations and is given in the Appendix.

The theorem can easily be generalized to SU(n), essentially by replacing 3 by n. It may be noted that for SU(2) any $m \neq 1,2$ is permitted because any prime $= 1$ mod 2 (except 2). Thus, the dihedral groups D_m are defined for all $m > 2$.

It is amusing to observe that as n increases the "dihedral" subgroups of SU(n) become "rarer."

It is a bit displeasing that we cannot give the possible m and a explicitly. There is however, one sequence of groups, namely the $G(1 + n + n^2,3,n)$, $n \in \mathbb{N}$ which always are SU(3) subgroups.

From (2.3) we immediately get the expression for the representations and characters:

$$3 \ \dim: [T^{[0,j]}(g^\alpha_\beta)]_{ja^t,ja^s} = \omega^{ja^{(s-\alpha)}}\delta^m_{ja^t,ja^{(s-\alpha)}}, \tag{2.4}$$

$$1 \ \dim: [T^{[k,0]}(g^\alpha_\beta)] = \epsilon^{k\alpha},$$

$$\chi^{[0,j]}(g^\alpha_\beta) = \sum_{s=0}^{2}\omega^{ja^s\beta}\delta^m_{j,ja^{(s-\alpha)}}. \tag{2.5}$$

The Clebsch–Gordon coefficients may be found in Ref. 7.

III. THE $\Delta(3n^2)$ AND $\Delta(6n^2)$ GROUPS

In the last section we introduced the dihedrallike subgroups of SU(3) and pointed out how their properties can be investigated. But the methods applied to them can also be used to analyze[5,6] the sequences $\Delta(3n^2)$, $\Delta(6n^2)$. These groups display a more complicated structure and should reflect the properties of SU(3) in more detail; nevertheless we will show that the representations and Clebsch–Gordon coefficients can again be given in an analytical form.

The application of the methods used in Ref. 7 depends on the observation that the $\Delta(3n^2)$ and $\Delta(6n^2)$ groups can be written as

$$\Delta(3n^2) = (\mathbb{Z}_n \otimes \mathbb{Z}_n) \text{ s } \mathbb{Z}_3, \tag{3.1a}$$

$$\Delta(6n^2) = (\mathbb{Z}_n \otimes \mathbb{Z}_n) \text{ s } S_3. \tag{3.1b}$$

S_3 itself is a semidirect product[7] of the form $\mathbb{Z}_3 \text{ s } \mathbb{Z}_2$, and its elements can be labelled by g^r_γ, $r \in \mathbb{Z}_2$, $\gamma \in \mathbb{Z}_3$. $\mathbb{Z}_3$ is a subgroup of S_3 with $r \equiv 0$ (or omitting r). If we identify $\mathbb{Z}_n \otimes \mathbb{Z}_n$ with the two-dimensional module $\mathbb{Z}_n^2$ over $\mathbb{Z}_n$, spanned by $\binom{1}{0}$ and $\binom{0}{1}$, the automorphisms defining the semidirect products in (3.1) can be given in matrix form as follows:

$$(\gamma,r) \to M^\gamma N^r, \tag{3.2}$$

with

$$M=\begin{pmatrix}-1 & -1\\ 1 & 0\end{pmatrix},\quad N=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}.$$

M and N form just the two-dimensional representation of S_3 over Z_n, clearly $M^3=1$, $N^2=1$. We will denote the transpose of M by $\bar{M}$.

The elements of the $\Delta\,(6n^2)$ are labelled by

$$g_{\mathbf{p}}^{r,\gamma},\quad r\in\mathbb{Z}_2,\ \gamma\in\mathbb{Z}_3,\ \mathbf{p}\in\mathbb{Z}_n^2, \tag{3.3}$$

and the product rule is

$$g_{\mathbf{p}}^{r,\gamma}\cdot g_{\mathbf{p}'}^{r',\gamma'}=g_{\mathbf{p}+M^{\gamma}N^{r}\mathbf{p}'}^{r+r',\gamma+(-1)^{r}\gamma'}. \tag{3.4}$$

For $\Delta\,(3n^2)$ we have simply

$$g_{\mathbf{p}}^{\gamma}\,g_{\mathbf{p}'}^{\gamma'}=g_{\mathbf{p}+M^{\gamma}\mathbf{p}'}^{\gamma+\gamma'}. \tag{3.5}$$

From (3.3) and (3.5) we see that if $n=2$, the groups $\Delta\,(3n^2)$ and $\Delta\,(6n^2)$ are the tetrahedral group T and the octahedral group O, respectively.

The representations can be constructed using the method of induced representations.[9] We do it for the $\Delta\,(6n^2)$ groups and specialize to the $\Delta\,(3n^2)$ groups in the end. (We give the results without derivation; the method is described in Ref. 9).

First some notation: For an element $\mathbf{m}=\begin{pmatrix}m_1\\ m_2\end{pmatrix}\in\mathbb{Z}_n^2$ the set

$$\begin{aligned}\{\mathbf{m}\}&=\{\mathbf{m}'|\mathbf{m}=\bar{M}^{\gamma}N^{r}\mathbf{m},\ \gamma\in\mathbb{Z}_3,\ r\in\mathbb{Z}_2\}\\ &=\left\{\begin{pmatrix}m_1\\ m_2\end{pmatrix},\begin{pmatrix}m_2\\ m_1\end{pmatrix},\begin{pmatrix}m_2-m_1\\ -m_1\end{pmatrix},\begin{pmatrix}m_1-m_2\\ -m_2\end{pmatrix},\right.\\ &\quad\left.\begin{pmatrix}-m_2\\ m_1-m_2\end{pmatrix},\begin{pmatrix}-m_1\\ m_2-m_1\end{pmatrix}\right\}\end{aligned} \tag{3.6}$$

is called the orbit of $\mathbf{m}$ with respect to S_3; its number of elements, or length, will be denoted by $l_{\mathbf{m}}$ (obviously, $l_{\mathbf{m}}$ is at most 6). Two orbits are either identical or disjoint.

We can introduce an ordering on the module $\mathbb{Z}_n^2$ with respect to S_3. It allows one to select, in a canonical way, a set of representatives for all orbits in $\mathbb{Z}_n^2$ with respect to S_3. Let this set be $\mathcal{O}$. To label all representations, we also need to introduce the orbits in $\mathbb{Z}_3$ with respect to $\mathbb{Z}_2$. There are only two orbits,[7] $0=\{0\}$ and $1=\{1,2\}$ of length 1 and 2, respectively. Let R be the set of these two orbits.

The representations will be denoted by

$$[k_2,k_3,j,\mathbf{m}],\quad k_i\in\mathbb{Z}_i,\ j\in R,\ \mathbf{m}\in\mathcal{O}, \tag{3.7}$$

and k_2, k_3 depend on the j, m (see below). We can put $j=0$ if $\mathbf{m}\neq 0$ and $\mathbf{m}=0$ if $j\neq 0$. The dimension of (3.7) is $l_{\mathbf{m}}$ or l_j. As mentioned $l_j=1,2$ and $l_{\mathbf{m}}=6,3$ or 1 in general; if $n=3\cdot z$, $z\in\mathbb{N}$, an orbit (in $\mathbb{Z}_n^2$) of length 2 occurs. It is easy to check that

$$l_0=1,\quad l_{\binom{m}{0}}=3,\quad l_{\binom{n/3}{2n/3}}=2,\quad n=3z.$$

The number of representations is thus[6]:

$n=3x$	$\frac{1}{6}n(n-3)$,		six-dimensional rep.,
	$2(n-1)$,		three-dimensional rep.,
	4,		two-dimensional rep.,
	2,		one-dimensional rep.,
$n\neq 3x$	$\frac{1}{6}(n-2)(n-1)$,		six-dimensional rep.,
	$2(n-1)$,		three-dimensional rep.,
	1,		two-dimensional rep.,
	2,		one-dimensional rep.

The basis vectors in the representation space of $[k_1,k_2,j,\mathbf{m}]$ are written as

$$e_{\bar{M}^{a}N^{b}\mathbf{m};(-1)^{c}j},\quad a\in\mathbb{Z}_3,\ b\in\mathbb{Z}_2,\ c\in\mathbb{Z}_2. \tag{3.8}$$

The assignment to a standard Euclidean basis is made according to the ordering introduced on the module $\mathbb{Z}_n^2$ and on Z_3.

The number of basis vectors is obviously the length of the orbits j or $\mathbf{m}$. The matrix representations are now

$$\begin{aligned}&[T^{[k_2,k_3,j,\mathbf{m}]}(g_{\mathbf{p}}^{r\gamma})]_{\bar{M}^{a'}N^{b'}\mathbf{m};(-1)^{c'}j,\bar{M}^{a}N^{b}\mathbf{m};(-1)^{c}j}\\ &\quad=\omega_3^{l_{\mathbf{m}}k_3(-1)^{b}\gamma}\cdot(-1)^{l_{\mathbf{m}}k_2r}\cdot\omega_3^{j\gamma(-1)^{c+r}}\\ &\quad\times\omega_n^{\bar{M}^{a}N^{b}\mathbf{m}\cdot N^{r}M^{-\gamma}\mathbf{p}}\,\delta^3_{j(-1)^{c'},j(-1)^{c+r}}\,\delta^r_{\bar{M}^{a'}N^{b'}\mathbf{m},\bar{M}^{(-1)^{r}a-\gamma}N^{b+r}\mathbf{m}},\end{aligned} \tag{3.9}$$

with $\omega_j=e^{2\pi i/j}$, and the dot means the standard scalar product. Although this expression looks somewhat complicated, it simplifies considerably if it is used for the particular representations. For the representations with $j\neq 0$ we can neglect a, a', b, b'; if $l_{\mathbf{m}}=3(2)$ we omit c, c' and b, $b'(a,a')$. Only for the six-dimensional one's we must keep a, a', b 'b' (but omit c, c'). Notice that for $l_{\mathbf{m}}=3$ we must take $\mathbf{m}=(m,m)$.

For the $\Delta\,(3n^2)$ the expressions simplify considerably. We can omit all Z_2 indices and also j, c, c'. Eq. (3.9) then reduces to

$$[T^{[k,\mathbf{m}]}(g_{\mathbf{p}}^{\gamma})]_{\bar{M}^{a'}\mathbf{m},\bar{M}^{a}\mathbf{m}}=\omega_3^{l_{\mathbf{m}}k\gamma}\,\omega_n^{\bar{M}^{a}\mathbf{m}\cdot M^{-\gamma}\mathbf{p}}\delta^n_{\bar{M}a'\mathbf{m},\bar{M}^{a-\gamma}\mathbf{m}}. \tag{3.10}$$

$\mathbf{m}$ is now to be taken from a set of representatives of the orbit decomposition of $\mathbb{Z}_n^2$ with respect to $\mathbb{Z}_3$. $l_{\mathbf{m}}$ is 1 or 3, $k\in Z_3$, and we have[6]

$n\neq 3x$:	$\frac{1}{3}(n^2-1)$,	three-dimensional rep.,
	3,	one-dimensional rep.,
$n=3x$:	$\frac{1}{3}(n^2-3)$,	three-dimensional rep.,
	9,	one-dimensional rep.

Equations (3.9) and (3.10) yield the following for the characters:

$$\Delta\,(3n^2):\quad \chi^{[k,\mathbf{m}]}(g_{\mathbf{p}}^{\gamma})=\sum_{a\in Z_3}\omega_3^{k\gamma}\,\omega_n^{\mathbf{m}M^{a-\gamma}\mathbf{p}}\,\delta^n_{\bar{M}^{a}\mathbf{m},\bar{M}^{a-\gamma}\mathbf{m}}, \tag{3.11}$$

$$l_{\mathbf{m}}=3:\quad \chi^{[k,\mathbf{m}]}(g_{\mathbf{p}}^{\gamma})=\delta^3_{\gamma,0}\,(\omega_n^{\mathbf{mp}}+\omega_n^{\mathbf{m}M\mathbf{p}}+\omega_n^{\mathbf{m}M^2\mathbf{p}}), \tag{3.12}$$

$$l_{\mathbf{m}}=1:\quad \chi^{[k,\mathbf{m}]}(g_{\mathbf{p}}^{\gamma})=\omega_n^{\mathbf{mp}}\cdot\omega_3^{k\gamma}, \tag{3.13}$$

$\Delta\,(6n^2)$:

$$l_{\mathbf{m}}=6,\ j=0:\quad \chi^{[0,0,0,\mathbf{m}]}(g_{\mathbf{p}}^{r\gamma})=\sum_{\substack{a\in Z_3\\ b\in Z_2}}\omega_n^{\mathbf{m}M^{a}N^{b}\mathbf{p}}\cdot\delta^2_{r0}\cdot\delta^3_{\gamma 0}, \tag{3.14}$$

$$\begin{aligned}l_{\mathbf{m}}=3,\ j=0:\quad &\chi^{[k_2,0,0,\mathbf{m}]}(g_{\mathbf{p}}^{r\gamma})=\sum_{a\in Z_3}(-1)^{k_2r}\,\omega^{\bar{M}^{a}\mathbf{m}\cdot\mathbf{p}}\\ &\times(\delta_{r,0}\,\delta_{\gamma,0}+\delta_{r,1}\,\delta_{\gamma,a-1}),\end{aligned} \tag{3.15}$$

$$l_{\mathbf{m}}=2:\quad \chi^{[0,k_3,0,\mathbf{m}]}(g_{\mathbf{p}}^{r\gamma})=\sum_{b\in Z_2}\omega_3^{(-1)^{b}k_3\gamma}\,\omega^{\mathbf{m}N^{b}\mathbf{p}}\delta_{r,0}$$

$$j=0$$
$$\mathbf{m}=\binom{n/3}{-n/3}$$
$$= 2\delta_{r,0}\cos\tfrac{2}{3}\pi(k\gamma + p_1 - p_2), \quad (\mathbf{p} = p_1,p_2), \tag{3.16}$$

$$l_{\mathbf{m}} = 0: \quad \chi^{[0,0,1,\mathbf{0}]}(g_{\mathbf{p}}^{r\gamma}) = \delta_{r,0}(3\delta_{\gamma,0} - 1), \tag{3.17}$$
$$j=1$$

$$l_{\mathbf{m}} = 0: \quad \chi^{[k_2,0,0,\mathbf{0}]}(g_{\mathbf{p}}^{r\gamma}) = \omega_3^{k_2 r}. \tag{3.18}$$
$$j=0$$

The above forms of the representations can be used to calculate the Clebsch–Gordon coefficients [and higher-order coupling coefficients (6-j etc.)].

We illustrate the procedure for the case of the $\Delta\,(3n^2)$ groups. We follow the standard method of constructing projection operators onto the irreducible components in the tensor products.[10]

Reducing the product representation $[k',\mathbf{m}'] \otimes [k'',\mathbf{m}'']$, we construct the vectors

$$P_a^{[k,\mathbf{m}]}(e_{\overline{M}^{a'}\mathbf{m}'} \otimes e_{\overline{M}^{a''}\mathbf{m}''}) = \sum_{\gamma,\mathbf{p}} [T^{[k,\mathbf{m}]}(g_{\mathbf{p}}^{\gamma})]^*_{\overline{M}^a\mathbf{m},\mathbf{m}}$$
$$\cdot (T^{[k',\mathbf{m}']} \otimes T^{[k'',\mathbf{m}'']})\,(g_{\mathbf{p}}^{\gamma}) e_{\overline{M}^{a'}\mathbf{m}'} \otimes e_{\overline{M}^{a''}\mathbf{m}''}$$
$$= \sum_{\gamma,\mathbf{p}} \omega_3^{(l_{\mathbf{m}'}k' + l_{\mathbf{m}''}k'' - l_{\mathbf{m}}k)} \delta_{\overline{M}^{a}_{\mathbf{m}},\overline{M}^{\;\;\gamma}\mathbf{m}}\, \omega_n^{(\overline{M}^{a'\;\;\gamma}\mathbf{m}' + \overline{M}^{a''\;\;\gamma}\mathbf{m}'' - \overline{M}^{a}\mathbf{m})\mathbf{p}}$$
$$\cdot e_{\overline{M}^{a'\;\;\gamma}\mathbf{m}'} \otimes e_{\overline{M}^{a''\;\;\gamma}\mathbf{m}''}$$
$$= \sum_{\gamma=0}^{3/l_{\mathbf{m}}-1} \omega_3^{(l_{\mathbf{m}'}k' + l_{\mathbf{m}''}k'' - l_{\mathbf{m}}k)(\gamma - a)} \delta_{\mathbf{m},\overline{M}^{a'}\mathbf{m}' + \overline{M}^{a''}\mathbf{m}''}$$
$$\cdot e_{\overline{M}^{a'+a\;\;\gamma}\mathbf{m}'} \otimes e_{\overline{M}^{a''+a\;\;\gamma}\mathbf{m}''}.$$

Thus taking into account that summation over γ occurs only if $[k,m]$ is one-dimensional, we get

$$P_a^{[k,m]}\, P_0^{[k,m]}(e_{\overline{M}^{a'}\mathbf{m}'} \otimes e_{\overline{M}^{a''}\mathbf{m}''})$$
$$= \sum_{\gamma=0}^{3/l_{\mathbf{m}}-1} \omega_3^{(l_{\mathbf{m}'}k' + l_{\mathbf{m}''}k'' - l_{\mathbf{m}}k)(\gamma - a)} \delta_{\mathbf{m},\overline{M}^{a}(\mathbf{m}' + \overline{M}^{a'' \; a}\mathbf{m}'')}$$
$$\cdot\, e_{\overline{M}^{a'+a\;\;\gamma}\mathbf{m}'} \otimes e_{\overline{M}^{a''+a\;\;\gamma}\mathbf{m}''}. \tag{3.19}$$

We notice that if $l_{\mathbf{m}} = 1$ we can add to a' and a'' some common integer, say s, and obtain proportional vectors. One of them can be selected by restricting a' by

$$0 \leqslant a' < \min(l_{\mathbf{m}''}, l_{\mathbf{m}}),$$

where we assumed $l_{\mathbf{m}'} \geqslant l_{\mathbf{m}''}$ without restriction of generality. From (3.19) the Clebsch–Gordon coefficients follow:

$$C^{[k,\mathbf{m}]\eta,a}_{[k',\mathbf{m}']a',[k'',\mathbf{m}'']a''}$$
$$= \left(\frac{l_{\mathbf{m}}}{3}\right)^{1/2} \sum_{r=0}^{l_{\mathbf{m}'}-1} \sum_{\gamma=0}^{3/l_{\mathbf{m}}-1} \omega_3^{(l_{\mathbf{m}'}k' + l_{\mathbf{m}''}k'' - l_{\mathbf{m}}k)(\gamma - a)}$$
$$\times \delta^n_{\mathbf{m},\overline{M}^{\eta}(\mathbf{m}' + \overline{M}^{r}\mathbf{m}'')}\, \delta^{l_{\mathbf{m}'}}_{\eta + a - \gamma, a'}\, \delta^{l_{\mathbf{m}''}}_{r + \eta + a - \gamma, a''}. \tag{3.20}$$

The parameter η counts the multiplicity of the representation $[k,m]$ in $[k',m'] \otimes [k'',m'']$. It allows one to label in a group theoretical way the several copies of $[k,m]$ occuring in the decomposition of $[k'm'] \otimes [k'',m'']$. The possibility to define such a label is closely related to the structure of the group as a semidirect product as discussed in Ref. 7.

To illustrate (3.20) consider some special cases:

$$l_{\mathbf{m}'} = l_{\mathbf{m}''} = l_{\mathbf{m}} = 3,$$
$$C^{[0,\mathbf{m}]\eta,a}_{[0,\mathbf{m}']a';[0,\mathbf{m}'']a''} = \sum_{u\in Z_3} \delta^n_{\mathbf{m},\overline{M}^{\eta}(\mathbf{m}' + \overline{M}^{u}\mathbf{m}'')}\, \delta^3_{\eta + a,a'}\, \delta^3_{u+\eta+a,a''}.$$
$$l_{\mathbf{m}'} = l_{\mathbf{m}''} = 3, \quad l_{\mathbf{m}} = 1,$$
$$C^{[k,\mathbf{m}]}_{[0,\mathbf{m}']a';[0,\mathbf{m}'']a''} = \frac{1}{\sqrt{3}} \sum_{\substack{u\in Z_3\\ \gamma\in Z_3}} \omega_3^{-k\gamma}\, \delta^n_{\mathbf{m},\mathbf{m}' + \overline{M}^{u}\mathbf{m}''}\, \delta^3_{a',-\gamma}\, \delta^3_{u-\gamma,a''}.$$
$$l_{\mathbf{m}'} = 3; l_{\mathbf{m}''} = 1, \quad l_{\mathbf{m}} = 3,$$
$$C^{[0,\mathbf{m}],a}_{[0,\mathbf{m}']a';[k'',\mathbf{m}'']} = \omega_3^{-k''a} \delta^n_{\mathbf{m},\mathbf{m}'+\mathbf{m}''}\, \delta^3_{a,a'}.$$
$$l_{\mathbf{m}'} = l_{\mathbf{m}''} = l_{\mathbf{m}} = 1,$$
$$C^{[k,\mathbf{m}]}_{[k',\mathbf{m}'],[k'',\mathbf{m}'']} = \delta^3_{k,k'+k''}\, \delta^n_{\mathbf{m},\mathbf{m}'+\mathbf{m}''}.$$

The coupling coefficients of the $\Delta\,(6n^2)$ have been found in an analogous way.[11]

All of the above is applicable to any SU(N). For example, we could define the $\Delta\,(N!n^{N-1}) = (\mathbb{Z}_n \otimes \mathbb{Z}_n \otimes \cdots \otimes \mathbb{Z}_n)\,\mathrm{s}\,S_N$ as subgroups of SU(N).

ACKNOWLEDGMENTS

We thank V. Rittenberg for suggesting that some of the groups of Ref. 7 might be SU(3) subgroups and for discussions. We thank M. Marcu for many most helpful comments.

APPENDIX

In this appendix we wish to prove the theorem in Sec. II.

We first show the following lemmas:

Lemma 1: If $P = p_1^{\beta_1}\cdots p_S^{\beta_S}$, with $p_i = 3z_i + 1$, there exists $a\in Zp$ such that $1 + a + a^2 = 0 \bmod P$. To show this we observe that there exist a_i such that $1 + a_i + a_i^2 = 0 \bmod p_i$. In fact, from $p_i = 3z_i + 1$ it follows that there exists a number, say b_i such that $3z_i$ is the smallest solution to $b_i^{\,x} = 0 \bmod p_i$. (b_i is a so called primitive root[12] which exists for all prime numbers) Now set $a_i = b_i^{\,z_i}$. Since $a_i \neq 1 \bmod p_i$, the equation $(a_i^3 - 1) = (a_i - 1)(1 + a_i + a_i^2) = 0 \bmod p_i$ implies $1 + a_i + a_i^2 = 0 \bmod p_i$ (because p_i is prime).

Now, using the theory of congruences,[12] it is easy to show that there exists an a with the property stated in Lemma 1. Furthermore, once a set of a_i's is known, it is possible to construct a in a step-by-step fashion.

Lemma 2: If $m = Q\cdot P$ or $m = 3QP$, P as in Lemma 1, $Q = q_1^{\alpha_1}\cdots q^{\alpha_r}$, $q_i \neq 3z_i + 1$ and $q_i \neq 3$ and $a^3 = 1 \bmod m$, then $G(m,3,a) = G(P,3,a) \otimes \mathbb{Z}_Q$ or $G(m,3,a) = G(P,3,a) \otimes \mathbb{Z}_{3Q}$. $a^3 = 1 \bmod m$ implies $a^3 = 1 \bmod P$ and $a^3 = 1 \bmod Q$. But, due to the assumption, this implies $a = 1 \bmod Q$. The Lemma is now a direct consequence of a factorization theorem stated in Ref. 7.

It remains to consider the case where m contains the factor 3 more than once.

Lemma 3: Let $m = 3^i\cdot Q\cdot P$, $i > 1$. Then there exists a such that $(a - 1)(1 + a + a^2) = 0 \bmod m$ and $(a - 1)a \neq a - 1 \bmod m$. The first part of the lemma is trival; for the second observe that $(a - 1)a = (a - 1) \bmod m$ implies

$(a-1)a = (a-1) \bmod p$; thus $a+1 = 2a \bmod p$ and so $1 + a + a^2 = 3 \bmod p$. This implies $a = 1 \bmod P$. (See Lemma 1.) But according to Lemma 1 there exists $a_p \neq 1 \bmod P$ and $a_p{}^3 = 1 \bmod P$. Lemma 2 states that the only solution to $a^3 = 1 \bmod Q$ is $1 \bmod Q$. The congruence $a^3 = 1 \bmod 3^i$ has a solution, say a_3. Applying now a theorem from the theory of congruences, the so-called Chinese remainder theorem,[12] we can show that there exists an a satisfying $a^3 = 1 \bmod 3^i QP$ such that $a = a_p \bmod P$, $a = 1 \bmod Q$, $a = a_3 \bmod 3^i$; thus $a \neq 1 \bmod P$ and we see that Lemma 3 holds.

We are now ready to show for what values of m we obtain subgroups of SU(3):

Therorem: Let $m = 3^i \cdot Q \cdot P$, Q, P, defined above, $P \neq 1$. Then there exists an a such that the group $G(m,3,a)$ is a subgroup of SU(3). Furthermore, $G(m,3,a) = G(3^i P,3,a) \otimes \mathbb{Z}_Q$ if $i > 1$ and $G(m,3,a) = G(P,3,a) \otimes \mathbb{Z}_{3Q}{}^i$ if $i = 1,0$.

Proof: If $i \leqslant 1$ the theorem follows immediately from Lemmas 1 and 2 considering the representation $[0,1]$. If $i > 1$, we consider the representation $[0,a-1]$ and observe that by Lemma 3 the orbit of $a-1$, that is $a-1$, $a(a-1)$, $a^2(a-1)$ has length 3 [that is $a-1$, $a(a-1)$ and $a^2(a-1)$ are all different], thus, $[0,a-1]$ has dimension 3 and determinant $(a-1)(1+a+a^2) = 1$, which proves the theorem.

[1]M. Creutz, L. Jacobs, and C. Rebbi, Phys. Rev. Lett. **42**, 1390 (1979); M. Creutz, Phys. Rev. Lett. **43**, 553 (1979).
[2]See, e.g., J. Kogut, Rev. Mod. Phys. **51**, 659 (1979).
[3]C. Rebbi, Phys. Rev. D **21**, 3350 (1980).
[4]D. Petcher and D. H. Weingarten, Indiana University Preprint IUHET-53, 1980.
[5]G. A. Miller, H. E. Dickson, and L. E. Blichfeldt, *Theory and Applications of Finite Groups* (G. E. Stechart, New York, 1938).
[6]W. M. Fairbairn, T. Fulton, and W. H. Klink, J. Math. Phys. **5**, 1038 (1964).
[7]A. Bovier, M. Lüling, D. Wyler, Bonn-Preprint-HE-80-10, Universität Bonn, 1980.
[8]In particular, we were able to solve the "multiplicity" problem; that is if in a Clebsch–Gordon series the same irreducible representation occurs several times, there is a group theoretical way to label them.
[9]C. Curtis and I. Reiner, *Representation Theory of Finite Groups and Associative Algebras* (Interscience, New York, 1962); G. W. Mackey, *Induced Representations of Groups and Quantum Mechanics* (Benjamin, New York, 1968); see also Ref. 7.
[10]See for example G. Ljubarskii, *The Theory of Groups and Its Applications to Physics* (Pergamon, Oxford, 1960).
[11]The expressions are very complicated and will not be given here.
[12]See for example B. M. Stewart, *Theory of Numbers* (Macmillan, New York, 1964).

Semiunitary projective representations of the complete Galilei group [a)]

José F. Cariñena
Departamento de Física Teórica, Facultad de Ciencias, Universidad de Zaragoza, Zaragoza, Spain
Mariano Santander [b)]
Departamento de Física Teórica, Facultad de Ciencias, Universidad de Valladolid, Valladolid, Spain

(Received 15 April 1980; accepted for publication 14 November 1980)

A systematic study of the semiunitary projective representations of the Galilei group including reflections is presented. They are found by means of the semiunitary representations of a representation group.

PACS numbers: 02.20.Qs

1. INTRODUCTION

This paper deals with an important application of group theory to "nonrelativistic" quantum physics; that of how to determine the semiunitary projective representations (hereafter SUPR's) of the kinematical group describing the Newtonian universe, namely the complete Galilei group $\mathscr{G}$.

Why the physical interest of this mathematical problem? From the classical works of Wigner[1] and Bargmann[2] it is well known that a quantum system which is invariant under a group G of spacetime transformations is described by means of the representation space of a semiunitary (unitary/antiunitary) projective representation of the group G. So in order to be able to solve this problem several techniques have been developed according to the specific structure of the group G, mainly when it is a finite group,[3–5] a connected Lie group[6,7] or a more general kind of topological group.[8,9] Papers devoted to the study of the corresponding problem when G is a nonconnected Lie group are seldom found even if discrete symmetries have played a very important role in quantum physics; for instance, the first paper where the problem of the complete Poincaré group is considered, is the one of Wigner[10] although the results of his paper had been previously announced by Wightman.[11] However, the method which was used by him is very cumbersome and it cannot be considered as a general method but a very "*ad hoc*" one. The more recent papers by Shaw and Lever[12] and that of Ebner[13] are also worthy of note.

In spite of the great importance of the complete Galilei group there are very few papers devoted to the study of the SUPR's of $\mathscr{G}$, the oldest being that of Brennich[14] to the best of our knowledge. However, Brennich studied only some of the representations of $\mathscr{G}$ and this through the multiplier representations of an auxiliary group. Cattaneo[15] has shown how it is possible to determine all the SUPR's of $\mathscr{G}$ by means of the SUR's of a "representation group" for $\mathscr{G}$. In this paper we deal with the method of determining such a group as well as carrying out all calculations which are needed in order to describe the SUPR's of $\mathscr{G}$.

The first point to be clarified is that it is not fully correct to say "semiunitary representation of G" but "semiunitary representation of the pair (G,G_+)", where G_+ is the subgroup (of index one or two) of the elements of G represented by unitary operators: therefore the subgroup G_+ is to be specified. In the case of the complete Galilei group $\mathscr{G}$, the only four possibilities for a closed subgroup of index one or two are

(i) $\mathscr{G}_+ = \mathscr{G}_0 \cup \mathscr{G}_0 P$,

(ii) $\mathscr{G}_+ = \mathscr{G}_0 \cup \mathscr{G}_0 T$,

(iii) $\mathscr{G}_+ = \mathscr{G}_0 \cup \mathscr{G}_0 PT$,

and the trivial case $\mathscr{G}_+ = \mathscr{G}$, where $\mathscr{G}_0$ is the connected component of the identity.

The first possibility seems to be the only appropriate one due to analogy with the relativistic case of the complete Poincaré group, but in this case there is an additional reason[16]: with a different choice for $\mathscr{G}_+$, the restriction of any SUPR of $(\mathscr{G},\mathscr{G}_+)$ to $\mathscr{G}_0$ would correspond to a "massless representation" of $\mathscr{G}_0$ which is unphysical[17] and hence only the first possibility will be considered.

We remind that a representation group $\overline{G}$ for (G,G_+) is an extension of G by a well-defined abelian group A[5,8]; the elements of $\overline{G}$ may be denoted as pairs (a,g) where $a \in A$ and $g \in G$. There is an epimorphism $p: \overline{G} \to G$, $p[(a,g)] = g$, and therefore $\ker p$ is isomorphic to A. Let $\overline{G}_+$ be given by $\overline{G}_+ = p^{-1}(G_+)$. The link between SUPR's of (G,G_+) and SUR's of $(\overline{G},\overline{G}_+)$ is as follows: if $\mathscr{U}$ is a SUR of $(\overline{G},\overline{G}_+)$ such that its restriction to A is a (multiple of a) one dimensional unitary representation of A, we define the associated SUPR $\overline{\mathscr{U}}$ of (G,G_+) by $\overline{\mathscr{U}} = \pi \cdot \mathscr{U} \cdot s$, where s is an arbitrary section $s: G \to \overline{G}, p \cdot s = \mathrm{id}_G$.

Notice that the subgroup A is isomorphic to the dual (relative to some appropriated topology) of the second cohomology group $H^2_*(G,\mathbf{T})$. The subscript * is used to denote the action of G on $\mathbf{T}$ given by $\lambda^g = \lambda$ whenever $g \in G_+$ and $\lambda^g = \lambda^{-1}$ if $g \notin G_+$.

When the trivial action of G on an abelian group is considered, the corresponding cohomology groups will be denoted by the subscript 0. Finally, when the action of G on the abelian group A is $a^g = a$ if $g \in G_+$, $a^g = a^{-1}$ if $g \notin G_+$. The corresponding cohomology groups will carry the subscript $-$ [e.g., $Z^2_-(G,A)$].

The organization of this paper is as follows: in Sec. 2 we find one representation group $(\overline{\mathscr{G}},\overline{\mathscr{G}}_+)$ to be used in the derivation of the SUPR's of $(\mathscr{G},\mathscr{G}_+)$ and its structure is studied.

[a)]This work has been supported by Instituto de Estudios Nucleares.
[b)]In part from the Ph.D. thesis.

 © 1981 American Institute of Physics

In Sec. 3 we determine the SUR's of $(\bar{\mathscr{G}},\bar{\mathscr{G}}_+)$ which may be gotten from the UR's of $\bar{\mathscr{G}}_+$ by means of a well-known technique; a great part of these calculations are given in Appendices A and B. In Sec. 4 we study the SUPR's of $(\mathscr{G},\mathscr{G}_+)$, that is to say, we study the classes of pseudoequivalence of the SUR's of $(\bar{\mathscr{G}},\bar{\mathscr{G}}_+)$. Finally, Sec. 5 is a short comment of the relation between this work and the precedent ones, as well as the physical significance of some of the results which are obtained.

2. THE STRUCTURE OF A REPRESENTATION GROUP FOR THE COMPLETE GALILEI GROUP

In this section we study the structure of a representation group for the complete Galilei group $\mathscr{G}$. The group $\mathscr{G}$ is a semidirect product of the restricted group $\mathscr{G}_0$ by the inversions group $V = \{I,P,T,PT\}$ (notations as in Lévy-Leblond's paper[18]). The action of the subgroup V on $\mathscr{G}_0$ is the usual one, namely, b changes its sign under T and PT, etc. We will denote (g,α) the elements of $\mathscr{G}$, with $g\in\mathscr{G}_0$ and $\alpha\in V$; g^α means the image of g by α. Sometimes we will use the notation $(b,\mathbf{a},\mathbf{v},A)^\alpha = (b^\alpha,\mathbf{a}^\alpha,\mathbf{v}^\alpha,A)$, bearing in mind the different meanings of b^α, $\mathbf{a}^\alpha$, and $\mathbf{v}^\alpha$.

As indicated in the introduction, we shall only study those SUPR's of $\mathscr{G}$ where $\mathscr{U}(P)$ is unitary and $\mathscr{U}(T)$ antiunitary,[16] so that our choice for $\mathscr{G}_+$ is $\mathscr{G}_+ = \{(g,\alpha), g\in\mathscr{G}_0, \alpha = I,P\}$. In this case the representation groups have been given by Santander[9] and Cattaneo.[15] There are three nonisomorphic representation groups. Now we shall give the necessary details about the structure and the construction of one of these groups. The notations and procedures are similar to those used in Ref. 5 which deals with the case of finite groups.

The first step is to know $H^2_{\bullet}(\mathscr{G},\mathbf{T})$, where $\mathscr{G}$ acts on $\mathbf{T}$ via V, i.e. I and P act as the identity while T and PT as the inversion. The group $H^2_*(\mathscr{G},\mathbf{T})$ is isomorphic to $\mathbb{R}\otimes C_2\otimes V$, and the generic element of this group will be denoted by $[M,l,m,n]$, $M\in\mathbb{R}, l,m,n\in\{-1,1\}$. The isomorphism can be derived as in Ref. 15, or in a more pedestrian way by a simple analysis of the structure of the factor system of a semidirect product following the pattern of Mackey's result[19,20] in an adequately generalized way in order to allow a nontrivial action.[9,16,21] We refrain from giving the unnecessary details. The representation group we are going to use is obtained as follows: we choose a homomorphic section $s:H^2_*(\mathscr{G},\mathbf{T})\to Z^2_*(\mathscr{G},\mathbf{T})$, given by

$$s[M,l,m,n]\to\{(g',\alpha';g,\alpha)\\ \to\exp[iM(b^{\alpha'}\mathbf{v}'^2+\mathbf{v}'R'\mathbf{a}^{\alpha'})]\xi_l(R',R)\omega_{mn}(\alpha',\alpha)\}, \tag{2.1}$$

where ξ_l is a lifting of $[l]\in H^2_0(\mathrm{SO}(3),\mathbf{T})$ with $\xi_1(R',R)=1$, and $\xi_{-1}(R',R)$ takes only the values ± 1 [for example, the one obtained from a section σ: SO(3)$\to$SU(2) as $\sigma(R')\cdot\sigma(R)=\xi_{-1}(R',R)\sigma(R',R)$], and finally, ω_{mn} is the lifting of $[m,n]\in H^2_{\bullet}(V,\mathbf{T})$ given by $\omega_{mn}=(m,n,mn,n)$ in the notation of Ref. 5 (Sec. 9.2) for the elements of $Z^2_{\bullet}(V,\mathbf{T})$, i.e., explicitly

ω_{mn}	P	T	PT
P	$m\cdot n$	1	$m\cdot n$
T	1	n	n
PT	$m\cdot n$	n	m

(2.2)

Now let us define the applications

$$W(g',\alpha';g,\alpha):H^2_{\bullet}(\mathscr{G},\mathbf{T})\to\mathbf{T},$$

$$W(g',\alpha';g,\alpha):[M,l,m,n]\to(s[M,l,m,n])(g',\alpha';g,\alpha). \tag{2.3}$$

The product of the canonical topology on $\mathbb{R}$ and the discrete ones on each C_2 endows $H^2_{\bullet}$ with a locally compact topology making continuous all the maps $W(g',\alpha';g,\alpha)$. The dual space of $H^2_{\bullet}(\mathscr{G},\mathbf{T})$ is isomorphic to $\mathbb{R}\otimes C_2\otimes V$ and is generated by $\theta\in\mathbb{R}$ and by λ,μ,ν given by

$$\lambda[0,l,m,n]=l,\quad \mu[0,l,m,n]=m,$$

$$\nu[0,l,m,n]=n,\quad \theta[M,l,m,n]=e^{i\theta M}. \tag{2.4}$$

Hence, $W\in Z^2_-(\mathscr{G},\widehat{H^2_{\bullet}})$ and the corresponding extension is the representation group we are looking for. The action of $(g,\alpha)\in\mathscr{G}$ in $\widehat{H}^2_{\bullet}$ is the identity or the inversion according to $\alpha\in V_+$ or $\alpha\notin V_+$; explicitly, (β and γ are always of order two),

$$(g,\alpha):(\theta,\beta,\gamma)\to(\theta^\alpha,\beta,\gamma),\quad \theta\in\mathbb{R}, \beta\in C_2, \gamma\in V, \tag{2.5}$$

and $\theta^\alpha=\theta$ or $-\theta$ according to $\alpha\in V_+$ or $\alpha\notin V_+$ (i.e., θ transforms as time). In order to obtain an intrinsic characterization of W, the relation (2.3) can be rewritten, using (2.1) and (2.4) as follows:

$$W(g',\alpha';g,\alpha)\\ =(\tfrac{1}{2}b^{\alpha'}\mathbf{v}'^2+\mathbf{v}'\cdot R'\mathbf{a}^{\alpha'},\Xi(R',R),W(\alpha',\alpha)), \tag{2.6}$$

where Ξ is the lifting of the nontrivial element of H^2 [SO(3),$\widehat{H}^2$(SO(3),$\mathbf{T}$)] given by

$$\Xi(R',R)=\begin{cases}1 & \text{if}\quad \xi_{-1}(R',R)=1,\\ \lambda & \text{if}\quad \xi_{-1}(R',R)=-1,\end{cases} \tag{2.7}$$

and where $W(\alpha',\alpha)$ is given by a table like (2.2) with the replacement $m\to\mu$ and $n\to\nu$.

Finally, we obtain the representation group $\bar{\mathscr{G}}$ defined as the set $\mathbb{R}\times C_2\times V\times\mathscr{G}$, with composition law

$$(\theta',\beta',\gamma',g',\alpha')(\theta,\beta,\gamma,g,\alpha)=(\theta'+\theta^{\alpha'}+\tfrac{1}{2}b^{\alpha'}\mathbf{v}'^2+\mathbf{v}'R'\mathbf{a}^{\alpha'},\\ \beta'\beta\Xi(R',R),\gamma'\gamma W(\alpha',\alpha),g'g^{\alpha'},\alpha'\alpha), \tag{2.8}$$

and it is to be endowed with the unique locally compact topology making this extension a topological one. Notice that with a different choice of the homomorphic section in Eq. (2.1) we would get a different representation group where the subgroup $\{(0,0,\gamma,1,\alpha)\}$ would not be an abelian subgroup, while in the other aspects it seems much the same as this group $\bar{\mathscr{G}}$.

Next we analyze more carefully the structure of $\bar{\mathscr{G}}$. First of all, the subgroups $\{(\theta,1,\gamma,(b,\mathbf{a},\mathbf{v},I),\alpha)\}$ and $\{(0,\beta,1,(0,0,0,R),1)\}$ determine a semidirect structure, (β,R) acting on the former via the homomorphism $(\beta,R)\to R$. From (2.7) it is clear the $\{(\beta,R)\}$ is topologically isomorphic to SU(2). Denoting by A the elements of SU(2), and by $A\mathbf{a}, A\mathbf{v},\cdots$ the images of $\mathbf{a},\mathbf{v},\cdots$, respectively, under the rotation associated to A, and with a little reordering we will rewrite

(2.8) in the form

$$(\theta',b',\mathbf{a}',\mathbf{v}',A',\gamma',\alpha')(\theta,b,\mathbf{a},\mathbf{v},A,\gamma,\alpha)$$
$$= (\theta' + \theta^{\alpha'} + \tfrac{1}{2}b^{\alpha'}\mathbf{v}'^2 + \mathbf{v}'\cdot A'\mathbf{a}^{\alpha'}, b' + b^{\alpha'}, \mathbf{a}' + A'\mathbf{a}^{\alpha'} + \mathbf{v}b^{\alpha'}, \mathbf{v}' + A\mathbf{v}^{\alpha'}, A'A, \gamma'\gamma\, W(\alpha',\alpha), \alpha'\alpha) . \quad (2.9)$$

We must remark that the natural appearance of SU(2) instead of SO(3) is a direct consequence of the construction, and not an "a priori" substitution as in Refs. 12 and 14. Next, notice that the universal covering group $\mathscr{G}_0^*$ of $\mathscr{G}_0$ which is identified with the subset $\{(0,b,\mathbf{a},\mathbf{v},A,1,1)\}$ is not a subgroup of $\bar{\mathscr{G}}$, but it can be identified with a factor group with respect to the subgroup $\{(\theta,\gamma,\alpha)\}$. On the other hand, an invariant subgroup of $\bar{\mathscr{G}}$, namely $\bar{\mathscr{G}}_0 = \{(\theta, b, \mathbf{a}, \mathbf{v}, A, 1, 1)\}$ can be canonically identified with the "projective covering group" of $\mathscr{G}_0$. The subgroups $\bar{\mathscr{G}}_0$ and $\bar{V} = \{(0,0,0,0,I,\gamma,\alpha)\}$ determine a semidirect structure for $\bar{\mathscr{G}}$, $\bar{\mathscr{G}} = \bar{\mathscr{G}}_0 \odot \bar{V}$ with action [see (2.5)]

$$(\gamma,\alpha) : (\theta,b,\mathbf{a},\mathbf{v},A) \mapsto (\theta^{\alpha},b^{\alpha},\mathbf{a}^{\alpha},\mathbf{v}^{\alpha},A) .$$

It is easy to show that $\bar{V}$ is isomorphic to $C_4 \otimes C_4$ generated by $(1,P)$ and $(1,T)$, so that if we forget momentarily the Galilean structure implying the appearance of θ, we may say that with respect to inversions, the transition from $\mathscr{G}$ to $\bar{\mathscr{G}}$ involves making tetracyclic the inversions $(1,P)$ and $(1,T)$, just as the "double group" trick does the linearization of all UPR's of the rotation group and its finite subgroups by making rotations of angle π tetracyclic rather than involutive. But in general a similar case does not work, as the elements appearing in $\bar{\mathscr{G}}$ for the Galilean case show.

3. THE IRREDUCIBLE SEMIUNITARY REPRESENTATIONS OF $(\bar{\mathscr{G}}, \bar{\mathscr{G}}_+)$

A. The irreducible unitary representations of $\bar{\mathscr{G}}_+$

The ISUR's of $(\bar{\mathscr{G}}, \bar{\mathscr{G}}_+)$ are to be found from the irreducible unitary representations of the subgroup $\bar{\mathscr{G}}_+$ (cf. part B in this section) and hence the first step is to know the IUR's of $\bar{\mathscr{G}}_+$. In order to find them, it is advisable to "descompose" $\bar{\mathscr{G}}_+$ as a semidirect product with abelian kernel (in that case it is very easy to apply Mackey's theory) and fortunately such a decomposition is possible. Let $\bar{T}_4$ and $\bar{K}_+$ be the following subsets of $\bar{\mathscr{G}}_+$

$$\bar{T}_4 = \{(\theta,b,\mathbf{a},0,I,1,1)\}, \quad \bar{K}_+ = \{(0,0,0,\mathbf{v},A,\gamma,\alpha), \alpha = 1,P\} .$$

Then $\bar{T}_4$ is an invariant subgroup of $\bar{\mathscr{G}}_+$, the "extended" spacetime translations subgroup which is topologically isomorphic to $\mathbb{R}^5$. The subset $\bar{K}_+$ is also a subgroup, and both $\bar{T}_4$ and $\bar{K}_+$ determine a semidirect structure $\bar{\mathscr{G}}_+ = \bar{T} \odot \bar{K}_+$; the action of $\bar{K}_+$ on $\bar{T}_4$ is given by

$$(\mathbf{v}, A,\gamma,\alpha):(\theta,b,\mathbf{a}) \mapsto (\theta + \tfrac{1}{2}b\mathbf{v}^2 + \mathbf{v}\cdot A\mathbf{a}^{\alpha}, b, A\mathbf{a}^{\alpha} + b\mathbf{v}) ,$$

and it is regular in Mackey's meaning. The whole standard theory of induced representations[22,23] may be used.

The dual space $\bar{T}_4$ is topologically isomorphic to $\mathbb{R}^5$ too. Its elements will be denoted $(M,E,\mathbf{p})$ according to

$$(M,E,\mathbf{p}):(\theta,b,\mathbf{a}) \mapsto \exp\{i(M\theta + Eb - \mathbf{pa})\} .$$

The natural action of $\bar{K}_+$ on $\bar{T}_4$ is

$$(\mathbf{v},A,\gamma,\alpha):(M,E,\mathbf{p}) \mapsto (M,E + \mathbf{v}^{\alpha}\cdot A\mathbf{p} + \tfrac{1}{2}M\mathbf{v}^2, A\mathbf{p}^{\alpha} + M\mathbf{v}) ,$$

and the orbits are naturally classified in three strata, as follows:

$$\text{Orbits } Z_{m,\rho} = \{(M,E,\mathbf{p})/M = m,\ 2mE - \mathbf{p}^2 = \rho\}, \quad m,\rho\in\mathbb{R},\ m\neq 0 ,$$
$$\text{Orbits } Z_{0,\rho} = \{(0,E,\mathbf{p})/\mathbf{p}^2 = \rho\}, \quad \rho\in\mathbb{R},\ \rho>0 ,$$
$$\text{Orbits } Z_{00E} = \{(0,E,0)\}, \quad E\in\mathbb{R} .$$

This orbit structure is the same as the one obtained in the absence of P; this result was physically foreseeable. Next, we must determine for each orbit Z, (a) the little group, and (b) the element $\mathscr{L}(x)$ (arbitrarily chosen) of $\bar{K}_+$ mapping a fixed point $x_0\in Z$ in the generic one $x\in Z$.

Orbits $Z_{m,\rho}$, $m\neq 0$: This orbit is a three-dimensional "paraboloid" of revolution, lying in the hyperplane $M = m$, and with its defining equation $E = \rho/2m + \mathbf{p}^2/2m$. Taking $\mathbf{p}\in\mathbb{R}^3$ as the representative of the point $(m, (\rho + \mathbf{p}^2)/2m, \mathbf{p})$, the invariant measure in $Z_{m,\rho}$ goes to the measure $d^3\mathbf{p}$ in $\mathbb{R}^3$. Here, the action is

$$(\mathbf{v},A,\gamma,\alpha):\mathbf{p} \mapsto A\mathbf{p}^{\alpha} + m\mathbf{v} ,$$
$$(\mathbf{v},A,\gamma,\alpha)^{-1}:\mathbf{p} \mapsto A^{-1}(\mathbf{p} - m\mathbf{v})^{\alpha} .$$

A natural choice for x_0 is $\mathbf{p}_0 = \mathbf{0}$, $x_0 = (m,\rho/2m,\mathbf{0})$. Then the little group is characterized by $m\mathbf{v} = 0$ and hence it is $\mathrm{SU}(2)\otimes\bar{V}_+$. Finally, an element $\mathscr{L}(\mathbf{p})$ of $\bar{K}_+$ mapping $\mathbf{p}_0$ into $\mathbf{p}$ is naturally selected as the pure Galilei transformation of speed $\mathbf{p}/m$, that is, $\mathscr{L}(\mathbf{p}) = (\mathbf{p}/m,I,1,1)$.

Orbits $Z_{0,\rho}$, $\rho>0$: This orbit is a "cylinder" with the base the sphere of radius ρ, S^2_ρ, $(\mathbf{p}^2 = \rho)$, and axis E. The invariant measure is $d\Omega_{\mathbf{p}}\, dE$ and the action of $\bar{K}_+$ on $Z_{0,\rho}$ is given by

$$(\mathbf{v},A,\gamma,\alpha):(E,\mathbf{p}) \mapsto (E,\mathbf{v}\cdot A\mathbf{p}^{\alpha}, A\mathbf{p}^{\alpha}) ,$$
$$(\mathbf{v},A,\gamma,\alpha)^{-1}:(E,\mathbf{p}) \mapsto (E - \mathbf{v}\cdot\mathbf{p}, A^{-1}\mathbf{p}^{\alpha}) .$$

For the fixed point x_0 in the orbit, we will take $E_0 = 0$ and for $\mathbf{p}_0$, the "north pole" of the sphere S^2_ρ, say, $\mathbf{p}_0 = (\sqrt{\rho})\mathbf{u}_z$. This little group is characterized by $A\mathbf{u}_z = \mathbf{u}_z^{\alpha}$ and $\mathbf{u}_z\cdot\mathbf{v} = 0$, and then it is:

$$G_{Z_{0,\rho}} = \{(\mathbf{v}_{xy}, A_z,\gamma,1), (\mathbf{v}_{xy},A_z\mathrm{S},\gamma,P)\} ,$$

where $\mathbf{v}_{xy}$ denotes a vector $\mathbf{v}$ contained in the "equatorial" Oxy plane, A_z denotes an element of SU(2) which is canonically projected on a rotation around Oz, and S is a fixed element of SU(2) corresponding to a rotation of angle π around some axis contained in the Oxy plane. A possible choice which we shall take is $\mathrm{S} = i\sigma_y$. The structure of this little group is given in Appendix A, and let us only quote here that it also appears in the relativistic case.[24] Finally, for $\mathscr{L}(E,\mathbf{p})$ we will choose

$$\mathscr{L}(E,\mathbf{p}) = (E\mathbf{p}/\rho, L(\mathbf{p}),1,1) ,$$

where

$$L(\mathbf{p}) = \begin{cases} \dfrac{I + (\sigma\cdot\mathbf{p}/\sqrt{\rho})\,\sigma_z}{[2(1 + p_z/\sqrt{\rho})]^{1/2}} & \text{for } \mathbf{p}\neq(0,0,-\sqrt{\rho}) , \\ -\mathrm{S} & \text{for } \mathbf{p} = (0,0,-\sqrt{\rho}) , \end{cases} \quad (3.1)$$

is the usual boost of SU(2) in S^2_ρ; the only discontinuity is in the "south pole" and the election of $-\mathrm{S}$ in that point is in order to preserve continuity along the "meridian" $\varphi_{\mathbf{p}} = 0$. It is not possible to select $\mathscr{L}(\mathbf{p})$ depending on $\mathbf{p}$ continuously over all S^2_ρ.[25]

Orbits Z_{00E}: This orbit reduces to one point, so that the little group is $\bar{K}_+$.

Now we have all the information needed in order to apply Mackey's theory. The irreducible representations of every little group are collected in Appendix A. The explicit expressions of the operators in a given representation do not depend on more choices than the ones previously made. In each case we give the operator $U(P) = U(0,0,0,I,1,P)$ an explicit expression which is very relevant for the induction of the ISUR's of $(\overline{\mathscr{G}}, \overline{\mathscr{G}}_+)$ at a later stage. The labelling adopted for the IUR's of $\overline{\mathscr{G}}_+$ follows the pattern of the notations of Inonu and Wigner[17] and Lévy-Leblond[18] for the IUR's of the proper Galilei group.

IUR's *of* $\overline{\mathscr{G}}_+$ *associated with the orbit* $Z_{m,\rho}$: The family $D_{j,\epsilon_1,\epsilon_2,\epsilon}$ of representations of the little group gives rise to a "kind" of IUR's of $\overline{\mathscr{G}}_+$ denoted by $m[\rho,j,\epsilon_1,\epsilon_2,\epsilon]$, $m\in\mathbb{R}$, $m\neq 0$, $\rho\in\mathbb{R}$, $2j\in\mathbb{N}$, $\epsilon_1,\epsilon_2,\epsilon\in\{1,-1\}$. The carrier space is $\mathscr{L}^2(\mathbb{R}^3\mapsto\mathbb{C}^{2j+1}, d^3\mathbf{p})$, and the operators themselves are

$$[U(\theta,b,\mathbf{a},\mathbf{v},A,\gamma,\alpha)\psi](\mathbf{p}) = \exp\{i(m\theta + [(\rho+\mathbf{p}^2)/2m]\,b - \mathbf{pa})\} \times \mathscr{D}_j(A)\,\Delta_{\epsilon_1,\epsilon_2,\epsilon}(\gamma,\alpha)\,\psi(A^{-1}(\mathbf{p}-m\mathbf{v})^\alpha). \tag{3.2}$$

For the parity

$$[U(P)\psi](\mathbf{p}) = \epsilon\psi(-\mathbf{p}) \quad \text{if } \epsilon_1\cdot\epsilon_2 = +1,$$
$$[U(P)\psi](\mathbf{p}) = i\epsilon\psi(-\mathbf{p}) \quad \text{if } \epsilon_1\cdot\epsilon_2 = -1. \tag{3.2a}$$

IUR's of $\overline{\mathscr{G}}_+$ *associated with the orbit* $Z_{0,\rho}$: Corresponding to the three series of IUR's of the little group, I, II, and III, we have here three series, which will be denoted $\mathrm{II_I}$, $\mathrm{II_{II}}$, and I respectively.

$\mathrm{II_I}(\rho,\epsilon_1,\epsilon_2,\epsilon)$: $\rho\in\mathbb{R}$, $\rho>0$, $\epsilon_1,\epsilon_2,\epsilon\in\{1,-1\}$. The carrier space is $\mathscr{L}^2(\mathbb{R}\times S^2_\rho\mapsto\mathbb{C}, dE\,d\Omega_\mathbf{p})$, and the action is

$$[U(\theta,b,\mathbf{a},\mathbf{v},A,\gamma,\alpha)\psi](E,\mathbf{p}) = \exp\{i(Eb-\mathbf{pa})\}\,\Delta_{\epsilon_1,\epsilon_2,\epsilon}(\gamma,\alpha)\,\psi(E-\mathbf{vp}, A^{-1}\mathbf{p}^\alpha) \tag{3.3}$$

and

$$[U(P)\psi](E,\mathbf{p}) = \begin{cases} \epsilon\psi(E,-\mathbf{p}) \text{ if } \epsilon_1\cdot\epsilon_2 = 1, \\ i\epsilon\psi(E,-\mathbf{p}) \text{ if } \epsilon_1\cdot\epsilon_2 = -1. \end{cases}$$

$\mathrm{II_{II}}(\rho,h,\epsilon_1,\epsilon_2)$: $\rho\in\mathbb{R}$, $\rho>0$, $2h\in\mathbb{N}$, $\epsilon_1,\epsilon_2\in\{1,-1\}$. The carrier space is $\mathscr{L}^2(\mathbb{R}\times S^2_\rho\to\mathbb{C}^2, dE\,d\Omega_\mathbf{p})$ and the representation is

$$[U(\theta,b,\mathbf{a},\mathbf{v},A,\gamma,\alpha)\psi](E,\mathbf{p}) = \exp\{i(Eb-\mathbf{p}\cdot\mathbf{a})\} \times D_{h\epsilon_1\epsilon_2}(L^{-1}(\mathbf{p})\,AL\,(A^{-1}\mathbf{p}^\alpha),\gamma,\alpha)\,\psi(E-\mathbf{vp},A^{-1}\mathbf{p}^\alpha). \tag{3.4}$$

Particularization for P here is slightly involved, and it gives (Appendix B)

$$[U(P)\psi](E,\mathbf{p}) = \begin{bmatrix} 0 & e^{2ih\varphi_\mathbf{p}} \\ (-1)^{2h}\epsilon_1\epsilon_2\,e^{-2ih\varphi_\mathbf{p}} & 0 \end{bmatrix},$$

where $\varphi_\mathbf{p}$ is the 'azimuthal" polar angle of $\mathbf{p}$ given in Appendix B (see the comments after B).

$\mathrm{I}(\rho,x,\eta,\epsilon_1,\epsilon_2)$: $\rho,x\in\mathbb{R}$, $\rho,x>0$, $\eta\in\{1,i,-1,-i\}$, $\epsilon_1,\epsilon_2\,\{1,-1\}$. These are the most complex because of the infinite dimensionality of the representations of the little group. Through obvious identifications the carrier space is (isometric to)

$$\mathscr{L}^2(S^1_x\times\mathbb{R}\times S^2_\rho, d\varphi_\mathbf{k}\,d\Omega_\mathbf{p}\,dE)$$

and the expression for the operators is

$$[U(\theta,b,\mathbf{a},\mathbf{v},A,\gamma,\alpha)\psi](\mathbf{k},E,\mathbf{p}) = \exp\{i(L(\mathbf{p})\,\mathbf{k}\cdot\mathbf{v} + Eb - \mathbf{pa})\} \times\Delta_{\eta\epsilon_1\epsilon_2}(\mathscr{L}^{-1}(\mathbf{k})\,L^{-1}(\mathbf{p})\,AL\,(A^{-1}\mathbf{p}^\alpha) \times\mathscr{L}([L^{-1}(\mathbf{p})\,AL\,(A^{-1}\mathbf{p}^\alpha)]^{-1}(\mathbf{k}^\alpha)),\gamma,\alpha) \times\psi([L^{-1}(\mathbf{p})\,AL\,(A^{-1}\mathbf{p}^\alpha)]^{-1}\mathbf{k}^\alpha, E-\mathbf{vp}, A^{-1}\mathbf{p}^\alpha), \tag{3.5}$$

where L is given by (3.1), and $\mathscr{L}$ by (A6). Particularization for P is not trivial here, either, because of the intricate nature of the element of the little group to be represented. The explicit calculation (Appendix B) gives

$$[U(P)\psi](\varphi,E,\mathbf{p}) = \delta_{\eta\epsilon_1\epsilon_2}(\varphi,\varphi_\mathbf{p})\,\psi(2\varphi_\mathbf{p}-\varphi,E,-\mathbf{p}), \tag{3.5a}$$

where $\mathbf{k}\in S^1_x$ is biunivocally represented by its azimuthal angle and the function $\delta_{\eta\epsilon_1\epsilon_2}(\varphi,\varphi_\mathbf{p})$ is defined in (B4).

IUR's *of* $\overline{\mathscr{G}}_+$ *from the orbits* Z_{00E}: Corresponding to the three series of IUR's of the little group, I, II, and III, we have here three series, called IV, $\mathrm{III_I}$, and $\mathrm{III_{II}}$, respectively.

$\mathrm{IV}(E,j,\epsilon_1,\epsilon_2,\epsilon)$: $E\in\mathbb{R}$, $2j\in\mathbb{N}$, $\epsilon_1,\epsilon_2,\epsilon\in\{1,-1\}$. The carrier space is $\mathbb{C}^{2j+1}$ and the operators are

$$U(\theta,b,\mathbf{a},\mathbf{v},A,\gamma,\alpha) = e^{iEb}\,\mathscr{D}_j(A)\,\Delta_{\epsilon_1\epsilon_2\epsilon}(\gamma,\alpha), \tag{3.6}$$

$$U(P) = \begin{cases} \epsilon & \text{if } \epsilon_1\cdot\epsilon_2 = 1, \\ i\epsilon & \text{if } \epsilon_1\cdot\epsilon_2 = -1. \end{cases} \tag{3.6a}$$

$\mathrm{III_I}(E,x,\epsilon_1,\epsilon_2,\epsilon)$: $E\in\mathbb{R}$, $x\in\mathbb{R}$, $x>0$, $\epsilon_1,\epsilon_2,\epsilon\in\{1,-1\}$. The carrier space is $\mathscr{L}^2(S^2_x\to\mathbb{C}, d\Omega_\mathbf{k})$ and the action

$$[U(\theta,b,\mathbf{a},\mathbf{v},A,\gamma,\alpha)\psi](\mathbf{k}) = \exp\{i(Eb+\mathbf{kv})\}\,\Delta_{\epsilon_1\epsilon_2\epsilon}(\gamma,\alpha)\,\psi(A^{-1}\mathbf{k}^\alpha), \tag{3.7}$$

and

$$[U(P)\psi](\mathbf{k}) = \begin{cases} \epsilon\psi(-\mathbf{k}) & \text{if } \epsilon_1\cdot\epsilon_2 = 1, \\ i\epsilon\psi(-\mathbf{k}) & \text{if } \epsilon_1\cdot\epsilon_2 = -1. \end{cases} \tag{3.7a}$$

$\mathrm{III_{II}}(E,x,h,\epsilon_1,\epsilon_2)$: $E\in\mathbb{R}$, $0<x\in\mathbb{R}$, $2h\in\mathbb{N}$, $\epsilon_1,\epsilon_2\in\{1,-1\}$. The support space is $\mathscr{L}^2(S^2_x\to\mathbb{C}^2, d\Omega_\mathbf{k})$. The operators themselves are

$$[U(\theta,b,\mathbf{a},\mathbf{v},A,\gamma,\alpha)\psi](\mathbf{k}) = \exp\{i(Eb+\mathbf{kv})\}\,D_{h\epsilon_1\epsilon_2}(L^{-1}(\mathbf{k}) \times AL\,(A^{-1}\mathbf{k}^\alpha),\gamma,\alpha)\,\psi(A^{-1}\mathbf{k}^\alpha). \tag{3.8}$$

For P, matters are similar to case $\mathrm{II_{II}}$, and here

$$[U(P)\psi](\mathbf{k}) = \begin{bmatrix} 0 & e^{2ih\varphi_\mathbf{k}} \\ (-1)^{2h}\epsilon_1\epsilon_2\,e^{-2ih\varphi_\mathbf{k}} & 0 \end{bmatrix}.$$

These seven kinds of representations given by (3.2)–(3.8) exhaust all IUR's of $\overline{\mathscr{G}}_+$. When restricted to the "projective covering" for the connected Galilei group, only types $\mathrm{II_{II}}$ and $\mathrm{III_{II}}$ reduce in a direct sum of two, whereas all others remain irreducible, so that an effect of including P is the mixing of helicities h and $-h$.

B. The irreducible semiunitary representations of $\overline{\mathscr{G}}$

A basis-free version of the well-known Wigner's process[26] of determining the ISUR's of a given group from the IUR's of the "unitary" subgroup (of index two) has been given by Shaw and Lever,[27] and we shall follow this method without further comments. Notice that Shaw and Lever study the more general problem of obtaining *multiplier* ISUR's; with our method multipliers do not appear because

TABLE I. Wigner types of $m(\rho, j, \epsilon_T, \epsilon_{PT}, \epsilon)$ representations of $\overline{\mathscr{G}}_+$.

$\epsilon_T \epsilon_{PT}$	ϵ_1	ϵ_2	$[U(T)]^2$	$[U(PT)]^2$	$U \sim U^{PT}$	$\mathbb{V}$	$\mathbb{V}^2$	Wigner type	Doubling
$+\ +$	$(-)^{2j}$	$(-)^{2j}$	$(-)^{2j}$	$(-)^{2j}$	$\sim$	$\mathscr{S}_j(\mathbb{S})$	$(-)^{2j}$	I	no
$+\ -$	$-(-)^{2j}$	$(-)^{2j}$	$(-)^{2j}$	$-(-)^{2j}$	$\nsim$	$-$	$-$	III	yes
$-\ +$	$(-)^{2j}$	$-(-)^{2j}$	$-(-)^{2j}$	$(-)^{2j}$	$\nsim$	$-$	$-$	III	yes
$-\ -$	$-(-)^{2j}$	$-(-)^{2j}$	$-(-)^{2j}$	$-(-)^{2j}$	$\sim$	$\mathscr{S}_j(\mathbb{S})$	$(-)^{2j}$	II	yes

of the use of a representation group.

The application of the method consists of, first, the selection of a fixed element $g_0 \in \overline{\mathscr{G}} - \overline{\mathscr{G}}_+$, second, the classification of IUR's of $\overline{\mathscr{G}}_+$ in Wigner types according to the relationship of between U and U^{g_0} where $U^{g_0}(h) = U(g_0^{-1} h\, g_0)$, and third, the construction, for each U, and depending on its type, of the associated semiunitary representation $\mathscr{U}$ (Ref. 27, Theorem B). We only quote the criterion of classification in Wigner types and the expressions for the associated SUR, $\mathscr{U}$, for the elements $h \in \overline{\mathscr{G}}_+$ and the element $g_0 \in \overline{\mathscr{G}} - \overline{\mathscr{G}}_+$. Of course, $\mathscr{U}(hg_0) = \mathscr{U}(h)\, \mathscr{U}(g_0)$ in all cases.

Type I: U and U^{g_0} are antiunitarily equivalent, and the antiunitary operator $\mathbb{V}$, realizing the equivalence, verifies $\mathbb{V}^2 = +U(g_0^2)$. In this case $\mathscr{U}(h) = U(h)$ and $\mathscr{U}(g_0) = \mathbb{V}$.

Type II: U and U^{g_0} are antiunitarily equivalent, and $\mathbb{V}^2 = -U(g_0^2)$. Here,

$$\mathscr{U}(h) = \begin{bmatrix} U(h) & 0 \\ 0 & U(h) \end{bmatrix}$$

while

$$\mathscr{U}(g_0) = \begin{bmatrix} 0 & \mathbb{V} \\ -\mathbb{V} & 0 \end{bmatrix}.$$

Type III: U and U^{g_0} are antiunitarily inequivalent. Now,

$$\mathscr{U}(h) = \begin{bmatrix} U(h) & 0 \\ 0 & \mathbb{K}U^{g_0}(h)\mathbb{K}^{-1} \end{bmatrix}$$

and

$$\mathscr{U}(g_0) = \begin{bmatrix} 0 & U(g_0^2)\,\mathbb{K}^{-1} \\ \mathbb{K} & 0 \end{bmatrix},$$

where $\mathbb{K}$ is an arbitrary but fixed antiunitary operator. We will choose the complex conjugation: in this case $\mathbb{K} = \mathbb{K}^{-1}$.

Different selections of the arbitrary element $g_0 \in \overline{\mathscr{G}} - \overline{\mathscr{G}}_+$ would lead to equivalent results. There is a particularly suitable choice, namely $g_0 = PT = (0,0,0,0,I,1,PT)$. In this case the automorphism $h \to g_0^{-1} h g_0$ of $\overline{\mathscr{G}}_+$ is given by $(\theta, b, \mathbf{a}, \mathbf{v}, A, \gamma, \alpha) \to (-\theta, -b, -\mathbf{a}, \mathbf{v}, A, \gamma, \alpha)$, that is to say, PT acts as the inversion on $\overline{T}_4$ and as the identity on $\overline{K}_+$ [for the calculation notice that $g_0^{-1} = (0,0,0,0,I,\mu,PT)$]. Then U and U^{PT} differ systematically through a complex conjugation in the characters and coincide in the corresponding little group representation. Each antiunitary operator can be factorized as the product of a fixed antiunitary invertible operator (just we take $\mathbb{K}$) and some unitary operator V, say $\mathbb{V} = V{\cdot}\mathbb{K}$, so that antiunitary equivalence/inequivalence of U and U^{PT} translates to unitary equivalence of U and $\mathbb{K}U^{PT}\mathbb{K}$. But U and $\mathbb{K}U^{PT}\mathbb{K}$ coincide in the characters as well as in the transformation of the arguments due to our particular choice; the investigation about the Wigner type of each U is very easy because U and $\mathbb{K}U^{PT}\mathbb{K}$ differ only in the little group representation, say, D and $\mathbb{K}D\mathbb{K}$. Notice that $U_{\cdots\epsilon_1\epsilon_2\cdots}(PT)^2 = \epsilon_1$ in every representation.

In order to give a unified form to the following results, the substitution of ϵ_1, ϵ_2, by a new set of indces ϵ_T and ϵ_{PT} turns out to be convenient. The relation between both pairs is

$$\epsilon_T = (-)^{2j}[U_{\cdots j\epsilon_1\epsilon_2\cdots}(T)]^2 = (-)^{2j}\epsilon_2,$$

$$\epsilon_{PT} = (-)^{2j}[U_{\cdots j\epsilon_1\epsilon_2\cdots}(PT)]^2 = (-)^{2j}\epsilon_1,$$

where j is the spinlike index of the representation, that is, $j = j$ for m representations, $j = h$ in II_{11} and III_{11}, and $j = 0$ otherwise. From now on we shall use ϵ_T and ϵ_{PT} as representation indices while ϵ_1, ϵ_2 will be understood to be auxiliary variables, defined in the representation $U_{\cdots j\epsilon_1\epsilon_2\cdots}$ by $\epsilon_1 = (-)^{2j}\epsilon_{PT}$ and $\epsilon_2 = (-)^{2j}\epsilon_T$. The couple $(\epsilon_T, \epsilon_{PT})$ will be called the type of the corresponding ISUR, $\mathscr{U}$, of $(\overline{\mathscr{G}}, \overline{\mathscr{G}}_+)$ and it is to not be confused with the Wigner type of U.

Now we shall study each case separately.

$m(\rho, j, \epsilon_T, \epsilon_{PT}, \epsilon)$: Here the representation $\mathscr{D}_j \otimes \Delta_{\epsilon_1\epsilon_2\epsilon}$ and $\mathscr{D}_j^* \otimes \Delta^*_{\epsilon_1\epsilon_2\epsilon}$ are to be compared. They are unitarily

TABLE II. Wigner types of $\mathrm{II}_{11}(\rho, h, \epsilon_T, \epsilon_{PT}, \epsilon)$ representations of $\overline{\mathscr{G}}_+$.

ϵ_T	ϵ_{PT}	ϵ_1	ϵ_2	$[U(T)]^2$	$[U(PT)]^2$	V	$\mathbb{V}^2$	Wigner type	Doubling
$+$	$+$	$(-)^{2h}$	$(-)^{2h}$	$(-)^{2h}$	$(-)^{2h}$	$\begin{bmatrix} 0 & 1 \\ (-)^{2h} & 0 \end{bmatrix}$	$(-)^{2h}$	I	no
$+$	$-$	$-(-)^{2h}$	$(-)^{2h}$	$(-)^{2h}$	$-(-)^{2h}$	$\begin{bmatrix} 0 & 1 \\ -(-)^{2h} & 0 \end{bmatrix}$	$-(-)^{2h}$	I	no
$-$	$+$	$(-)^{2h}$	$-(-)^{2h}$	$-(-)^{2h}$	$(-)^{2h}$	$\begin{bmatrix} 0 & 1 \\ -(-)^{2h} & 0 \end{bmatrix}$	$-(-)^{2h}$	II	yes
$-$	$-$	$-(-)^{2h}$	$-(-)^{2h}$	$-(-)^{2h}$	$-(-)^{2h}$	$\begin{bmatrix} 0 & 1 \\ (-)^{2h} & 0 \end{bmatrix}$	$(-)^{2h}$	II	yes

TABLE III. Wigner types of $\mathrm{I}(\rho,x,\eta,\epsilon_T,\epsilon_{PT})$ representations of $\overline{\mathscr{G}}_+$.

η	ϵ_T	ϵ_{PT}	ϵ_1	ϵ_2	$[U(T)]^2$	$[U(PT)]^2$	$U\sim U^{PT}V$		$\mathbb{V}^2$	Wigner type	Doubling
$+1$	$+$	$+$	$+$	$+$	$+$	$+$	$\sim$	π_x	1	I	no
$+1$	$+$	$-$	$-$	$+$	$+$	$-$	$\sim$	π_x	1	II	yes
$+1$	$-$	$+$	$+$	$-$	$-$	$+$	$\sim$	π_x	1	I	no
$+1$	$-$	$-$	$-$	$-$	$-$	$-$	$\sim$	π_x	1	II	yes
$\pm i$	ϵ_T	ϵ_{PT}	ϵ_{PT}	ϵ_T	ϵ_T	ϵ_{PT}	$\nsim$	$-$	$-$	III	yes

equivalent iff $\epsilon_1\epsilon_2=\epsilon_T\,\epsilon_{PT}=1$, with $V=\mathscr{D}_j(\mathbb{S})$, while they are inequivalent if $\epsilon_1\,\epsilon_2=\epsilon_T\,\epsilon_{PT}=-1$. The situation is depicted in Table I, where $\sim$ means antiunitary equivalence.

$\mathrm{II_I}(\rho,\epsilon_T,\epsilon_{PT},\epsilon)$: The comparison being between $\Delta_{\epsilon_1\epsilon_2\epsilon}$ and $\Delta^*_{\epsilon_1\epsilon_2\epsilon}$, the results are the same ones as those obtained in the m-case with $j=0$.

$\mathrm{II_{II}}(\rho,h,\epsilon_T,\epsilon_{PT},\epsilon)$: In this case the representations $D_{h\epsilon_1\epsilon_2}$ and $D^*_{h\epsilon_1\epsilon_2}$ are always equivalent; we can take

$$\mathrm{V}=\begin{bmatrix}0 & 1\\ (-)^{2h}\,\epsilon_T\,\epsilon_{PT} & 0\end{bmatrix},$$

then $\mathbb{V}^2=(-)^{2h}\,\epsilon_T\,\epsilon_{PT}$. The results are given in Table II.

$\mathrm{I}(\rho,x,\eta,\epsilon_T,\epsilon_{PT})$: If $\eta=+i$ or $\eta=-i$ the representations $D_{x\eta\epsilon_T\epsilon_{PT}}$ and $\mathbb{K}D_{x\eta\,\epsilon_T\epsilon_{PT}}\mathbb{K}$ are not unitarily equivalent; on the other hand, when $\eta=+1$ or $\eta=-1$ they are equivalent with V being the operator Π_x defined by $[\Pi_x f](\mathbf{k})=f(-\mathbf{k})$. In this case $\mathbb{V}^2=I$ (see Table III).

$\mathrm{IV}(j,\epsilon_T,\epsilon_{PT},\epsilon)$: The results are identical to those obtained for the case $m(\rho,j,\epsilon_T,\epsilon_{PT},\epsilon)$

$\mathrm{III_I}(E,x,\epsilon_T,\epsilon_{PT},\epsilon)$: The results are the same ones as in the case $\mathrm{II_I}(\rho,\epsilon_T,\epsilon_{PT},\epsilon)$.

$\mathrm{III_{II}}(E,x,h,\epsilon_T,\epsilon_{PT})$: The results are identical to those obtained for the case $\mathrm{II_{II}}(\rho,h,\epsilon_T,\epsilon_{PT})$.

C. Explicit form of the irreducible semiunitary representations of $(\overline{\mathscr{G}},\overline{\mathscr{G}}_+)$

With the preceding results we are able to give the unitary equivalence classes of the irreducible semiunitary representations $\mathscr{U}$ of $(\overline{\mathscr{G}},\overline{\mathscr{G}}_+)$ explicitly. Taking into account the relation $(\theta,b,\mathbf{a},\mathbf{v},A,\gamma,\alpha)=(\theta,b,\mathbf{a},\mathbf{v},A,\gamma,1)\cdot(0,0,0,0,I,1,\alpha)$, it will be sufficient to give the unitary/antiunitary operators $\mathscr{U}(\alpha)$ because $\mathscr{U}(\theta,b,\mathbf{a},\mathbf{v},A,\gamma,1)$ will be given by

$$\mathscr{U}(\theta,b,\mathbf{a},\mathbf{v},A,\gamma,1)=\begin{cases}U(\theta,b,\mathbf{a},\mathbf{v},A,\gamma,1) & \text{if } U \text{ belongs to type I}\\ (U\oplus U)(\theta,b,\mathbf{a},\mathbf{v},A,\gamma,1) & \text{if } U \text{ belongs to type II}\\ (U\oplus U^{PT})(\theta,b,\mathbf{a},\mathbf{v},A,\gamma,1) & \text{if } U \text{ belongs to type III}\end{cases}$$

More details are shown in Table IV.

Notice that the element $(0,0,0,0,I,1,T)$ can be factorized as $(0,0,0,0,I,1,T)=(0,0,0,0,I,\mu\nu,P)\cdot(0,0,0,0,I,1,PT)$ and hence $\mathscr{U}(T)=\mathscr{U}(\mu\nu,P)\cdot\mathscr{U}(PT)$. Furthermore, P is in $\bar{K}_+$ and then $U^{PT}(P)=U(P)$, so that in the following table we have omitted the ineffective PT in $\mathscr{U}(P)$ for type III representations.

The application of the process on every representation leads to the results displayed inTable V. The operator $\mathbb{K}$ is the complex conjugation $(\mathbb{K}f)(x)=f^*(x)$ for the functions defined in each orbit.

Finally we may state our result. All classes of unitary equivalence of ISUR's of $(\overline{\mathscr{G}},\overline{\mathscr{G}}_+)$ are contained in Table V. We refer to all preceding results (3.2)–(3.8) for the explicit expression of a representation in each class.

4. THE SEMIUNITARY PROJECTIVE REPRESENTATIONS OF $(\mathscr{G},\mathscr{G}_+)$

Each ISUPR of $(\mathscr{G},\mathscr{G}_+)$ can be lifted to some ISUR of $(\overline{\mathscr{G}},\overline{\mathscr{G}}_+)$, but two SUR's of $(\overline{\mathscr{G}},\overline{\mathscr{G}}_+)$ can give rise to unitarily projective equivalent ISUPR's of $(\mathscr{G},\mathscr{G}_+)$. So, for a classification of the unitary projective equivalence classes of ISUPR's of $(\mathscr{G},\mathscr{G}_+)$ we must answer to the following question: when do two ISUR's of $(\overline{\mathscr{G}},\overline{\mathscr{G}})$ lead to unitarily projective equivalent ISUPR's of $(\mathscr{G},\mathscr{G}_+)$?

From the practical viewpoint, the answer to this question is contained in the following lemma, whose easy verification is left to the reader.

Lemma: For the unitary projective equivalence of two ISUPR's of $(\mathscr{G},\mathscr{G}_+)$ obtained from two ISUR's $\mathscr{U}$ and $\mathscr{U}'$ of $(\overline{\mathscr{G}},\overline{\mathscr{G}}_+)$, a necessary and sufficient condition is the unitary equivalence of $\mathscr{U}$ and $\Gamma\otimes\mathscr{U}'$ for some continuous crossed homomorphism Γ.

The group $Z^1_*(\overline{\mathscr{G}},\mathbf{T})$ of all continuous crossed homomorphisms $\overline{\mathscr{G}}\to\mathbf{T}$ is easily calculated. It is closely related to the set of semiunitary one dimensional representations of $(\overline{\mathscr{G}},\overline{\mathscr{G}}_+)$: it is obtained by removing a complex conjugation from the latter. We have $Z^1_*(\overline{\mathscr{G}},\mathbf{T})\approx\mathbf{T}\otimes\mathbb{R}\otimes C_2$. The crossed homomorphism $(e^{i\omega},E,\epsilon)$ with $E\in R$, $\epsilon\in\{1,-1\}$ is given by $\Gamma(\theta,b,\mathbf{a},\mathbf{v},A,\gamma,1)=e^{iEb}$, $\Gamma(P)=\epsilon$, $\Gamma(T)=e^{i\omega}\epsilon$,

TABLE IV. Induction of a ISUR $\mathscr{U}$ of $(\overline{\mathscr{G}},\overline{\mathscr{G}}_+)$ from a IUR U of $\overline{\mathscr{G}}_+$.

$[U]$	Wigner Type	$\mathscr{U}(P)$	$\mathscr{U}(T)$	$\mathscr{U}(PT)$
$U\overset{\mathbb{K}}{\sim}U$	I $\mathbf{V}^2=[U(PT)]^2$	$U(P)$	$U(\mu\nu)U(P)\mathbf{V}$	$\mathbf{V}$
$U^{\mathbb{K}}=\mathbf{V}^{-1}U\mathbf{V}$	II $\mathbf{V}^2=-[U(PT)]^2$	$\begin{pmatrix}1&0\\0&1\end{pmatrix}U(P)$	$\begin{pmatrix}0&1\\-1&0\end{pmatrix}U(\mu\nu,P)\mathbf{V}$	$\begin{pmatrix}0&1\\-1&0\end{pmatrix}\mathbf{V}$
$\{U,U^{\mathbb{K}}\}$	III	$\begin{pmatrix}U(P)&0\\0&\mathbf{K}U(P)\mathbf{K}\end{pmatrix}$	$\begin{pmatrix}0&U(\nu,P)\\U(\mu\nu)\mathbf{K}U(P)\mathbf{K}&0\end{pmatrix}\mathbf{K}$	$\begin{pmatrix}0&U(\mu)\\1&0\end{pmatrix}\mathbf{K}$

TABLE V. The ISUR's of $(\mathscr{G}, \mathscr{G}_{+})$

Representation class	$\{U\}$ or $\{U, U^{\aleph_0}\}$	$\mathscr{U}(P)$	$\mathscr{U}(T)$	$\mathscr{U}(PT)$	Observations
	$m\rho j++\epsilon$	$\epsilon\Pi_m$	$\epsilon\mathscr{D}_j(\mathbb{S})\Pi_m\mathbf{K}$	$\mathscr{D}_j(\mathbb{S})\mathbf{K}$	
	$m\rho j+-+;\ m\rho j+--$	$\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}i\Pi_m$	$\begin{bmatrix}0 & (-)^{2j}\\ 1 & 0\end{bmatrix}i\Pi_m\mathbf{K}$	$\begin{bmatrix}0 & (-)^{2j}\\ 1 & 0\end{bmatrix}\mathbf{K}$	
$m(\rho,j,\epsilon_T,\epsilon_{PT},\epsilon)$	$m\rho j-++;m\rho j-+-$	$\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}i\Pi_m$	$\begin{bmatrix}0 & (-)^{2j}\\ 1 & 0\end{bmatrix}i\Pi_m\mathbf{K}$	$\begin{bmatrix}0 & (-)^{2j}\\ 1 & 0\end{bmatrix}\mathbf{K}$	$[\Pi_m f](\mathbf{p}) = f(-\mathbf{p})$
	$m\rho j--\epsilon$	$\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\epsilon\Pi_m$	$\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}\epsilon\mathscr{D}_j(\mathbb{S})\Pi_m\mathbf{K}$	$\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}\mathscr{D}_j(\mathbb{S})\mathbf{K}$	$\mathbb{S} = \begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}$
	$\rho++\epsilon$	$\epsilon\Pi_{\rm II}$	$\epsilon\Pi_{\rm II}\mathbf{K}$	$\mathbf{K}$	
$\mathrm{II}_{\rm I}(\rho,\epsilon_T,\epsilon_{PT},\epsilon)$	$\rho+-+;\rho+--$	$\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}i\Pi_{\rm II}$	$\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}i\Pi_{\rm II}\mathbf{K}$	$\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}\mathbf{K}$	$\Pi_{\rm II} f(E,\mathbf{p})$
	$\rho-++;\rho-+-$	$\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}i\Pi_{\rm II}$	$\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}i\Pi_{\rm II}\mathbf{K}$	$\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\mathbf{K}$	$= f(E,-\mathbf{p})$
	$\rho--\epsilon$	$\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\epsilon\Pi_{\rm II}$	$\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}\epsilon\Pi_{\rm II}\mathbf{K}$	$\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}\mathbf{K}$	
	$\rho h++$	$\begin{bmatrix} & E_h^+\\ (-)^{2h}E_h & \end{bmatrix}$	$\begin{bmatrix}E_h^+ & \\ & E_h\end{bmatrix}(-)^{2h}\mathbf{K}$	$\begin{bmatrix} & 1\\ (-)^{2h} & \end{bmatrix}\mathbf{K}$	$[E_h^+ f](E,\mathbf{p})$
	$\rho h+-$	$\begin{bmatrix} & E_h^+\\ -(-)^{2h}E_h^- & \end{bmatrix}$	$\begin{bmatrix}E_h^+ & \\ & E_h^-\end{bmatrix}(-)^{2h}\mathbf{K}$	$\begin{bmatrix} & 1\\ -(-)^{2h} & \end{bmatrix}\mathbf{K}$	$= e^{\pm 2ih\varphi_{\mathbf{p}}} f(E,-\mathbf{p})$
$\mathrm{II}_{\rm II}(\rho,h,\epsilon_T,\epsilon_{PT})$					
	$\rho h-+$	$\begin{bmatrix} & E_h^+ & & \\ -(-)^{2h}E_h^- & & & \\ & & & E_h^+\\ & & -(-)^{2h}E_h^- & \end{bmatrix}$	$\begin{bmatrix} & & E_h^+ & \\ & & & E_h\\ -E_h^+ & & & \\ & -E_h^- & & \end{bmatrix}(-)^{2h}\mathbf{K}$	$\begin{bmatrix} & & & 1\\ & & -(-)^{2h} & \\ & -1 & & \\ (-)^{2h} & & & \end{bmatrix}\mathbf{K}$	
	$\rho h--$	$\begin{bmatrix} & E_h^+ & & \\ (-)^{2h}E_h & & & \\ & & & E_h^+\\ & & (-)^{2h}E_h & \end{bmatrix}$	$\begin{bmatrix} & & E_h^+ & \\ & & & E_h\\ -E_h^+ & & & \\ & -E_h^- & & \end{bmatrix}(-)^{2h}\mathbf{K}$	$\begin{bmatrix} & & & 1\\ & & (-)^{2h} & \\ & -1 & & \\ -(-)^{2h} & & & \end{bmatrix}\mathbf{K}$	
	$\rho x\eta++$	$\eta\Pi_t$	$\eta\Pi_t\Pi_x\mathbf{K}$	$\Pi_x\mathbf{K}$	$[\Pi_t f](\varphi_{\mathbf{k}},E,\mathbf{p})$
	$\rho x\eta+-$	$\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\eta S$	$\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}\eta S\Pi_x\mathbf{K}$	$\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}\Pi_x\mathbf{K}$	$= f(2\varphi_{\mathbf{p}}-\varphi_{\mathbf{k}},E,-\mathbf{p})$
	$\eta=\pm 1$				
	$\rho x\eta-+$	ηS	$-\eta S\Pi_x\mathbf{K}$	$\Pi_x\mathbf{K}$	$[\Pi_x f](\varphi_{\mathbf{k}},E,\mathbf{p})$
	$\rho x\eta--$	$\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\eta\Pi_t$	$\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}\eta\Pi_t\Pi_x\mathbf{K}$	$\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}\Pi_x\mathbf{K}$	$= f(\varphi_{\mathbf{k}}+\pi,E,\mathbf{p})$
$\mathrm{I}(\rho,x,\eta,\epsilon_T,\epsilon_{PT})$					
	$\rho,x,i,+,+;\rho,x,-i,+,+$	$\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}iS$	$\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}iS\mathbf{K}$	$\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\mathbf{K}$	$[Sf](\varphi_{\mathbf{k}},E,\mathbf{p})$
	$\rho,x,i,+,-;\rho,x,-i,+,-$	$\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}i\Pi_t$	$\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}i\Pi_t\mathbf{K}$	$\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}\mathbf{K}$	$= \Pi(\varphi_{\mathbf{k}},\varphi_{\mathbf{p}})f(2\varphi_{\mathbf{p}}-\varphi_{\mathbf{k}},$
	$\rho,x,i,-,+;\rho,x,-i,-,+$	$\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}i\Pi_t$	$\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}i\Pi_t\mathbf{K}$	$\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\mathbf{K}$	$E,-\mathbf{p})$
	$\rho,x,i,-,-;\rho,x,-i,-,-$	$\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}iS$	$\begin{pmatrix}0 & -1\\ -1 & 0\end{pmatrix}iS\mathbf{K}$	$\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}\mathbf{K}$	
$\mathrm{IV}\,(E,j,\epsilon_T,\epsilon_{PT},\epsilon)$	Identical expressions to those of class $m\ (\rho,j,\epsilon_T,\epsilon_{PT},\epsilon)$ with $\Pi_m\to\Pi_{\rm IV}$				$\Pi_{\rm IV} = \mathrm{Id}_{\mathbb{C}^{2j+1}}$
$\mathrm{III}_{\rm I}(E,x,\epsilon_T,\epsilon_{PT},\epsilon)$	Identical expressions to those of classes				$[\Pi_{\rm III} f](\mathbf{k})i = f(-\mathbf{k})$
$\mathrm{III}_{\rm II}(E,x,h,\epsilon_T,\epsilon_{PT})$	$\begin{Bmatrix}\mathrm{II}_{\rm I}(E,x,\epsilon_T,\epsilon_{PT},\epsilon)\\ \mathrm{II}_{\rm II}(E,x,h,\epsilon_T,\epsilon_{PT})\end{Bmatrix}$ with $\begin{Bmatrix}\Pi_{\rm II}\to\Pi_{\rm III}\\ E_h^{\pm}\to F_h^{\pm}\end{Bmatrix}$				$[F_h^{\pm} f](\mathbf{k}) = e^{\pm 2ih\varphi_{\mathbf{k}}} f(\mathbf{k})$

$\Gamma(PT) = e^{i\omega}$.

The crossed homomorphisms $(e^{i\omega},0,1)$ are principal homomorphisms, so that only those of the form $(1, E, \epsilon)$ are effective in making unitarily projectively equivalent two ISUPR's of $(\mathcal{G}, \mathcal{G}_+)$ obtained from two unitarily inequivalent ISUR's, $\mathcal{U}$ and $\mathcal{U}'$ of $(\overline{\mathcal{G}}, \overline{\mathcal{G}}_+)$.

All these crossed homomorphisms are generated by $\Gamma_{E,1}$ and $\Gamma_{0,-1}$; we will study for each class $\mathcal{U}$ of ISUR's of $(\overline{\mathcal{G}}, \overline{\mathcal{G}}_+)$ the products $\Gamma_{E,1} \otimes \mathcal{U}$ and $\Gamma_{0,-1} \otimes \mathcal{U}$ as well as giving the operators W realizing the equivalence $\Gamma_{E,\epsilon} \otimes \mathcal{U} = W^{-1}\mathcal{U}'W$, when $\mathcal{U}'$ and $\Gamma_{E,\epsilon} \otimes \mathcal{U}$ are equivalent.

$m(\rho, j, \epsilon_T, \epsilon_{PT}, \epsilon)$: Here the representation $\mathcal{D}_j \otimes \Delta_{\epsilon_1\epsilon_2\epsilon}$ and $\mathcal{D}_j^* \otimes \Delta^*_{\epsilon_1\epsilon_2\epsilon}$ are to be compared. They are unitarily equivalent iff $\epsilon_1\epsilon_2 = \epsilon_T\, \epsilon_{PT} = 1$, with $V = \mathcal{D}_j(\mathbb{S})$, while they are inequivalent if $\epsilon_1\, \epsilon_2 = \epsilon_T\, \epsilon_{PT} = -1$. The situation is depicted in Table I, where $\sim a$ means antiunitary equivalence. $\Gamma_{0,-1} \otimes m(\rho, j, \pm, \pm, \epsilon) = m(\rho, j, \pm, \pm, -\epsilon)$ and hence the eventual index ϵ also appears when considering projective equivalence. All other indices remain because of the relations $\Gamma_{0,-1} \otimes m(\rho, j, +, -) \approx m(\rho, j, +, -)$ and $\Gamma_{0,-1} \otimes m(\rho, j, -, +) \approx m(\rho, j, -, +)$; the W's are equal to

$$\begin{bmatrix} 0 & -(-)^{2j} \\ 1 & 0 \end{bmatrix},$$

and

$$\begin{bmatrix} 0 & (-)^{2j} \\ 1 & 0 \end{bmatrix},$$

respectively. The corresponding ISUPR's of $(\mathcal{G}, \mathcal{G}_+)$ are characterized by the indices $m(j, \epsilon_T, \epsilon_{PT})$ and will be denoted as $\overline{m}(j, \epsilon_T, \epsilon_{PT})$.

$\mathrm{II_I}(\rho, \epsilon_T, \epsilon_{PT}, \cdots)$: Now $\Gamma_{E',1} \otimes \mathrm{II_I}(\rho, \epsilon_T, \epsilon_{PT}, \cdots) \approx \mathrm{II_I}(\rho, \epsilon_T, \epsilon_{PT}, \cdots)$, with W being given by $[W_{E'} f](E, \mathbf{p}) = f(E + E', \mathbf{p})$. For $\Gamma_{0,-1}$ the situation is similar to that of the preceding case, with $j = 0$. We obtain the projective representations labelled as follows: $\overline{\mathrm{II_I}}\ (\rho, \epsilon_T, \epsilon_{PT})$.

$\mathrm{II_{II}}(\rho, h, \epsilon_T, \epsilon_{PT})$: for $\Gamma_{E,0}$ the situation is analogous to the case $\mathrm{II_I}$. Now $\Gamma_{0,-1} \otimes \mathrm{II_{II}}(\rho, h, \epsilon_T, \epsilon_{PT}) \approx \mathrm{II_{II}}(\rho, h, \epsilon_T, \epsilon_{PT})$ with

$$W = \begin{bmatrix} ia & 0 \\ 0 & -ia \end{bmatrix}, \quad a \in \mathbb{R} \ \text{ for } \epsilon_T = +1$$

and

$$W = \begin{bmatrix} \alpha\sigma_3 & \beta\sigma_3 \\ \beta^*\sigma_3 & -\alpha^*\sigma_3 \end{bmatrix}, \quad \alpha, \beta \in \mathbb{C} \ \text{ for } \epsilon_T = -1.$$

We obtain the projective representations $\overline{\mathrm{II_{II}}}\ (\rho, h, \epsilon_T, \epsilon_{PT})$.

$\mathrm{I}(\rho, x, \eta, \epsilon_T, \epsilon_{PT})$: The operator $W_{E'}$ defined by $[W_{E'} f](\varphi_{\mathbf{k}}, E, \mathbf{p}) = f(\varphi_{\mathbf{k}}, E + E', \mathbf{p})$ realizes the equivalence between $\Gamma_{E',1} \otimes \mathrm{I}(\rho, x, \eta, \epsilon_T, \epsilon_{PT})$ and $I(\rho, x, \eta, \epsilon_T, \epsilon_{PT})$. For remaining crossed homomorphism $\Gamma_{0,-1}$ we have the following results: a) when $\epsilon = \eta = \pm i$, $\Gamma_{0,-1} \otimes \mathrm{I}(\rho, x, \epsilon, \epsilon_T, \epsilon_{PT}) = \mathrm{I}(\rho, x, -\epsilon, \epsilon_T, \epsilon_{PT})$, so that $\eta = 1$ and $\eta = -1$ lead to projectively equivalent ISUR's of $(\mathcal{G}, \mathcal{G}_+)$; b) when $\eta = \pm i$, then $\Gamma_{0,-1} \otimes \mathrm{I}(\rho, x, \eta, \epsilon_T, \epsilon_{PT}) \approx \mathrm{I}(\rho, x, \eta, \epsilon_T, \epsilon_{PT})$ with

$$W = \begin{bmatrix} 0 & -a + ib \\ a + ib & 0 \end{bmatrix}, \quad a, b \in \mathbb{R} \text{for } \epsilon_{PT} = -1.$$

and

$$W = \begin{bmatrix} 0 & -ai + b \\ ai + b & 0 \end{bmatrix}, \ a, b \in \mathbb{R} \text{for } \epsilon_{PT} = +1.$$

The projective equivalence classes of ISUPR's of $(\mathcal{G}, \mathcal{G}_+)$ will be denoted $\overline{\mathrm{I}}(\rho, x, \eta, \epsilon_T, \epsilon_{PT})$ with $\eta = 1, i$.

$\mathrm{IV}(E, j, \epsilon_T, \epsilon_{PT}, \cdots)$: Now $\Gamma_{E',1} \otimes \mathrm{IV}(E, j, \epsilon_T, \epsilon_{PT}, \cdots) = \mathrm{IV}(E + E', j, \epsilon_T, \epsilon_{PT}, \cdots)$. For $\Gamma_{0,-1}$ we obtain the same situation as in the case $m(\rho, j, \epsilon_T, \epsilon_{PT}, \cdots)$. The ISUPR's obtained will be denoted $\overline{\mathrm{IV}}(j, \epsilon_T, \epsilon_{PT})$.

$\mathrm{III_I}(E, x, \epsilon_T, \epsilon_{PT}, \cdots)$ and $\mathrm{III_{II}}(E, x, h, \epsilon_T, \epsilon_{PT})$: The product by $\Gamma_{E',1}$ "shifts" by E', as in the preceding case. For $\Gamma_{0,-1}$ we have the same results as in case $\mathrm{II_I}$ and $\mathrm{II_{II}}$. The corresponding ISUPR's of $(\mathcal{G}, \mathcal{G}_+)$ are $\overline{\mathrm{III_I}}(x, \epsilon_T, \epsilon_{PT})$ and $\overline{\mathrm{III_{II}}}\ (x, h, \epsilon_T, \epsilon_{PT})$.

Finally, we give in Table VI the classes of projective equivalence of ISUPR's of $(\mathcal{G}, \mathcal{G}_+)$. However, we must remind you that the knowledge of all its irreducible components up to projective equivalence such as given in the first column of Table VI is not sufficient for a complete specification of a decomposable SUPR $\mathcal{V}$ of $(\mathcal{G}, \mathcal{G}_+)$. The use of a representation group permits us to enounce this fact in a clear way: let $\overline{\mathcal{V}}$ be a SUR of $(\overline{\mathcal{G}}, \overline{\mathcal{G}}_+)$ lifting $\mathcal{V}$. Then $\overline{\mathcal{V}}$ (and hence $\mathcal{V}$) can be completely specified (up to equivalence) by giving all its irreducible constituents (which are given in Table V). Hence in Table VI we have paid attention to the disappearing indices which act as "relative phases" between the different irreducible constituents of a given decomposable SUPR of $(\mathcal{G}, \mathcal{G}_+)$. From this viewpoint the role of the internal energy was stressed by Lévy-Leblond.[18] We point out that when inversions are taken into account there are in some cases indices (like ϵ) which appear as relative parities and therefore they are physically relevant for nonelementary systems.

5. COMMENTS AND CONCLUSIONS

First of all, let us remark that in spite of Bargmann's

TABLE VI. The irreducible semiunitary projective representations of $(\mathcal{G}, \mathcal{G}_+)$.

Projective equivalence class	Disappearing indices
$\overline{m}(j, \epsilon_T, \epsilon_{PT})$	ρ, ϵ when $\epsilon_T\, \epsilon_{PT} = 1$
$\overline{\mathrm{I}}(\rho, x, \eta, \epsilon_T, \epsilon_{PT})$, $\eta = 1, i$	η when η is real
$\overline{\mathrm{II_I}}\ (\rho, \epsilon_T, \epsilon_{PT})$	ϵ when $\epsilon_T\, \epsilon_{PT} = 1$
$\overline{\mathrm{II_{II}}}\ (\rho, h, \epsilon_T, \epsilon_{PT})$	
$\overline{\mathrm{III_I}}\ (x, \epsilon_T, \epsilon_{PT})$	ϵ when $\epsilon_T\, \epsilon_{PT} = 1$
$\overline{\mathrm{III_{II}}}\ (x, h, \epsilon_T, \epsilon_{PT})$	
$\overline{\mathrm{IV}}(j, \epsilon_T, \epsilon_{PT})$	E, ϵ when $\epsilon_T\, \epsilon_{PT} = 1$

comment (Ref. 2, p. 2) the problem of adjoining the inversions to a given kinematical group (either Poincaré, Galilei, or any other) in order to study its ISUPR's is far from trivial. In fact, the first systematic study of this problem in the case of Poincaré group $\mathscr{P}$ was carried out by Wigner[10] by a direct method which is rather involved (essentially an induction from the representations of the connected proper group $\mathscr{P}_0$). As Wigner said "the amount of computation that was necessary··· (only in order to include P)···is surprising. I do not know how this calculation could be simplfied." The same problem was later studied using, in all cases, various auxiliary groups by Shirokov,[28] Parthasarathy[29], and Shaw and Lever.[12] The Galilei group plus inversions was also studied by Brennich.[14]

The new method we have employed needs only one auxiliary group (the representation group) and the reduction of the projective problem allows a full use of our background knowledge of the theory of linear representations, making unnecessary the direct study of the projective or multiplier case. The Poincaré group has also been studied by us from this viewpoint.[24]

Brennich's interest was limited to the physical classes, m and II representations. We feel it useful to spend a little time in comparing his methods and results with ours. Brennich's construction of the ISUPR's of $(\mathscr{G}, \mathscr{G}_+)$ proceeds via the multiplier representations of an auxiliary group which is introduced *a priori*. This group, called by him $\widetilde{\text{FIGG}}$, is arbitrarily selected to be one of the eight candidates to universal covering of $\mathscr{G}$ (see also Ref. 13 in connection with this for the case of $\mathscr{P}$). The group $\widetilde{\text{FIGG}}$ is only the factor group of our $\overline{\mathscr{G}}$ by the subgroup $\{(\theta, \gamma)\}$ corresponding to only a part of the kernel $\widehat{H}^2_*(\mathscr{G}, \mathbf{T})$. Thus, Breenich's method only "weakens" the projective character of the pertinent representations because of $\widetilde{\text{FIGG}}$ being in some sense an "intermediate group" between $\mathscr{G}$ and $\overline{\mathscr{G}}$.

The multiplier representations of $\widetilde{\text{FIGG}}$ have factor systems which have lost the part $\xi_l(R', R)$, because of the replacement of SO(3) by SU(2). In our method this replacement follows as a natural consequence, and furthermore, it is clear that the extension from $\mathscr{G}$ to $\overline{\mathscr{G}}$ is a minimal one in order to make the projective character of the representations fully disappear. It has also been shown elsewhere[7] that the transition from some connected Lie groups to their representation groups does not always reduce to the replacement of the group for is universal covering group, although this replacement works for some splitting groups which are easier to handle than representation groups.

Our intention has been to show how the method of obtaining all ISUPR's from a representation group works in an interesting but not fully studied case. To carry out the comparison of the representations in classes m and II, we intend our representation to be multiplier representations of $\widetilde{\text{FIGG}}$ by means of the section $\widetilde{\text{FIGG}} \to \overline{\mathscr{G}}$ given by $(U,\mathbf{v};\mathbf{a},\ddot{a};\epsilon_s,\epsilon_t) \to (0, a, \mathbf{a}, \mathbf{v}, U, I, \alpha)$, where on the left-hand side Brennich's notation is used. On the right-hand side we use those of this paper; α is the inversion corresponding to the pair (ϵ_s, ϵ_t).

Then the representation we have denoted $m(\cdots j, \epsilon_T, \epsilon_{PT}, \cdots)$ (resp. $\mathrm{II}(\cdots, h, \epsilon_T, \epsilon_{PT})$ is pseudoequivalent[8] to Brennich's one $(\cdots j)_m^{x\epsilon}$ (resp. $(\cdots h)_0^{x\epsilon}$, the linking of the notations being $\epsilon \longleftrightarrow (-1)^{2j}\epsilon_T$, $x \leftrightarrow \epsilon_T \epsilon_{PT}$.

Brennich's multiplier $(\cdots, j, \cdots)_m^{x\epsilon}$ is $\omega_m^{x,(-1)^{2j}\epsilon}$ (see Sec. 18 and 19 of Ref. 14). Cumbersome but straightforward computations show the already indicated pseudoequivalence $\mathscr{U}(g,\alpha) = \lambda(g,\alpha)\, V\, \mathscr{U}_B(g,\alpha)\, V^{-1}$, with λ depending only on α, and (λ, V) is given for each case in Table VII. In this table $\lambda(1) = 1$ and $\lambda(PT) = \lambda(P)\cdot\lambda(T)$.

Now we shall discuss the structure $\overline{\mathscr{G}} = \overline{\mathscr{G}}_0 \odot \overline{V}$ of the representation group a little. The reason for the analogy with $\mathscr{G} = \mathscr{G}_0 \odot V$ is because of $H^2_*(\mathscr{G}, \mathbf{T}) = H^2_0(\mathscr{G}_0, \mathbf{T}) \otimes H^2_*(V, \mathbf{T})$, or in more pictorical terms, due to the absence of an "interaction part" [as $\Lambda(g, \alpha)$ in theorem 5.6 of Ref. 20] in the expression of the factor system of the complete group. This structure which also arises in the relativistic case[24] explains why the consideration of group V alone, succeeds in

TABLE VII. Pseudoequivalence of Brennich's representations with ours.

$\mathscr{U}$	$\mathscr{U}_{\text{Brennich}}$	V	$\lambda(P)$	$\lambda(T)$
$m(\rho, j, +, +, \epsilon)$	$(m, j, \rho)_\epsilon^{+\,+}$	$\begin{cases} 1 & \text{if } \epsilon = 1 \\ -1 & \text{if } \epsilon = -1 \end{cases}$	1	1
$m(\rho, j, +, -)$	$(m, j, \rho)^{+\,(-)^{2j}}$	$\begin{bmatrix} 1 & 0 \\ 0 & \mathscr{D}_j(S) \end{bmatrix}$	i	i
$m(\rho, j, -, +)$	$(m, j, \rho)^{-\,(-)^{2j}}$	$\begin{bmatrix} 1 & 0 \\ 0 & \mathscr{D}_j(S) \end{bmatrix}$	i	i
$m(\rho, j, -, -, \epsilon)$	$(m, j, \rho)_\epsilon^{-}$	σ_x if $\epsilon = 1$; σ_y if $\epsilon = -1$	1	1
$\mathrm{II}_{11}(\rho, h, +, +)$	$(0, h, \rho)^{+\,+}$	$\begin{bmatrix} 1 & 0 \\ 0 & (-)^{2h} \end{bmatrix}$ if h integer	1	1
1		$i\begin{bmatrix} 1 & 0 \\ 0 & (-)^{2h} \end{bmatrix}$ if h half odd		
$\mathrm{II}_{11}(\rho, h, +, -)$	$(0, h, \rho)^{-\,+}$	$\begin{bmatrix} 1 & 0 \\ 0 & i(-)^{2h} \end{bmatrix}$	i	$(-)^{2h}$
$\mathrm{II}_{11}(\rho, h, -, +)$	$(0, h, \rho)$	$\begin{bmatrix} 1 & 0 & & \\ 0 & i(-)^{2h} & & \\ & & -1 & 0 \\ & & 0 & i(-)^{2h} \end{bmatrix}$	i	$(-)^{2h}$
$\mathrm{II}_{11}(\rho, h, -, -)$	$(0, h, \rho)^{+}$	$\begin{bmatrix} 1 & 0 & & \\ 0 & (-)^{2h} & & \\ & & 1 & 0 \\ & & 0 & (-)^{2h} \end{bmatrix}$	1	$(-)^{2h}$

giving all "types". But the point to be stressed here is that this result could have been different (for instance, when the interaction part is not trivial). Even in this case a separate consideration of $\bar{\mathscr{G}}_0$ and $\bar{V}$ is not fully satisfactory, because there are some ISUPR's of $(\mathscr{G}, \mathscr{G}_+)$ whose projective equivalence class is not specified by its restriction to $\bar{\mathscr{G}}_0$ and its type (viz., class I; index η can attain two different values).

For physical representations, i.e., classes m and II, the situation is very similar to the relativistic case.[24] The elements of $H^2_*(\mathscr{G}, \mathbf{T})$ can be interpreted as originating superselection rules, according to the well-known Bargmann's argument about mass in Galilean quantum mechanics. So, now we have mass, univalence, and "type" superselection rules. These questions have been discussed by Brennich and we do not insist upon them.

Now we can understand the role of the missing indices in the transition from the equivalence classes to pseudoequivalence classes of ISUR's. For instance, index ϵ which is present for some classes is a "relative parity" of a system composed of two elementary systems. Everything is very similar for the relativistic case. It is also to be remarked that relative parity does not appear for massive particles with "types" $+\ -$ or $-\ +$.

Finally, let us indicate that for kinematical groups[30] other than the Poincaré or the Galilei group, a representation group for the complete group is found in a very similar way. The results obtained are the following ones: if G is a "relative time" group, then $\bar{G}$ is $G^*_0 \odot \bar{V}$, just as in the Poincaré case, while if G is an "absolute time" group, then $\bar{G}$ is $\bar{G}_0 \odot \bar{V}$ similarly to the case of the Galilei group, but where the $M = 1$ nontrivial factor system $\frac{1}{2}\mathbf{b}\mathbf{v}'^2 + \mathbf{v}'\cdot R\,'\mathbf{a}$ of the Galilei group is replaced by the $M = 1$ corresponding factor system of each group. So, the case of the Newton–Hooke group has been studied[31] and the results are very similar to the Galilei case. A similar study for symmetry groups in one and two space dimensions is in course of development.

APPENDIX A: STRUCTURE AND REPRESENTATIONS OF THE LITTLE GROUPS

Orbits $Z_{m,p}$: This little group is $SU(2) \otimes \bar{V}_+$, with $\bar{V}_+ = \{(\gamma, \alpha), \alpha = 1, P\}$. It is easy to show that $\bar{V}_+ \simeq Z_4 \otimes Z_2$, with generators $(1,P)$ and $(\mu,1)$ respectively. Its irreducible representations are called $\Delta_{\epsilon_1\epsilon_2\epsilon}$ with $\epsilon_1, \epsilon_2, \epsilon \in \{1, -1\}$ and are given by

$$\Delta_{\epsilon_1\epsilon_2\epsilon}(1,P) = \begin{cases} \epsilon & \text{if} \quad \epsilon_1\epsilon_2 = 1, \\ i\epsilon & \text{if} \quad \epsilon_1\epsilon_2 = -1, \end{cases} \quad \begin{aligned} \Delta_{\epsilon_1\epsilon_2\epsilon}(\mu,1) &= \epsilon_1, \\ \Delta_{\epsilon_1\epsilon_2\epsilon}(\nu,1) &= \epsilon_2 \end{aligned} \tag{A1}$$

Orbits $Z_{0,p}$: This little group is one of the extensions of $\bar{V}_+$ by the first covering of the dimensional Euclidean group $1 \to E(2) \to G_{Z_{0,p}} \to \bar{V}_+ \to 1$ with the action

$$(\gamma,1) = (v_x, v_y, A_z) \to (v_x, v_y, A_z),$$

$$(\gamma,P) = (v_x, v_y, A_z)(v_x, -v_y, A_z^{+}).$$

The factor system relative to the natural section $(\gamma,1) \to (0,I,\gamma,1)$, $(\gamma,P) \to (0,\mathbb{S},\gamma,P)$, is $\omega(\gamma',P;\gamma,P) = (0,0,-I)$ while the other values of ω are equal to $(0,0,I)$. The subgroup $G_{Z_{0,p}}$ may be written as a semidirect product $G_{Z_{0,p}}$

$= T_2 \odot \widetilde{\widetilde{SO(2)}}_+$, where

$T_2 = \{(v_x, v_y, I, 1, 1)\}$,

$\widetilde{\widetilde{SO(2)}}_+ = \{(0,0,A_z,\gamma,1), (0,0,A_z\mathbb{S},\gamma,P)\}$.

It is regular in Mackey's meaning. We omit the details of the construction of the corresponding induced representations and quote only the result

$I(\epsilon_1, \epsilon_2, \epsilon)$: $\epsilon_1, \epsilon_2, \epsilon \in \{1, -1\}$: Carrier space $\mathbb{C}$. The representation is given by

$$\begin{aligned} D_{\epsilon_1\epsilon_2\epsilon}(\mathbf{v}, A_z, \gamma, 1) &= \Delta_{\epsilon_1\epsilon_2\epsilon}(\gamma,1), \\ D_{\epsilon_1\epsilon_2\epsilon}(\mathbf{v}, A_z, \gamma, P) &= \Delta_{\epsilon_1\epsilon_2\epsilon}(\gamma,P). \end{aligned} \tag{A2}$$

$II(h, \epsilon_1, \epsilon_2)$: $2h \in \mathbb{N}$, $\epsilon_1, \epsilon_2 \in \{1, -1\}$. Carrier space $\mathbb{C}^2$

$$D_{h\epsilon_1\epsilon_2}(\mathbf{v}, A, 1, 1) = \begin{bmatrix} e^{ih\varphi} & 0 \\ 0 & e^{-ih\varphi} \end{bmatrix},$$

$$D_{h\epsilon_1\epsilon_2}(0, \mathbb{S}, 1, P) = \begin{bmatrix} 0 & 1 \\ (-1)^{2h}\epsilon_1\epsilon_2 & 0 \end{bmatrix},$$

$$D_{h\epsilon_1\epsilon_2}(0, I, \mu, 1) = \epsilon_1 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

$$D_{h\epsilon_1\epsilon_2}(0, I, \nu, 1) = \epsilon_2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{A3}$$

where φ is the rotation angle (mod 4π) of $A_z(\varphi)$.

$III((x, \eta, \epsilon_1, \epsilon_2)$: $x \in \mathbb{R}$, $x > 0$, $\eta \in \{1, i, -1, -i\}$. The carrier space is $\mathscr{L}^2(S^1_x \to \mathbb{C}, d\varphi_{\mathbf{k}})$ and the representation is given by

$$\begin{aligned} &[D_{x,\eta,\epsilon_1,\epsilon_2}(\mathbf{v}, A, \gamma, \alpha) f](\mathbf{k}) \\ &\quad = e^{i\mathbf{k}\mathbf{v}} \Delta_{\eta\epsilon_1\epsilon_2}[\mathscr{L}^{-1}(\mathbf{k}) A \mathscr{L}(A^{-1}\mathbf{k}^\alpha), \gamma, \alpha] f(A^{-1}\mathbf{k}^\alpha), \end{aligned} \tag{A4}$$

where $\Delta_{\eta\epsilon_1\epsilon_2}$ is a representation of the group $Z_4 \otimes Z_2 \otimes Z_2$ [generated by $(\mathbb{S},1,P)$, $(I,\mu,1)$, and $(I,\nu,1)$] which is given by

$$\Delta_{\eta\epsilon_1\epsilon_2}(\mathbb{S},1,P) = \eta, \quad \Delta_{\eta\epsilon_1\epsilon_2}(I,\mu,1) = \epsilon_1, \quad \Delta_{\eta\epsilon_1\epsilon_2}(I,\nu,1) = \epsilon_2, \tag{A5}$$

and

$$\mathscr{L}(\mathbf{k}) = s(\tfrac{1}{2}\varphi_{\mathbf{k}})[\cos\tfrac{1}{2}\varphi_{\mathbf{k}} - i\sigma_z \sin\tfrac{1}{2}\varphi_{\mathbf{k}}]. \tag{A6}$$

The fixed point is $\mathbf{k}_0 = (x,0)$ and the function s is defined by

$$s(\varphi) = \begin{cases} 1 & \text{if} \quad \varphi \in [0,\pi) \quad (\text{mod } 2\pi) \\ -1 & \text{if} \quad \varphi \in [\pi,2\pi) \quad (\text{mod } 2\pi). \end{cases}$$

Orbits $Z_{0,0,E}$: This little group $\bar{K}_+$ admits a regular semidirect structure $G_{Z_{0,0,E}} = T_3 \odot (SU(2) \otimes \bar{V}_+)$, the action is (A, γ, α): $\mathbf{v} \to A\mathbf{v}^\alpha$ The representation of $\bar{K}_+$ are the following ones: $I(j, \epsilon_1, \epsilon_2, \epsilon)$, $2j \in \mathbb{N}$, $\epsilon_1, \epsilon_2, \epsilon \in \{1, -1\}$. The carrier space is $\mathbb{C}^{2j+1}$. The representation is

$$D_{j\epsilon_1\epsilon_2\epsilon}(\mathbf{v}, A, \gamma, \alpha) = \mathscr{D}_j(A)\,\Delta_{\epsilon_1\epsilon_2\epsilon}(\gamma, \alpha). \tag{A7}$$

$II(x, \epsilon_1, \epsilon_2, \epsilon)$: $x \in \mathbb{R}$, $x > 0$, $\epsilon_1, \epsilon_2, \epsilon \in \{1, -1\}$. The carrier space is $\mathscr{L}^2(S^2_x \to \mathbb{C}, d\Omega_{\mathbf{k}})$. The representation is given by

$$\begin{aligned} &[D_{x\epsilon_1\epsilon_2\epsilon}(\mathbf{v}, A, \gamma, \alpha) f](\mathbf{k}) \\ &\quad = e^{i\mathbf{k}\mathbf{v}} \Delta_{\epsilon_1\epsilon_2\epsilon}(\gamma, \alpha) f(A^{-1}\mathbf{k}^\alpha). \end{aligned} \tag{A8}$$

$III(x, h, \epsilon_1, \epsilon_2)$: $x \in \mathbb{R}$, $x > 0$, $2h \in \mathbb{N}$, $\epsilon_1, \epsilon_2 \in \{1, -1\}$. The carrier space is $\mathscr{L}^2(S^2_x \to \mathbb{C}^2, d\Omega_{\mathbf{k}})$. The representation is given by

$$\begin{aligned} &[D_{xh\epsilon_1\epsilon_2}(\mathbf{v}, A, \gamma, \alpha) f](\mathbf{k}) \\ &\quad = e^{i\mathbf{k}\mathbf{v}} D_{h\epsilon_1\epsilon_2}(L^{-1}(\mathbf{k}) A L(A^{-1}\mathbf{k}^\alpha), \gamma, \alpha) f(A^{-1}\mathbf{k}^\alpha), \end{aligned}$$

where $L(k)$ is given by (3.1) (with the change $\rho \longleftrightarrow x$) and

$D_{h\epsilon_1\epsilon_2}$ is the representation (A3).

APPENDIX B: PARTICULARIZATION TO P OF THE REPRESENTATIONS II_{11} AND III_{11}

II_{11}: In this case

$$[U(P)\psi](E,\mathbf{p}) = D_{h\epsilon_1\epsilon_2}\{L^{-1}(\mathbf{p})L(-\mathbf{p}),1,P\}\,\psi(E,-\mathbf{p}). \quad \text{(B1)}$$

Now we have $L^{-1}(\mathbf{p})\cdot L(-\mathbf{p}) = A_z(\varphi(\mathbf{p}))\cdot\mathrm{S}$ for an angle $\varphi(\mathbf{p})$ (mod 4π). Then $A_z(\varphi(\mathbf{p})) = -L^{-1}(\mathbf{p})\cdot L(-\mathbf{p})\cdot\mathrm{S}$, and by making use of (3.1) we obtain

$$\cos\tfrac{1}{2}\varphi(\mathbf{p}) - i\sigma_z\sin\tfrac{1}{2}\varphi(\mathbf{p}) = (-p_x + i\sigma_z p_y)/(\rho - p_z^2)^{1/2},$$
$$\mathbf{p}\notin\{(0,0,\sqrt{\rho}),(0,0,-\sqrt{\rho})\};$$

when we introduce the usual spherical angles $\theta_{\mathbf{p}}$, $\varphi_{\mathbf{p}}$, this result reads $\varphi(\mathbf{p}) = 2\varphi_{\mathbf{p}}$ (if $\theta_{\mathbf{p}}\notin\{0,\pi\}$. If $\mathbf{p}$ is on either of the poles, the same procedure leads to

$$\varphi(\mathbf{p}) = \begin{cases} 2\pi & \text{for} \quad \theta_{\mathbf{p}} = 0, \\ 0 & \text{for} \quad \theta_{\mathbf{p}} = \pi. \end{cases}$$

In order to avoid particular specifications, we shall take for granted that in the poles, where the azimuthal angle is ill defined, we take the values $\varphi_{\mathbf{p}} = \pi$ for $\theta_{\mathbf{p}} = 0$ and $\varphi_{\mathbf{p}} = 0$ for $\theta_{\mathbf{p}} = \pi$ in such a way that $\varphi(\mathbf{p}) = 2\varphi_{\mathbf{p}}$ will always be true. The final expression is

$$[U(P)\psi](E,\mathbf{p}) = \begin{pmatrix} 0 & e^{2ih\varphi_{\mathbf{p}}} \\ (-)^{2h}\epsilon_1\epsilon_2 e^{-2ih\varphi_{\mathbf{p}}} & 0 \end{pmatrix}\psi(E,-\mathbf{p})$$

III_{11}: Now,

$$\begin{aligned}[U(P)\psi](E,\mathbf{p}) &= \boldsymbol{\Delta}_{\eta\epsilon_1\epsilon_2}(\mathscr{L}^{-1}(\mathbf{k})L^{-1}(\mathbf{p}) \\ &\quad\times\mathscr{L}([L^{-1}(\mathbf{p})L(-\mathbf{p})]^{-1}(-\mathbf{k})),1,P) \\ &\quad\times\psi([L^{-1}(\mathbf{p})L(-\mathbf{p})]^{-1}(-\mathbf{k}),E,-\mathbf{p}). \end{aligned} \quad \text{(B2)}$$

The azimuthal polar angle of the vector $[L^{-1}(\mathbf{p})\cdot L(-\mathbf{p})]^{-1}(-\mathbf{k})$ is $2\varphi_{\mathbf{p}} - \varphi$. The first argument in $\boldsymbol{\Delta}_{\eta\epsilon_1\epsilon_2}$ in (B2) can be calculated by making use of (A6) as follows:

$$\begin{aligned}\mathscr{L}^{-1}(\mathbf{k})\cdots &= s(\tfrac{1}{2}\varphi)_1\{\cos(\tfrac{1}{2}\varphi)_1 + i\sigma_z\sin(\tfrac{1}{2}\varphi)_1\} \\ &\quad\times\{\cos\varphi_{\mathbf{p}} - i\sigma_z\sin\varphi_{\mathbf{p}}\}(-i\sigma_y)s(\varphi_{\mathbf{p}} - (\tfrac{1}{2}\varphi)_2) \\ &\quad\times\{\cos(\varphi_{\mathbf{p}} - (\tfrac{1}{2}\varphi)_2) - i\sigma_z\sin(\varphi_{\mathbf{p}} - (\tfrac{1}{2}\varphi)_2)\},\end{aligned}$$

where $(\frac{1}{2}\varphi)_1$ and $(\frac{1}{2}\varphi)_2$ are two arbitrary independent determinations of half the angle φ. A straightforward calculation leads to

$$\begin{aligned}\mathscr{L}^{-1}(\mathbf{k})\cdots &= s((\tfrac{1}{2}\varphi)_1)\,s(\varphi_{\mathbf{p}} - (\tfrac{1}{2}\varphi)_2) \\ &\quad\times\{\cos[(\tfrac{1}{2}\varphi)_1 - (\tfrac{1}{2}\varphi)_2] - i\sigma_z\sin[(\tfrac{1}{2}\varphi)_1 - (\tfrac{1}{2}\varphi)_2]\}\,\mathrm{S},\end{aligned}$$

which does not depend on the determinations we had previously chosen. In particular when the two determinations coincide,

$$\mathscr{L}^{-1}(\mathbf{k})\cdots = s(\tfrac{1}{2}\varphi)\,s(\varphi_{\mathbf{p}} - \tfrac{1}{2}\varphi)\,\mathrm{S}.$$

Notice that $s(\frac{1}{2}\varphi)$ has no sense by itself, but the product $\pi(\varphi,\varphi_{\mathbf{p}}) = s(\frac{1}{2}\varphi)\cdot s(\varphi_{\mathbf{p}} - \frac{1}{2}\varphi)$ is well-defined provided that the same determination for $\frac{1}{2}\varphi$ be used in both factors. In order to simplify the notation let us define

$$\delta_{\eta\epsilon_1\epsilon_2}(\varphi,\varphi_{\mathbf{p}}) = \boldsymbol{\Delta}_{\eta\epsilon_1\epsilon_2}(\pi(\varphi,\varphi_{\mathbf{p}})\,\mathrm{S},1,P). \quad \text{(B3)}$$

The final expression is

$$[U(P)\psi](\varphi,E,\mathbf{p}) = \delta_{\eta\epsilon_1\epsilon_2}(\varphi,\varphi_{\mathbf{p}})\,\psi(2\varphi_{\mathbf{p}} - \varphi,E,-\mathbf{p}). \quad \text{(B4)}$$

According to the value of $\pi(\varphi,\varphi_{\mathbf{p}})$ being either 1 or -1 the value of $\delta_{\eta\epsilon_1\epsilon_2}(\varphi,\varphi_{\mathbf{p}})$ is η or $\eta^3\epsilon_1\epsilon_2$ respectively. A check for the calculation is $[U(P)]^2 = \epsilon_1\epsilon_2$.

[1]E. P. Wigner, Ann. Math. **40**, 149 (1939).
[2]V. Bargmann, Ann. Math. **59**, 1 (1954).
[3]I. Schur, J. Reine Angew. Math. **127**, 20 (1904); **132**, 85 (1907).
[4]T. Janssen, J. Math. Phys. **13**, 342 (1972).
[5]J. F. Cariñena and M. Santander, J. Math. Phys. **20**, 2168 (1979).
[6]J. F. Cariñena and M. Santander, J. Math. Phys. **16**, 1416 (1975).
[7]J. F. Cariñena and M. Santander, J. Math. Phys. **21**, 440 (1980).
[8]U. Cattaneo, J. Math. Phys. **19**, 452 (1978).
[9]M. Santander, "Linealización de representaciones proyectivas semiunitarias," Ph.D. thesis, Universidad de Valladolid, 1977.
[10]E. P. Wigner, "Unitary representations of the inhomogeneous Lorentz group including reflections," in *Group Theoretical Concepts and Methods in Elementary Particle Physics*, edited by F. Gursey (Gordon and Breach, New York, 1964).
[11]S. Wightman, "L'invariance dans la Mecanique quantique relativiste," in *Relations de dispersion et particules elementaires*, edited by C. de Witt and R. Omnes (Hermann, Paris, 1960).
[12]R. Shaw and J. Lever, Commun. Math. Phys. **38**, 279 (1974).
[13]D. Ebner, Gen. Relativ. Gravit. **8**, 15 (1977).
[14]R. H. Brennich, Ann. Inst. Henri Poincaré A **13**, 137 (1970).
[15]U. Cattaneo, J. Math. Phys. **19**, 767 (1978).
[16]J. F. Cariñena and M. Santander, Int. J. Theor. Phys. (to appear).
[17]E. Inonu and E. P. Wigner, Nuovo Cimento **9**, 705 (1952).
[18]J. M. Lévy-Leblond, "Galilei group and Galilean Invariance" in *Group Theory and Its Applications*, Vol. II, edited by E. M. Loebl (Academic, New York, 1971).
[19]G. W. Mackey, Acta Math. **99**, 265 (1958), Theorem 9.4.
[20]K. R. Parthasarathy, *Multiplier on Locally Compact Groups*, Lecture Notes in Physics No. 93 (Springer, New York, 1969), Theorem 5.6.
[21]P. M. van den Broek, J. Phys. A: Math. Gen. **9**, 855 (1976).
[22]G. W. Mackey, Ann. Math. **55**, 101 (1952).
[23]D. J. Simms, *Lie Groups and Quantum Mechanics*, Lecture Notes in Mathematics No. 52 (Springer, 1968).
[24]J. F. Cariñena and M. Santander, "Projective Semiunitary Representations of the Complete Poincaré Group" in *Algunas Cuestiones de Física Teórica, 1975*, GIFT 4/76.
[25]L. J. Boya, J. F. Cariñena, and M. Santander, Commun. Math. Phys. **37**, 331 (1974).
[26]E. P. Wigner, *Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra* (Academic, New York, 1959), Sec. 26.
[27]R. Shaw and J. Lever, Commun. Math. Phys. **38**, 257 (1974).
[28]Y. M. Shirokov, Soviet Physics JETP **36**, 620 (1959); Nucl. Phys. **15**, 1 (1960); **15**, 13 (1960).
[29]K. R. Parthasarathy, Commun. Math. Phys. **15**, 305 (1969).
[30]H. Bacry and J. M. Lévy-Leblond, J. Math. Phys. **9**, 1605 (1968).
[31]M. Valls de Quintana, "El groupo de Hooke completo," Tesina de Licenciatura, Universidad de Valladolid, 1977.

A unified treatment of the representation functions of SO(n,1), SO($n+1$), and ISO(n)

M.K.F. Wong
Fairfield University, Fairfield, Connecticut 06430

Hsin-Yang Yeh
Moorhead State University, Moorhead, Minnesota 56560

(Received 15 July 1980; accepted for publication 25 September 1980)

An explicit expression is obtained for all the representation functions of SO(n,1), SO($n+1$), and ISO(n). It is found that the representation functions of SO(n,1) and SO($n+1$) are basically expressible as hypergeometric functions ${}_2F_1$ with arguments $1-e^{-2\zeta}$ and $1-e^{2i\delta}$, respectively, for $n>2$, multiplied by Weyl coefficients of SO(p), $p=3, 4, ..., n$. The representation functions of ISO(n) are then obtained from those of SO(n,1) or SO($n+1$) by contraction. They are expressible as sums over a confluent hypergeometric function with argument $2i\gamma\xi$, multiplied by Weyl coefficients of SO(p), $p=3, 4, ..., n$. This provides an interesting alternate form for the representation functions of ISO(n) obtained previously by Wong and Yeh as a sum over Bessel functions.

PACS numbers: 02.20.Qs

I. INTRODUCTION

Recently[1] we have obtained an explicit expression for the representation functions of IU(n) through the contraction of the representation functions of U($n+1$) or U(n, 1). In doing so we have demonstrated that there is a unified approach for the explicit evaluation of the representation functions of the groups U(n, 1), U($n+1$) and IU(n). This approach is possible because the relevant matrices $b(\theta)$ for both U(n, 1) and U($n+1$) can be written in the form

$$b(\theta)=b(I+te_{n,n-1})\,b_0(\theta)\,b(I\pm te_{n-1,n}), \qquad (1.1)$$

where $b_0(\theta)$ is a diagonal matrix. The explicit expression for the representation functions of $b(I+te_{n,n-1})$ and $b(I\pm te_{n-1,n})$ have been obtained by Gel'fand and Graev.[2] The representation functions of IU(n) are then obtained from those of U($n+1$) or U(n, 1) by contraction. It is interesting to note that in the process of contraction all terms containing infinity disappear. Thus a unified approach exists for the explicit evaluation of the representation functions of U(n, 1), U($n+1$), and IU(n).

A similar question may be asked for the groups SO(n, 1), SO($n+1$), and ISO(n). In order to answer this question, one must start with the explicit evaluation of the representation functions of SO(n, 1) which can be analytically continued to those of SO($n+1$). Since the principal series of SO(n, 1) can be analytically continued to the general irreducible representations of SO($n+1$), we restrict ourselves to the consideration of the principal series of SO(n, 1) only. If such an explicit expression exists, then the representation functions of ISO(n) can be obtained from it through contraction, provided terms containing infinity disappear. Fortunately there is a formula for the representation functions of SO(n, 1) which can be analytically continued to SO($n+1$). This formula was obtained by Wolf[3] through the theory of induced representations. The corresponding formula for SO($n+1$) was obtained earlier by Vilenkin.[4] So far, however, we know of no explicit evaluation of Wolf's formula for $n>3$. It is our purpose to show in this paper that an explicit evaluation of Wolf's formula can be carried out by means of the Weyl coefficients of SO(n) discussed by Wong.[5] The result is that all representation functions of SO(n, 1), $n>2$, can be expressed basically as a sum over a hypergeometric function ${}_2F_1$ with argument $1-e^{-2\zeta}$, multiplied by Weyl coefficients of SO(p), $p=3,4,\ldots,n$. The corresponding representation functions of SO($n+1$) can then be obtained immediately from those of SO(n, 1) by analytic continuation. Finally, the representation functions of ISO(n) can be obtained from those of SO(n, 1) or SO($n+1$) by contraction. We find that in the contraction process all terms containing infinity disappear. The result is that the representation functions of ISO(n), $n>2$, are expressible as sums over a confluent hypergeometric function ${}_1F_1$ with argument $2i\gamma\xi$, multiplied by Weyl coefficients of SO(p). This provides an interesting alternate form for the representation functions of ISO(n), as obtained previously by Wong and Yeh[6] in terms of Bessel functions with argument $\gamma\xi$, multiplied by Clebsch–Gordan (CG) coefficients of SO(n) involving the most degenerate representation $[k,\dot{0}]$. We wish to note that whereas the previous expression in terms of Bessel functions is quite compact, the explicit evaluation of the CG coefficients of SO(n) for the representation $[k,\dot{0}]$ has not yet been carried out for general n and k, $n>4$. Therefore, in a sense, our present expression is even more explicit, since every term can be written down explicitly.

Some interesting questions arise from our work. We know that the two expressions for the representation functions of ISO(n), one in terms of Bessel functions and the other in terms of confluent hypergeometric functions, are equal. However, a direct proof that these two expressions are equal is very difficult, even in the case of ISO(3) (though an indirect proof exists from the work of Smorodinskii and Shepelev,[7] Wong and Yeh,[8] or Rashid[9]). A solution of this problem for general n would mean that an explicit evaluation of the CG coefficients of SO(n) for the most degenerate representation $[k,\dot{0}]$ can be carried out.

In Sec. II, we present our method for a unified treatment of the representation functions of SO(n, 1), SO($n+1$), and ISO(n). The general theory developed in Sec. II is then applied to special cases in Secs. III and IV. In Sec. III, we discuss the explicit evaluation of the representation functions of SO(3, 1), SO(4), and ISO(3). In Sec. IV, we discuss the explicit evaluation of the representation functions of SO(4, 1), SO(5), and ISO(4). In Sec. V, we

 0022-2488/81/081559-07$1.00 © 1981 American Institute of Physics

discuss the explicit evaluation of the representation functions of SO(n, 1), SO(n+1), and ISO(n).

II. UNIFIED APPROACH FOR THE REPRESENTATION FUNCTIONS OF SO(n, 1), SO(n+1), AND ISO(n)

Before presenting the general theory, we would like to establish some notational convention. The representation functions of SO(n, 1), SO(n+1), and ISO(n) shall be denoted by ${}^Pd$, d, and ${}^Id$, respectively. The UIR label of SO(n) will be either written out explicitly or represented by (m_n), where

$$(m_n) = m_{1n}, m_{2n}, \ldots, m_{[n/2],n}. \tag{2.1}$$

The d-functions of SO(n, 1), SO(n+1), and ISO(n) depend on four UIR labels: (m_{n+1}), (m_n), (m'_n), and (m_{n-1}). We shall write, as an example, the d-functions of SO(n+1) as

$$d^{(m_{n+1})}_{(m_n)(m'_n)\,(m_{n-1})},$$

where it is understood that the d-function, regarded as a matrix, is diagonal with respect to the label of (m_{n-1}). Let us now proceed to the theory.

As we have mentioned in the introduction, a unified approach for the explicit evaluation of the representation functions of SO(n, 1) SO(n+1), and ISO(n) must start from an explicit evaluation of the representation functions of SO(n, 1). This is because one can obtain the representation functions of SO(n+1) from analytic continuation of those of SO(n, 1), but not necessarily the other way round. This point is illustrated by considering the case of SO(4). If one writes the d-function of SO(4) as

$$d^{m_{14}m_{24}}_{m_{13}m'_{13}\,m_{12}}(\theta) = \sum_m [(2m_{13}+1)(2m'_{13}+1)]^{1/2} \times \begin{pmatrix} \frac{1}{2}(m_{14}+m_{24}) & \frac{1}{2}(m_{14}-m_{24}) & m_{13} \\ m & m_{12}-m & -m_{12} \end{pmatrix} \times \begin{pmatrix} \frac{1}{2}(m_{14}+m_{24}) & \frac{1}{2}(m_{14}-m_{24}) & m'_{13} \\ m & m_{12}-m & -m_{12} \end{pmatrix} e^{-2im\theta}, \tag{2.2}$$

then one cannot directly continue to SO(3, 1), as shown by Wong and Yeh,[8] and also by Rashid,[9] since the d-function of SO(3, 1) contains two terms, while (2.2) contains only one term.

For the d-functions of SO(n, 1) we use the formula obtained by Wolf [Ref. 3, Eqs. (5.11) and (5.12)]:

$${}^Pd^{\lambda L}_{JL'J'}(\zeta) = \frac{(\dim_n J \dim_n J')^{1/2}\,\Gamma(n/2)}{\dim_{n-1} L \dim_{n-1} L'\, \pi^{1/2}\,\Gamma(\frac{1}{2}(n-1))} \times \sum_M \dim_{n-2} M \times \int_0^\pi \sin^{n-2}\theta\, d\theta\, \overline{d^J_{LML'}(\theta)} \left(\frac{\sin\theta}{\sin\theta'}\right)^\lambda d^{J'}_{LML'}(\theta'), \tag{2.3}$$

where

$$\sin\theta/\sin\theta' = \cosh\zeta - \cos\theta \sinh\zeta, \tag{2.4}$$

$$\cos\theta' = \frac{\cos\theta\cosh\zeta - \sinh\zeta}{\cosh\zeta - \cos\theta\sinh\zeta}. \tag{2.5}$$

For the compact group SO(n+1), one replaces ζ by $-i\delta$ in Eq. (2.3). Therefore

$$\sin\theta/\sin\theta' = \cos\delta + i\cos\theta\sin\delta, \tag{2.6}$$

$$\cos\theta' = \frac{\cos\theta\cos\delta + i\sin\delta}{\cos\delta + i\cos\theta\sin\delta}. \tag{2.7}$$

Also, to continue from SO(n, 1) to SO(n+1), one should multiply by the phase factor W_1 of Maekawa[10] to the d-functions of SO(n, 1).

The d-functions on the right-hand side of Eq. (2.3) are those of SO(n). We write those in terms of the Weyl coefficients discussed by Wong[5] and the d-functions of SO(3). Thus, for example, if, say, $n=5$, then we write

$$D_{45}(\theta) = W^{-1}_{34}W^{-1}_{45}W^{-1}_{23}W^{-1}_{34}D_{23}(\theta)W_{34}W_{23}W_{45}W_{34}, \tag{2.8}$$

where

$$W_{ij} = D_{ij}(\pi/2) \quad\text{and}\quad W^{-1}_{ij} = D_{ij}(-\pi/2). \tag{2.9}$$

Thus the θ-dependent part in Eq. (2.3) can be explicitly taken out and integrated for all SO(n, 1). We show in Sec. IV and V that this results in having all the representation functions of SO(n, 1) expressible as a sum over a hypergeometric function ${}_2F_1$ with argument $1-e^{2\zeta}$, multiplied by Weyl coefficients of SO(p), $p=3, 4, \ldots, n$. Note that because of the numerous relations between the solutions of the hypergeometric equation, the ${}_2F_1$ function can be expressed in different forms, e.g., with argument $1-e^{-2\zeta}$.

The representation functions of SO(n+1) are immediately obtained from those of SO(n, 1) by analytic continuation.

The representation functions of ISO(n) are obtained from those of SO(n, 1) or SO(n+1) by contraction. The contraction process goes as follows:

$$\text{for SO}(n,1)\text{:}\quad \lambda \to i\infty, \quad i\lambda\zeta = \gamma\xi, \tag{2.10}$$

$$\text{for SO}(n+1)\text{:}\quad m_{1n+1} \to \infty, \quad m_{1n+1}\delta = \gamma\xi. \tag{2.11}$$

As a result of the contraction process, we find that the representation functions of ISO(n) are expressible as a summation over a confluent hypergeometric function ${}_1F_1$ with argument $2i\gamma\xi$, mulitiplied by Weyl coefficients of SO(p), $p=3, 4, \ldots, n$. This is a surprising result since, to our knowledge, no one has indicated previously that the representation functions of ISO(n) are connected with confluent hypergeometric functions.

In a previous paper,[6] we obtained the representation functions of ISO(n) in terms of Bessel functions and Clebsch–Gordan coefficients of SO(n) involving the most degenerate representation $[k, 0]$ of SO(n). This expression, though in a very compact form, is nevertheless not completely explicit, because the CG coefficients of SO(n), for general n and k, have not yet been explicitly obtained. In a sense, therefore, our present expression is more explicit because every term can be explicitly written down.

III. UNIFIED APPROACH FOR THE RPRESENTATION FUNCTIONS OF SO(3, 1), SO(4), AND ISO(3)

The representation functions of SO(3, 1) have been obtained by many authors.[11] We shall use the result obtained by Makarov and Shepelev in Ref. 11, Eq. (3). Thus the d-function of SO(3, 1) is expressed as

$$
{}^{P}d^{\lambda\, m_{24}}_{m_{13} m'_{13} \atop m_{12}}(\zeta) = (2m_{13}+1)^{1/2}(2m'_{13}+1)^{1/2}
$$

$$
\times \left(\frac{(m_{13}+m_{12})!\,(m'_{13}+m_{12})!\,(m_{13}-m_{12})!\,(m'_{13}-m_{12})!}{(m_{13}-m_{24})!\,(m'_{13}-m_{24})!\,(m_{13}+m_{24})!\,(m'_{13}+m_{24})!} \right)^{1/2}
$$

$$
\times\ e^{\zeta(\lambda - m_{24} - m_{12})}(-1)^{2m_{13}}
$$

$$
\times \sum_{k=\max(m_{24},m_{12})}^{m'_{13}} \sum_{p=\max(m_{24},m_{12})}^{m_{13}} (-1)^{p+k} \binom{m_{13}+p}{p-m_{12}} \binom{m_{13}-m_{24}}{p-m_{24}}
$$

$$
\times \binom{m'_{13}+k}{k-m_{12}} \binom{m'_{13}-m_{24}}{k-m_{24}} B(m_{24}+m_{12}+1,\ p+k+1-m_{24}-m_{12})
$$

$$
\times\ {}_2F_1(k-\lambda,\ 1+m_{24}+m_{12},\ p+k+2;\ 1-e^{-2\zeta}), \tag{3.1}
$$

where

$$
B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} = B(y,x). \tag{3.2}
$$

The d-function of SO(4) is obtained from (3.1) by the following substitution: $\zeta \to -i\delta$, $\lambda \to m_{14}$, and multiplication by the phase factor W_1 of Maekawa.[10]

It has been explicitly demonstrated by Smorodinskii and Shepelev,[7] Wong and Yeh,[8] and Rashid,[9] that the d-function of SO(4) so obtained is equal to the d-function in Eq. (2.2).

Barut and Wilson[12] have shown that the d-function of SO(4) with $m'_{13} = m_{12}$ is expressible as a hypergeometric function with argument $1-e^{-2i\alpha}$ with no sums. They obtained this result by solving a second order partial differential equation. We show here that this is a direct consequence of Eq. (3.1), i.e., the sums over k and p can be carried out, resulting in a hypergeometric function with no sums. To show this we first observe that the sum over k is reduced to one term: $k = m_{12}$. Then the hypergeometric function is expanded in terms of a summation over r

$$
{}_2F_1(m_{12}-\lambda,\ 1+m_{24}+m_{12};\ p+m_{12}+2:\ 1-e^{2i\delta})
$$

$$
= \sum_r \frac{\Gamma(m_{12}-\lambda+r)\Gamma(1+m_{24}+m_{12}+r)\Gamma(p+m_{12}+2)(1-e^{2i\delta})^r}{\Gamma(m_{12}-\lambda)\Gamma(1+m_{24}+m_{12})\Gamma(p+m_{12}+2+r)\,r!\,(m_{12}-\lambda+r)} \tag{3.3}
$$

The summation over p is contained in the following terms:

$$
\sum_{p=m_{12}}^{m_{13}} \frac{(-1)^p (p+m_{13})!}{(m_{13}-p)!\,(p-m_{12})!\,(p+m_{12}+1-r)!}
$$

$$
= (-m_{12}-m_{13}-1)!\,(m_{12}+m_{13}+1)!\,(-1)^{m_{12}}
$$

$$
\times \sum_{p=0}^{m_{13}-m_{12}} \frac{1}{p'!\,(m_{13}-p'-m_{12})!\,(p'+2m_{12}+r+1)!}
$$

$$
\times \frac{1}{(-m_{12}-m_{13}-p'-1)!}
$$

$$
= \frac{(-1)^{m_{12}} r!\,(m_{12}+m_{13}+1)!}{(m_{13}-m_{12})!\,(r+m_{12}-m_{13})!\,(r+m_{12}+m_{13}+1)!}. \tag{3.4}
$$

Thus the summation over p and k has been carried out. The hypergrometric function so obtained is transformed to the form of Barut and Wilson through Euler's identity

$$
{}_2F_1(a,b;c;z) = (1-z)^{c-a-b}\ {}_2F_1(c-a,\ c-b;\ c;z). \tag{3.5}
$$

Our final result agrees with Barut and Wilson by identifying δ with $-\alpha$.

Let us now apply the contraction process to Eq. (3.1). We have $\lambda \to i\infty$, $i\lambda\zeta = \gamma\xi$,

$$
e^{\zeta(\lambda - m_{24} - m_{12})} \to e^{-i\gamma\xi}, \tag{3.6}
$$

$$
{}_2F_1(k-\lambda,\ m_{12}+m_{24}+1,\ p+k+2,\ 1-e^{-2\zeta}) = \sum_n \frac{(k-\lambda)_n (m_{12}+m_{24}+1)_n}{(p+k+2)_n\, n!},
$$

$$
(1-e^{-2\zeta})^n = \sum_n \frac{\Gamma(k-\lambda+n)(m_{12}+m_{24}+1)_n}{\Gamma(k-\lambda)(p+k+2)_n\, n!}(2\zeta)^n \to \sum_n \frac{(m_{12}+m_{24}+1)_n}{(p+k+2)_n\, n!}(2i\gamma\xi)^n = {}_1F_1(m_{24}+m_{12}+1,\ p+k+2,\ 2i\gamma\xi). \tag{3.7}
$$

Therefore the d-function of ISO(3) can be written as

$$
{}^{I}d^{\gamma\, m_{24}}_{m_{13} m'_{13} \atop m_{12}}(\xi)
$$

$$
= N(m_{13}, m'_{13}, m_{12}, m_{24})\, e^{-i\gamma\xi} \sum_{p,k} (-1)^{p+k} \frac{(m_{13}+p)!\,(m'_{13}+k)!\,(m_{12}+m_{24})!\,(p+k-m_{12}-m_{24})!}{(m_{13}-p)!\,(p-m_{12})!\,(p-m_{24})!\,(m'_{13}-k)!\,(k-m_{12})!\,(k-m_{24})!\,(p+k+1)!}
$$

$$
\times\ {}_1F_1(m_{12}+m_{24}+1,\ p+k+2,\ 2i\gamma\xi), \tag{3.8}
$$

where

$$
N(m_{13}, m'_{13}, m_{12}, m_{24}) = (-1)^{2m_{13}} \left((2m_{13}+1)(2m'_{13}+1) \frac{(m_{13}-m_{12})!\,(m_{13}-m_{24})!\,(m'_{13}-m_{12})!\,(m'_{13}-m_{24})!}{(m_{13}+m_{12})!\,(m_{13}+m_{24})!\,(m'_{13}+m_{12})!\,(m'_{13}+m_{24})!} \right)^{1/2}. \tag{3.9}
$$

As far as we know, this expression has not been found in the literature before, since previously all representation functions of ISO(n) have been expressed in terms of Bessel functions. We shall show in subsequent sections that the representation functions of ISO(n) can all be expressed as a summation over a confluent hypergeometric function with argument $2i\gamma\xi$, multiplied by Weyl coefficients of SO(p), $p = 3, 4, \ldots, n$.

It has been shown previously [Ref. 13, Eq. (6.14)] that the d-function of ISO(3) is

$$
{}^{I}d^{\gamma\, m_{24}}_{\substack{m_{13} m'_{13} \\ m_{12}}}(\xi) = [(2m_{13}+1)(2m'_{13}+1)]^{1/2} \sum_{L=0}^{\infty} i^{L} j_{L}(\gamma\xi)
$$

$$
\times C\begin{pmatrix} m'_{13} & L & m_{13} \\ m_{24} & 0 & m_{24} \end{pmatrix} C\begin{pmatrix} m'_{13} & L & m_{13} \\ m_{12} & 0 & m_{12} \end{pmatrix}. \tag{3.10}
$$

From the derivation of (3.8) as well as the work of the authors in Refs. 7–9, it is clear that (3.8) is equal to (3.10). It is tempting to ask whether a direct proof of the equality of these two expressions can be attained. This is not a trivial problem, since the solution of the problem might lead to new results in groups higher than ISO(3).

IV. UNIFIED APPROACH FOR THE REPRESENTATION FUNCTIONS OF SO(4, 1), SO(5), AND ISO(4)

An explicit expression for the representation function of SO(5) has been given by Holman[14] through the method of boson operators and Wong[5] through the method of Weyl coefficients. Holman[15] has also discussed the d-function of SO(4, 1) through spinor calculus. We[6] have obtained an explicit expression for the representation function of ISO(4) in terms of Bessel functions and $6j$ symbols of SO(3).

The representation function of ISO(4) should be derivable from that of either SO(5) or SO(4, 1) by contraction. However, we find it difficult to derive the result of ISO(4) from Holman's expression for either SO(5) or SO(4, 1) by contraction, because terms containing infinity do not disappear. Thus we are led to a further search for an explicit expression for the representation functions of SO(5) and SO(4, 1) such that they are contractable to the representation function of ISO(4). If such an expression can be found, then it would lead to a unified treatment for the representation functions of SO(4, 1), SO(5), and ISO(4). In what follows we show that this unified treatment is indeed possible.

As we have mentioned before, this unified treatment must start with the representation function of SO(4, 1). Again we use Wolf's formula as given in Eq. (2.3). For the d-functions of SO(4) on the right-hand side, we use the Weyl coefficients of SO(4) as discussed by Wong[5] and the d-function of SO(3). Thus we write

$$
D_{34}(\theta) = W_{23}^{-1} W_{34}^{-1} D_{23}(\theta) W_{34} W_{23}. \tag{4.1}
$$

Next we use Vilenkin's formula (Ref. 16, p. 117) for the d-function of SO(3) (with a change of phase so as to make the d-function real)

$$
d^{J}_{mn}(z) = (-1)^{-m}\left[\frac{(J-m)!\,(J-n)!}{(J+m)!\,(J+n)!}\right]\left(\frac{1+z}{1-z}\right)^{(1/2)(m+n)}
$$

$$
\times \sum_{j} \frac{(-1)^{j}(J+j)!}{(J-j)!\,(j-m)!\,(j-n)!}\left(\frac{1-z}{2}\right)^{j}, \tag{4.2}
$$

where $z = \cos\theta$.

For the integration over θ, we use the following transformation:

$$
\begin{aligned}
1+z' &= \exp(-\zeta)\,(1+z)(\cosh\zeta - z\sinh\zeta)^{-1}, \\
1-z' &= \exp(\zeta)\,(1-z)\,(\cosh\zeta - z\sinh\zeta)^{-1}, \\
(\cosh\zeta - z\sinh\zeta)^{-1} &= \sinh^{-1}\zeta\left(\frac{1+\exp(-2\zeta)}{1-\exp(-2\zeta)} - z\right)^{-1}.
\end{aligned} \tag{4.3}
$$

Finally, putting $z = -y + 1$, we can evaluate the integral by Eq. (3.259.2) of Gradshteyn and Ryzhik[17]

$$
\int_0^u y^{\nu-1}(u-y)^{\mu-1}(y^m+\beta^m)^{\lambda}\,dy = \beta^{m\lambda}u^{\mu+\nu-1}B(\mu,\nu)
$$

$$
\times {}_{m+1}F_{m}(-\lambda, \nu/m, \ldots, (\nu+m-1)/m;\ (\mu+\nu)/m,
$$

$$
\ldots, (\mu+\nu+m-1)/m;\ (-u/\beta)^m) \tag{4.4}
$$

Since $m = 1$ in (4.4), we obtain a hypergeometric function ${}_2F_1$ with argument $1-e^{2\zeta}$. This ${}_2F_1$ function can be transformed to other ${}_2F_1$ functions with different arguments. Among other cases, we can obtain a hypergeometric function ${}_2F_1$ with argument $1-e^{-2\zeta}$.

The explicit expression for the representation function of SO(4, 1) is as follows

$$
{}^{P}d^{\lambda\, m_{25}}_{\substack{m_{14} m_{24} m'_{14} m'_{24} \\ m_{13}}}(\zeta)
$$

$$
= \frac{[(m_{14}+m_{24}+1)(m_{14}-m_{24}+1)(m'_{14}+m'_{24}+1)(m'_{14}-m'_{24}+1)]^{1/2}}{(2m_{13}+1)(2m_{25}+1)\,\Gamma(\tfrac{3}{2})\,\pi^{1/2}} \sum_{m_{12}} \int_0^{\pi} \sin^2\theta\, d\theta\, \overline{d^{\,m_{14} m_{24}}_{\substack{m_{25} m_{13} \\ m_{12}}}(\theta)} \left(\frac{\sin\theta}{\sin\theta'}\right)^{\lambda} d^{\,m'_{14} m'_{24}}_{\substack{m_{25} m_{13} \\ m_{12}}}(\theta') \tag{4.5}
$$

$$
= N \sum_{\substack{m_{12}, m'_{12}, m''_{12}, m'''_{12} \\ m''''_{12}, m'_{25}, m''_{25}}} \overline{(W_{23}^{-1})^{m_{25}}_{m_{12} m'_{12}} (W_{34}^{-1})^{m_{14} m_{24}}_{\substack{m_{25} m'_{25} \\ m'_{12}}}}\ \overline{(W_{34})^{m_{14} m_{24}}_{\substack{m'_{25} m_{13} \\ m''_{12}}} (W_{23})^{m_{13}}_{m''_{12} m_{12}}}\ (W_{23}^{-1})^{m_{25}}_{m_{12} m'''_{12}} (W_{34}^{-1})^{m'_{14} m'_{24}}_{\substack{m_{25} m''_{25} \\ m'''_{12}}} (W_{34})^{m'_{14} m'_{24}}_{\substack{m''_{25} m_{13} \\ m''''_{12}}}
$$

$$
\times (W_{23})^{m_{13}}_{m''''_{12} m_{12}} (-1)^{-m'_{12}-m'''_{12}} \sum_{j,j'} \left(\frac{(m'_{25}-m'_{12})!\,(m'_{25}-m''_{12})!\,(m''_{25}-m'''_{12})!\,(m''_{25}-m''''_{12})!}{(m'_{25}+m'_{12})!\,(m'_{25}+m''_{12})!\,(m''_{25}+m'''_{12})!\,(m''_{25}+m''''_{12})!}\right)^{1/2}
$$

$$
\times \frac{(-1)^{j+j'}(m'_{25}+j)!\,(m''_{25}+j')!}{(m'_{25}-j)!\,(j-m'_{12})!\,(j-m''_{12})!\,(m''_{25}-j')!\,(j'-m'''_{12})!\,(j'-m''''_{12})!}\, 2^2
$$

$$
\times \exp(\lambda - m'''_{12} - m''''_{12})\zeta\, \frac{\Gamma(\tfrac{3}{2}+m'_{12}+m''_{12}+m'''_{12}+m''''_{12})\,\Gamma(\tfrac{3}{2}-m'_{12}-m''_{12}-m'''_{12}-m''''_{12}+j+j')}{\Gamma(3+j+j')}
$$

$$
\times {}_2F_1(j'-\lambda, \tfrac{3}{2}+\tfrac{1}{2}(m'_{12}+m''_{12}+m'''_{12}+m''''_{12});\ 3+j+j';\ 1-e^{-2\zeta}), \tag{4.6}
$$

where

$$N=\frac{[(m_{14}+m_{24}+1)(m_{14}-m_{24}+1]^{1/2}}{(2m_{13}+1)(2m_{25}+1)\Gamma(\frac{3}{2})\pi^{1/2}}[(m'_{14}+m'_{24}+1)(m'_{14}-m'_{24}+1)]^{1/2} \tag{4.7}$$

and W_{ij}, W^{-1}_{ij} are the Weyl coefficients discussed by Wong.[5] For W_{34}, e.g., see Eq. (4.1) of Ref. 5.

Let us now make an important observation with regard to (4.6). The only terms containing λ, which will go to infinity in the contraction process, are found in the exponential function and the first term of the ${}_2F_1$ function. Thus the contraction process will give meaningful results.

From (4.6) we obtain an explicit evaluation of the representation functions of SO(5) and ISO(4). The procedure is as follows.

For the representation function of SO(5): Replace λ by m_{15}, ζ by $-i\delta$, and multiply (4.6) by W_1, i.e.,

$$\sum_{j=1}^{2}\left[\frac{\Gamma(m_{15}+m_{j4}+4-j)\ \Gamma(m'_{j4}-m_{15}-j+1)}{\Gamma(m_{j4}-m_{15}-j+1)\ \Gamma(m_{15}+m'_{j4}+4-j)}\right]^{1/2}.$$

The explicit expression for the d-function of SO(5) is

$$\begin{aligned} d^{m_{15}m_{25}}_{m_{14}m_{24}m'_{14}m'_{24}\,m_{13}}(\delta) = N &\sum_{\substack{m_{12},m'_{12},m''_{12},m'''_{12},\\ m''''_{12},m'_{25},m''_{25}}} W_1 WM\exp[-i(m_{15}-m'''_{12}-m''''_{12})]\delta \\ &\times\frac{\Gamma(\frac{3}{2}+m'_{12}+m''_{12}+m'''_{12}+m''''_{12})\ \Gamma(\frac{3}{2}-m'_{12}-m''_{12}-m'''_{12}-m''''_{12}+j+j')}{\Gamma(3+j+j')} \\ &\times{}_2F_1(j'-m_{15},\ \tfrac{3}{2}+\tfrac{1}{2}(m'_{12}+m''_{12}+m'''_{12}+m''''_{12});\ 3+j+j';\ 1-e^{2i\delta}) \end{aligned} \tag{4.8}$$

where

$$W=\overline{(W^{-1}_{23})^{m_{25}}_{m_{12}m'_{12}}\ (W^{-1}_{34})^{m_{14}m_{24}}_{m_{25}m'_{25}\,m'_{12}}\ (W_{34})^{m_{14}m_{24}}_{m_{25}m'_{25}\,m''_{12}}\ (W_{23})^{m_{13}}_{m''_{12}m_{12}}}\ (W^{-1}_{23})^{m_{25}}_{m_{12}m'''_{12}}\ (W^{-1}_{34})^{m'_{14}m'_{24}}_{m_{25}m''_{25}\,m'''_{12}}\ (W_{34})^{m'_{14}m'_{24}}_{m''_{25}m_{13}\,m''''_{12}}\ (W_{23})^{m_{13}}_{m''''_{12}m_{12}}, \tag{4.9}$$

$$\begin{aligned} M &= (-1)^{-m'_{12}-m'''_{12}} \\ &\times\sum_{j,j'}\left[\frac{(m'_{25}-m'_{12})!\,(m'_{25}-m''_{12})!\,(m''_{25}-m'''_{12})!\,(m''_{25}-m''''_{12})!}{(m'_{25}+m'_{12})!\,(m'_{25}+m''_{12})!\,(m''_{25}+m'''_{12})!\,(m''_{25}+m''''_{12})!}\right]^{1/2}\frac{(-1)^{j+j'}(m'_{25}+j)!\,(m''_{25}+j')!}{(m'_{25}-j)!\,(j-m'_{12})!\,(j-m''_{12})!\,(m''_{25}-j')!\,(j'-m'''_{12})!\,(j'-m''''_{12})!}\,2^2. \end{aligned} \tag{4.10}$$

For the representation function of ISO(4): Replace $\exp(\lambda-m'''_{12}-m''''_{12})\zeta$ by $\exp(-i\gamma\xi)$, and replace ${}_2F_1(j'-\lambda,\ \frac{3}{2}+\frac{1}{2}(m'_{12}+m''_{12}+m'''_{12}+m''''_{12});\ 3+j+j';\ 1-e^{-2\zeta})$ by ${}_1F_1(\frac{3}{2}+\frac{1}{2}(m'_{12}+m''_{12}+m'''_{12}+m''''_{12});\ 3+j+j';2i\gamma\xi)$. The explicit expression for the d-function of ISO(4) is as follows:

$$\begin{aligned} {}^{I}d^{\gamma\,m_{25}}_{m_{14}m_{24}m'_{14}m'_{24}\,m_{13}}(\xi) = N &\sum_{\substack{m_{12},m'_{12},m''_{12},m'''_{12},\\ m''''_{12},m'_{25},m''_{25}}} WM\exp(-i\gamma\xi)\ \frac{\Gamma(\frac{3}{2}+m'_{12}+m''_{12}+m'''_{12}+m''''_{12})\,\Gamma(\frac{3}{2}-m'_{12}-m''_{12}-m'''_{12}-m''''_{12}+j+j')}{\Gamma(3+j+j')} \\ &\times{}_1F_1(\tfrac{3}{2}+\tfrac{1}{2}(m'_{12}+m''_{12}+m'''_{12}+m''''_{12});\ 3+j+j';\ 2i\gamma\xi). \end{aligned} \tag{4.11}$$

If one compares the results for SO(4, 1), SO(5), and ISO(4) with those of SO(3, 1), SO(4), and ISO(3) obtained in the previous section, one finds already a certain regularity in all the expressions. That is, the d-functions of SO(3, 1) and SO(4, 1) are expressible as a summation over ${}_2F_1$ functions with argument $1-e^{-2\zeta}$. The d-functions of SO(4) and SO(5) are expressible as a summation over ${}_2F_1$ functions with argument $1-e^{2i\delta}$. Finally the d-functions of ISO(3) and ISO(4) are expressible as a summation over ${}_1F_1$ functions with argument $2i\gamma\xi$. We shall show in the following section that this result can be generalized to all SO(n, 1), SO($n+1$), and ISO(n).

We showed in a previous paper[6] that the d-function of ISO(4) can be written as

$$\begin{aligned} {}^{I}d^{\gamma\,m_{25}}_{m_{14}m_{24}m'_{14}m'_{24}\,m_{13}}(\xi) &= \sum_{n=0}^{\infty}(-1)^{m'_{14}+m'_{24}+n+m_{14}-m_{24}+m_{25}+m_{13}} \\ &\times[(m_{14}+m_{24}+1)(m_{14}-m_{24}+1)(m'_{14}+m'_{24}+1)(m'_{14}-m'_{24}+1)]^{1/2} \\ &\times 2\pi(n+1)(\gamma\xi)^{-1} \\ &\times J_{n+1}(\gamma\xi)\begin{Bmatrix}\frac{1}{2}(m'_{14}-m'_{24}) & \frac{1}{2}(m_{14}-m_{24}) & n/2\\ \frac{1}{2}(m_{14}+m_{24}) & \frac{1}{2}(m'_{14}+m'_{24}) & m_{25}\end{Bmatrix} \\ &\times\begin{Bmatrix}\frac{1}{2}(m'_{14}-m'_{24}) & \frac{1}{2}(m_{14}-m_{24}) & n/2\\ \frac{1}{2}(m_{14}+m_{24}) & \frac{1}{2}(m'_{14}+m'_{24}) & m_{13}\end{Bmatrix}. \end{aligned} \tag{4.12}$$

Again we can ask the question: How can one show directly that (4.11) is equal to (4.12)? This seems to be quite a difficult problem. So far we have not been able to find a direct proof.

V. GENERALIZATION TO SO(n,1), SO($n+1$), AND ISO(n)

We shall now generalize the procedures used in the previous sections to obtain an explicit expression for the representation functions of all SO(n, 1), SO($n+1$), and ISO(n).

We start with the d-function of SO(n, 1) as given in Eq. (2.3). For the two d-functions of SO(n) on the right-hand side, we use the Weyl coefficients of SO(p), $p=3, 4, \ldots, n$, and the d-functions of SO(3). Thus

$$D_{n-1,n}(\theta) = W^{-1}_{n-2,n-1} W^{-1}_{n-1,n} \cdots W^{-1}_{23} W^{-1}_{34} D_{23}(\theta) W_{34} W_{23} \cdots W_{n-1n} W_{n-2,n-1}. \tag{5.1}$$

In terms of state labels, it is understood that each term on the right-hand side of Eq. (5.1) is a matrix, and the multiplication of these matrices follows the standard procedure of matrix multiplication. As an example, see Eq. (4.6) or Eq. (4.8) and (4.9).

For the integration over θ, we find that the case treated in Sec. IV for SO(4, 1) is already general enough to be directly applicable for all SO(n, 1), $n \geq 4$. The only minor modification one has to make is the following:

For SO(4, 1)	For SO(n, 1)
2^2	2^{n-2}
$\dfrac{\Gamma(\frac{3}{2}+a)\Gamma(\frac{3}{2}+b)}{\Gamma(3+j+j')}$	$\dfrac{\Gamma(\frac{1}{2}(n-1)+a)\Gamma(\frac{1}{2}(n-1)+b)}{\Gamma(n-1+j+j')}$
${}_2F_1(j'-\lambda, \frac{3}{2}+a/2; 3+j+j'; 1-e^{-2\zeta})$	${}_2F_1(j'-\lambda, \frac{1}{2}(n-1)+a/2; n-1+j+j'; 1-e^{-2\zeta})$
N	appropriate normalization factor in accordance with (2.3), i.e., the factor in front of the summation sign on the right-hand side of (2.3)
$D_{34}(\theta)$	$D_{n-1,n}(\theta)$ as in (5.1). (5.2)

With the replacement given in (5.2), we find that the explicit expression for the representation functions of SO(n, 1) are basically of the same form as (4.6). The evaluation of the representation function of SO(n, 1) can then be achieved by an iteration process as follows: We assume that the representation function of SO($n-1$, 1) is known. From it we obtain the representation function of SO(n). Therefore the Weyl coefficients of SO(p), $p=3, 4, \ldots, n$ are all known. This is all the information we need to evaluate Eq. (2.3). But we certainly know the representation function of SO(3, 1), i.e., Eq. (3.1). Therefore the iteration process can be started and carried through for all n.

After the representation function of SO(n, 1) has been obtained, the representation function of SO($n+1$) is obtained from it by analytic continuation. Finally, the representation function of ISO(n) is obtained from either of the two above by contraction. This contraction process is meaningful only if terms containing infinity disappear. We find that this is indeed the case in the present situation. Thus we have obtained a unified treatment for an explicit evaluation of the representation functions of all SO(n, 1), SO($n+1$), and ISO(n). We summarize our results as follows:

(1). The representation functions of SO(n, 1) are expressible as a sum over hypergeometric functions with argument $1-e^{-2\zeta}$, multiplied by Weyl coefficients of SO(p), $p=3, 4, \ldots, n$. The explicit expression is given in (4.6) with modification given in (5.2).

(2). The representation functions of SO($n+1$) are obtained from those of SO(n, 1) by the replacement

$\zeta \to i\delta$, $\lambda \to m_{1n+1}$, and multiplication of W_1, i.e.,

$$\left(\prod_{j=1}^{[(n-1)/2]} \frac{\Gamma(\lambda+m_{j,n-1}+n-j-1)\Gamma(m'_{jn-1}-\lambda-j+1)}{\Gamma(m_{jn-1}-\lambda-j+1)\Gamma(\lambda+m'_{jn-1}+n-j-1)}\right)^{1/2}$$

(3). The representation functions of ISO(n) are obtained from those of SO(n, 1) by the replacement

$$\exp(\lambda - m''_{12} - m'''_{12})\zeta \to \exp(-i\gamma\xi)$$

and

$${}_2F_1(j'-\lambda, a; b; 1-e^{-2\zeta}) \to {}_1F_1(a; b; 2i\gamma\xi).$$

ACKNOWLEDGMENT

We wish to thank Moorhead State University for financial support of this work.

[1] M. K. F. Wong and H. Y. Yeh (to be published).
[2] I. M. Gel'fand and M. I. Graev, Izv. Akad. Nauk SSSR Mekh. Mashinostr. 29, 1329 (1965) [Am. Math. Soc. Transl. Ser. 2, 65, 116 (1967)].
[3] K. B. Wolf, J. Math. Phys. 12, 197 (1971).
[4] N. Ja Vilenkin, Dokl. Akad. Nauk SSSR 113, 16 (1957).
[5] M. K. F. Wong, J. Math. Phys. 19, 703 (1978).
[6] M. K. F. Wong and H. Y. Yeh, J. Math. Phys. 21, 1 (1980).
[7] Ya. A. Smorodinskii and G. I. Shepelev, Yad. Fiz. 13, 441 (1971) [Sov. J. Nucl. Phys. 13, 248 (1971)].
[8] M. K. F. Wong and H. Y. Yeh, J. Math. Phys. 18, 1768 (1977).

[9]M. A. Rashid, J. Math. Phys. **20**, 1514 (1979).
[10]T. Maekawa, J. Math. Phys. **16**, 2372 (1975).
[11]A. A. Makarov and G. I. Shepelev, Yad. Fiz. **12**, 1092 (1970) [Sov. J. Nucl. Phys. **12**, 596 (1971)], and references therein.
[12]A. O. Barut and R. Wilson, J. Math. Phys. **17**, 900 (1976).
[13]See, e.g., W. Miller, Jr., *Lie Theory and Special Functions* (Academic, New York, 1968).
[14]W. J. Holman, III, J. Math. Phys. **10**, 1710 (1969).
[15]W. J. Holman, III, J. Math. Phys. **10**, 1888 (1969).
[16]N. Ja Vilenkin, *Special Functions and the Theory of Group Representations*, *Mathematical Monographs Vol. 22*, translated by V. N. Singh (American Mathematical Society, Providence, RI, 1968).
[17]I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic, New York, 1965).

Tensor products of positive energy representations of $\tilde{S}O(3,2)$ and $\tilde{S}O(4,2)$

W. Heidenreich
Max-Planck-Institut zur Erforschung der Lebensbedingungen der wissenschaftlich-technischen Welt, D-813 Starnberg, West Germany

(Received 27 May 1980; accepted for publication 26 November 1980)

The Clebsch–Gordan series of the tensor product of a wide class of unitary irreducible positive energy representations of the universal covering groups of SO(3,2) and SO(4,2) are calculated by comparing weight diagrams.

PACS numbers: 02.20.Qs, 02.20.Rt

I. INTRODUCTION

The assignment of elementary particles to representations of the Poincaré group is generally accepted. In contrast, proposals to describe the space-time behavior of particles with representations of the universal covering of the de Sitter group SO(3,2) or the conformal group SO(4,2) are less well known.

To obtain experimentally-testable predictions on, for example, the S-matrix of a theory with one of these invariance groups, one needs the Clebsch–Gordan (CG) series of the tensor product of the physically relevant representations. Only very few cases are known explicitly.[1–3] This paper attempts to close the gap.

To achieve this I use the method of comparing weight diagrams. In order to illustrate it I will consider the simple case of the group SU(2). The third component of the "angular momentum" generates a U(1) subgroup of SU(2). Any representation D^l of SU(2) can be decomposed into a sum of U(1) representations $e^{i\alpha m}$, abbreviated by the weight (m), which has multiplicity one in this case:

$$D(-l) \equiv \sum_{m=-l}^{+l} (m).$$

The lowest weight $(-l)$ labels the representations uniquely. The weight diagram of the tensor product of two representations is

$$D(-l) \otimes D(-l') \equiv \sum_{m=-l}^{+l} (m) \otimes \sum_{m'=-l'}^{+l'} (m') = \sum_{m=-l}^{+l} \sum_{m'=-l'}^{+l'} (m+m'),$$

where the CG series of U(1) is used: $(m) \otimes (m') = (m+m')$. The lowest weight $(-l-l')$ must be the lowest weight of a representation of the CG series. Taking away all weights of this representation and looking for the lowest weight of the rest, one gets successively the whole CG series. That is an easy way to calculate particular cases.

The general formula can be obtained by changing the sum-indices

$$\sum_{m=-l}^{+l} \sum_{m'=-l'}^{+l'} (m+m') = \sum_{L=|l-l'|}^{l+l'} \sum_{m+m'=-L}^{+L} (m+m').$$

So $\Sigma^{l+l'}_{L=|l-l'|} D^L$ has the same weight diagram as $D^l \times D^{l'}$.

The above step-by-step construction shows that the reduction of the weight diagram is unique; therefore it must be the CG series. In the following the notation will not distinguish between representation D^l and weight diagram $D(-l)$.

Preconditions for this method are the existence of a unique lowest weight for representations of a certain type, and the knowledge of all weights and their multiplicities of all those representations which can occur in the CG series. If unknown representations occur in the direct product, the method will exhibit this fact.

The finiteness of the representations is not a precondition. For the unitary irreducible representations with positive energy of $\tilde{S}O(3,2)$ and $\tilde{S}O(4,2)$ these conditions are fulfilled. Positive energy here means positive eigenvalues of the SO(2) generator in the maximal essentially compact subgroup. These include the candidates for physical interpretations.[3,4]

The rest of the paper is organized as follows:

In Sec. IIA all weights of these representations of $\tilde{S}O(3,2)$ will be quoted in a form which is adjusted to the present problem.

In Sec. IIB and IIC more and more complex tensor products will be reduced.

I try to sketch the graphical method which in all cases lead to a guess for the reduction, which is proved algebraically. I don't believe that the algebraic proof is much help for finding the solutions.

In Sec. III the same scheme is followed for $\tilde{S}O(4,2)$.

II. TENSOR PRODUCTS OF $\tilde{S}O(3,2)$-REPRESENTATIONS

A. All unitary irreducible representations with positive energy

Evans[5] lists all representations of this type. I repeat them here in a form which is adapted to the present purposes. (E,l) symbolizes the $(2l+1)$-dim space of a $\tilde{S}O(2) \times SU(2)$ representation; $D(E,l)$ is the space of a $\tilde{S}O(3,2)$ representation with the lowest weight (E,l). Lowest weight here means the $\tilde{S}O(2) \times SU(2)$ representation which contains the lowest weight $(E, l_3 = -l)$. In analogy to $\tilde{S}O(4,2)$, I call $t = E - l$ the twist of the representation.

The twist 1/2 representations are the Dirac singletons[6,7]:

$$D(\tfrac{1}{2},0) = \sum_{S=0}^{\infty} (\tfrac{1}{2}+S,S),$$

$$D(1,\tfrac{1}{2}) = \sum_{S=0}^{\infty} (1+S,\tfrac{1}{2}+S). \tag{1}$$

 0022-2488/81/081566-09$01.00 © 1981 American Institute of Physics

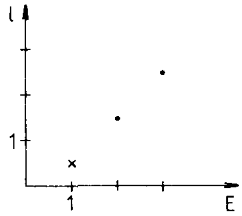


FIG. 1. Weights of $D(1,\frac{1}{2})$.

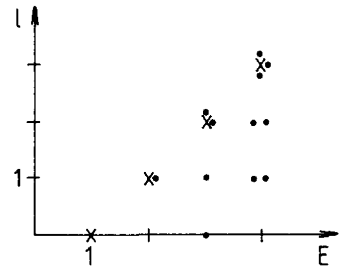


FIG. 3. Weights of $D(\frac{1}{2},0)\otimes D(\frac{1}{2},0)$.

For the other weight diagrams I use the abbreviations

$$\Delta(E,l)=\sum_{S=0}^{\infty}L(E+S,l+S)=\sum_{S=0}^{\infty}\sum_{n=0}^{\infty}(E+S+n,l+S),$$
$$\Delta'(E,l)=\sum_{S=0}^{\infty}L'(E+S,l+S)=\sum_{S=0}^{\infty}\sum_{n=0}^{\infty}{}'(E+S+n,l+S), \tag{2}$$

where

$$\sum_{n=c}^{\infty}{}'a_n=\sum_{n=c}^{\infty}a_{2n-c}.$$

Now the twist-1 representations have the following weights:

$$D(1,0)=\Delta'(1,0),\qquad D(l+1,l)=\Delta(l+1,l)\quad(l=\tfrac{1}{2},1,\tfrac{3}{2},\cdots). \tag{3}$$

For $l=0,1/2$ and $t>1/2$ Evans gets:

$$D(E,0)=\Delta'(E,0),\ D(E,\tfrac{1}{2})=\Delta(E,\tfrac{1}{2}), \tag{4}$$

and for $l=1,3/2,\cdots,t>1$ he has for

$$l\text{ integer, }D(E,l)=\sum_{S=0}^{l-1}\Delta(E+S,l-S)\oplus\Delta'(E+l,0),$$
$$l\text{ half-integer, }D(E,l)=\sum_{S=0}^{l-\frac{1}{2}}\Delta(E+S,l-S). \tag{5}$$

For two typical weight diagrams see Figs. 1 and 2, where the lowest weights are given by crosses ($\times$), whereas the rest of the weights are given by dots ($\cdot$) or circles (O), classified by their triangle (Δ) structure [see Eq. (5)].

B. Tensor products with Dirac singletons

I use one of these cases to demonstrate how the method works for infinite representations. Using the CG series of SU(2), one gets

$$D(\tfrac{1}{2},0)\otimes D(\tfrac{1}{2},0)=\sum_{S=0}^{\infty}\sum_{S'=0}^{\infty}\sum_{l=|S-S'|}^{S+S'}(S+S'+1,l). \tag{6}$$

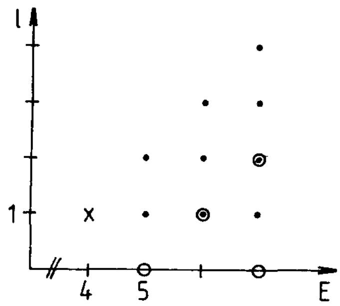


FIG. 2. Weights of $D(4,1)$.

Some of the weights are shown in Fig. 3. The critical reader is invited to check how the first few terms of the CG series are obtained. The lowest weight must be the lowest weight of the first term. Take away all weights of this representation and start again. The lowest weights of the representations obtained (crosses in Fig. 3) lie on the line with twist 1. The full series of the tensor product of two twist 1/2 representations is described by the following

Theorem 1:

$$D(\tfrac{1}{2},0)\otimes D(\tfrac{1}{2},0)=\sum_{S=0}^{\infty}D(S+1,S),$$
$$D(1,\tfrac{1}{2})\otimes D(\tfrac{1}{2},0)=\sum_{S=0}^{\infty}D(S+\tfrac{3}{2},S+\tfrac{1}{2}), \tag{7}$$
$$D(1,\tfrac{1}{2})\otimes D(1,\tfrac{1}{2})=D(2,0)\oplus\sum_{S=1}^{\infty}D(S+1,S).$$

This is the result of Flato and Fronsdal.[2]

Proof (by induction):

For $D(\frac{1}{2},0)\otimes D(\frac{1}{2},0)$.

The starting values $E=1,2$ are already proved graphically in Fig. 3.

Left-hand side (lhs):

$$D(\tfrac{1}{2},0)\otimes D(\tfrac{1}{2},0)=\sum_{N=0}^{\infty}\sum_{S-S'=-N}^{+N}\sum_{j=|S-S'|}^{N}(N+1,j)\text{ with }N=S+S'.$$

In going from E to $E+2$ (N to $N+2$) the following weights in j and their multiplicities also occur:

$$(N+1)(N+3,N+1)\oplus(N+3)(N+3,N+2).$$

Right-hand side (rhs):

$$\sum_{S=0}^{\infty}D(S+1,S)=\sum_{j=0}^{\infty}\sum_{n=0}^{\infty}{}'(j+n+1,j)\oplus\sum_{S=1}^{\infty}\sum_{j=S}^{\infty}\sum_{n=0}^{\infty}(j+n+1,j)$$
$$=\Big(\sum_{N=0}^{\infty}{}'\sum_{j=0}^{N}\oplus\sum_{N-1=0}^{\infty}{}'\sum_{j=1}^{N}\Big)(N+1,j)\oplus\sum_{N=1}^{\infty}\sum_{j=1}^{N}j(N+1,j).$$

From this one gets for the additional terms in j the same terms as for the lhs.

The other cases are proved similarly.

To reduce the tensor product of a Dirac singleton and any other representation I found it most convenient to decompose the latter into triangles Δ, and lines L, as was done in Sec. IIA.

Lemma 2: The tensor product of a Dirac singleton and a line has the weight diagram

$$
\begin{aligned}
&D(\tfrac{1}{2},0)\otimes L(E,l)=D(E+\tfrac{1}{2},l)\oplus D(E+\tfrac{3}{2},l),\\
&D(\tfrac{1}{2},0)\otimes L'(E,l)=D(E+\tfrac{1}{2},l),\\
&D(1,\tfrac{1}{2})\otimes L(E,l)=D(E+1,l+\tfrac{1}{2})\oplus D(E+1,l-\tfrac{1}{2}),\\
&D(1,\tfrac{1}{2})\otimes L'(E,l)=\Delta'(E+1,l+\tfrac{1}{2})\oplus D(E+1,l-\tfrac{1}{2}).
\end{aligned}
\tag{8}
$$

If the SU(2) eigenvalue of a term is smaller than 0 omit it.

A more general formula will be proved in Sec. IIC. From Lemma 2 one gets easily

Theorem 3: The CG series of the tensor product of a twist 1/2 and a twist-1 representation is for $l>0$,

$$
\begin{aligned}
&D(\tfrac{1}{2},0)\otimes D(l+1,l)\\
&\quad=\sum_{S=0}^{\infty}\{D(l+S+\tfrac{3}{2},l+S)\oplus D(l+S+\tfrac{5}{2},l+S)\},\\
&D(1,\tfrac{1}{2})\otimes D(l+1,l)\\
&=\sum_{S=0}^{\infty}\{D(l+S+2,l+S+\tfrac{1}{2})\oplus D(l+S+2,l+S-\tfrac{1}{2})\},
\end{aligned}
\tag{9}
$$

and for $E=1$,

$$
\begin{aligned}
&D(\tfrac{1}{2},0)\otimes D(E,0)=\sum_{S=0}^{\infty}D(E+S+\tfrac{1}{2},S),\\
&D(1,\tfrac{1}{2})\otimes D(E,0)=\sum_{S=0}^{\infty}D(E+S+1,S+\tfrac{1}{2}).
\end{aligned}
\tag{10}
$$

The CG series includes in the case $l=0$ only twist 3/2 representations, in the other cases twist 3/2 and twist 5/2.

Proof: The first three formulas are straightforward, e.g., the first,

$$
\begin{aligned}
D(\tfrac{1}{2},0)\otimes\Delta(E,l)&=\sum_{S=0}^{\infty}\{D(\tfrac{1}{2},0)\otimes L(E+S,l+S)\}\\
&=\sum_{S=0}^{\infty}\{D(E+S+\tfrac{1}{2},l+S)\oplus D(E+S+\tfrac{3}{2},l+S)\}.
\end{aligned}
$$

Just put $E=l+1$.

The last formula needs some rearranging of Δ'

$$
\begin{aligned}
D(1,\tfrac{1}{2})\otimes\sum_{S=0}^{\infty}L'(E+S,S)&=\sum_{S=0}^{\infty}\{\Delta'(E+S+1,S+\tfrac{1}{2})\\
&\oplus\sum_{j=0}^{S-1}[\Delta'(E+S+j+1,S-j-\tfrac{1}{2})\\
&\oplus\Delta'(E+S+j+2,S-j-\tfrac{1}{2})]\}\\
&=\sum_{S=0}^{\infty}D(E+S+1,S+\tfrac{1}{2}).
\end{aligned}
$$

Substituting $E=1$ by $E>1/2$ in Eq. (10), one gets the general product of a Dirac singleton, and a spinless representation. A similar shift of the energy in Eq. (9) yields the result for $l=1/2$. For higher spins one has to use the decomposition [Eq. (5)] into traiangles. A straightforward calculation gives for $l\geqslant 1$, $E>l+1$,

$$
D(\tfrac{1}{2},0)\otimes D(E,l)=\left.\begin{matrix}\sum_{j=0}^{l-1}\\ \sum_{j=0}^{l-\frac{1}{2}}\end{matrix}\right\}\sum_{S=0}^{\infty}\{D(E+j+S+\tfrac{1}{2},l-j+S)\oplus D(E+j+S+\tfrac{3}{2},l-j+S)\}\left\{\begin{matrix}\oplus\sum_{S=0}^{\infty}D(E+l+S+\tfrac{1}{2},S)\\ \oplus 0\end{matrix}\right.,
$$

$$
\begin{aligned}
&D(1,\tfrac{1}{2})\otimes D(E,l)\\
&=\left.\begin{matrix}\sum_{j=0}^{l-1}\\ \sum_{j=0}^{l-\frac{1}{2}}\end{matrix}\right\}\sum_{S=0}^{\infty}\{D(E+j+S+1,l-j+S-\tfrac{1}{2})\oplus D(E+j+S+1,l-j+S+\tfrac{1}{2})\}\left\{\begin{matrix}\oplus\sum_{S=0}^{\infty}D(E+l+S+1,S+\tfrac{1}{2})\\ \oplus 0\end{matrix}\right.,
\end{aligned}
\tag{11}
$$

where the upper part of each formula holds for l integer, the lower for l half-integer.

The CG series contains twists between $t=E-l+1/2$ and $t=E+l+1/2$. Fig. 4 shows a typical example of this type. The tensor product of three Dirac singletons can be reduced by combining Eq. (7) with Eqs. (9) and (10). Again only representations with twist 3/2 and 5/2 occur, most of them more than once.

$$
\begin{aligned}
&[D(\tfrac{1}{2},0)]^3=\sum_{S=0}^{\infty}\{(S+1)D(S+\tfrac{3}{2},S)\oplus(S+1)D(S+\tfrac{5}{2},S)\},\\
&[D(\tfrac{1}{2},0)]^2\otimes D(1,\tfrac{1}{2})\\
&=\sum_{S=0}^{\infty}\{(S+1)D(S+2,S+\tfrac{1}{2})\oplus(S+1)D(S+3,S+\tfrac{1}{2})\},\\
&D(\tfrac{1}{2},0)\otimes(D(1,\tfrac{1}{2}))^2\\
&\quad=\sum_{S=0}^{\infty}\{(S+1)D(S+\tfrac{5}{2},S)\oplus(S+1)D(S+\tfrac{5}{2},S+1)\},
\end{aligned}
\tag{12}
$$

$$
\begin{aligned}
&[D(1,\tfrac{1}{2})]^3\\
&=\sum_{S=0}^{\infty}\{(S+1)D(S+3,S+\tfrac{3}{2})\oplus(S+2)D(S+3,S+\tfrac{1}{2})\}.
\end{aligned}
$$

The CG series for the tensor product of four Dirac singletons can be obtained from Eq. (12) and Eq. (11). The first few terms in one case are shown in Fig. 5. All twists $\geqslant 2$ occur. The multiplicity of the representations diverges for $E\to\infty$ for both t or l fixed.

C. The tensor product of twist > 1/2 representations

To reduce the tensor product of two twist 1 representations, decompose one into lines, as was done in Sec. II A, the other one into diagonals

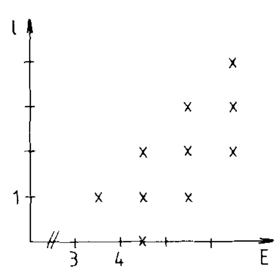


FIG. 4. Lowest weights in $D(\tfrac{1}{2},0)\otimes D(3,1)$.

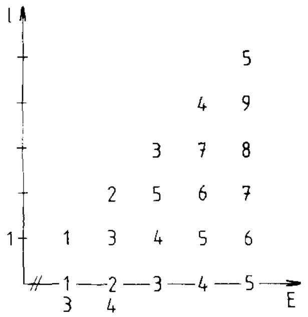


FIG. 5. Multiplicity of the lowest weights in the product $[D(\frac{1}{2},0)]^2 \otimes [D(1,\frac{1}{2})]^2$.

$$S(E,l) = \sum_{S=0}^{\infty} (E+S, l+S), \tag{13}$$

that is for

$$l>0,\ D(l+1,l) = \sum_{n=0}^{\infty} S(l+1+n,n,l),$$

$$D(1,0) = \sum_{n=0}^{\infty}{}' S(l+1+n,l).$$

For the tensor product of a diagonal and a line I found

Lemma 4:

$$\begin{aligned}
S(E,l)\otimes L(E',l') &= \sum_{j=l-l'}^{l+l'} \Delta(E+E',j) \quad \text{for } l>l',\\
&= \sum_{j=l'-l+2}^{l+l'} \Delta(E+E',j)\\
&\quad \oplus D(E+E',l'-l+1)\oplus D(E+E',l'-l) \quad \text{for } 0<l\leqslant l',\\
&= D(E+E',l')\oplus D(E+E'+1,l') \quad \text{for } l=0;
\end{aligned} \tag{14}$$

$$\begin{aligned}
S(E,l)\otimes L'(E',l') &= \sum_{j=l-l'}^{l+l'} \Delta'(E+E',j) \quad \text{for } l>l',\\
&= \sum_{j=l'-l+1}^{l+l'} \Delta'(E+E',j)\oplus D(E+E',l'-l) \quad \text{for } l\leqslant l'.
\end{aligned} \tag{15}$$

The special cases with $S(\frac{1}{2},0)$ and $S(1,\frac{1}{2})$ were used in Sec. IIB.

Proof: It consists in writing down the sum of weights for both sides and changing the sum indices to make the expressions identical.

lhs:

$$S(E,l)\otimes L'(E',l') = \sum_{n=0}^{\infty}{}' \sum_{S=0}^{\infty} \sum_{j=|S+l-l'|}^{S+l+l'} (E+E'+S+n,j).$$

rhs:

$$\sum_{j=l-l'}^{l+l'} \Delta'(E+E',j) = \sum_{n=0}^{\infty}{}' \sum_{S=0}^{\infty} \sum_{j=S+l-l'}^{S+l+l'} (E+E'+S+n,j)$$

for $l>l'$,

and for $l\leqslant l'$

$$\sum_{j=l'-l+1}^{l'+l} \Delta'(E+E',j)\oplus D(E+E',l'-l) \equiv A.$$

If $l-l'$ is half-integer, the sum means

$$\sum_{l-l'}^{l+l'} a_n = \sum_{n=l-l'-1/2}^{l+l'-1/2} a_{n+1/2}.$$

Decomposing Δ' and D into weights (E,l) (distinguishing between $l'-l$ integer or half-integer), rearranging them, using

$$\sum_{S=0}^{\infty} \sum_{j=0}^{a} = \sum_{j+S=0}^{\infty} \sum_{S-j=|S+j-a|-a}^{S+j}{}',$$

and in the case $l'-l$ half-integer

$$\sum_{j=|S-a+1/2|}^{S+a-1/2}{}' \oplus \sum_{j=|S-a-1/2|}^{S+a+1/2} = \sum_{j=|S-a|-1/2}^{S+a+1/2} \quad \text{for } a\geqslant 0$$

one gets finally,

$$A = \sum_{n=0}^{\infty}{}' \sum_{S'=0}^{\infty} \sum_{j'=|S'-l'+l|}^{S'+l'+l} (E+E'+S'+n,j').$$

So lhs = rhs.

$S\otimes L$ is obtained from this, using $L(E,l) = L'(E,l) \oplus L'(E+1,l)$, $\Delta(E,l) = \Delta'(E,l)\oplus\Delta'(E+1,l)$, and $\Delta(E,l)\oplus D(E+1,l-1) = D(E,l)$.

Graphically I got from Lemma 4 the CG series of the tensor product of two twist one representations very easily. The algebraic proof is somewhat lengthy.

Theorem 5: For $l,l'>0$,

$$\begin{aligned}
D(l+1,l)\otimes D(l'+1,l') &= \sum_{j=|l-l'|}^{l+l'-1} \sum_{n=0}^{\infty} D(l+l'+2+n,j)\\
&\oplus \sum_{S=0}^{\infty} \Big\{ D(l+l'+2+S,l+l'+S)\\
&\oplus \sum_{n=1}^{\infty} 2D(l+l'+2+S+n,l+l'+S)\Big\}; \text{ for } l'>0,
\end{aligned} \tag{16}$$

$$D(1,0)\otimes D(l'+1,l') = \sum_{S=0}^{\infty}\sum_{n=0}^{\infty} D(l'+2+S+n,l'+S),$$

$$D(1,0)\otimes D(1,0) = \sum_{S=0}^{\infty}\sum_{n=0}^{\infty}{}' D(2+S+n,S).$$

The representations have multiplicities 1 or 2, and all twists $\geqslant 2$ occur (Fig. 6).

Proof: I prove a more general formula, with the energy shifted, by substituting E' for $(l'+1)$ and Δ for the D on the left side. It can always be achieved that $l\leqslant l'$. As a first step I calculate the product of a diagonal and a Δ:

$$\begin{aligned}
&S(E,l)\otimes\Delta(E',l')\\
&= \sum_{S=0}^{\infty} \Big\{ \sum_{j=S}^{S+2l-2} \Delta(E+E'+S,l'-l+j+2)\\
&\quad \oplus D(E+E'+S,l-l'+S+1)\\
&\quad \oplus D(E+E'+S,l'-l+S) \Big\}.
\end{aligned}$$

Changing the sum indices of the first term into $j+s$ (odd or even) and $j-s$, and using formulas of the type

$$\sum_{S=0}^{n} \Delta(E+S,l-S)\oplus D(E+n+1,l-n-1) = D(E,l),$$

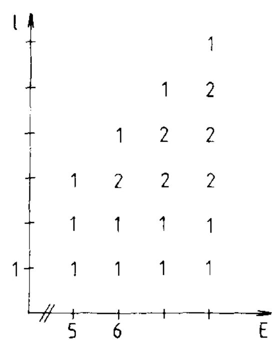


FIG. 6. Multiplicity of the lowest weights in the product $D(3,2)\otimes D(2,1)$.

one gets

$$(S\otimes\Delta) = \sum_{j=0}^{2l} D(E+E',l'-l+j)$$
$$\oplus \sum_{S=0}^{\infty}\{D(E+E'+S+1,l'+l+S+1)$$
$$\oplus D(E+E'+S+1,l'+l+S)\}.$$

Using the decomposition of Δ into diagonals $\Delta(E,l) = \Sigma_{n=0}^{\infty} S(E+n,l)$ one obtains for $0<l\leqslant l'$,

$$\Delta(E,l)\otimes\Delta(E',l') = \sum_{j=0}^{2l-1}\sum_{n=0}^{\infty} D(E+E'+n,l'-l+j)$$
$$\oplus\sum_{S=0}^{\infty}\Big\{D(E+E'+S,l+l'+S)$$
$$\oplus\sum_{n=1}^{\infty} 2D(E+E'+S+n,l+l'+S)\Big\}. \quad (17)$$

So the first case in the theorem is proved.

The two others are straightforward applications of Eqs. (14) and (15).

The general product $D(E,l)\otimes D(E',l')$ can easily be obtained from Eq. (16) shifted in E [see Eq. (17)] and Eq. (5).

III. TENSOR PRODUCTS OF $\widetilde{\mathrm{SO}}(4,2)$ REPRESENTATIONS

A. All unitary irreducible representations with positive energy

All lowest weights (E,J_1,J_2) of the representations discussed here with respect to the maximal essentially compact subgroup $\tilde{\mathrm{U}}(1)\times \mathrm{SU}(2)\times\mathrm{SU}(2)$ have been given by Mack.[8] Again lowest weight here means the $\tilde{\mathrm{U}}(1)\times\mathrm{SU}(2)\times\mathrm{SU}(2)$ representation which contains the lowest weight $(E,J_1^3 = -J_1,J_2^3 = -J_2)$. All weights and their multiplicities of most of the representations were calculated by Yao.[9] However, he missed the representations with twist 2 and $J_1\cdot J_2 = 0$, where the twist is defined as $t = E-(J_1+J_2)$. The weights of these "limit-representations" can be guessed from Mack's analysis of their Poincaré content and from analogy to the $\widetilde{\mathrm{SO}}(3,2)$ case; this guess will be confirmed by the tensor product calculations (see Sec. IIIB). A similar reasoning yields the weights of representations with noninteger twist, which do not occur in Yao's discussion, as he restricts himself to the spin-covering group SU(2,2) of SO(4,2). Combining all these informations, the weight diagrams are

for twist 1,

$$D(1+J_1,J_1,0) = \sum_{S=0}^{\infty}\left(1+J_1+S,J_1+\frac{S}{2},\frac{S}{2}\right),$$
$$D(1+J_2,0,J_2) = \sum_{S=0}^{\infty}\left(1+J_2+S,\frac{S}{2},J_2+\frac{S}{2}\right), \quad (18)$$

and for twist 2,

$$D(J_1+J_2+2,J_1,J_2) = \Delta(J_1+J_2+2,J_1,J_2),$$
$$\Delta(E,J_1,J_2) \equiv \sum_{r=-2J_1}^{2J_2}\Delta'\left(E+|r|,J_1+\frac{r}{2},J_2-\frac{r}{2}\right), \quad (19)$$
$$\Delta'(E,j_1,j_2) \equiv \sum_{S=0}^{\infty} L'\left(E+S,j_1+\frac{S}{2},j_2+\frac{S}{2}\right)$$
$$\equiv \sum_{S=0}^{\infty}\sum_{n=0}^{\infty}{}'\left(E+S+n,j_1+\frac{S}{2},j_2+\frac{S}{2}\right),$$

where $J_1,J_2 = 0,\frac{1}{2},1,\cdots$.

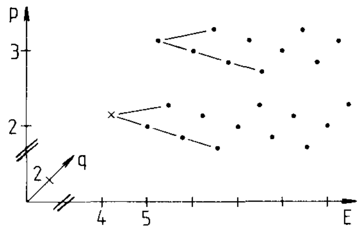


FIG. 7. Weights of $D(4,\frac{3}{2},\frac{1}{2})$ in a three-dimensional weight diagram, where $p = j_1+j_2$, $q = j_1-j_2$.

For a typical example of this type see Fig. 7.

All other representations can be built up easily from the $\Delta(E,J_1,J_2)$. For $t>1$ and $J_1\cdot J_2 = 0$ they are

$$D(E,J_1,J_2) = \Delta(E,J_1,J_2), \quad (20)$$

and for $t>2$, $J_1\cdot J_2\neq 0$,

$$D(E,J_1,J_2) = \sum_{S=0}^{2\min(J_1,J_2)}\Delta\left(E+S,J_1-\frac{S}{2},J_2-\frac{S}{2}\right). \quad (21)$$

The Δ-structure of $D(6,2,1)$ can be seen in Fig. 8. In the limit $t\to 2$ only the highest Δ survives; all weights get the multiplicity 1. The spin spectrum vanishes.[8]

B. The tensor product of two twist 1 representations

The CG series for these products were already given by Castell[1]:

Theorem 6:

$$D(J_1+1,J_1,0)\otimes D(J_2+1,0,J_2)$$
$$= \sum_{S=0}^{\infty} D\left(J_1+J_2+2+S,J_1+\frac{S}{2},J_2+\frac{S}{2}\right),$$
$$D(J_1+1,J_1,0)\otimes D(J_2+1,J_2,0)$$
$$= \sum_{S=|J_1-J_2|}^{J_1+J_2-1} D(J_1+J_2+2,S,0)$$
$$\oplus\sum_{S=0}^{\infty} D\left(J_1+J_2+2+S,J_1+J_2+\frac{S}{2},\frac{S}{2}\right). \quad (22)$$

In order to obtain $D(J_1+1,0,J_1)\otimes D(J_2+1,0,J_2)$ exchange the two SU(2) eigenvalues.

The proof of this theorem, using the weight diagrams for $t=2$, $J_1\cdot J_2\neq 0$, yields a confirmation for the latter.

Proof (by induction): I demonstrate the proof of the (more difficult) second Eq. of (22). As in the case of Theorem 1 it is shown that the same new weights in the SU(2) eigenvalues occur, when going from E to $E+2$ on both sides of the

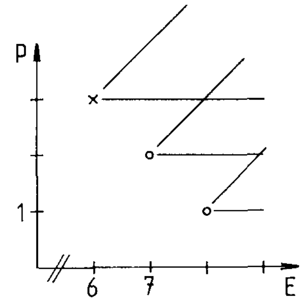


FIG. 8. Some features of $D(6,2,1)$.

equation. Using Eq. (18) and the SU(2) CG series one gets the weights of the left side:

$$D(J_1+1,J_1,0)\otimes D(J_1+1,J_2,0) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{r=|J_1-J_2+(i-j)/2|}^{J_1+J_2+(i+j)/2}\ \sum_{S=|(i-j)/2|}^{(i+j)/2}(J_1+J_2+2+i+j,r,S),$$

and using Eqs. (20) and (21) those of the right side:

$$\sum_{n=0}^{\infty}{}'\sum_{j=0}^{\infty}\Big\{\sum_{S=|J_1-J_2|}^{J_1+J_2-1}\ \sum_{r=0}^{2S}\Big(J_1+J_2+2+j+r+n,\ S+\frac{j-r}{2},\frac{j+r}{2}\Big) \oplus \sum_{S=0}^{\infty}\ \sum_{r=-S}^{2(J_1+J_2)+S}\Big(J_1+J_2+2+S+j+|r|+n,\ J_1+J_2+\frac{S+j-r}{2},\frac{S+j+r}{2}\Big)\Big\}.$$

It's easy to show lhs = rhs for the starting values $E=J_1+J_2+2$, J_1+J_2+3. Changing the sum indices to $i'=i+j$, $j'=i-j$, the left-hand side becomes

$$\sum_{i'=0}^{\infty}\ \sum_{j'=-i'}^{+i'}{}'\ \sum_{r=|J_1-J_2+j'/2|}^{J_1+J_2+i'/2}\ \sum_{S=|j'/2|}^{i'/2}(J_1+J_2+2+i',r,S).$$

Now the new weights which occur in the step from i' to $i'+2$ can be given:

$$\sum_{j'=-i'}^{+i'}{}'\ \sum_{S=|j'/2|}^{\frac{i'}{2}}\Big(J_1+J_2+4+i',J_1+J_2+\frac{i'}{2}+1,S\Big) \oplus \sum_{j'=-i'-2}^{i'+2}\ \sum_{r=|J_1-J_2+j'/2|}^{J_1+J_2+i'/2+1}\Big(J_1+J_2+4+i',r,\frac{i'}{2}+1\Big). \tag{23}$$

In the expression for the right-hand side, the new weights are just those with $n=0$. After some changing of indices I get for the new weights at $i'+2$,

$$\sum_{k=0}^{i'+2}\ \sum_{r=|i'/2-J_1-J_2-k+1|}^{i'/2+J_1+J_2+1}\Big(J_1+J_2+4+i',r,\frac{i'}{2}+1\Big) \oplus \sum_{k=0}^{i'+1}\ \sum_{S=|i'/2-k+1|}^{i'/2}\Big(J_1+J_2+4+i',J_1+J_2+\frac{i'}{2}+1,S\Big). \tag{24}$$

Equation (23) and Eq. (24) can be shown to be equal by distinguishing all the cases, where the expression whose absolute values occur, are positive respectively negative. The first Eq. in Theorem 6 is proved analogously. The last step, the equality of the sums at the level of Eqs. (23) and (24) is much easier in this case.

C. The tensor product of twist 1 and twist 2 representation

As in the case of $\tilde{S}O(3,2)$ I found it convenient to split the calculation and to first look for an expression for an expression for the product of a Twist 1 representation and an arbitrary line $L'(E,j_1,j_2)$.

See Fig. 9 for an example to the following

Lemma 7:

$$D(\lambda+1,\lambda,0)\otimes L'(E,j_1,j_2)$$

for $\lambda \leqslant j_1$

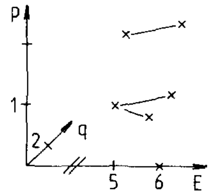


FIG. 9. The reduction of $D(\frac{3}{2},\frac{1}{2},0)\otimes L'(\frac{7}{2},1,\frac{1}{2})$.

$$= \sum_{S=0}^{2j_2}\ \sum_{r=j_1-\lambda+1}^{j_1+\lambda}\Delta'\Big(E+\lambda+1+S,r+\frac{S}{2},j_2-\frac{S}{2}\Big) \oplus D(E+\lambda+1,j_1-\lambda,j_2),$$

and for $\lambda > j_1$

$$= \sum_{S=0}^{2j_2}\ \sum_{r=|j_1-\lambda|}^{j_1+\lambda}\Delta'\Big(E+\lambda+S+1,r+\frac{S}{2},j_2-\frac{S}{2}\Big). \tag{25}$$

Proof: It consists in comparing the weights and their multiplicities. Using Eqs. (18) and (19) and the CG series of SU(2), the weights of the left side are

$$D(\lambda+1,\lambda,0)\otimes L'(E,j_1,j_2) = \sum_{n=0}^{\infty}{}'\sum_{S=0}^{\infty}\ \sum_{r=|\lambda+S/2-j_1|}^{\lambda+j_1+S/2}\ \sum_{t=|S/2-j_2|}^{S/2+j_2}(E+\lambda+1+S+n,r,t).$$

The right side is for $\lambda > j_1$:

$$\sum_{n=0}^{\infty}{}'\sum_{t=0}^{\infty}\sum_{S=0}^{2j_2}\ \sum_{r=\lambda-j_1}^{\lambda+j_1}\Big(E+\lambda+1+S+t+n,\ r+\frac{S+t}{2},j_2+\frac{t-S}{2}\Big).$$

First, changing the sum indices to $s'=s+t$, $t'=t-s$, one gets

$$\sum_{n=0}^{\infty}{}'\ \sum_{S'=0}^{\infty}\ \sum_{t'=|S'-2j_2|-2j_2}^{S'}\ \sum_{r=\lambda-j_1}^{\lambda+j_1}\Big(E+\lambda+1+S'+n,\ r+\frac{S'}{2},j_2+\frac{t'}{2}\Big),$$

then shifting to $r'=r+\frac{1}{2}S'$, $\bar{t}=j_2+\frac{1}{2}t'$, one obtains for the right-hand side

$$\sum_{n=0}^{\infty}{}'\ \sum_{S'=0}^{\infty}\ \sum_{\bar{t}=|S'/2-j_2|}^{S'/2+j_2}\ \sum_{r'=\lambda-j_1+S'/2}^{\lambda+j_1+S'/2}(E+\lambda+1+S'+n,r',\bar{t}).$$

This is obviously equal to the left-hand side.

The other case $\lambda \leqslant j_1$ is more space consuming, but straightforward. One decomposes the Δ' and D in the right-hand side of Eq. (25) into a sum of weights, and changes the sum indices to bring the weights into the form $(E+\lambda+1+s+n,\ r,\ t)$. Then the sums can be seen to be equal to those of the left-hand side.

Now, by decomposing the twist 2 representations into lines as in Sec. IIIA and using Lemma 7, the CG series of the tensor product of a twist 1 and a twist 2 representation can be found.

Theorem 8:

$$D(\lambda+1,\lambda,0)\otimes D(J_1+J_2+2,J_1,J_2) = \sum_{r=1}^{2J_1}\ \sum_{j=-\min(r,2\lambda)}^{\infty}D\Big(E+j+r,J_1+\lambda+\frac{j-r}{2},J_2+\frac{i+r}{2}\Big) \oplus \sum_{r=0}^{2J_2}\ \sum_{j=0}^{\infty}D\Big(E+j+r,J_1+\lambda+\frac{j+r}{2},J_2+\frac{j-r}{2}\Big) \tag{26}$$

with $E=(\lambda+J_1+J_2+3)$.

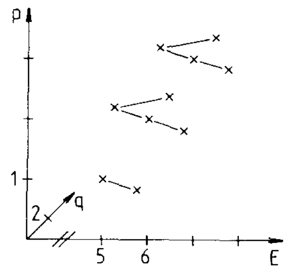


FIG. 10. The first terms in the CG series of $D(\frac{3}{2},\frac{1}{2},0)\otimes D(\frac{7}{2},1,\frac{1}{2})$. (x) are the lowest weights.

In order to obtain $D(\lambda+1,0,\lambda)\otimes D(J_1+J_2+2,J_2,J_1)$ exchange the two SU(2) eigenvalues. An important feature of this formula is, that the twist in the CG series can only have values between

$$3\leqslant t\leqslant 3+2\max(J_1,J_2). \tag{27}$$

For an easy example see Fig. 10.

The spin 0 case has been treated by solving differential equation.[3] For a proof see the appendix.

D. General tensor products

Putting $E=\lambda+1+E'$ in the right side of Eq. (26), one obtains the reduction of the weight diagram of

$$D(\lambda+1,\lambda,0)\otimes\Delta(E',J_1,J_2).$$

The CG series for the tensor product of a twist 1 and a $t>1$, $J_1\cdot J_2=0$ representation can be obtained from this directly, using Eq. (20). For the tensor product of a twist 1 and a $t>2$, $J_1\cdot J_2\neq 0$ representation it's a straightforward calculation from Theorem 8 to the CG series, using Eq. (21). In this case representations with twist t' occur,

$$t+1\leqslant t'\leqslant t+1+2(J_1+J_2). \tag{28}$$

The general tensor product of representations with twist >1 can in principle be reduced with the present method. As in the case of $\tilde{S}O(3,2)$ one has to decompose the triangles Δ of one representation into diagonals S, the others into lines L'. Then a generalized form of Lemma 7 for $S\otimes L'$ is needed. From the physical point of view these CG series are less interesting, as from spin 0 calculations, and from the analogy to $\tilde{S}O(3,2)$ one expects that all twists occur.

Some physical applications of the CG series found here will be published together with L. Castell,[10] with whom many discussions are acknowledged. He suggested the graphical method for the reduction in the above noncompact cases.

APPENDIX

Proof of Theorem 8: Using Lemma 7 and Eqs. (19) and (21), both sides can be decomposed into sums of Δ'. The left side is for $j<2(\lambda-J_1)+r\equiv a$.

$$\sum_{r=-2J_2}^{2J_1}\sum_{j=0}^{a-1}\sum_{S=0}^{2J_1+j}\sum_{t=0}^{r\,2J_2+j+r}\Delta'\Big(E+j+|r|+t,\lambda-J_1+\frac{r-j+t}{2}+S,J_2+\frac{j+r-t}{2}\Big), \tag{A1}$$

with $E=\lambda+J_1+J_2+3$,

and for $j\geqslant a$

$$\sum_{r=-2J_2}^{2J_1}\sum_{j=\max(0,a)}^{\infty}\Big\{\sum_{S=0}^{2\lambda-1}\sum_{t=0}^{2J_2+j+r}\Delta'\Big(E+j+|r|+t,J_1-\lambda+\frac{j-r+t}{2}+1+S,J_2+\frac{r+j-t}{2}\Big) \tag{A2}$$

$$\oplus\sum_{S=0}^{\min(j-a,2J_2+j+r)}\sum_{t=-(2J_2+r+j-S)}^{j-S-a}\Delta'\Big(E+j+|r|+S+|t|,J_1-\lambda+\frac{j-r-S-t}{2},J_2+\frac{j+r-S+t}{2}\Big)\Big\}. \tag{A3}$$

The decomposition of the right side yields

$$\sum_{r=0}^{2J_2}\sum_{j=0}^{\infty}\sum_{S=0}^{2\min[J_1+\lambda+(j+r)/2,\,J_2+(j-r)/2]}$$

$$\times\sum_{t=-2[J_2+(j-r-S)/2]}^{2(J_1+\lambda)+j+r-S}\Delta'\Big(E+j+r+S+|t|,J_1+\lambda+\frac{j+r-S-t}{2},J_2+\frac{j-r-S+t}{2}\Big) \tag{A4}$$

$$\oplus\sum_{r=1}^{2J_1}\sum_{j=-\min(r,2\lambda)}^{\infty}\sum_{S=0}^{2\min[J_1+\lambda+(j-r)/2,\,J_2+(j+r)/2]}$$

$$\times\sum_{t=-(2J_1+j+r-S)}^{2(J_1+\lambda)+j-r-S}\Delta'\Big(E+j+r+S+|t|,J_1+\lambda+\frac{j-r-S-t}{2},J_2+\frac{j+r-S+t}{2}\Big). \tag{A5}$$

Next all Δ' shall be arranged into structures of the form

$$\sum_{t'}\Delta'\Big(E+|r|+S'+t',J_1+\lambda+\frac{j'-r+t'}{2},J_2+\frac{j'+r-t'}{2}\Big).$$

For Eqs. (A1) and (A2) this is achieved by changing the sum indices appropriately:

$$\sum_{r=-2J_2-j}^{2J_1}\;\sum_{r=-2J_1-S/2}^{a-1}\;\sum_{[j/2]}^{a-j/2-1}\;\sum_{t=(j-S)/2}^{r+2J_2+j}\Delta'\Big(E+|r|+S+t,J_1+\lambda+\frac{t-r+j}{2},J_2+\frac{r+j-t}{2}\Big), \tag{A1a}$$

$$\sum_{r=-2J_2}^{2J_1}\ \sum_{j=\max(0,a)+1-2\lambda}^{\infty}\ \sum_{S/2=|j-\max(0,a)|/2+|\max(0,a)|/2}^{\frac{j}{2}+2\lambda-1}$$

$$\times \sum_{t=(j-S)/2}^{2J_2+j+r} \Delta'\Big(E+|r|+S+t),J_1+\lambda+\frac{j+t-r}{2},J_2+\frac{j+r-t}{2}\Big). \tag{A2a}$$

In Eqs. (A3)–(A5), terms of the type $\Sigma_t(E'+|t|,j_1-t/2,j_2+t/2)$ must be rearranged. To demonstrate the necessary steps for Eq. (A3) I introduce $j'=j-s$, $s'=j+s$ and get

$$\sum_{r=-2J_2}^{2J_1}\ \sum_{j'=\max(a,-2J_1-r)}^{\infty}\Big\{\sum_{S'/2=c}^{\infty}\ \sum_{t=0}^{2J_2+j'+r}\Delta'\Big(E+|r|+S'+t,J_1-\lambda+\frac{j'-r+t}{2},J_2+\frac{j'+r-t}{2}\Big)$$

$$\oplus\sum_{t=1}^{j'-a}\ \sum_{S'/2=c}^{\infty}\Delta'\Big(E+|r|+S'+t,J_1-\lambda+\frac{j'-r-t}{2},J_2+\frac{j'+r+t}{2}\Big)\Big\}, \tag{A3a}$$

with

$$c=|j'/2-\max(0,a)/2|+\max(0,a)/2.$$

After changing the sum-indices of the second term in braces $\{\ \}$ to $t'=-t$, $\bar S=s'+2t$, the two terms can be formulated in a single expression

$$\{\ \}=\sum_{t'=a-j'}^{2J_2+j'+r}\ \sum_{\frac{\bar S}{2}=c-\min(0,t')}^{\infty}\Delta'\Big(E+|r|+\bar S+t',J_1-\lambda+\frac{j'-r+t}{2},J_2+\frac{j+r-t}{2}\Big).$$

Reversing the t' and $\bar S$-sums and shifting j', t', and $(-\bar S)$ by 2λ one gets the required form,

$$\sum_{r=-2J_2}^{2J_1}\ \sum_{j'=\max(r-2J_1-r,\,-2(J_2+\lambda))}^{\infty}\ \sum_{S/2=c'+\lambda}^{\infty}\ \sum_{t=\max\left[r-j'-2(J_1+\lambda),\,-\frac{S}{2}+c'-\lambda\right]}^{2J_2+j'+r}\Delta'(E+|r|+S+t,J_1+\lambda$$

$$+\frac{j-r+t}{2},J_2+\frac{j+r-t}{2}\Big), \tag{A3b}$$

where $c'=|j'/2-\max(0,a)/2+\lambda|+\max(0,a)/2$.

Similar calculations give for Eqs. (A4) and (A5)

$$\sum_{r=0}^{2J_1}\ \sum_{j'=\max(r-2J_1-r,\,-2(J_1+\lambda))}^{\infty}\ \sum_{S'/2=|j'/2|}^{\infty}\ \sum_{t'=\max\left[-r-j'-2(J_1+\lambda),\,-\frac{S'}{2}+|\frac{j'}{2}|\right]}^{2J_2+j-r'}\Delta'(E+|r|+S'+t',J_1+\lambda+\frac{j'+r+t'}{2},$$

$$J_2+\frac{j'-r-t'}{2}\Big), \tag{A4a}$$

$$\sum_{r=1}^{2J_1}\ \sum_{j'=\max[r-2(J_1+r),\,-r-2J_2]}^{\infty}\ \sum_{S'/2=d}^{\infty}\ \sum_{t'=\max[r-j'-2(J_1+\lambda),\,-S'/2+d]}^{2J_2+j'+r}\Delta'(E+|r|+S'+t',J_1+\lambda+\frac{j'-r+t'}{2},$$

$$J_2+\frac{j'+r-t'}{2}\Big), \tag{A5a}$$

where

$$d=\left|\frac{j'}{2}+\frac{\min(r,2\lambda)}{2}\right|-\frac{\min(r,2\lambda)}{2}.$$

Now lhs = rhs can be seen almost by inspection for $j'>0$, which corresponds to $j_1+j_2>J_1+J_2+\lambda$ in Fig. 10. For $j'\leqslant 0$ the r-sums have to be included into the rearrangement. Exchanging the r and j' sums in Eq. (A4a) and (A5a) yields for the right side

$$\sum_{j'=\max[-2J_2,\,-(J_1+J_2+\lambda)]}^{\infty}\ \sum_{r=\max[0,\,-j'-2(J_1+\lambda)]}^{\min(j'+2J_2,2J_1)}\ \sum_{S/2=|j'/2|}^{\infty}\ \sum_{t'=\max\left[-r-j'-2(J_1+\lambda),\,|\frac{j'}{2}|-\frac{S}{2}\right]}^{2J_2+j'-r}\Delta'(\ ), \tag{A4b}$$

$$\otimes\sum_{j'=\max[-2(J_1+J_2),\,1-2(J_1+\lambda),\,-(J_1+J_2+\lambda)]}^{\infty}\ \sum_{r=\max(1,\,-j'-2J_2)}^{\min(2J_1,j'+2(J_1+\lambda))}\ \sum_{\frac{S'}{2}=d}^{\infty}\ \sum_{t'=\max[r-j'-2(J_1+\lambda),\,d-S'/2]}^{2J_2+j+r}\Delta'(\ ). \tag{A5b}$$

In the expressions for the lhs, in addition to this, the sum indices have to be shifted by

$$r=r'-2\lambda,\quad t=t'-2\lambda,\quad S=S'-|r'-2\lambda|+|r'|+2\lambda$$

for $j\leqslant -2\lambda$.

Equations (A1a) and (A2a) do not contribute and Eq. (A3b) gives

$$\sum_{j=\max[-2(J_1+J_2),\,-(J_1+J_2+\lambda)]}^{-2\lambda}\ \sum_{r'=-j-2J_2}^{j+2(J_1+\lambda)}\ \sum_{S'/2=-j/2+|r'/2-\lambda|-|r'/2|-\lambda}^{\infty}$$

$$\sum_{t'=\max\left[r'-j-2(J_1+\lambda),\,-\frac{S'}{2}-\frac{j}{2}+|\frac{r'}{2}-\lambda|-|\frac{r'}{2}|-\lambda\right]}^{2J_2+j+r}$$

$$\Delta'\left(E+|r|+S+t', J_1+\lambda+\frac{j+t'-r}{2}, J_2+\frac{j-t'+r}{2}\right). \tag{A3c}$$

Again lhs = rhs can be seen directly.

For $-2\lambda < j \leqslant 0$ the necessary shifts are:

$$r=r'+j, t=t'+j, S=S'-|r'+j|+|r'|-j.$$

In this case all three terms of the lhs, coming from Eqs. (A1a), (A2a), and (A3b) have to be collected, and compared with Eqs. (A4b) and (A5b), which is a lengthy but straightforward calculation.

[1]L. Castell, in Symposium on de Sitter and conformal groups, University of Colorado, 1970.
[2]M. Flato and C. Fronsdal, Lett. Math. Phys. **2**, 421 (1978).
[3]L. Castell, Phys. Rev. D **2**, 1161 (1970).
[4]C. Fronsdal, Phys. Rev. D **12**, 3819 (1975).
[5]N. T. Evans, J. Math. Phys. **8**, 170 (1967).
[6]E. Majorana, Nuovo Cimento **9**, 335 (1932).
[7]P.A.M. Dirac, J. Math. Phys. **4**, 901 (1963).
[8]G. Mack, Commun. Math. Phys. **55**, 1 (1977).
[9]Tsu Yao, J. Math. Phys. **8**, 1931 (1967); **9**, 1615 (1968). In his Theorem 3, p. 1950, left column, the fifth line from bottom must be $M(j,k,\lambda)=\min(t+1, 2K_m+1, 2J+s-\gamma+1)$. The new term comes from his Appendix E (Theorem of Harish-Chandra and Godement).
[10]L. Castell and W. Heidenreich, "Restrictions of $\gamma\pi$, γp, $\bar{\nu}p$, νp scattering from conformal invariance," "SO(3,2) invariant scattering and Dirac Singletons," Phys. Rev. D (to be published).

Sturm–Liouville eigenproblems with an interior pole

John P. Boyd
Department of Atmospheric and Oceanic Science, University of Michigan, Ann Arbor, Michigan 48109

(Received 17 September 1979; accepted for publication 18 March 1981)

The eigenvalues and eigenfunctions of self-adjoint Sturm–Liouville problems with a simple pole on the *interior* of the interval $[A, B]$ are investigated. Three general theorems are proved and it is shown that as $n \to \infty$, the eigenfunctions more and more closely resemble those of an ordinary Sturm–Liouville problem and $\lambda_n \sim -m^2\pi^2/(B-A)^2$, just as if there were no singularity. The low-order modes, however, differ drastically from those of a nonsingular eigenproblem in that (i) both eigenvalues and eigenfunctions are *complex* (despite the fact the problem is self-adjoint), (ii) the real and imaginary parts of the nth eigenfunction may both have ever-increasing numbers of interior zeros as $B \to \infty$, instead of just $(n-1)$ zeros, and (iii) as $B \to \infty$, the eigenvalues for all small n may cluster about a common value in contrast to the widely separated eigenvalues of the corresponding nonsingular problem. These results are general, but in order to present quantitative solutions for the low-order modes, too, special attention is given to the particular case

$$u'' + (1/x - \lambda)u = 0, \tag{1}$$

with $u(A) = u(B) = 0$ where λ is the eigenvalue and A and B are of *opposite* signs. For small n, one can obtain the approximation

$$\lambda_n \sim \exp[(1 + 3^{1/2}i)d_n/(2B^{1/3})]/B, \tag{2}$$

where d_n is the nth root of the Airy function $\mathrm{Ai}(-z)$. The imaginary part of (2) shows explicitly how profoundly the interior pole has modified the structure of the eigenproblem.
The WKB method, which was used to derive (2), is shown to be accurate for all n. The WKB analysis is of some interest in and of itself. Although the number of WKB "transition" points is the same as for the half-century old quantum harmonic oscillator (two), the substitution of the interior pole for one of the turning points has a profound (and fascinating) impact on both the WKB formalism and the numerical results. Thus, although this problem was motivated by the physics of hydrodynamic waves, it is also an extension to both classical Sturm–Liouville theory and to the WKB treatment of eigenvalue problems.

PACS numbers: 02.30.Hq

1. INTRODUCTION

Normal self-adjoint Sturm–Liouville (SL) eigenproblems on an interval $[A, B]$ fall into two classes: those whose equations have no singularities on $[A, B]$ and those which are singular only on the boundaries. The theory of the diurnal ocean tide introduced a third class: equations which would otherwise be described by the classical SL theory except for having so-called "apparent" singularities in the interior of the domain. Although the tidal equation was derived by Laplace in the eighteenth century and despite the fact that the eigenfunctions themselves are analytic everywhere in the interior of the domain, the mathematical problems of this example of the third class were not resolved until 1970, ending a long history of confusion, controversy, and many published blunders.[1,2] The goal of the present work is to study the simplest example of a fourth class of SL eigenproblems in which the eigenfunctions themselves, as well as the differential equation, are singular in the interior of the interval.

Up to now, this fourth class of self-adjoint SL problems has been completely ignored, and small wonder. Such problems seem bizarre and outrageous: what physical theory could lead to equations whose solutions are singular inside the physical domain? In reality, such interior singularities or "critical surfaces" arise as naturally in fluid waves as kittens from cats. Physically, the singularity is removed by *friction*, which shifts it into the complex plane. In the real world, there is always at least a little friction, so the actual fluid waves are always finite and well-behaved, as one would expect. Because the dissipation is so weak, however, it is a good approximation to take the inviscid limit so as to eliminate the friction as an explicit parameter, and this will be done in most of the paper. In the next two sections, however, the friction is temporarily restored to a finite value to show how the singularity should be interpreted when making this approximation. (In brief, the conclusion is that the eigenfunctions should be made single-valued by a branch cut in the *upper* half-plane.)

Although some attention will be given to a general class of problems, for simplicity and for the sake of giving explicit results instead of vague generalizations, most of our attention will be focused on the particular example

$$u_{xx} + (1/x - \lambda)u = 0, \tag{1.1}$$

$$u(A) = u(B) = 0, \tag{1.2}$$

where λ is the eigenvalue. If A and B are of the same sign, then (1.1) and (1.2) are merely a normal, self-adjoint Sturm–

 0022-2488/81/081575-16$01.00 © 1981 American Institute of Physics

Liouville problem of the first kind with no singularities on $[A, B]$. Here, however, A and B will be of *opposite* signs so that both (1.1) and the eigenfunctions are singular in the interior of the interval $[A, B]$. None of the usual theorems of conventional SL theory apply because the interior singularity violates the conditions of the theorems, and most are no longer true. In particular, the eigenfunctions and eigenvalues are complex.

Thus, one has no choice but to regard (1.1) and (1.2) as a new species, a "Sturm–Liouville eigenproblem of the fourth kind," when the singularity is in the interior. The problem is not a lack of self-adjointness (it is well known that non-self-adjoint equations may have complex eigenvalues); actually, (1.1) is *self-adjoint*. The rub is solely that the differential equation has a *pole* on the *interior* of the domain.

In several years of searching, it has not been possible to locate a single paper other than this one which attempts a systematic attack on such "fourth kind" eigenproblems, but there have been three precursors. Dickinson[3] and Tung[4,5] analyzed waves with "critical latitudes" using the continuous spectrum approach discussed in Appendix B. This work is complementary to that reported here, and some of Dickinson's WKB analysis can be carried over. Simmons[6,7] is the only author besides Boyd[8] to have previously computed discrete, singular eigenfunctions, but his calculations are strictly numerical and limited to more complicated equations than (1.1).

This present work has three principal goals: (i) to prove some simple theorems about the general SL problem of the fourth kind, (ii) to obtain analytic approximations to the high- and low-order eigenvalues of (1.1) in particular, and (iii) to describe the WKB treatment of an eigenvalue problem with a turning point and a simple pole. The reasons for investigating "fourth kind" SL problems have already been explained above and also in Boyd.[8] The purpose of studying (1.1) is to understand the general class by thoroughly examining a particular example.

The WKB analysis has several motives. First, it is a straightforward and familiar method for obtaining asymptotic approximations to the solution of (1.1). In addition, however, the WKB analysis is of interest in itself. Generations of budding physicists have studied the quantum mechanical harmonic oscillator from a WKB viewpoint. Here, however, although the number of WKB "transition points" is the same (two), the replacement of a turning point by a simple pole profoundly alters the solution, and it is fascinating to see how the application of such familiar ideas can lead to such radically different conclusions.

The plan of the paper is as follows. The next section proves three theorems for general Sturm–Liouville problems of the fourth kind. Section 3 gives the exact analytic solution of (1.1) in terms of Whittaker functions and also the rather unorthodox choice of branch cut which is physically appropriate for making the eigenfunctions single-valued. The next two sections discuss the eigenvalues and eigenfunctions in the limits $n \to \infty$ and $n \to 0$, respectively. Sections 6 and 7 analyze the WKB method and its accuracy. The eighth section is a case study of the complete spectrum for a particular choice of parameters, paying particular attention to modes of intermediate n. The final section summarizes the similarities and differences between normal Sturm–Liouville eigenproblems and those of the singular fourth kind discussed here. The three appendices discuss the Whittaker functions, discrete versus continuous eigenvalues, and Chebyshev approximations for the eigenvalues, respectively.

The theorems of Sec. 2 and the asymptotic $n \to \infty$ approximations of Sec. 4 [Eqs. (4.6) and (4.10)] are applicable to general Sturm–Liouville eigenproblems of the fourth kind. Most of the remaining results are quantitatively applicable only to the particular example (1.1), but the methods used to derive them are general also.

2. THREE THEOREMS

In this section, some simple results will be proved for an equation more general than (1.1). To interpret the singularity of (2.1), the friction ϵ is explicitly included. As noted in the introduction, ϵ is normally so small that it is good approximation to take the limit $\epsilon \to 0$, which reduces the number of parameters from three (ϵ, A, B) to two (A and B).

Theorems

Let $u_m(x)$ and $u_n(x)$ be eigenfunctions of the differential equation

$$u_{xx} + [r(x)/(x - i\epsilon) + p(x) - \lambda]u = 0, \tag{2.1}$$

with

$$u(A) = u(B), \tag{2.2}$$

where $p(x)$ and $r(x)$ are real and analytic on $[A, B]$, λ is the eigenvalue, A and B are of opposite signs, and $\epsilon > 0$ is a real constant. Then in the limit $\epsilon \to 0$, one can prove

Theorem 1: If $\lambda_m \neq \lambda_n$, the eigenfunctions are *orthogonal*, i. e.,

$$\int_A^B u_m u_n \, dx = 0. \tag{2.3}$$

Theorem 2: Letting $\mathrm{Im}(\lambda_n)$ denote the imaginary part of the eigenvalue,

$$\mathrm{Im}(\lambda_n) \int_A^B |u_n| \, dx = \pi |u_n(0)|^2 r(0). \tag{2.4}$$

Theorem 3: The eigenvalue λ is always in the upper half-plane, i. e.,

$$\mathrm{Im}(\lambda_n) \geqslant 0 \quad \text{for all } n, \tag{2.5}$$

if $r(x) > 0$.

Proofs

The demonstration of Theorem 1 is identical with the proof of orthogonality for orthodox Sturm–Liouville problems as given in Morse and Feshbach,[2] for example. Let

$$q(x) \equiv r(x)/(x - i\epsilon) + p(x). \tag{2.6}$$

Writing the differential equations satisfied by $u_m(x)$ and $u_n(x)$ after multiplication by the other mode gives (letting primes denote differentiation)

$$u_m (u_n'' + q u_n - \lambda_n u_n) = 0, \tag{2.7}$$

$$u_n (u_m'' + q u_m - \lambda_m u_m) = 0. \tag{2.8}$$

Subtracting (2.8) from (2.7) gives

$$u_m u_n'' - u_n u_m'' + (\lambda_m - \lambda_n) u_m u_n = 0. \tag{2.9}$$

The offending singular term $q(x)$ has already disappeared through subtraction, and the remaining steps—rewriting the first two terms in (2.9) as a perfect derivative, integrating from A to B, and invoking the homogeneous boundary conditions—give

$$(\lambda_m - \lambda_n) \int_A^B u_m(x) u_n(x)\, dx = 0, \tag{2.10}$$

from which (2.3) is obvious.

The steps in the proof of the second theorem are formally identical to those for the first except that $u_m(x)$ is replaced by $u_n(x)^*$ where the asterisk denotes the complex conjugate. Since $u_n(x)$ cannot be orthogonal to its own complex conjugate, this argument is used in formal Sturm–Liouville theory to show that $\lambda_n^* = \lambda_n$, i. e., all the eigenvalues are real. For the singular class examined here, the rub is that because of the pole (and the friction ϵ), $q(x) \neq q(x)^*$, so the singular terms do *not* cancel out and the equivalent of (2.9) is

$$u_n u_n^{*\prime\prime} - u_n^* u_n'' + (\lambda_n - \lambda_n^*)|u_n|^2 = (q - q^*)|u_n|^2. \tag{2.11}$$

Following through the remaining steps gives

$$(\lambda_n - \lambda_n^*) \int_A^B |u_n|^2\, dx = \int_A^B \frac{|u_n|^2 2i\epsilon r(x)}{x^2 + \epsilon^2}\, dx. \tag{2.12}$$

Carrier, Krook, and Pearson[9] show that

$$\lim_{\epsilon \to 0} \frac{\epsilon}{x^2 + \epsilon^2} = \pi \delta(x). \tag{2.13}$$

Substituting this into (2.12) and performing the integration on the right-hand side gives (2.4).

The third theorem follows trivially from the second. Since all the other quantities in (2.4) are absolute values and therefore positive semidefinite, it follows that $\mathrm{Im}(\lambda_n)$ must be as well.

A few remarks are in order. First, only Theorem 2 actually requires $\epsilon \to 0$; the third theorem can be proved directly from (2.12).

Second, Theorem 1 shows that the eigenfunctions are mutually orthogonal among themselves. The orthogonality relation (2.3) does not involve the complex conjugates of the eigenfunctions nor is (2.3) a biorthogonality equation involving inner products of the eigenfunctions paired with those of the adjoint. Despite its complex eigenvalues, (1.1) with (1.2) is *self-adjoint* and the form of the orthogonality relation, Theorem 1, reflects this.

The second theorem shows that $\mathrm{Im}(\lambda_n) = 0$ only when

$$u_n(0) = 0, \tag{2.14}$$

i. e., in the very special case that $u_n(x)$ is nonsingular. When $p(x) = 0$ this means $u_n(x)$ is proportional to $M_{-\kappa,\frac{1}{2}}(-x/\kappa)$, which always has a zero at $x = 0$. Since (2.14) plus (2.2) are equivalent to imposing *three* boundary conditions on a *second* order differential equation, Theorem 2 implies that λ_n is *real* on a set of *measure zero*. In other words, there are certain sets of values of (A, B, n) for which λ_n is real, but if one chooses A and B at random, the odds are infinitesimally small that any of the modes will have a real eigenvalue (although the imaginary parts of some may be very small).

Theorem 3 states what will be assumed in later sections in working out the WKB formalism: that λ is always in the upper half-plane and κ [defined by (3.4) below] therefore, always in the fourth quadrant. The physical significance (and necessity!) of this are discussed in Boyd.[8] The condition that $r(x)$ be positive is equivalent to satisfying the well-known Rayleigh–Kuo criterion for barotropic stability, and is almost always true in the upper atmosphere. It is automatically satisfied by the linear wind shear model [$r(x) = \text{const}$] that will be considered in the rest of this paper.

3. THE EXACT SOLUTION OF THE MODEL PROBLEM

The general problem

$$u_{zz} + (\alpha/z - \lambda')u = 0, \tag{3.1}$$

$$u(A') = u(B') = 0, \tag{3.2}$$

can be reduced to the canonical form (1.1) and (1.2) through the substitutions

$$x = \alpha z, \tag{3.3a}$$

$$A = \alpha A', \tag{3.3b}$$

$$B = \alpha B', \tag{3.3c}$$

$$\lambda = \lambda'/\alpha^2. \tag{3.3d}$$

Equation (1.1) is a special case of Whittaker's equation, which in turn is merely a transformed version of the confluent hypergeometric equation. Defining (principal branch)

$$\kappa \equiv 1/2\lambda^{1/2}, \tag{3.4}$$

the linearly independent solutions may be taken as

$$u_1(x, \lambda) = M_{-\kappa,\frac{1}{2}}(-x/\kappa), \tag{3.5}$$

$$u_2(x, \lambda) = \Gamma(1 + \kappa) W_{-\kappa,\frac{1}{2}}(-x/\kappa). \tag{3.6}$$

The power series for these Whittaker functions and their relations to the usual M and U confluent hypergeometric functions are given in Appendix A. The minus signs in (3.5) and (3.6) are a convention introduced by Dickinson to ensure that the Whittaker functions have different asymptotic be-

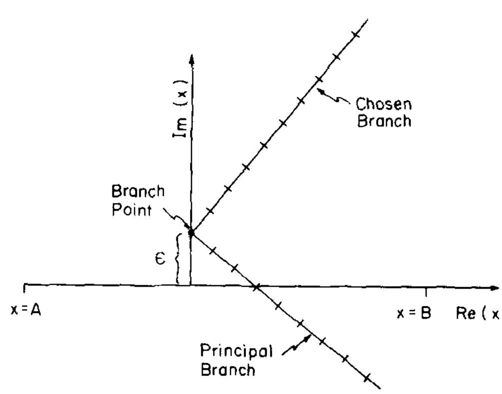


FIG. 1. Two possible branch cuts for the solution of Eq. (2.7) with friction coefficient ϵ and complex eigenvalue λ. The principal branch of the Whittaker function cuts the real axis between the boundaries A and B, which would make the solution discontinuous. The chosen branch is convenient and avoids this discontinuity. Any other branch cut which avoids the real axis would be acceptable, however.

havior ($M_{-\kappa,\frac{1}{2}}$ blows up and $W_{-\kappa,\frac{1}{2}}$ decays) as $x \to -\infty$. With this convention, the lowest few eigenfunctions are approximately proportional to the W function alone, as explained in the next section, which is a great simplification.

$M_{-\kappa,\frac{1}{2}}(y)$ is an entire function, but $W_{-\kappa,\frac{1}{2}}(y)$ has a branch point at $y = 0$. The obvious choice is to take the principal branch of the function, but this is not physically allowed. If one inserts a small amount of dissipation with friction coefficient ϵ (with the understanding that $\epsilon \to 0$ in the end), (1.1) becomes

$$u_{xx} + [1/(x - i\epsilon) - \lambda]u = 0, \tag{3.7}$$

and the singularity is shifted into the upper half-plane. If one uses the fact that κ lies always in the fourth quadrant (proved in Sec. 2), then the branch cut for the principal branch of $W_{-\kappa,\frac{1}{2}}(-x/\kappa)$ would cross the real x axis as shown schematically in Fig. 1, which is absurd. The simplest allowable choice is to place the branch cut along the ray

$$\arg y = -\pi/2, \tag{3.8}$$

where (note the sign difference between y and x)

$$y \equiv -x/\kappa. \tag{3.9}$$

Dickinson[3] made the same choice. Any branch cut which lies above the real x axis is permitted, however, and in fact the different choice $\arg x = \pi/2$ is made in Fig. 5 for the sake of clarity. Since there is always (weak) damping in a real fluid, such frictional arguments have been used to choose the proper branch in fluid mechanics for a very long time.

Unfortunately, this nonstandard choice of branch implies that the usual textbook asymptotic formulas for $W_{-\kappa,\frac{1}{2}}(y)$ cannot be directly applied to our Whittaker function when y is in the third quadrant. However, it is a property of logarithmic solutions to linear, second-order differential equations that the coefficient of the logarithm is always proportional to that solution of the equation—in this case, $M_{-\kappa,\frac{1}{2}}(y)$—which is analytic at $y = 0$. If one defines $\ln(y)$ to be the logarithm with branch cut at $\arg y = -\pi/2$ and $\ln^{(P)}(y)$ to be the principal branch of the logarithm, then

$$\ln(y) = \begin{cases} \ln^{(P)}(y), & -\pi/2 \leqslant \arg y \leqslant \pi \\ \ln^{(P)}(y) + 2\pi i, & \pi < \arg y \leqslant 3\pi/2 \end{cases}. \tag{3.10}$$

From this it follows that

$$W_{-\kappa,\frac{1}{2}}(y) = \begin{cases} W^{(P)}_{-\kappa,\frac{1}{2}}(y), & -\pi/2 \leqslant \arg y \leqslant \pi \\ W^{(P)}_{-\kappa,\frac{1}{2}}(y) + \dfrac{2\pi i\kappa}{\Gamma(1+\kappa)} M_{-\kappa,\frac{1}{2}}(y), & \pi < \arg y \leqslant 3\pi/2 \end{cases}. \tag{3.11}$$

The most efficient way to evaluate the Whittaker functions is by numerical integration of (1.1), using the power series for $M_{-\kappa,\frac{1}{2}}(y)$ and $W_{-\kappa,\frac{1}{2}}(y)$ to initialize the calculation for small y. Even though (1.1) is "stiff" in the parlance of numerical analysis, an ordinary fourth-order Runge–Kutta program gave high accuracy even for large x, and was used to compute the "exact" results presented in later sections.

Letting

$$u(x,\lambda) = \alpha u_1(x,\lambda) + \beta u_2(x,\lambda), \tag{3.12}$$

the boundary conditions (1.2) can be written in the form of a 2×2 matrix equation whose determinant is

$$\Delta(x,\lambda) = u_1(A,\lambda)u_2(B,\lambda) - u_1(B,\lambda)u_2(A,\lambda). \tag{3.13}$$

The eigenrelation is then

$$\Delta(\lambda) = 0. \tag{3.14}$$

Once the eigenvalues have been determined from (3.14), it is trivial to solve the matrix equation for α and β in (3.12) to obtain the eigenfunctions.

4. HIGH-ORDER MODES

In the limit $|y| \to \infty$ with κ fixed, the Whittaker functions have the familiar asymptotic approximations

$$M_{-\kappa,\frac{1}{2}}(y) = S[\sin(\kappa\pi)/\kappa\pi]\Gamma(1+\kappa)W^{(P)}_{-\kappa,\frac{1}{2}}(y) + \frac{e^{y/2}y^{\kappa}}{\Gamma(1+\kappa)}\left(1 - \frac{\kappa(1-\kappa)}{y} - \frac{\kappa(1-\kappa)(1-\kappa)(2-\kappa)}{y^2} + \cdots\right), \tag{4.1}$$

$$\Gamma(1+\kappa)W^{(P)}_{-\kappa,\frac{1}{2}}(y) = \frac{\Gamma(1+\kappa)e^{-y/2}}{y^{\kappa}}\left(1 - \frac{\kappa(1+\kappa)}{y} + \frac{\kappa(1+\kappa)(1+\kappa)(2+\kappa)}{y^2} - \cdots\right), \tag{4.2}$$

where

$$S = \begin{cases} -e^{i\pi\kappa} & \mathrm{Im}(y) > 0, \\ -e^{-i\pi\kappa} & \mathrm{Im}(y) < 0, \end{cases} \tag{4.3}$$

and where the superscript (P) denotes the principal branch of the Whittaker function as before. The asymptotic approximation to our Whittaker function of unorthodox branch can be obtained from (4.1) and (4.2) via (3.11).

In the limit $\lambda \to \infty$, $\kappa \to 0$ along the negative imaginary axis, and (4.1) and (4.2) simplify to

$$M_{-\kappa,\frac{1}{2}}(-x/\kappa) \approx -2i\sin(|\lambda|^{\frac{1}{2}}x), \tag{4.4}$$

$$\Gamma(1+\kappa)W_{-\kappa,\frac{1}{2}}(-x/\kappa) \approx e^{i|\lambda|^{1/2}x}, \tag{4.5}$$

for fixed x (at either sign) with relative error $O(1/4\lambda x)$ where the Whittaker function has a branch cut at $\arg y = -\pi/2$. Substituting (4.4) and (4.5) into (3.14) gives

$$\lambda_n = -\pi^2 m^2/(B-A)^2, \quad n\to\infty. \tag{4.6}$$

Several features of this eigenrelation deserve comment.

First, the integer m that appears in (4.6) is not necessarily equal to the mode number n when the modes are ordered according to $|\lambda|$. A counterexample where $m = n+2$ is given in Table IV of Sec. 8.

Second, (4.6) implies that as was assumed in obtaining it, $\lambda_n \to -\infty$ as $n\to\infty$. Thus, the derivation of (4.4) through (4.6) is consistent.

Third, if we generalize (1.1) to

$$u_{xx} + [1/x - \lambda + p(x)]u = 0, \tag{4.7}$$

as done in the theorems of Sec. 2, where $p(x)$ is analytic on $[A,B]$, then

$$p(x) \ll \lambda \tag{4.8}$$

uniformly on $[A,B]$ in the limit that λ is large. Thus, the function $p(x)$ is only a small perturbation to the eigenmodes of (1.1) for sufficiently large n. Therefore, (4.6) is a valid approximation to the large eigenvalues of (4.7) for *general* bounded $p(x)$—though of course the approximation is more accurate (for a given n) when $p(x)=0$ than when it is nonzero. One could presumably correct for $p(x)\neq 0$ along the lines of the usual Rayleigh–Schrödinger perturbation theory, but (4.6) will suffice for the present.

Fourth, (4.6) is identical with the eigenrelation of the same problem with the *pole removed*, i. e.,

$$u_{xx} - \lambda u = 0, \tag{4.9}$$

with the usual boundary conditions (1.2). Further, the eigenfunctions of (4.9) are given by a linear combination of the trigonometric eigenfunctions of (4.4) and (4.5),

$$u_n(x) \sim \sin(m\pi x/[B-A]). \tag{4.10}$$

Thus, for the *high n* modes of an equation with an interior pole, the singularity is essentially irrelevant. The solutions differ from those of (4.9) only in two small ways.

First, λ_n always has a small imaginary part $\lambda_{\rm im}$ which appears to decrease roughly as $O(1/n)$.[10] Second, the approximation (4.10) breaks down in an internal boundary layer of width $O(1/\lambda)$ about the singularity at $x=0$, where the full Whittaker functions must be used. Since both $\lambda_{\rm im}$ and the width of the internal boundary layer decrease as $n\to\infty$, however, it still remains true that the singularity has little effect on the higher-order modes.

5. LOW-ORDER MODES

When λ is small, the internal boundary layer in which the asymptotic series (4.1) and (4.2) are inaccurate includes the whole of $[A,B]$, and more powerful, (and alas, more complicated) methods are needed. There is, however, one powerful simplification that we can make before applying them.

When n is large, λ hugs the negative real axis and the eigenfunctions are sinusoidal as shown explicitly by (4.4) and (4.5). For the low-order modes, however, λ is complex with either a large imaginary part or a positive real part; and then the eigenfunction must decay exponentially on $[A,0]$ away from the pole.

The reason for this decay is most easily seen by assuming λ is real and positive (as it is in the limiting case) and looking at the equation to which (1.1) reduces for large $|x|$:

$$u_{xx} - \lambda u = 0. \tag{5.1}$$

In order to satisfy the boundary condition of vanishing at $x=A$ where A is negative, $u(x)$ must be of the form

$$u = (\text{const})\,(e^{-\lambda^{1/2}|x|} - e^{-2\lambda^{1/2}|A|}e^{\lambda^{1/2}|x|}), \tag{5.2}$$

which is approximately

$$u(x) \approx (\text{const})e^{-\lambda^{1/2}|x|} \tag{5.3}$$

everywhere on $[A,B]$ except in a narrow boundary layer of width $O(1/\lambda^{\frac{1}{2}})$ near $x=A$ where the growing exponential is significant. In this boundary layer, however, $u(x)$ is exponentially small in comparison to its value at $x=0$ (by a factor of $e^{-\lambda^{1/2}|A|}$), so the *absolute* error in replacing the exact solution (5.2) by the approximation (5.3) is exponentially small *everywhere* on $[A,B]$.

In general, of course λ is complex rather than real and we want to solve (1.1) rather than (5.1), but these complications do not affect the basic argument in the least. The sign of $1/x$, like that of $-\lambda$, is negative for $x<0$, so the pole merely makes the two linearly independent solutions grow or decay faster. The complexity of λ will cause *oscillatory* growth or decay, but the growth or decay is still there unless λ is negative real—as is approximately true for the high order modes discussed in the previous section.

Thus, the qualitative behavior of the solutions of (5.1) is identical with that of the low order eigenfunctions of (1.1). From the asymptotic approximations (4.1) and (4.2), one sees that $M_{-\kappa,\frac{1}{2}}(-x/\kappa)$ is analogous to the positive exponential in (5.2) while $W_{-\kappa,\frac{1}{2}}(-x/\kappa)$ decays exponentially away from the pole. (These asymptotic approximations may not be *numerically* accurate for the small λ we are interested in here, but they do indicate the correct exponential growth-/decay behavior as one can verify from the more powerful WKB approximations of the next section). Thus, it must be approximately true, in analogue to (5.3), that

$$u(x) \sim W_{-\kappa,\frac{1}{2}}(-x/\kappa) \tag{5.4}$$

—in words, that the low order eigenfunction is proportional to the W-function alone.

This approximation, which is equivalent to setting

$$A = -\infty \tag{5.5}$$

since (5.3) and (5.4) are exact in this limit, is justified provided

$$e^{-2\lambda^{1/2}A} \ll 1. \tag{5.6}$$

In the next section, we will assume (5.5) and then check *a posteriori* that (5.6) is in fact satisfied for small n and not-too-small A and B.

This assumption (5.5) and the reasoning behind it is important both physically and mathematically. Physically, the argument is important because it tells us that the low-order eigenfunction has nothing except an exponentially decaying tail to the left of $x=0$—in startling contrast to the high n

modes, which are oscillatory on *both* sides of $x = 0$. Figure 2 compares the amplitudes for a typical low-order and a typical high-order mode. (To avoid repeating "small n modes" and "large n modes" *ad nauseam*, it is convenient to introduce the terms "monokeric,"—literally, "one-sided"—for a model which has only an exponentially decaying tail to the left of $x = 0$ as in the top of Fig. 2, and "dikeric"—"two-sided"—for a mode which is sinusoidal on both sides of $x = 0$ as in the bottom of Fig. 2.) Mathematically (5.5) is significant because it reduces the number of parameters from two (A and B) down to one (B alone).

Turning to the eigenvalues, we show in the next section that for small n and moderate or large A and B, i. e., a "monokeric" small λ mode

$$\lambda_n \sim (1/B)\exp[(1 + 3^{1/2}i)|d_n|/(2B^{1/3})], \tag{5.7}$$

where

$$d_n = \left| \tfrac{3}{8}\pi(4n-1)\right|^{2/3} \tag{5.8}$$

is the nth root of Airy's function $\mathrm{Ai}(-x)$. The hodograph of the product of B with the exact λ_1 (as determined by numerical integration) in the complex plane is shown in Fig. 3. The approximation (5.7) is a good qualitative description of the entire graph.

As $B \to \infty$, (5.7) becomes exact and

$$\lambda_n \sim 1/B \tag{5.9}$$

independently of n. This clustering of eigenvalues for large B is in sharp contrast to ordinary SL theory, where for the boundary conditions (1.2), one can prove that the eigenvalues must always be distinct and well-separated.

For finite B, (5.7) and the hodograph show that the eigenvalue is complex even though the differential equation (1.1) is both real and self-adjoint. This again would be provably impossible for a real, self-adjoint Sturm–Liouville problem of the usual classes.

The hodograph of λ is shown only for the lowest mode because one can prove from (5.7) that

$$\lambda_n(B) = \frac{9}{(4n-1)^2}\lambda_1\left[\frac{9B}{(4n-1)^2}\right]. \tag{5.10}$$

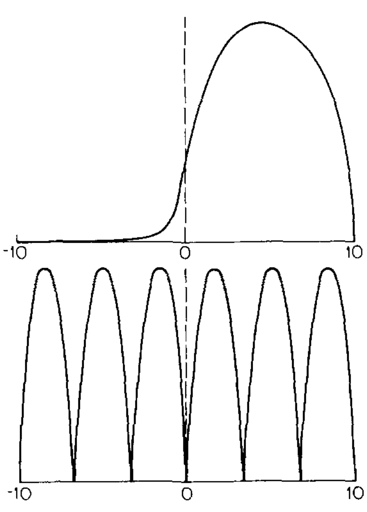


FIG. 2. A comparison of the absolute value of a low order, small $|\lambda|$, monokeric mode with that of a high order, large $|\lambda|$, dikeric mode.

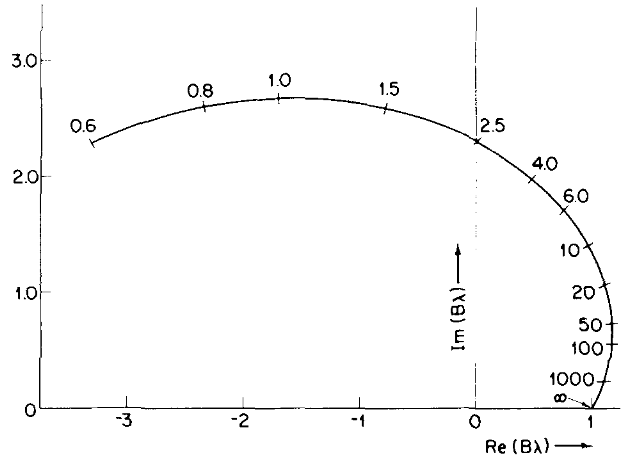


FIG. 3. The hodograph in the complex plane of the lowest eigenvalue for $A = -\infty$. The numbers labelling the curve give the values of B. Note that $B\lambda$, rather than λ itself, is the quantity plotted.

Thus, to within the accuracy of (5.7), the hodograph for λ_1 will apply to all the low-order modes with appropriate rescaling of the axes and tic-marks.

As $B \to 0$, or equivalently as $n \to \infty$ for fixed B, one can see from Fig. 2 that λ is tending towards the negative imaginary axis. This, of course, is what has been already been shown by (4.6). Thus, the high n and low n modes blend smoothly into one another.

For intermediate values of n neither (4.6) nor (5.7) is a good approximation, and the eigenfunctions are hybrids of the two extreme forms shown in Fig. 2. Nonetheless, enough has been obtained to give a good qualitative picture of the whole spectrum. In the next section, we will explore the bizarre behavior of the low-order modes via WKB, derive (5.7), and discuss its accuracy.

6. WKB

A. The method of matched asymptotic expansions

The grand strategy of this section is to derive asymptotic approximations by combining the WKB method with the method of matched asymptotic expansions (MMAE).

Although the WKB method itself is of ancient lineage, this pairing with the MMAE technique has been widely used only in the last decade. Historically, the WKB "connection formulas" were derived through a variety of coordinate transformations, integral representations, and other arguments. The books by Heading,[11] Dingle,[12] and Olver[13] describe this line of WKB development and extensions to higher order. After the MMAE method had been developed to a high art for boundary layer problems in fluid mechanics, however, it was recognized that it could be applied to a huge variety of other problems including WKB. The recent books by Bender and Orszag[14] and Nayfeh[15] present this "revisionist" derivation of WKB as well as a thorough treatment of the method of matched asymptotic expansions and its many uses. Because of its versatility and its familiarity to fluid dynamicists the WKB/MMAE approach is adopted here.

B. "Transition" points

Over most of the complex plane, Dickinson[3] showed that the WKB approximation to the general solution of (1.1) is given by an arbitrary linear combination of $W_-(x)$ and $W_+(x)$, where

$$W_-(x) = -iQ(x)^{-1/4}\exp[-2\kappa i\phi(\lambda x) + i\pi/4], \quad (6.1)$$

$$W_+(x) = Q^{-1/4}(x)\exp[2\kappa i\phi(\lambda x) - i\pi/4], \quad (6.2)$$

where, as defined in (3.4), $\kappa = \frac{1}{2}\lambda^{-1/2}$ and

$$Q(x) = 1/x - \lambda, \quad (6.3)$$

$$\phi(y) = \int_0^y \left(\frac{1}{x} - 1\right)^{1/2} dx \quad (6.4)$$

$$= \sin^{-1} y^{1/2} + y^{1/2}(1-y)^{1/2}. \quad (6.5)$$

The exceptional regions are the neighborhoods of the "transition points," which are defined to be the points where $Q(x)$ is either 0 or ∞—both make the WKB approximation singular.

The transition points thus play a central role in the analysis. Indeed, one can classify WKB problems according to the number and type of transition points in the same spirit in which one can classify a linear differential equation according to the number and type of its singularities.

The Whittaker equation (1.1) has two transition points: a simple pole at $x = 0$ and a "turning point" at

$$x_t = 1/\lambda. \quad (6.6)$$

The quantum harmonic oscillator, which is used as an example by most physics texts also has two transition points, but both are turning points.

In the parlance of matched asymptotics, the neighborhoods of the transition points constitute internal boundary layers. The WKB approximation using (6.1) and (6.2) is the "outer" solution; in the "inner" regions surrounding the transition points, $u(x)$ must be approximated using transcendentals more complicated than the exponentials appearing in (6.1) and (6.2). By matching the inner and outer solutions together and using the boundary conditions, one obtains a complete approximation to the problem.

When the differential equation has two transition points, however, there are two ways to carry out this recipe. The first is to define the inner region so that it simultaneously encloses *both* transition points. In this case, the inner approximation involves a sum of Whittaker functions (one turning point and one pole) or parabolic cylinder functions (two turning points), since these are the simplest functions with the required number of transition points. This would seem to send us round in circles when we attempt to solve (1.1) itself, but to apply asymptotic matching to fully determine the outer (WKB) solution of (1.1) and the eigenvalue, we need only the asymptotic expansions of the Whittaker functions given by (4.1) and (4.2) above, not the Whittaker functions themselves. Requiring that the WKB (outer) solution vanish at $x = B$ (and at $A = -\infty$) then gives the eigenrelation (7.1) below.

The alternative is to define two separate inner regions, one around each transition point. In this case, the inner solutions both involve Bessel functions of different orders—order one near the pole and order one-third (Airy functions) near the turning point. Though seemingly more complicated than the jointly matched or "Whittaker" matching described above, this separate or "double Bessel" matching has powerful advantages. First, because Bessel functions are simpler transcendentals than Whittaker functions, the double-inner-region method gives a simpler eigenrelation (7.1 is a function of two parameters, 7.4 only of one). Second, the use of separate inner approximations permits deeper and more precise insight into $u(x)$ instead of lumping both near-the-pole and near-the-turning point behavior together and hiding them behind the mysterious, inscrutable veil of a Whittaker function. Consequently, it is upon this "double Bessel" matching that our discussion will center.

Since the local analysis and the matching of inner and outer solutions has already been done—for the pole, by Dickinson,[3] and for the turning point by a number of independent workers more than a half a century ago—we shall merely quote their results. The challenge is to fit these two local analyses together with the boundary conditions to obtain a global description of the solution. The principle obstacle in completing this jigsaw puzzle is that while the "outer," WKB solution is always a sum of $W_+(x)$ and $W_-(x)$, the coefficients of the sum are *different* in different portions of the complex plane—Stokes' phenomenon. Thus, in order to make the final answer intelligible, it is necessary to digress briefly and explain this.

C. Stokes' phenomenon

If the WKB solutions $W_+(x)$ and $W_-(x)$ are written in the symbolic form

$$W(x) = Q^{-1/4}(x)e^{P(x)}, \quad (6.7)$$

then the Stokes lines are defined by,[16]

$$\mathrm{Im}[P(x)] = \text{const}, \quad (6.8)$$

and the anti-Stokes lines by

$$\mathrm{Re}[P(x)] = \text{const}. \quad (6.9)$$

On the Stokes lines, which will be indicated on the graphs below by *solid* lines, the WKB solutions grow or decay *exponentially* without change of phase. The anti-Stokes lines are curves of purely sinusoidal behavior: $W(x)$ oscillates without change of amplitude. To emphasize the oscillatory character of the WKB solutions upon them, the *anti-Stokes* curves will be graphed as *wavy* lines.

The heart of Stokes' phenomenon is that while $u(x)$ can always be represented as

$$u(x) \sim aW_-(x) + bW_+(x) \quad (6.10)$$

(except near a transition point), the coefficients must be different in different sectors of the complex plane. Within the sector bounded by adjoining anti-Stokes lines A_1 and A_2, one WKB solution (let it be $W_-(x)$ for definiteness) will be exponentially large ("dominant") in comparison to the other, which is said to be "subdominant" in that sector. It then follows that b in (6.10), because of the smallness of $W_+(x)$, can be arbitrary without violating the formal asymptotic equality because exponentially small quantities are com-

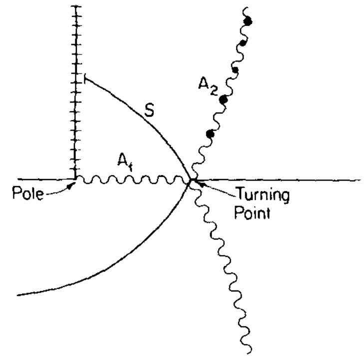


FIG. 4. The Stokes lines (solid) and anti-Stokes lines (wavy) for $\lambda = 1/100$. The branch line is marked with crosscuts. Black dots mark the zeros of the Whittaker function.

pletely ignored in Poincaré's definition of asymptotic relations. On the anti-Stokes lines, however, b must assume definite (and usually different) values because W_+ is the same magnitude as W_- upon them. Stokes established the convention[17] that the coefficient of the subdominant solution jumps from $b(A_1)$ to $b(A_2)$ as one crosses the Stokes line between them. This convention ensures that (6.10) will be numerically accurate near, as well as on, A_1 and A_2, and also, since W_+ is smallest in comparison to W_- on the Stokes line, that when b jumps, the corresponding jump in $u(x)$ is as small as possible.

The Stokes and anti-Stokes lines for the solutions of (1.1) for λ positive and real are shown in Fig. 4. Three Stokes and three anti-Stokes lines radiate from the turning point, but one of each ends on the branch line, so only two Stokes and two anti-Stokes lines radiate to infinity. Their number (two of each) is consistent with what one would have deduced directly from

$$u_{xx} - \lambda u = 0, \tag{6.11}$$

which approximates (1.1) as $|x| \to \infty$; parenthetically, we note that only these surviving pairs are relevant when performing the joint or "Whittaker" matching described above.

Making the simplifying assumption $A = -\infty$, justified previously, let us look first at the Stokes line radiating from the pole leftward to $x = -\infty$. Since $W_+(x)$ blows up exponentially along this Stokes line, b in (6.10) must be zero and $u(x)$ proportional to W_- alone, so that the boundary condition at $x = -\infty$ can be satisfied.

Since the coefficient of $W_+(x)$ can only jump to a nonzero value on a Stokes line, it follows that

$$u(x) \sim W_-(x) \quad \text{on } A_1, \tag{6.12}$$

which is the anti-Stokes line connecting the two points. The argument of the exponential in (6.1) is now pure imaginary, implying sinusoidal behavior. Dickinson[3] shows that, physically, (6.12) correponds to a Rossby wave propagating towards the pole and being *absorbed* there. So far so good, but (6.12) brings us face to face with an apparent paradox: how can a single complex exponential ever satisfy the boundary condition?

The answer is that it cannot; Stokes' phenomenon saves the day by forcing b to jump to a new nonzero value on the Stokes line S in Fig. 4. The new value of b is determined in two steps. First, the proper Airy function approximation to $u(x)$ in the vicinity of the turning point is found by demanding that it asympototically match to (6.12) along the anti-Stokes line A_1. Then, $b(A_2)$ is determined by matching the "inner" Airy approximation to the "outer" WKB solution along the anti-Stokes line A_2.

The matching is routine, but the result is not. The Airy functions Ai(x) and Bi(x) are both standing waves for x negative and real. This is fine for the quantum harmonic oscillator problem in which the other transition point is also a turning point; the two turning points reflect the wave back and forth between them to create the standing wave. Here, however, as shown by Dickinson,[3] the transition point at $x = 0$ is a perfect *absorber*.

The function that correctly matches to (6.12) is $\mathrm{Ai}(ze^{2\pi i/3})$ where $z \equiv \lambda^{2/3}(x - x_t)$, which has the asymptotic approximations

$$\begin{aligned}\mathrm{Ai}(ze^{2\pi i/3}) &\sim \frac{e^{i\pi/12}}{2\pi^{1/2}|z|^{1/4}}\, e^{i\frac{2}{3}|z|^{3/2}}, \quad \arg z = \pi[\text{on } A_1],\\ &\sim \frac{1}{\pi^{1/2}|z|^{1/4}} \cos[\tfrac{2}{3}|z|^{3/2} - \pi/4],\\ &\qquad \arg z = \frac{\pi}{3}[\text{on } A_2].\end{aligned} \tag{6.13}$$

The reader can easily verify that these large $|x|$ limits of the "inner" solution are identical with the $|x - x_t| \to 0$ limits of the WKB approximations along A_1 and A_2, where the former is given by (6.12) and the latter is found by matching with (6.13) to be

$$u(x) \sim Q(x)^{-1/4} e^{-i\pi\kappa} \cos[2\kappa\phi(\lambda x) - \kappa\pi - \pi/4]. \tag{6.14}$$

This plainly has an infinite number of zeros along the anti-Stokes line A_2 which are schematically denoted by the black dots in Fig. 4.

Unfortunately, when λ is real (as assumed for clarity above), all theses zeros are in the upper half-plane and are perfectly useless for satisfying the boundary condition at $x = B$ on the real axis. One can see now why the eigenvalue must be complex: when λ is moved into the upper half-plane (as consistent with Theorem 3 above), the turning point x_t is moved into the lower half-plane. It is then possible to make one of the zeros along A_2 coincide with the real axis.

Figure 5 shows the Stokes and anti-Stokes lines of the fourth mode for $B = 100$. The Whittaker function has three roots below the real axis, an infinite number above, and its fourth root along A_2 is real and satisfies the boundary condition at $x = B$.

We will see in the next section that "double Bessel" matching gives extremely accurate approximations to the low-order eigenvalues and eigenfunctions, but (6.14) must fail as $n \to \infty$ for fixed B, because as we have already seen $|\lambda| \to \infty$ in this limit. In turn, this implies that $|x_t| \to 0$, and when the turning point and the pole become too close together, it is no longer sensible—either physically or mathematically—to separate near-the-pole behavior from near-the-turning point behavior. The "Whittaker" matching is free from this defect and can in fact reproduce all the results of

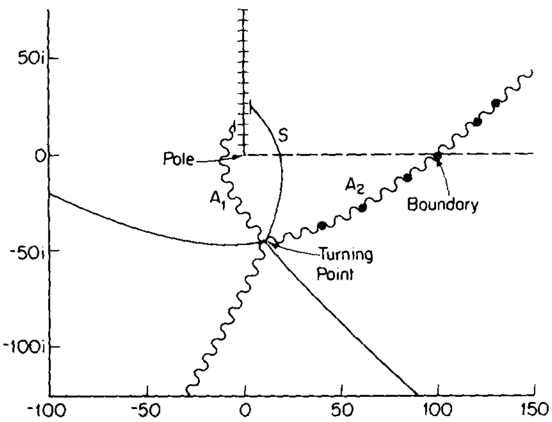


FIG. 5. The Stokes lines (solid) and anti-Stokes lines (wavy) for the fourth mode for $B = 100$ ($\lambda = 0.0050 + 0.0208i$). As in Fig. 4, the branch line is marked with crosscuts, and the zeros of the Whittaker function with black dots. The fourth root is on the real axis at $x = B$ so that the boundary condition is satisfied.

Sec. 4 on high-order modes if one relaxes the assumption $A = -\infty$. In practice, however, as shall be seen in the next section, the "double Bessel" matching gives acceptable accuracy when $|\lambda| < 1$, which turns out to include the range of n and B which is of primary physical interest.

D. Simplification of the "double Bessel" eigenrelation

The vanishing of (6.13) at $x = B$ is equivalent to the eigenrelation

$$2\kappa\phi(\lambda B) - \kappa\pi = (n - \tfrac{1}{4})\pi, \qquad (6.15)$$

where n is a positive integer, the mode number. One can eliminate the $\sin^{-1}$ implicit in $\phi(y)$ by letting

$$\lambda = \sin^2\tau/B, \qquad (6.16)$$

which transforms (6.14) to

$$\tau + \sin\tau\cos\tau - \tfrac{1}{2}\pi = [(n - \tfrac{1}{4})/B^{1/2}]\pi\sin\tau. \qquad (6.17)$$

What is striking about (6.16) is τ is not a function of B or n alone, but is rather a function of the single parameter

$$q \equiv (n - \tfrac{1}{4})/B^{1/2}. \qquad (6.18)$$

This implies that, just as with (5.7), the solutions of (6.17) are identical for all modes with appropriate rescaling of axes, i. e.,

$$\lambda_n(B) = \frac{9}{(4n-1)^2}\lambda_1\left[\frac{9B}{(4n-1)^2}\right]. \qquad (6.19)$$

E. The "Airy" approximation

Equation (6.17) has the drawback that it is only an *implicit* equation for λ. When the parameter

$$\sigma \equiv e^{\pi i/3} d_n / B^{1/3} \qquad (6.20)$$

is small, however, where

$$d_n \equiv [\tfrac{3}{8}\pi(4n - 1)]^{2/3}, \qquad (6.21)$$

one can solve (6.17) by a power series in σ to obtain

$$\lambda = (1/B)(1 + \sigma + \cdots), \qquad (6.22)$$

or in exponential form

$$\lambda = e^{\sigma}/B. \qquad (6.23)$$

Equation (6.23) is the "Airy approximation" given in (5.7) and the abstract; empirically (not systematically) it was found that the exponential form was much more accurate than the power series (6.22) for moderate σ, but both are *exact* in the limit $B\to\infty$ for fixed n, i. e., the limit $\sigma\to 0$.

The reason for the name "Airy approximation" is that in the limit $B\to\infty$ (5.7) shows that $\lambda\to 0$, implying that $|x_t|\to\infty$. Thus, the turning point and the pole move away from each other in this limit, and the radius over which the inner approximation, i. e.,

$$u(x) \sim \mathrm{Ai}(ze^{2\pi i/3}), \qquad (6.24)$$

where

$$z = \lambda^{2/3}(x - x_t), \qquad (6.25)$$

is valid, becomes larger and larger (in terms of $|z|$). Thus, the first few zeros of (6.13) are really the first few zeros of the Airy function (6.24). These are known constants, however, and the d_n given by (6.21) are in fact the nth roots of $\mathrm{Ai}(-z)$.[18] The approximation (6.22) is precisely what one would obtain by determining λ so as to make $x = B$ coincide with the nth root of the Airy function (6.24)—hence the name "Airy approximation" for (6.22) and (6.23), the latter being the form we shall actually use.

7. ACCURACY OF WKB

Much of the books on asymptotic approximations by Dingle and Olver is almost morbidly concerned with formal error terms and bounds, but this elaborate machinery is not useful here. The error in our approximate eigenmodes is not merely due to truncating an asymptotic series at lowest order but also depends on the accuracy of the value of λ which is used to evaluate the WKB expression. In turn, the error in λ may be large or small in comparison to the accuracy of the WKB approximation at the boundaries. Thus, the simplest and most reliable way to see how well WKB works is to compare the approximate results with the exact answers obtained by brute force numerical solution of (1.1).

The three eigenvalue approximations compared are

$$\tau + \sin\tau\cos\tau - \tfrac{1}{2}\pi = \left[\frac{(n-\frac{1}{4})}{B^{1/2}}\right]\pi\sin\tau - i\ln\left[\Omega\left(\frac{B^{1/2}}{2\sin\tau}\right)\right] \quad \begin{bmatrix}\text{Whittaker}\\ \text{matching}\end{bmatrix}, \qquad (7.1)$$

where

$$\Omega(\kappa) = (2\pi)^{1/2}\kappa^{1/2+\kappa}e^{-\kappa}/\Gamma(1+\kappa); \qquad (7.2)$$

$$\lambda = \frac{\sin^2\tau}{B}\tau, \qquad (7.3)$$

$$\tau + \sin\tau\cos\tau - \tfrac{1}{2}\pi = \left\{\frac{(n-\frac{1}{4})}{B^{1/2}}\right\}\pi\sin\tau \quad \begin{bmatrix}\text{double}\\ \text{Bessel}\\ \text{matching}\end{bmatrix}$$

with λ again related to τ through (7.3); and (7.4)

$$\lambda = (1/B)e^{\sigma} \quad \text{(Airy approximation)}, \qquad (7.5)$$

$$\sigma = \frac{e^{\pi i/3}}{B^{1/3}}[\tfrac{3}{8}\pi(4n - 1)]^{2/3}. \qquad (7.6)$$

TABLE I. A comparison of the exact and approximate eigenvalues λ for the lowest mode with $A = -\infty$ and various B. "Whittaker" refers to the WKB approximation with coefficients determined by matching with the asymptotics of the Whittaker function; "Double Bessel" is the WKB determined through matching the two local Bessel function approximations.

	$\mathrm{Re}(\lambda)$	$\mathrm{Im}(\lambda)$	$\lvert\lambda\rvert$	Phase	Relative $\lvert\lambda\rvert$	Errors Phase
$B = 100$	$\lvert\sigma\rvert = 0.50$					
Exact	0.011 58	0.005 50	0.012 81	25.407		
Whittaker	0.011 56	0.005 46	0.012 78	25.270	0.26%	0.54%
Double Bessel	0.011 61	0.005 47	0.012 83	25.231	0.14%	0.69%
Airy	0.011 66	0.005 39	0.012 84	24.804	0.19%	2.4 %
$B = 40$	$\lvert\sigma\rvert = 0.68$					
Exact	0.028 75	0.019 89	0.034 96	34.674		
Whittaker	0.028 69	0.019 73	0.034 82	34.525	0.40%	0.43%
Double Bessel	0.028 92	0.019 83	0.035 07	34.444	0.31%	0.66%
Airy	0.029 21	0.019 45	0.035 10	33.664	0.40%	2.9 %
$B = 10$	$\lvert\sigma\rvert = 1.08$					
Exact	0.095 96	0.1400	0.1696	55.660		
Whittaker	0.095 21	0.1389	0.1684	55.571	0.68%	0.16%
Double Bessel	0.097 62	0.1420	0.1723	55.494	1.6 %	0.29%
Airy	0.1021	0.1376	0.1713	53.439	1.1 %	4.0 %
$B = 4$	$\lvert\sigma\rvert = 1.46$					
Exact	0.1218	0.4960	0.5107	76.203		
Whittaker	0.1200	0.4919	0.5063	76.288	0.86%	0.11%
Double Bessel	0.1252	0.5200	0.5349	76.465	4.7 %	0.34%
Airy	0.1559	0.4952	0.5192	72.527	1.7 %	4.8 %
$B = 1$	$\lvert\sigma\rvert = 2.32$					
Exact	− 1.691	2.675	3.164	122.311		
Whittaker	− 1.712	2.644	3.150	122.927	0.46%	1.1%
Double Bessel	− 2.269	3.311	4.014	124.426	27. %	3.7%
Airy	− 1.355	2.888	3.190	115.130	0.82%	12. %
$B = 0.4$	$\lvert\sigma\rvert = 3.15$					
Exact	− 12.46	3.241	12.88	165.420		
Whittaker	− 12.59	3.146	12.98	165.966	0.77%	0.33%
Double Bessel	− 22.73	8.838	24.39	158.753	89. %	45. %
Airy	− 11.05	4.861	12.07	156.255	6.2 %	63. %

The first two approximations are *implicit* and (7.1) and (7.4) must be solved analytically or by perturbation theory; the Airy approximation is *explicit*. The Whittaker matching eigenrelation differs from that from double Bessel matching by only a single term, but that term causes the solution of (7.1) to depend on n and B *independently* instead of through a *single* parameter formed of B and n. Thus, (5.10) is true of the second and third approximations above (7.4) and (7.6) but not the first (7.1).

Physically, one is primarily interested in $n \leqslant 3$ and $B \geqslant 4$ since smaller values of B would correspond to unrealistically large (supersonic) winds, and $n > 3$ is rarely observed in the stratosphere. Dickinson[3] thoroughly discusses the physics of the atmospheric wave problem that motivated this work. Some controversies have arisen and it has been argued that Dickinson's WKB reasoning is rubbish because WKB is not sufficiently accurate to handle such singular SL problems of the fourth kind. It is thus a matter of physics—not merely numerical analysis—to examine the accuracy of our approximations.

Tables I through III compare the exact and approximate eigenvalues for the lowest three modes. The Whittaker-matched eigenrelation is the numerical star; the relative error is no worse than 1.1% for any of the values tabulated. The price is greater complexity (a Γ function of complex argument) and loss of insight because the near-turning-point and near-the-pole behaviors are lumped together into a single inner solution, and also because of the loss of (5.10) which shows that the curves $\lambda_n(B)$ all have similar shape for small n.

The double Bessel-matched approximation, though poorer, is still quite acceptable. In the range of physical interest, $n \leqslant 3$ and $B \geqslant 4$, the error is no worse than 10% in absolute value and 5% in phase. Both this and the Airy approximation—but not (7.1)—lose accuracy as σ (and therefore $\lvert\lambda\rvert$) increase where σ is defined by (6.20). For fixed B, σ increases as n increases as noted in the tables; so the tables for $n = 2$ and $n = 3$ are shorter than that for $n = 1$ to remind us that (7.4) and (7.5) are useful for an ever narrower range of B as the mode number becomes larger.

The Airy approximation (7.5) is the crudest of all, but it is still amazing that an explicit approximation of this simplicity can work so well for a problem whose differential equation is singular. For $n = 1$, the errors are less than 12% even for $B = 1$, so (7.5) is a good description of the entire hodograph in Fig. 3.

The approximate and exact eigenfunctions for the lowest mode are compared in Figs. 6, 7, and 8. Again, accuracy improves as B increases just as for λ, but the agreement is still remarkable.

Why does WKB work so well? The method of multiple scales,[14,15] which is one of many alternative ways of justify-

TABLE II. A comparison of the exact and approximate eigenvalues λ for the second mode with $A = -\infty$ and various B.

	Re(λ)	Im(λ)	$\|\lambda\|$	Phase	Relative $\|\lambda\|$	Errors Phase
$B = 100$	$\|\sigma\| = 0.88$					
Exact	0.010 97	0.010 99	0.015 52	45.047		
Whittaker	0.010 95	0.010 95	0.015 49	45.001	0.24%	0.10%
Double Bessel	0.010 99	0.010 98	0.015 33	44.972	0.04%	0.17%
Airy	0.011 23	0.010 71	0.015 52	43.635	0.00%	3.1 %
$B = 40$	$\|\sigma\| = 1.19$					
Exact	0.021 59	0.040 24	0.045 66	61.785		
Whittaker	0.021 56	0.040 18	0.045 60	61.789	0.14%	0.34%
Double Bessel	0.021 70	0.040 42	0.045 88	61.769	0.47%	0.02%
Airy	0.023 23	0.039 01	0.045 41	59.222	0.56%	4.1 %
$B = 10$	$\|\sigma\| = 1.89$					
Exact	-0.050 09	0.2729	0.2775	100.400		
Whittaker	-0.050 33	0.2725	0.2771	100.466	0.15%	0.08%
Double Bessel	-0.052 92	0.2789	0.2839	100.743	2.3 %	0.43%
Airy	-0.018 03	0.2572	0.2579	94.010	7.1 %	8.0 %
$B = 4$	$\|\sigma\| = 2.57$					
Exact	-0.8611	0.8200	1.189	136.398		
Whittaker	-0.8625	0.8183	1.189	136.507	0.01%	0.25%
Double Bessel	-0.9428	0.8704	1.283	137.286	7.9 %	2.0 %
Airy	-0.5516	0.7165	0.9043	127.590	24. %	20. %
$B = 1$	$\|\sigma\| = 4.08$					
Exact	-19.37	0.1890	19.37	179.441		
Whittaker	-19.38	0.1870	19.38	179.447	0.05%	1.1 %
Double Bessel	-24.44	3.376	24.68	172.136	27. %	1300. %
Airy	-7.11	-2.950	7.698	-157.463	60. %	Hopeless

ing WKB (away from transition points), provides an amusing and ironic answer.

In brief, the multiple scale argument states that the faster the eigenfunction oscillates, i. e., the greater the ratio of the "slow" scale on which the coefficients of the differential equation vary to the "fast" scale on which $u(x)$ itself is oscillating, the better the accuracy of the WKB approximation. The WKB eigencondition for a normal SL problem is that the total phase change on $[A,B]$ is $n\pi$, so the eigenfunction obviously oscillates more rapidly as n increases. In practice, this means that WKB is poor for the lowest mode, fair for moderate n, and superb for large n.

For the lowest mode of a singular SL problem of the fourth kind, however, the total phase change is usually greater than π and increases steadily with B. Figures 6 through 9 show that the real part of the lowest mode has no interior zeros for $B = 1$, one for $B = 5$, two for $B = 20$, and no fewer than four for $B = 100$. (The imaginary part oscillates similarly, but its roots coincide with those of the real part only at $x = B$). Because the eigenfunction graphed in Fig. 9 oscil-

TABLE III. A comparison of the exact and approximate eigenvalues λ for the third mode with $A = -\infty$ and various B.

	Re(λ)	Im(λ)	$\|\lambda\|$	Phase	Relative $\|\lambda\|$	Errors Phase
$B = 100$	$\|\sigma\| = 1.19$					
Exact	0.008 63	0.016 00	0.018 18	61.661		
Whittaker	0.008 70	0.016 04	0.018 25	61.525	0.41%	0.22%
Double Bessel	0.008 73	0.016 08	0.018 29	61.500	0.64%	0.26%
Airy	0.009 34	0.015 53	0.018 12	58.980	0.33%	4.3 %
$B = 40$	$\|\sigma\| = 1.61$					
Exact	0.005 20	0.058 13	0.058 36	84.886		
Whittaker	0.005 22	0.058 11	0.058 35	84.867	0.03%	0.02%
Double Bessel	0.005 22	0.058 48	0.058 71	84.904	0.60%	0.02%
Airy	0.009 68	0.055 16	0.056 01	80.048	4.0 %	5.7 %
$B = 10$	$\|\sigma\| = 2.56$					
Exact	-0.3559	0.3401	0.4923	136.300		
Whittaker	-0.3562	0.3398	0.4923	136.344	0.01%	0.03%
Double Bessel	-0.3698	0.3477	0.5076	136.766	3.1 %	1.1 %
Airy	-0.2169	0.2871	0.3598	127.068	27. %	21. %
$B = 4$	$\|\sigma\| = 3.48$					
Exact	-3.086	0.7583	3.178	166.194		
Whittaker	-3.087	0.7579	3.179	166.206	0.03%	0.01%
Double Bessel	-3.365	0.8694	3.475	165.513	9.4 %	4.9 %
Airy	-1.409	0.1866	1.421	172.457	55. %	45. %

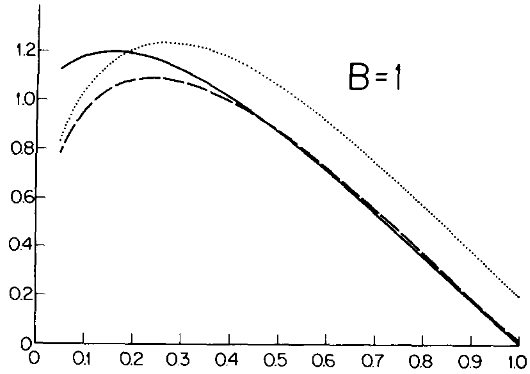


FIG. 6. A comparison of the exact (solid line), jointly (Whittaker) matched WKB (dashed line), and separately matched (double Bessel) WKB (dotted line) graphs for the real part of the lowest mode for $A = -\infty$, $B = 1$.

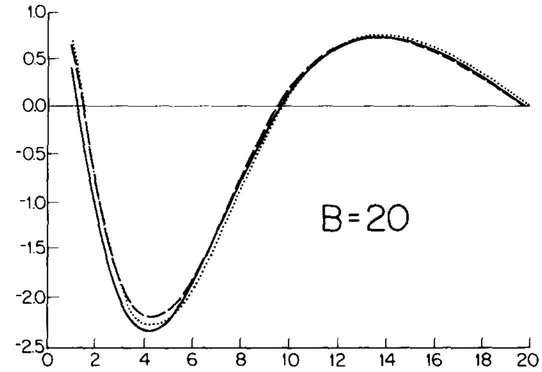


FIG. 8. A comparison of the exact (solid line), Whittaker-matched WKB (dashed line), and double Bessel-matched WKB (dotted line) graphs for the real part of the lowest mode for $A = -\infty$, $B = 20$.

lates as rapidly as the fifth mode (four interior zeros) of a normal SL problem, the WKB approximation to it has the same accuracy as for the fifth mode of a nonsingular equation—but it is the lowest mode nonetheless.

This increasing phase variation with B can be seen by noting that as $B \to \infty$ and $\lambda \to 0$ proportional to $1/B$ (from 7.5), one can approximate (1.1) over an increasingly large interval by

$$u_{xx} + (1/x)u = 0, \tag{7.7}$$

whose asymptotic approximation [matching to (6.12)] is proportional to

$$x^{1/4}e^{-2ix^{1/2}}. \tag{7.8}$$

The scale of the oscillation thus varies with x, but the total phase change on $[0,B]$ is obviously $O(2B^{1/2})$. [Using (6.1), one can show more precisely that the total phase change is $(\pi/2)B^{1/2}$ plus terms vanishing as $B \to \infty$.] Thus, WKB must inevitably improve for a given mode as B increases.

The Whittaker matched WKB, seen from the tables to be very good for small n, does but improve for large n; as noted earlier, it can—if we relax the restriction $A = -\infty$ inherent in (7.1)—reproduce all the results of Sec. 4 for higher-order modes as well. The double Bessel and Airy approximations fail for large n, but this is not the fault of the WKB per se. Rather, we have obtained (7.4) from (7.1) by replacing the complex gamma function by its approximation for large argument, and (7.5) from (7.4) by applying Taylor expansions in σ—both non-WKB simplifications.

Thus, the Whittaker-matched WKB works for all n here whereas WKB is successful only for large n for a Type I problem. Thus, we are led to an amusing and ironic conclusion: WKB actually works *better* for *singular* eigenproblems of Type IV than for the conventional nonsingular Sturm–Liouville equations of the classes so thoroughly studied in the past.

8. THE COMPLETE SPECTRUM

So far, we have looked at the small n and large n modes separately, the former with the additional assumption that $A = -\infty$. It is now appropriate to tie these ideas together by looking at a dozen chosen modes for a typical case ($A = -6$, $B = 6$). The eigenvalues are listed in Table 4, and Fig. 10 shows the lowest nine values of $\lambda^{1/2}$, which is graphed instead of λ itself for visual clarity. The modes can be grouped into three categories.

First, the four modes marked by *'s in the second column of the table (and by triangles in Fig. 10) are shining examples of the low-order modes discussed in Sec. 5. The ratio of the coefficient of $W_{-\kappa,\frac{1}{2}}$ to that of $M_{-\kappa,\frac{1}{2}}$, tabulated in the third column, is very large. The exact eigenvalues for

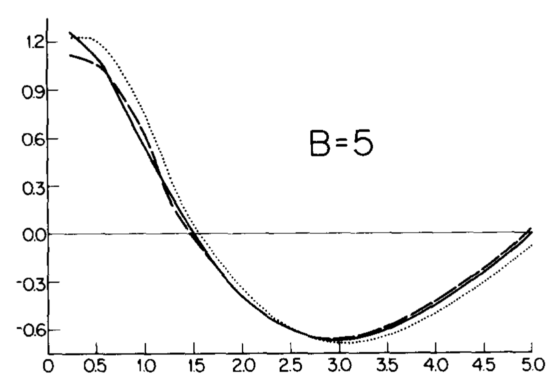


FIG. 7. A comparison of the exact (solid line), Whittaker-matched WKB (dashed line), and double Bessel-matched WKB (dotted line) graphs for the real part of the lowest mode for $A = -\infty$, $B = 5$.

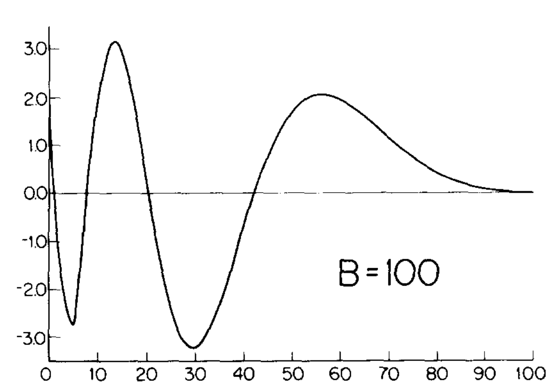


FIG. 9. The real part of the lowest mode for $B = 100$.

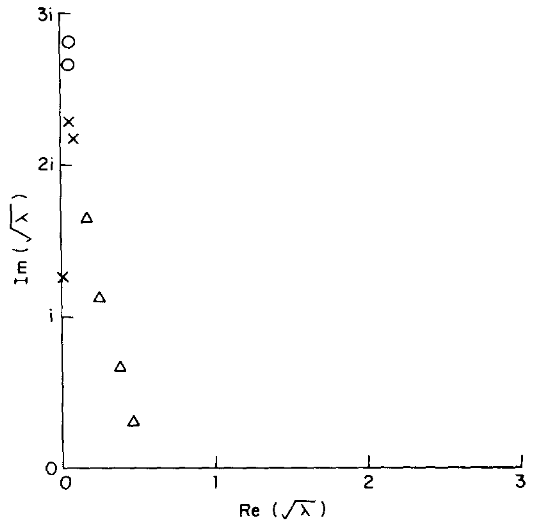


FIG. 10. The square root of λ (*not* λ itself) is shown for the lowest nine modes on $[-6,6]$. The four modes marked with triangles are well approximated by the corresponding eigenvalues for $[-\infty, 6]$. The two eigenvalues marked by circles are well approximated by (4.5). The crosses represent intermediate modes for which no simple approximation is known.

$A = -6$ are well approximated (to within 4%) by those for $A = -\infty$, which are given on the second line of each entry for these four asterisked modes.

The last five modes in the table (circles in Fig. 10) are examples of the large n dikeric modes discussed in Sec. 4. The second line of each entry for these five gives the approximate eigenvalues computed via (4.5) with that value of m which is given in the second column. Note that in this case $m = n + 2$, where n is the mode number determined by ordering the modes according to $|\lambda|$. For a normal SL problem, of course, $m = n$. Since (4.6) gives a purely real answer, the relative error in $\text{Im}(\lambda)$ is infinite, but the absolute errors in both the real and imaginary parts are small in comparison to $|\lambda_n - \lambda_{n+1}|$ and decrease *algebraically* [as does $\text{Im}(\lambda)$ itself as $n \to \infty$].

The three modes marked "Intermediate" in the table (crosses in Fig. 10) are hybrids of the two classes above. Whittaker-matched WKB would give an eigenrelation for them, but it would be both implicit and messy. As noted in the table, no simple explicit approximation is available for these modes.

Mode 4 is interesting because (i) it interrupts the pattern of the low-order monokeric modes which are well approximated by their counterparts for $A = -\infty$ and (ii) $\text{Im}(\lambda_4)$ is almost zero. As noted in Sec. 2, Theorem 3 shows that $\text{Im}(\lambda) = 0$ is possible only when $u(x)$ in effect satisfies three boundary conditions which can occur only on a set of measure zero in (A, B, n) parameter space. Here, the fourth mode—through sheer luck—happens to be close to one of these cases.

Modes 6 and 7 are interesting because, although they are nearly degenerate (i. e., $\lambda_6 \approx \lambda_7$) here—in contrast to the widely spaced eigenvalues of a normal one-dimensional SL eigenproblem with nonperiodic boundary conditions—they diverge wildly as the parameters are changed. For example, when $|A|$ is increased (with B fixed), the seventh mode—the one with the *higher* n and smaller ratio of W/M initially—rapidly becomes a pure monokeric mode like that illustrated in Fig. 12 (top). [At $A = 7$, $W/M = 13.4$ and $\lambda_7 = (-4.855, .5060)$, which differs little from its asymptotic ($A = -\infty$) value of $(-4.931, .4375)$.] The sixth mode, which is closer to a monokeric mode initially, behaves in a completely opposite fashion: the ratio W/M *decreases* very rapidly until it falls to *zero* at $A = -7.008$ where $\lambda_6 = (-3.743, 0)$. Thus, ($n = 6$, $A = -7.008$, $B = 6$) is one of the members of that set of measure zero, where λ is real and the eigenfunction, being proportional to $M_{-\kappa,\frac{1}{2}}$ alone, is an entire function.

Thus, these intermediate modes show that there is not a monotonic transition from the limiting behavior for small n to the limiting behavior for large n; rather, there can be some interleaving of the two. It is for this reason that the terms "monokeric" and "dikeric" were introduced earlier. Although the $n = 1$ mode is monokeric, i. e., exponentially decaying for $x < 0$ (unless A and B are both too small to be relevant to the original physical problem), and although one can prove that as $n \to \infty$ the modes must be dikeric, i. e., oscillatory on both sides of $x = 0$, modes of *moderate* n may resemble *either* graph in Fig. 2 or some hybrid of the two.

9. SUMMARY: A COMPARISON OF NORMAL AND SINGULAR STURM–LIOUVILLE EIGENPROBLEMS

The principal provable similarities between the first and fourth classes of Sturm–Liouville problems are the following. First, the eigenfunctions are orthogonal. Second, in the limit $n \to \infty$, the eigenfunctions and eigenvalues are essentially the same with or without the $1/x$ term in the differential equation. For finite n, there is (i) a small boundary layer about $x = 0$ and (ii) a nonzero imaginary part of the eigenvalue if the pole is present, but these disappear in the limit.

The principal differences are the following. First, the eigenvalues and eigenfunctions of a nonsingular, self-adjoint Sturm–Liouville eigenproblem are always real. Here, however, in spite of the fact that the problem is still self-adjoint, the eigenvalues and eigenfunctions are both *complex*.

Second, the modes of a Sturm–Liouville eigenproblem of the first kind can be characterized by their nodes: the nth mode has exactly $(n - 1)$ zeros on the interior of $[A,B]$.[2] Here, however, the real and imaginary parts of the low-order eigenfunctions have an ever increasing number of zeros as $B \to \infty$ with n fixed. The real part of the lowest mode for $B = 100$, illustrated in Fig. 9, has no fewer than four interior zeros, for example. Nor do the higher modes escape. The integer m which appears in the asymptotic ($n \to \infty$) eigenvalue formula (4.5) is generally different from the mode number n, where the latter is determined by ordering the eigenvalues according to $|\lambda|$. Thus, the $n = 8$ mode of Table IV has nine interior zeros instead of the expected seven. As explained in Sec. 7, this tendency of the singular modes of a given n to oscillate more rapidly than their counterparts for a nonsingular equation makes the WKB method actually work better, sometimes much better, for Sturm–Liouville problems of the fourth kind than for the nonsingular and seemingly more amenable equations of the first kind.

TABLE IV. The eigenvalues for $A = -6$, $B = 6$. The mode number n, the integer m which appears in (7.7) (if applicable), the absolute value of the ratio of the coefficient of $W_{-\kappa,\frac{1}{2}}$ to that of $M_{-\kappa,\frac{1}{2}}$, the approximate eigenvalues obtained by either setting $A = -\infty$ (for the purely singular modes) or using (4.5) for large n modes, and the relative errors of the approximations are also shown. Monokeric modes are indicated by asterisk in the second column.

n	m	W/M	$\mathrm{Re}(\lambda)$	$\mathrm{Im}(\lambda)$	*Relative Errors* $\mathrm{Re}(\lambda)$	$\mathrm{Im}(\lambda)$
1	*	1.06E5	0.1251	0.2850		
		$(A = -\infty)$	0.1251	0.2850	0.0%	0.0%
2	*	2.34E4	−0.2967	0.5203		
		$(A = -\infty)$	−0.2971	0.5204	0.2%	0.02%
3	*	141.4	−1.218	0.5634		
		$(A = -\infty)$	−1.225	0.5626	0.7%	0.12%
4	—	0.402	−1.602	0.0078		
			Intermediate—No Simple Approximation)			
5	*	18.6	−2.726	0.4900		
		$(A = -\infty)$	−2.781	0.5101	2.0%	4.1%
6	—	4.61	−4.717	0.3324		
			Intermediate—No Simple Approximation)			
7	—	2.48	−5.224	0.2186		
			Intermediate—No Simple Approximation)			
8	10	2.40	−7.156	0.2323		
		Eq. (4.5)	−6.85	0.0000	4.3%	∞
9	11	3.20	−7.996	0.3123		
		Eq. (4.5)	−8.29	0.0000	3.7%	∞
10	12	1.74	−10.14	0.1813		
		Eq. (4.5)	−9.87	0.0000	2.7%	∞
38	40	0.517	−109.791	0.0333		
		Eq. (4.5)	−109.662	0.0000	0.11%	∞
98	100	0.241	−685.454	0.0076		
		Eq. (4.5)	−685.389	0.0000	0.01%	∞

Third, the eigenvalues—all eigenvalues—of a normal one-dimensional Sturm–Liouville equation with nonperiodic boundary conditions are well separated. Here, however, $\lim_{B\to\infty} \lambda_n = 1/B$ for all fixed n (see Sec. 5) so that the eigenvalues of the lowest few modes cluster about a common value and become quasidegenerate. Furthermore, the proportionality to $1/B$ is different from the $1/B^2$ [strictly, $1/(B-A)^2$] for a given λ_n of a first kind eigenproblem in this same limit.

These differences and similarities are provocative, but a number of important questions remain for future research. First, completeness. It is plausible, especially in view of their asymptotic identity with ordinary sine functions, to suppose that the modes are complete at least for the original partial differential equation which gave rise to this problem. The possibility of expanding an *arbitrary* analytic function, however, in terms of a series of *singular* functions like the modes of (1.1) raises fascinating questions that I will not attempt to answer here.

Second, one may ask: would the conclusions given above all hold if the first-order pole in (1.1) were replaced by a second-order pole or some other species of singularity? (Olver[19] has made a start on this). Clearly, a rich harvest awaits the future in these Sturm–Liouville eigenproblems of the fourth kind.

ACKNOWLEDGMENTS

I would like to thank Robert E. Dickinson and K. K. Tung for inspiring this problem. Both of them and Richard S. Lindzen provided helpful discussions. I also wish to thank the referee for his comments which have greatly improved the readability of this paper and made it accessible to a wider audience. This work was supported by N. S. F. grant OCE-79-09191 and NASA grant NSG-7209 at The University of Michigan and by NASA grant NGL-22-007-228 at Harvard. I am appreciative of the hospitality extended to me by Professor Richard Lindzen at Harvard where this work was completed.

APPENDIX A: CONFLUENT HYPERGEOMETRIC FUNCTIONS

The Whittaker functions of Sec. 2 are related to the standard confluent hypergeometric functions by the identities

$$M_{-\kappa,\frac{1}{2}}(y) = e^{-y/2} y M(1+\kappa,2,y), \tag{A1}$$

$$W_{-\kappa,\frac{1}{2}}(y) = e^{-y/2} y U(1+\kappa,2,y), \tag{A2}$$

which have the power series representations

$$M(1+\kappa,2,y) = \sum_{m=0}^{\infty} \frac{(1+\kappa)_m}{(2)_m} \frac{y^m}{m!} = 1 + \frac{1+\kappa}{2} y + \frac{(1+\kappa)(2+\kappa)}{12} y^2 + \cdots, \tag{A3}$$

$$U(1+\kappa,2,y) = \frac{1}{\Gamma(1+\kappa)} \biggl(\frac{1}{y} + \kappa M(1+\kappa,2,y) \log y + \kappa \sum_{m=0}^{\infty} \frac{(1+\kappa)_m}{(2)_m} \frac{y^m}{m!} [\psi(1+\kappa+m) - \psi(1+m) - \psi(2+m)] \biggr), \tag{A4}$$

where

$$(x)_m = x(x+1)\cdots(x+m-1), \tag{A5}$$

and $\psi(x)$ is the logarithmic derivative of the gamma function ("digamma" function). The reason for the factor of $\Gamma(1+\kappa)$ in (3.6) is to eliminate the corresponding factor in (A4).

The corresponding asymptotic approximations for fixed κ, $y \to \infty$, are given by (4.1) and (4.2) above.

As a final note, the formulas of the paper require computing two transcendental functions—$\sin^{-1}(z)$ and $\Gamma(z)$—for *complex* argument. For the former, however, identity 4.4.37 of Abramowitz and Stegun[20] reduces the task to evaluating (i) the complex logarithm, which is a built-in library function on most computers and (ii) the arcsine function for a *real* argument between 0 and 1, which can be done via the polynomial approximation 4.4.46 of Abramowitz and Stegun.[20] The complex gamma function can be evaluated by using its well-known recursion relation $\Gamma(z+1) = z\Gamma(z)$ to march out to large z, using its known asymptotic expansion, and then marching back the same way. A FORTRAN program to do this is given by Lucas and Terril.[21]

APPENDIX B: THE DISCRETE AND CONTINUOUS SPECTRUM

There are two fundamentally different ways of analyzing the inviscid limit. The first, adopted by Dickinson[3] is the continuum modes approach. This has the great advantage that all the arithmetic is real, but it has the disadvantage that any physically realizable solution is an *integral* over the *real* eigenvalue λ. For the special case of a δ-function lower boundary forcing, he was able to perform the integrals via stationary phase.

Unfortunately, the need for λ integration implies that a continuum mode—i. e., a Whittaker function for some particular real value of λ—is never a legitimate solution of the original problem. [To put it another way, there is no sum of $M_{-\kappa,\frac{1}{2}}(-x/\kappa)$ and $W_{-\kappa,\frac{1}{2}}(-x/\kappa)$ which can satisfy both boundary conditions (1.2) with κ and λ real.] It is therefore exceedingly dangerous to infer the behavior of the integrated solution from that of a single continuum mode, and this has led to some confusion. For example, Dickinson proved that the momentum flux ($\overline{u'v'}$ in meteorological parlance) is everywhere constant except for a jump at the singularity, and is therefore nonzero on at least one boundary for a *single* continuum mode. Physically, however, this quantity must vary with latitude so as to vanish (like the wave itself) on *both* boundaries. Although this variability has been described[6] as "contrary to a conclusion of Dickinson," such criticism is a comparison of apples and oranges. When the λ integration is performed, mutual cancellation of different values of λ permits $\overline{u'v'}$ to vary and the integrated wave to satisfy the boundary conditions. Since the λ integration cannot be performed analytically, however, this need for integration limits the amount of insight that can be obtained from the continuum modes.

With friction, as in (2.7), be it ever so small, the continuum spectrum breaks up into discrete normal modes which have well-defined limits as the friction tends to zero. The two advantages of this second approach are first, each mode is an independent solution of the original problem so that no integration over λ is necessary. Second, numerical calculations normally incorporate weak dissipation to survive the singularity, so discrete normal modes are what the computer programs actually calculate as in Simmons[7] and Boyd.[9] The disadvantages are that now both the eigenvalues and eigenfunctions are complex and one must wrestle with Stokes' phenomenon.

If no additional approximations or assumptions are made, both approaches—in spite of their great dissimilarity in form—give the same numerical answer. Dickinson (private communication) has suggested a more familiar example that makes this numerical equality more plausible. The Fourier integral

$$I(x) = \int_{-\infty}^{\infty} \frac{e^{i\lambda x}\, d\lambda}{\cosh(\lambda x)} \tag{B1}$$

can be numerically evaluated by direct integration along the real λ axis via the trapezoidal rule. Alternatively, one can complete the contour via a semicircle of infinite radius in the upper half-plane and evaluate the integral as an infinite sum of the residues at the poles of the integrand on the positive imaginary λ axis. These two options are the same as for the singular eigenproblem: the *integral* over *real* λ or the infinite *sum* of discrete *complex* values of λ, and both give the same result.

This point, too, has caused confusion. Physically, verti-

TABLE V. The coefficients of the Chebyshev series for $B\lambda_n$ for the lowest three eigenvalues with $A = -\infty$. The argument of the polynomials is $x = 2^{5/3}/B^{1/3} - 1$. The approximations are accurate for $B \in [4, \infty]$.

Degree of Polynomial	*Mode Number* $n=1$ Real part	Imag. part	$n=2$ Real part	Imag. part	$n=3$ Real part	Imag. part
0	1.881 41	1.867 98	− 0.444 47	3.436 86	− 5.638 86	4.006 26
1	− .237 46	1.011 37	− 2.067 50	1.769 41	− 6.114 44	1.767 83
2	− .197 85	0.060 24	− 0.989 17	− 0.063 30	− 2.806 82	− 0.507 14
3	− .019 13	− 0.019 37	− 0.152 78	− 0.130 97	− 0.567 80	− 0.273 23
4	0.000 41	− 0.002 26	− 0.010 63	− 0.015 90	− 0.053 98	0.022 34
5	0.000 60	− 0.000 15	− 0.001 99	0.001 28	0.006 91	0.030 18

cally propagating waves must decay exponentially with height because of absorption at the latitude of the singularity. In the discrete modes procedure, the decay rate is dependent upon the imaginary part of λ_n. This might seem worrisome because Dickinson's formalism involves only real λ, but in fact his λ integrated solution[3] decays with height as it should.

Nonetheless, it is obviously desirable to incorporate this decay rate and other properties explicitly in the modes rather than in a λ integration which cannot be analytically performed. For this reason, the discrete modes approach has been adopted here. Because of its greater complexity (literally and figuratively), this procedure is complementary rather than competitive with the continuum modes approach of Dickinson[3] and others.

APPENDIX C: CHEBYSHEV EXPANSIONS FOR THE EIGENVALUES

Although the eigenvalue relation—even when simplified via the WKB method—cannot be solved in terms of any known transcendental, it is nonetheless possible to provide *analytic exact* solutions in the form of Chebyshev series in the parameters. The method is thoroughly explained in Boyd,[22] so it will not be repeated here. To provide a springboard for future work and a sample of the usefulness of the Chebyshev technique, Table V gives the first six expansion coefficients for the lowest three modes with $A = -\infty$.

The form of the approximation is

$$\lambda_n(B) = \frac{1}{B}\left(\frac{1}{2} a_0^{(n)} + \sum_{m=1}^{5} a_m^{(n)} T_m(x)\right), \tag{C1}$$

where

$$x = 2^{5/3}/B^{1/3} - 1. \tag{C2}$$

On the interval $B \in [4, \infty]$, the error in (C1) is at most one part in 4000 for $n = 1$, one part in 700 for $n = 2$, and one part in 200 for $n = 3$.

One can equally well obtain expansions accurate for small B. Accuracy for a given number of polynomials can be improved by choosing a meeting point between the large and small B approximations which increases with n, instead of taking $B = 4$ as the lower limit for all n as done here.

[1]The myth that as in a normal SL problem all the eigenvalues were of one sign persisted until about 1965, when it was discovered that there was in fact an infinite number of eigenvalues of the opposite sign. The eigenfunctions themselves fall into two classes: one class which is oscillatory between the apparent singularities and exponentially small at higher latitudes, and a second class which is oscillatory between the poles and apparent singularities and exponentially small near the equator. Because half the spectrum was left out, all atmospheric tidal calculations up to 1965 were completely wrong. Even then, doubts persisted about the completeness of the eigenfunctions that were not resolved until a rigorous completeness proof was given in 1970. The whole sordid mess is reviewed by R. S. Lindzen, Lect. Appl. Math. **14**, 293–362 (1971).
This history of confusion and error in such recent times for a relatively easy problem should convince the reader that the subject of the present work is far from trivial; because the eigenfunctions are analytic and the eigenvalues are real, an SL problem of the third kind like the tidal equation is much closer to normal SL problems of the first two classes than the singular fourth kind studied here.

[2]A compact and highly readable treatment of normal Sturm–Liouville theory is given in Chap. 6 of P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).

[3]R. E. Dickinson, J. Atmos. Sci. **25**, 984 (1968).
[4]K. K. Tung, "Stationary Atmospheric Long Waves and the Phenomena of Blocking and Sudden Warming," Ph. D. Thesis, Harvard University, 1977.
[5]K. K. Tung, Mon. Weather Rev. **107**, 751 (1979).
[6]A. J. Simmons, Q. J. Roy. Meteorol. Soc. **100**, 76 (1974).
[7]A. J. Simmons, Q. J. Roy. Meteorol. Soc. **104**, 595 (1978).
[8]J. P. Boyd, "Planetary Waves and the Semiannual Warming in the tropical Upper Stratosphere," Ph. D. Thesis, Harvard University, 1976; J. P. Boyd, J. Atmos. Sci., to be published (1981).
[9]G. F. Carrier, M. Krook, and C. E. Pearson, *Functions of a Complex Variable* (McGraw-Hill, New York, 1966), p. 319.

[10]An attempt was made to obtain an explicit approximation to $\lambda_{\rm im}$, but the lowest order result [to $O(\kappa)$], obtained via the algebraic manipulation language REDUCE2, contained no fewer than 37 terms—too many to be practical. The difficulty is too many parameters: A, B, $\ln(A)$, $\ln(B)$, κ, $\ln(\kappa)$, π, and Euler's constant γ all appear in the equation $\Delta(\kappa) = 0$. [The logarithms come from the $y^{\pm\kappa}$ terms in (4.1) and (4.2).] There is little one can do to simplify the mess because the terms are oscillatory. It can be seen in Table IV that the imaginary part of λ in fact fluctuates irregularly from mode to mode.

[11]J. Heading, *An Introduction to Phase-Integral Methods* (Wiley, New York, 1962).
[12]R. B. Dingle, *Asymptotic Expansions: Their Derivation and Interpretation* (Academic, New York, 1973).
[13]F. W. J. Olver, *Asymptotics and Special Functions* (Academic, New York, 1973).
[14]C. E. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers* (McGraw-Hill, New York, 1978).
[15]A. H. Nayfeh, *Perturbation Methods* (Wiley, New York, 1973).

[16]There is much confusion in the literature about these definitions. Bender and Orszag (Ref. 14) reverse the terminology without even noting that there is a controversy. The definitions (6.8) and (6.9) are in accord with Stokes' own as discussed in F. W. J. Olver (Ref. 13, p. 518) and also agree with those of J. Heading (Ref. 11) and R. B. Dingle (Ref. 12).

[17]I use the word "convention" because under Poincaré's rather forgiving definition of asymptoticity, any way of varying b across the sector is legitimate so long as $b(A_1)$ and $b(A_2)$ take their proper values. One could, for example, vary b linearly with $\arg x$. However, $W_{-\kappa,\frac{1}{2}}(y)$ has an exact integral representation, valid for $0 \leqslant \arg y \leqslant \pi$, which upon expansion gives $W_1(x)$ alone as its first term. Though one cannot be entirely comfortable with jumps in the representations of functions which may themselves be smooth or even entire, varying b linearly with $\arg y$ or otherwise using a convention different from Stokes' will generally only make the numerical error greater.

[18]Strictly speaking, (6.21) gives the large n asymptotic approximation to the roofs of $\mathrm{Ai}(-z)$ rather than the exact zeros, but since the error is only one part in 200 even for $n = 1$, I have ignored this largely irrelevant distinction in the body of the paper.

[19]F. W. J. Olver, Philos. Trans. R. Soc. London, Ser. A **289**, 501 (1978).
[20]M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965), pp. 80–81.
[21]C. W. Lucas, Jr. and C. W. Terril, Commun. ACM **14**, 48 (1971).
[22]J. P. Boyd, J. Math. Phys. **19**, 1445 (1978).

Convergence of Berryman's iterative method for some Emden–Fowler equations

C. D. Luning
Department of Mathematics, Sam Houston State University, Huntsville, Texas 77340
W. L. Perry
Department of Mathematics, Texas A&M University, College Station, Texas 77843

(Received 21 October 1980; accepted for publication 18 March 1981)

The equation $y'' + \lambda a(x) y^{\alpha}(x) = 0$, $0 < x < 1$, $y(0) = y(1) = 0$, $\alpha > 0$, arises in the study of nonlinear diffusion equations connected with crossfield diffusion in plasmas. We show that for a particular choice of starting iterate, the computational method which Berryman developed for this equation does converge, and furnishes upper and lower bounds for $y(x)$.

PACS numbers: 02.30.Hq

I. INTRODUCTION

In a sequence of papers[1–3] Berryman, and Berryman and Holland consider the nonlinear diffusion equation $(d/dx)[D(n)(dn/dx)] = f(x)\,(dn/dx)$ for $0 \leqslant x \leqslant 1$, where n is particle density, x is the spatial variable in one dimension, t is time, and $D(n)$, which is proportional to n^{δ} $(\delta > -1)$, is the diffusion coefficient. By looking for separable solutions, those authors were led to an investigation of the eigenvalue problem (EVP)

$$y''(x) + \lambda a(x) y^{\alpha}(x) = 0, \quad 0 < x < 1, \quad \alpha > 0,$$
$$y(0) = y(1) = 0. \tag{1}$$

We note that if a solution pair (λ, y) exists for (1), then for any $c > 0$, $(\lambda c^{1-\alpha}, cy)$ is also a solution pair. Thus, in order to specify a solution, some normalization must be chosen. We also note by means of the transformation $\tilde{y} = \lambda^{s} y$, where $s = 1/(\alpha - 1)$, that problem (1) is equivalent to the boundary value problem (BVP)

$$\tilde{y}''(x) + a(x)\tilde{y}^{\alpha}(x) = 0, \quad 0 < x < 1, \quad \alpha > 0,$$
$$\tilde{y}(0) = \tilde{y}(1) = 0. \tag{2}$$

Problem (2) has been extensively studied and, under certain conditions, uniqueness of the positive solution has been proved.[4,5] Thus if (2) has a unique solution, the shape of any solution to (1) is known within a multiplicative factor.

In this paper, the term "solution" as applied to the EVP (1) shall be used in the context of a particular specified normalization. In Refs. 1–3, Berryman chose the normalization $\max_{[0,1]} y(x) = 1$. In Ref. 1, Berryman developed a numerical method for finding an approximation to the solution of (1). The method consists of a Picard type iteration, but convergence of the iteration to an actual solution is not proved. The iteration may be described as follows:

(a) select as a starting iterate any continuous function $S_0(x)$ with $S_0(x) \geqslant 0$, $0 \leqslant x \leqslant 1$, and $\max_{[0,1]} S_0(x) = 1$;

(b) solve for $v_k(x)$, $k = 1,2,3\cdots$

$$v_k''(x) = -a(x) S_{k-1}^{\alpha}(x),$$
$$v_k(0) = v_k(1) = 0;$$

(c) set $\mu_k = \max_{[0,1]} (v_k(x))^{-1}$

$$S_k(x) = \mu_k v_k(x), \quad 0 \leqslant x \leqslant 1.$$

We will refer to this method as Berryman's scheme. In the Refs. 1–3, Berryman and Holland have applied Berryman's scheme to problems with various choices for α and $a(x)$. From their numerical results it appears that for some choices of α, $a(x)$, and $S_0(x)$, the Berryman iteration scheme is stable, convergent, and in some cases monotonically convergent. However, the proof of convergence and monotonicity are left as open questions. Another question left open is upper and lower bounds for the solution $y(x)$ in the important case when $a(x)$ is symmetric about $x = \frac{1}{2}$ and is nondecreasing on $(0, \frac{1}{2})$.

In this paper we provide some answers to these questions. We show that for certain $S_0(x)$ the sequence $\{S_k(x)\}$ is uniformly convergent to a solution of (1); that depending on the choice of $S_0(x)$, the sequence $\{S_k(x)\}$ may converge monotonically upward or monotonically downward to the solution; and finally we prove that in the symmetric case

$$2T(x) \leqslant y(x) \leqslant 4x(1-x), \tag{3}$$

where

$$T(x) = \begin{cases} x & 0 \leqslant x \leqslant \frac{1}{2} \\ 1 - x & \frac{1}{2} \leqslant x \leqslant 1 \end{cases}.$$

The iteration we use to prove these results is a Picard type iteration. However, it is different from the iteration in Berryman's scheme in that we use a different normalization. We do show the relationship between the two iterations.

II. ITERATIVE METHOD

We assume that $a(x) \geqslant 0$, $a(x) \not\equiv 0$, is continuous on $(0,1)$, that $\alpha > -1$ and that $\int_0^1 t^{\alpha} a(t)\,dt < \infty$. Let $u_0(x) = x$ and define the sequences $\{u_n(x)\}$, $\{\lambda_n\}$, $n = 1,2,3,\cdots$ by

$$u_n''(x) = -\lambda_n a(x) u_{n-1}^{\alpha}(x),$$
$$u_n(0) = u_n(1) = 0,$$
$$\lambda_n = 1\Big/\int_0^1 (1-\xi) a(\xi) u_{n-1}^{\alpha}(\xi)\,d\xi. \tag{4}$$

By considering the Green's function for the operator $Lu = u''$, $u(0) = 0$, $u(1) = 0$, it follows that the definition of λ_n in (4) forces the initial condition $u_n'(0) = 1$, $n = 1,2,3,\cdots$, which is the normalization that we use. In Ref. 6 we proved

 0022-2488/81/081591-05$01.00 © 1981 American Institute of Physics

that the sequences $\{\lambda_n\}$, $\{u_n\}$ defined by (4) converge to a solution pair $\{\lambda, y\}$ of (1). For the case $\alpha > 1$, another iteration scheme is proved in Ref. 7 to converge, however it is computationally significantly slower than the corresponding scheme in Ref. 6. The theorems proved in Ref. 6 are:

Theorem 1: If $u_0(x) = x$, if $a(x)$ is as above, if $\alpha > 0$, and if $\{u_n\}_{n=1}^{\infty}$, $\{\lambda_n\}_{n=1}^{\infty}$ are defined as in (4), then

$$0 < u_{n+1}(x) < u_n(x), \quad 0 < x < 1, \quad n = 0,1,2,\cdots,$$
$$0 < \lambda_n < \lambda_{n+1}, \; n = 1,2,\cdots. \tag{5}$$

Moreover, there is a positive solution pair $\{\lambda, y\}$ of (1) such that $\lim_{n\to\infty} \lambda_n = \lambda$ and $\lim_{n\to\infty} u_n(x) = y(x)$ uniformly on $[0,1]$.

Theorem 2: If $a(x)$ is as above, if $-1 < \alpha < 0$, and if $\{u_n(x)\}_{n=1}^{\infty}$, $\{\lambda_n\}_{n=1}^{\infty}$ are defined as in (4), then for $n \geqslant 1$,

$$0 < u_{2n-1}(x) < u_{2n+1}(x) < u_{2n}(x) < u_{2n-2}(x), \; 0 < x < 1,$$
$$\lambda_{2n} < \lambda_{2n+2} < \lambda_{2n+1} < \lambda_{2n-1}. \tag{6}$$

Moreover, there is a positive solution $\{\lambda, y\}$ of (1) such that $\lim_{n\to\infty} \lambda_n = \lambda$ and $\lim_{n\to\infty} u_n(x) = y(x)$ uniformly on $[0,1]$.

In order to show the convergence of Berryman's scheme, we give a simple relationship between our sequence $\{u_n(x)\}$ and his sequence $\{S_n(x)\}$ in the case where $S_0(x) = x$. Since $v_1''(x) + a(x)S_0^{\alpha}(x) = 0$ and $u_1''(x) + \lambda_1 a(x) u_0^{\alpha}(x) = 0$, it follows that $(\lambda_1^{-1} u_1)'' + a(x) u_0^{\alpha}(x) = 0$. Since $u_0(x) = S_0(x)$ and $\lambda_1^{-1} u_1$ and v_1 satisfy the same boundary conditions, we have $v_1 = \lambda_1^{-1} u_1$. Thus,

$$\mu = 1/\max_{0<x<1} v_1(x) = \lambda_1/\max_{0<x<1} u_1(x),$$
$$S_1(x) = \mu_1 v_1(x) = u_1(x)/\max_{0<x<1} u_1(x). \tag{7}$$

In the same manner it can be shown that

$$S_k(x) = u_k(x)/\max_{0<x<1} u_k(x), \qquad k = 2,3,4,\cdots. \tag{8}$$

From Theorem 1 and Theorem 2 there is a solution pair (λ, y) of (1) such that $u_n(x) \to y(x)$ uniformly on $[0,1]$ and $\lambda_n \to \lambda$. From (8) it now follows that there is a function $S(x)$ such that $S_k(x) \to S(x)$ uniformly on $[0,1]$. Moreover, there is a real number $c > 0$ such that $S(x) = cy(x)$, and thus $(\lambda c^{1-\alpha}, S(x))$ is also a solution pair of (1). We thus have that if $a(x)$ is as above, if $-1 < \alpha$, and if $S_0(x) = x$, then Berryman's iterative sequence converges to a solution of (1). Unless more is known about $a(x)$, we can not infer monotonicity of the sequence $\{S_k(x)\}$ from the monotonicity of the sequence $\{u_k(x)\}$. We also note that in general the convergence of Berryman's iteration for other choices of $S_0(x)$ is still an open question.

III. THE SYMMETRIC CASE

We now consider the case where $a(\frac{1}{2} - x) = a(\frac{1}{2} + x)$, $a(x)$ is nondecreasing on $(0,\frac{1}{2})$, and $\alpha > 0$. In this section we consider the two starting iterates $\check{u}_0(x) = x(1-x)$ and $\hat{u}_0(x) = T(x)$. The corresponding iterative sequences defined by (4) are denoted by $\{\check{\lambda}_k\}$, $\{\check{u}_k(x)\}$ and $\{\hat{\lambda}_k\}$, $\{\hat{u}_k(x)\}$, respectively. Before analyzing the sequences $\{\check{\lambda}_k\}$, $\{\check{u}_k(x)\}$, $\{\check{\lambda}_k\}$, and $\{\hat{u}_k(x)\}$, we give a lemma which will be used repeatedly.

Lemma 1: If $f(x), f'(x), f''(x)$ are continuous on $(0,1)$, $f(x)$ is symmetric about $x = \frac{1}{2}$, $f(0) = f'(0) = 0$, $f''(x) > 0$ on some interval $(0,\epsilon)$, and $f''(x)$ changes sign at most once on $(0,\frac{1}{2})$, then $f(x) > 0$ on $(0,1)$ and $f(x)$ is nondecreasing on $(0,\frac{1}{2})$.

The proof of Lemma 1 is straightfoward and is omitted. We point out that $f''(x)$ changes sign exactly once on $0 < x < \frac{1}{2}$ in order to maintain the continuity of $f'(x)$ at $x = \frac{1}{2}$. However this is not used in any of the arguments.

Theorem 3: The sequence $\{\check{u}_k(x)\}$ is monotone increasing for each x, and $\{\check{\lambda}_k\}$ is monotone decreasing. Moreover, there is a positive solution pair $\{\check{\lambda}, \check{u}(x)\}$ such that $\check{u}_k(x) \to \check{u}(x)$ uniformly on $[0,1]$ and $\check{\lambda}_k \to \check{\lambda}$.

Proof: For the sequence $\{\check{u}_k(x)\}$ we have from (4)

$$\check{u}_1''(x) - \check{u}_0''(x) = \frac{-a(x)\check{u}_0^{\alpha}(x)}{\displaystyle\int_0^1 (1-\xi) a(\xi) \check{u}_0^{\alpha}(\xi)\, d\xi} + 2. \tag{9}$$

Clearly $\check{u}_1''(x) - \check{u}_0''(x)$ is symmetric about $x = \frac{1}{2}$, and since $a(x)$ is bounded, $\check{u}_1''(x) - \check{u}_0''(x)$ is positive for x near zero. Since $a(x)\check{u}_0^{\alpha}(x)$ is nondecreasing on $0 < x < \frac{1}{2}$, $\check{u}_1''(x) - \check{u}_0''(x)$ can change sign at most once on $0 < x < \frac{1}{2}$. Also $\check{u}_1(0) - \check{u}_0(0) = 0$, $\check{u}_1'(0) - \check{u}_0'(0) = 0$. Thus by Lemma 1, $\check{u}_1(x) - \check{u}_0(x)$ is positive on $(0,1)$ and is nondecreasing on $(0,\frac{1}{2})$. It now follows from the definition of λ_k in (4) that $\check{\lambda}_2 < \check{\lambda}_1$. Of course $\check{u}_1(x)$ is symmetric about $x = \frac{1}{2}$. We also claim that if $0 < c_1 < c_0$, then $g_1(x) = -c_1\check{u}_1(x) + c_0\check{u}_0(x)$ can change sign at most once on $0 < x < \frac{1}{2}$. To see this we have $g_1''(x) = c_1\lambda_1 a(x)\check{u}_0^{\alpha}(x) - 2c_0$. Thus $g_1''(x) < 0$ for x near zero, and since $a(x)\check{u}_0^{\alpha}(x)$ is nondecreasing for $0 < x < \frac{1}{2}$, $g_1''(x)$ can change sign at most once on $(0,\frac{1}{2})$. Thus, in order to satisfy the boundary conditions $g_1(0) = 0$, $g_1'(0) > 0$, $g_1'(\frac{1}{2}) = 0$ we conclude $g_1(x)$ can change sign at most once on $(0,\frac{1}{2})$.

We now proceeed by induction. We assume $\check{u}_k(x)$ and $\check{u}_{k-1}(x)$ are symmetric about $x = \frac{1}{2}$, $\check{u}_k(x) - \check{u}_{k-1}(x)$ is positive and nondecreasing on $(0,\frac{1}{2})$, and for any $0 < c_k < c_{k-1}$, $g_k(x) = -c_k\check{u}_k(x) + c_{k-1}\check{u}_{k-1}(x)$ changes sign at most once on $(0,\frac{1}{2})$. We have by (4) that $\check{\lambda}_{k+1} < \check{\lambda}_k$. We now consider $g_{k+1}(x) = -c_{k+1}\check{u}_{k+1}(x) + c_k\check{u}_k(x)$, with $0 < c_{k+1} < c_k$.

$$g_{k+1}''(x) = a(x)(c_{k+1}\check{\lambda}_{k+1}\check{u}_k^{\alpha}(x) - c_k\check{\lambda}_k\check{u}_{k-1}^{\alpha}(x)). \tag{10}$$

Thus $g_{k+1}''(x) < 0$ for x near zero, and by the induction hypothesis, $g_{k+1}''(x)$ changes sign at most once on $(0,\frac{1}{2})$. Thus in order to satisfy the boundary conditions $g_{k+1}(0) = 0$, $g_{k+1}'(0) > 0$, $g_{k+1}'(\frac{1}{2}) = 0$, we conclude $g_{k+1}(x)$ can change sign at most once on $(0,\frac{1}{2})$. Now

$$\check{u}_{k+1}'' - \check{u}_k''(x) = -\lambda_{k+1} a(x)\check{u}_{k+1}^{\alpha}(x) + \check{\lambda}_k a(x)\check{u}_{k+1}^{\alpha}(x). \tag{11}$$

The function $f(x) = \check{u}_{k+1}(x) - \check{u}_k(x)$ is seen to satisfy the hypotheses of Lemma 1; that $f''(x)$ changes sign at most once on $(0,\frac{1}{2})$ is a consequence of the fact that $g_{k+1}(x)$ changes sign at most once on $(0,\frac{1}{2})$. Thus by Lemma 1, we have the desired result that $\check{u}_{k+1}(x) > \check{u}_k(x)$, $0 < x < 1$, and from (4), $\check{\lambda}_{k+2} < \check{\lambda}_{k+1}$.

We observe that since $\check{u}_n''(x) < 0$, $0 < x < 1$, $\check{u}_n'(0) = 1$, and $\check{u}_n(x)$ is symmetric about $x = \frac{1}{2}$, the sequence $\{\check{u}_n(x)\}$ is bounded above by $T(x)$. By considering the integral equation

$$\check{u}_{n+1}(x) = \check{\lambda}_{n+1}\left\{(1-x)\int_0^x \xi a(\xi)\check{u}_n^\alpha(\xi)\,d\xi + x \times \int_x^1 (1-\xi)a(\xi)\check{u}_n^\alpha(\xi)\,d\xi\right\} \quad (12)$$

equivalent to (4), it follows from equicontinuity and dominated convergence theorems that the sequence $\{\check{u}_n(x)\}$ converges uniformly to a solution $\check{u}(x)$ of (1) with eigenvalue $\check{\lambda} = \lim_{n\to\infty}\check{\lambda}_n$.

Theorem 4: The sequence $\{\hat{u}_k(x)\}$ is monotone decreasing for each x, and $\{\hat{\lambda}_k\}$ is monotone increasing. Moreover, there is a positive solution pair $\{\hat{\lambda}, \hat{u}(x)\}$ of (1) such that $\hat{u}_k(x)\to\hat{u}(x)$ uniformly on [0,1] and $\hat{\lambda}_k\to\hat{\lambda}$.

Proof: Since $\hat{u}_1(0) = \hat{u}_1(1) = 0$, $\hat{u}_1'(0) = -\hat{u}_1'(1) = 1$, $\hat{u}_1''(x) < 0$, $0<x<1$, and $\hat{u}_1(x)$ is symmetric about $x=\frac{1}{2}$, it follows that $\hat{u}_1(x) < x$, $0 < x \leqslant 1/2$, and $\hat{u}_1(x) < 1-x$, $\frac{1}{2}\leqslant x<1$; that is, $\hat{u}_1(x) < \hat{u}_0(x)$, $0<x<1$, and by (4) we conclude $\hat{\lambda}_1 < \hat{\lambda}_2$. An induction argument similar to that in the proof of Theorem 3 can be used to show $\hat{u}_{n+1}(x) < \hat{u}_n(x)$, $0<x<1$, and $\hat{\lambda}_{n+1} < \hat{\lambda}_{n+2}$. Since $\{\hat{u}_n(x)\}$ is a decreasing sequence which is bounded below (see Ref. 7), it has a limit $\hat{u}(x)$. By the arguments in Ref. 7, it follows that $\hat{u}(x) > 0$, $0<x<1$, and there is a $\hat{\lambda} > 0$ such that $\hat{\lambda}_n\to\hat{\lambda}$. As in the proof of Theorem 3, we conclude $\hat{u}_n(x)\to\hat{u}(x)$ uniformly on [0,1] and $\{\hat{\lambda},\hat{u}(x)\}$ is a solution pair of (1).

We note that by using the symmetry of $a(x)$ and of $\check{u}_n(x)$ or $\hat{u}_n(x)$ along with the integral equation

$$u_{n+1}(x) = \lambda_{n+1}\left\{(1-x)\int_0^x \xi a(\xi)u_n^\alpha(\xi)\,d\xi + x \times \int_x^1 (1-\xi)a(\xi)u_n^\alpha(\xi)\,d\xi\right\}, \quad (13)$$

which is valid for both $\{\hat{\lambda}_k,\hat{u}_k(x)\}$ and $\{\check{\lambda}_k,\check{u}_k(x)\}$, that $u_n'(x) > 0$, $0<x<\frac{1}{2}$ and $u_n'(x)<0$, $\frac{1}{2}<x<1$. Thus $u'(x)>0$, $0<x<\frac{1}{2}$ and $u'(x)<0$, $\frac{1}{2}<x<1$, where u_n, u can be $\check{u}_n$, $\check{u}$, or $\hat{u}_n$, $\hat{u}$.

What we have shown can so far can be summarized:

starting iterate	convergence to a solution	normalized by $u'(0) = 1$
$\check{u}_0(x) = x(1-x)$	$\{\check{u}_n\}$ monotone increasing,	$\{\check{\lambda}_n\}$ monotone decreasing
$\hat{u}_0(x) = T(x)$	$\{\hat{u}_n\}$ monotone decreasing,	$\{\hat{\lambda}_n\}$ monotone increasing

Bounds on $\{S_n(x)\}$: Using the fact that $\check{u}_n''(x) < 0$ along with the symmetry and boundary conditions, we have

$$x(1-x) = \check{u}_0(x) < \check{u}_n(x) < T(x), \quad 0<x<1. \quad (14)$$

In fact these same bounds also apply to $\hat{u}_n(x)$. We already have $\hat{u}_n(x) < T(x)$, $0<x<1$, so we only need to show $\hat{u}_n(x) > x(1-x)$, $0<x<1$. If we let $f(x) = \hat{u}_n(x) - x(1-x)$, we have found $f(0) = 0$, $f'(0) = 0$, $f(x)$ is symmetric about $x=\frac{1}{2}$, and $f''(x) = 2 - \hat{\lambda}_n a(x)\hat{u}_{n-1}^\alpha(x)$. Since $\hat{u}_{n-1}(0) = 0$ and $a(x)\hat{u}_n^\alpha(x)$ is nondecreasing on $(0,\frac{1}{2})$, we conclude that $f''(x) > 0$ in an interval $(0,\epsilon)$ and $f''(x)$ changes sign at most once in $(0,\frac{1}{2})$. Thus by Lemma 1 and symmetry we conclude $\hat{u}_n(x) > x(1-x)$, $0<x<1$. Thus, for either $\hat{u}_n(x)$ or $\check{u}_n(x)$, we have

$$\tfrac{1}{4} < u_n(\tfrac{1}{2}) = \max_{0<x<1} u_n(x) < \tfrac{1}{2} \quad (15)$$

We now consider Berryman's sequences $\{\hat{S}_n(x)\}$ and $\{\check{S}_n(x)\}$ which were shown in Eq. (8) to be given by

$$\check{S}_n(x) = \check{u}_n(x)/\max_{0<x<1}\check{u}_n(x) = \check{u}_n(x)/\check{u}_n(\tfrac{1}{2}),$$
$$\hat{S}_n(x) = \hat{u}_n(x)/\max_{0<x<1}\hat{u}_n(x) = \hat{u}_n(x)/\hat{u}_n(\tfrac{1}{2}). \quad (16)$$

By combining (15) and (16) we have the bounds on $y(x)$ stated in inequality (3). For either $\check{S}_n$ or $\hat{S}_n$ we claim

$$2T(x)\leqslant S_n(x)\leqslant 4x(1-x). \quad (17)$$

The lower bound is easily established since $S_N(x)$ is a concave down function for which $S_n(0) = S_n(1) = 0$ and for which $\max_{0<x<1} S_n(x) = S_n(\frac{1}{2}) = 1$. Thus $S_n(x)$ must be above the triangle function $2T(x)$. For the upper bound we consider

$$f(x) = S_n(x) - 4x(1-x) = u_n(x)/u_n(\tfrac{1}{2}) - 4x + 4x^2,$$

and by using Eq. (12) we have

$$f''(x) = u_n''(x)/u_n(\tfrac{1}{2}) + 8 =$$

$$= 8 - a(x)u_{n-1}^\alpha(x)\Big/\int_0^{\frac{1}{2}}\xi a(\xi)u_{n-1}^\alpha(\xi)\,d\xi. \quad (18)$$

Since $a(x)$ and $u_{n-1}(x)$ are increasing on $(0,\frac{1}{2})$, we conclude $f''(x) > 0$ on some interval $(0,\xi)$ and $f''(x)$ changes sign at most once on $(0,\frac{1}{2})$. Thus in order to satisfy the boundary conditions $f(0) = 0$, $f'(0) < 0$, $f(\frac{1}{2}) = 0$, we must have $f(x) < 0$, $0<x<\frac{1}{2}$. By also considering symmetry, we conclude $S_n(x)\leqslant 4x(1-x)$, $0<x<1$.

Thus we have

$$4G(x,\tfrac{1}{2}) = 2T(x)\leqslant S_n(x)\leqslant 4x(1-x), \quad 0\leqslant x\leqslant 1, \quad (19)$$

where

$$G(x,\xi) = \begin{matrix} x(1-\xi) & x<\xi \\ \xi(1-x) & x>\xi' \end{matrix}$$

proving the graphical result of Berryman.[1]

We now show the monotonicity of the sequence $\{S_n\}$ for either $\check{S}_n$ or $\hat{S}_n$.

Lemma 2: $\check{u}_1(x)/\check{u}_0(x)$ is an increasing function on $(0,\frac{1}{2})$, $\hat{u}_1(x)/\hat{u}_0(x)$ is a decreasing function on $(0,\frac{1}{2})$.

Proof: Using the integral representation of $\hat{u}_1(x)$ from Eq. (13) we have for $0<x<\frac{1}{2}$,

$$\frac{\hat{u}_1(x)}{\hat{u}_0(x)} = \hat{\lambda}_1\left[\frac{1}{x}\int_0^x a(\xi)\xi^{\alpha+1}\,d\xi - \int_0^x a(\xi)\xi^{\alpha+1}\,d\xi + \int_x^{\frac{1}{2}}(1-\xi)a(\xi)\xi^\alpha\,d\xi + \int_{\frac{1}{2}}^1 a(\xi)(1-\xi)^{\alpha+1}\,d\xi\right], \quad (20)$$

and thus

$$\left(\frac{\hat{u}_1(x)}{\hat{u}_0(x)}\right)' = -\frac{\hat{\lambda}_1}{x^2}\int_0^x a(\xi)\xi^{\alpha+1}\,d\xi < 0, \quad 0<x<\tfrac{1}{2}. \quad (21)$$

Similarly,

$$\frac{\check{u}_1(x)}{\check{u}_0(x)} = \check{\lambda}_1\Big[\frac{1}{x}\int_0^x \xi^{\alpha+1}(1-\xi)^\alpha a(\xi)\,d\xi,$$
$$+\frac{1}{(1-x)^2}\int_x^1 (1-\xi)^{\alpha+1}\xi^\alpha a(\xi)\,d\xi\Big], \qquad (22)$$

$$\Big(\frac{\check{u}_1(x)}{\check{u}_0(x)}\Big)' = \check{\lambda}_1\Big[-\frac{1}{x^2}\int_0^x \xi^{\alpha+1}(1-\xi)^\alpha a(\xi)\,d\xi$$
$$+\frac{1}{(1-x)^2}\int_x^1 (1-\xi)^{\alpha+1}\xi^\alpha a(\xi)\,d\xi\Big].$$

By using the symmetry of $a(x)$ and $\check{u}_0(x)$, we have

$$\Big(\frac{\check{u}_1(x)}{\check{u}_0(x)}\Big)' = \check{\lambda}_1\Big[\frac{1}{(1-x)^2}\int_x^{1-x} \xi^{\alpha+1}(1-\xi)^\alpha a(\xi)\,d\xi$$
$$-\frac{1-2x}{x^2(1-x)^2}\int_0^x \xi^{\alpha+1}(1-\xi)^\alpha a(\xi)\,d\xi\Big]. (23)$$

Thus $\check{u}_1(x)/\check{u}_0(x)$ is increasing on $0<x<\frac{1}{2}$ if and only if

$$x^2\int_x^{1-x} \xi^{\alpha+1}(1-\xi)^\alpha a(\xi)\,d\xi$$
$$>(1-2x)\int_0^x \xi^{\alpha+1}(1-\xi)^\alpha a(\xi)\,d\xi,\ 0<x<\tfrac{1}{2}. \qquad (24)$$

Since $a(x)$ and $u_0(x)=x(1-x)$ are increasing on $(0,\frac{1}{2})$ and symmetric about $x=\frac{1}{2}$, we have for $0<x<\frac{1}{2}$,

$$0<a(x)x^\alpha(1-x)^\alpha<a(\xi)\xi^\alpha(1-\xi)^\alpha,\ x<\xi<1-x,$$
$$0<a(\xi)\xi^\alpha(1-\xi)^\alpha<a(x)x^\alpha(1-x)^\alpha,\ 0<\xi<x. \qquad (25)$$

Thus

$$x^2\int_x^{1-x} \xi^{\alpha+1}(1-\xi)^\alpha a(\xi)\,d\xi$$
$$>a(x)x^{\alpha+2}(1-x)^\alpha\int_x^{1-x}\xi\,d\xi$$
$$=\frac{a(x)x^{\alpha+2}(1-x)^\alpha(1-2x)}{2}$$

and

$$(1-2x)\int_0^x \xi^{\alpha+1}(1-\xi)^\alpha a(\xi)\,d\xi$$
$$<(1-2x)a(x)x^\alpha(1-x)^\alpha\int_0^x\xi\,d\xi$$
$$=\frac{a(x)x^{\alpha+2}(1-x)^\alpha(1-2x)}{2}$$

from which the desired result now follows.

Lemma 3: $\check{u}_{n+1}(x)/\check{u}_n(x)$ is an increasing function on $(0,\frac{1}{2})$, $\hat{u}_{n+1}(x)/\hat{u}_n(x)$ is a decreasing function on $(0,\frac{1}{2})$.

Proof: The proof is by induction. By Lemma 2, the conclusion is true for $\check{u}_1/\check{u}_0$ and $\hat{u}_1/\hat{u}_0$. We assume the result to be true for $\check{u}_n/\check{u}_{n-1}$ and $\hat{u}_n/\hat{u}_{n-1}$. We have

$$(\check{u}_{n+1}(x)/\check{u}_n(x))'$$
$$=[\check{u}_n(x)\check{u}'_{n+1}(x)-\check{u}'_n(x)\check{u}_{n+1}(x)]/\check{u}_n^2(x).$$

Thus $\check{u}_{n+1}(x)/\check{u}_n(x)$ is increasing on $(0,\frac{1}{2})$ if and only if $\check{u}_n(x)\check{u}'_{n+1}(x)>\check{u}'_n(x)\check{u}_{n+1}(x)$, $0<x<\frac{1}{2}$. Since $\check{u}_n(x)$, $\check{u}'_n(x)$, $\check{u}_{n+1}(x)$, $\check{u}'_{n+1}(x)$ are positive on $(0,\frac{1}{2})$, this is equivalent to

$$\check{u}'_{n+1}(x)/\check{u}'_n(x)>\check{u}_{n+1}(x)/\check{u}_n(x),\ 0<x<\tfrac{1}{2}. \qquad (26)$$

Using the integral representation of $\check{u}_k(x)$ from Eq. (13) along with the symmetry of $a(x)$ and $\check{u}_{k-1}(x)$, we have

$$\check{u}_k(x)=\check{\lambda}_k\Big[\int_0^x\xi a(\xi)\check{u}^\alpha_{k-1}(\xi)\,d\xi+x\times\int_x^{1-x}\xi a(\xi)\check{u}^\alpha_{k-1}(\xi)\,d\xi\Big],$$
$$\check{u}'_k(x)=\check{\lambda}_k\int_x^{1-x}\xi a(\xi)\check{u}^\alpha_{k-1}(\xi)\,d\xi. \qquad (27)$$

Thus from (26) we conclude $\check{u}_{n+1}(x)/\check{u}_n(x)$ is increasing on $(0,\frac{1}{2})$ if and only if for $0<x<\frac{1}{2}$,

$$\frac{\int_x^{1-x}\xi a(\xi)\check{u}^\alpha_n(\xi)\,d\xi}{\int_x^{1-x}\xi a(\xi)\check{u}^\alpha_{n-1}(\xi)\,d\xi}>\frac{\int_0^x\xi a(\xi)\check{u}^\alpha_n(\xi)\,d\xi+x\int_x^{1-x}\xi a(\xi)\check{u}^\alpha_n(\xi)\,d\xi}{\int_0^x\xi a(\xi)\check{u}^\alpha_{n-1}(\xi)\,d\xi+x\int_x^{1-x}\xi a(\xi)\check{u}^\alpha_{n-1}(\xi)\,d\xi}. \qquad (28)$$

Equation (28) is true if and only if for $0<x<\frac{1}{2}$,

$$\frac{\int_0^x\xi a(\xi)\check{u}^\alpha_n(\xi)\,d\xi}{\int_0^x\xi a(\xi)\check{u}^\alpha_{n-1}(\xi)\,d\xi}<\frac{\int_x^{1-x}\xi a(\xi)\check{u}^\alpha_n(\xi)\,d\xi}{\int_x^{1-x}\xi a(\xi)\check{u}^\alpha_{n-1}(\xi)\,d\xi}. \qquad (29)$$

By the mean value theorem there exists ξ_1, $0<\xi_1<x$, such that

$$\frac{\int_0^x\xi a(\xi)\check{u}^\alpha_n(\xi)\,d\xi}{\int_0^x\xi a(\xi)\check{u}^\alpha_{n-1}(\xi)\,d\xi}=\frac{\int_0^x\xi a(\xi)\check{u}^\alpha_{n-1}(\xi)[\check{u}^\alpha_n(\xi)/\check{u}^\alpha_{n-1}(\xi)]\,d\xi}{\int_0^x\xi a(\xi)\check{u}^\alpha_{n-1}(\xi)\,d\xi}=\frac{\check{u}^\alpha_n(\xi_1)}{\check{u}^\alpha_{n-1}(\xi_1)}.$$

TABLE I. Summary of results.

Conditions on $a(x)$	$C[0,1]\not\equiv 0, \geqslant 0$ on $[0,1]$	$C[0,1]\geqslant 0, \not\equiv 0$, *symmetric about* $x=\frac{1}{2}$, nondecreasing on $[0,\frac{1}{2}]$	
Range of α	$\alpha>-1$	$\alpha>0$	
$S_0(x)$	x	$4x(1-x)$	$4G(x,\frac{1}{2})$
Monotonicity	not necessarily	yes	yes
		$S_{n+1}(x)<S_n(x)$ $\lambda_{n+1}>\lambda_n$	$S_{n+1}(x)>S_n(x)$ $\lambda_{n+1}<\lambda_n$
Error estimate	No unless alternating montonicity occurs	Yes, if solution is unique; *a posteriori* by comparing the increasing and decreasing iterative sequences at each stage.	

Similarly, by the mean value theorem and symmetry there exists ξ_2, $x < \xi_2 < \frac{1}{2}$ such that

$$\frac{\int_x^{1-x} \xi a(\xi) \check{u}_n^\alpha(\xi)\, d\xi}{\int_x^{1-x} \xi a(\xi) \check{u}_{n-1}^\alpha(\xi)\, d\xi} = \frac{\check{u}_n^\alpha(\xi_2)}{\check{u}_{n-1}^\alpha(\xi_2)}.$$

By the induction hypothesis $\check{u}_n(x)/\check{u}_{n-1}(x)$ is increasing on $(0, \frac{1}{2})$, and since $0 < \xi_1 < x < \xi_2 \leqslant \frac{1}{2}$, we conclude Eq. (29) is true and thus $\check{u}_{n+1}(x)/\check{u}_n(x)$ is increasing on $(0, \frac{1}{2})$. Similarly, we can conclude $\hat{u}_{n+1}(x)/\hat{u}_n(x)$ is decreasing on $(0, \frac{1}{2})$ by reversing the inequalities in the above argument.

Theorem 5: $\check{S}_{n+1}(x) < \check{S}_n(x)$ and $\hat{S}_{n+1}(x) > \hat{S}_n(x)$ for $0 < x < 1$.

Proof: By Lemma 3 $\check{u}_{n+1}(x)/\check{u}_n(x) < \check{u}_{n+1}(\frac{1}{2})/\check{u}_n(\frac{1}{2})$ and $\hat{u}_{n+1}(x)/\hat{u}_n(x) > \hat{u}_{n+1}(\frac{1}{2})/\hat{u}_n(\frac{1}{2})$, $0 < x < \frac{1}{2}$, from which along with symmetry, we conclude $\check{S}_{n+1}(x) < \check{S}_n(x)$ and $\hat{S}_{n+1}(x) > \hat{S}_n(x)$, $0 < x < 1$.

IV. SUMMARY

In Table I "solution" means normalized by having maximum on $(0,1)$ equal to one.

We note that if $-1 < \alpha < 0$ our sequence $\{u_n(x)\}$ generated by $u_0(x) = x$ is alternating monotone, converging to the unique[4] solution $u(x)$. Thus error estimates are available *a posteriori*. Then $S(x)$ can be calculated.

These results have not been established for the case in which $a(x)$ has an integrable singularity at an interior point $(0,1)$.

[1] J. G. Berryman, J. Math. Phys. **18**, 2108 (1977).

[2] J. G. Berryman and C. J. Holland, J. Math. Phys. **19**, 247 (1978).

[3] J. G. Berryman, J. Math. Phys. **21**, 1326 (1980).

[4] S. Taliaferro, Nonlinear Anal. **3**, 897 (1979). The case $-1 < \alpha < 0$ is covered here.

[5] J. S. W. Wong, SIAM Rev. **17**, 339 (1975). For example, if $a(x)$ is nondecreasing and continuous, $\alpha > 1$, then the positive solution is unique. For $0 < \alpha < 1$ (sublinear case) references are given: uniqueness is generally the case for $0 < \alpha < 1$.

[6] C. D. Luning and W. L. Perry, SIAM J. Math Anal., to appear.

[7] C. D. Luning and W. L. Perry, J. Differential Equations **33**, 359 (1979).

A direct study of a Marchenko fundamental equation with centripetal potential

Marcel Coz
University of Kentucky, Lexington, Kentucky 40506

(Received 20 May 1980; accepted for publication 20 August 1980)

A direct study of a class of singular ($\ell \neq 0$) fundamental equations is shown to be possible. The method used for this proof follows Marchenko's nonsingular ($\ell = 0$) approach, step by step. Throughout the paper the interest of a simultaneous study of the "Marchenko associated equation" is stressed. Also it will be shown why the Marchenko approach does not extend to complex interactions. In order to treat a complex interaction one must therefore resort to a new set of ideas.

PACS numbers: 02.30.Hq, 03.65.Ge

1. INTRODUCTION

As is well known, Marchenko solved[1] an inverse problem where the spectral elements were replaced by scattering data. Its solution was therefore welcomed in physicist circles. However it is less well known that Marchenko's method is a two step method and consequently may be difficult to apply. Indeed Marchenko begins by solving the inverse problem with no centripetal forces ($\ell = 0$; no singularity) and in a second step treats the general case with singularities ($\ell \neq 0$; centripetal forces present) via the method of special transformations, a method related to Crum's transformations.[2] The artifice has enough generality and is very ingenious since it permits the following statement, which we state as Faddeev does[3]: If $S(k)$ is the S-function for the $L^{(0)}$ operator with potential $q^{(0)}(x)$, then it is also the S-function for the $L^{(m)}$ operator (for any $m = 0,1,2,\cdots$) where the potential $q^{(m)}(x)$ behaves like $q^{(0)}(x)$ as $x \to 0$ and $x \to \infty$. Faddeev's statement concerns scalar cases but, as Marchenko shows, a similar statement is valid for matrix cases. In Faddeev's statement the $L^{(i)}$ operator is defined as follows:

$$L^{(i)}(x) = d^2/dx^2 - \ell_i(\ell_i + 1)/x^2 - q^{(i)}(x).$$

In spite of its appealing generality the method of special transformations is complicated. From our own studies[4] its generality means its validity even in cases where an integrable translation kernel does not exist: in fact Marchenko is careful to tell his reader that special transformations cannot, in general, be expressed in an integral transformation form. The price to pay for its generality is, consequently, its complexity. By assuming as a basis for our studies the existence of a translation kernel, we already have sacrificed upon the altar of generality. However with some additional assumptions, we were able to prove the existence of a fundamental equation for a singular ($\ell \neq 0$) system which in addition was not necessarily Hermitian.[5] The proof was obtained following exactly the steps of Marchenko in Ref. 1. We show, in the present paper, that the ($\ell \neq 0$) fundamental equation related to a Hermitian operator can be solved directly. Of course, Marchenko's generality is lost in the process: only potentials with high ($2\ell + 1$) moments can be treated this way. Some extensions exist, but their nature seems limited. To obtain the Marchenko generality one may be obliged to use the special transformation method. Nevertheless we have provided a very simple and direct treatment for an important class of $\ell \neq 0$ singular interactions. We also make clear that the existence of a solution for the $\ell \neq 0$ inverse problem is related to the existence of a solution for a "Marchenko ($\ell = 0$) associated equation".

While writing Ref. 5 where non-Hermitian potentials were considered, we became convinced that Hermitian and non-Hermitian potentials need to be distinguished not only when one discusses their relative spectra but also when one discusses their fundamental equations. By following Ref. 1 step by step we analyze which arguments that work for Hermitian systems do not apply to non-Hermitian systems. It will be clear that arguments valid in the Hermitian case are no longer valid when a discrete spectrum is present. There is no reason to discuss the spectral singularities (k real $\neq 0$) since they are absent in Hermitian systems.

2. NOTATIONS

Consideration of analytical properties of scalar systems and of matrix systems rests upon the same principles. Differences essentially occur when the analysis or the identification of the scattering data is attempted.

Since this paper is concerned with the discussion of the analytical properties of the spectral kernel, we deal explicitly with scalar systems. The extension to matrices will not, however, introduce any essential difficulty. To avoid any misconception we use the terms interaction, or potential, indifferently.

The concern of the paper is the two term Schrödinger equation

$$[-(\hbar^2/2m)\nabla^2 + V(r)]\psi(r) = E\psi(r), \qquad (1)$$

with a spherically symmetric real interaction $V(r)$. Equation (1) is separated into a set of partial wave equations corresponding to each angular momentum and we study one of these partial wave differential equations

$$[-(d^2/dx^2 - \ell(\ell + 1)/x^2) - k^2 + U(x)]u = 0. \qquad (2)$$

A fundamental assumption is made at this point: there exists a Marchenko representation. In other words there exists a bounded and continuous function $K(x,y)$ such that solutions $f_1(k,x), f_2(k,x)$ defined by their behavior at infinity exist, and have the integral representation (4) which follows:

 © 1981 American Institute of Physics

$$f_1(k,x) \simeq h_1(kx) \simeq \exp[i(kx - \ell\pi/2)], \tag{3a}$$

$$f_2(k,x) \simeq h_2(kx) \simeq \exp[-i(kx - \ell\pi/2)]. \tag{3b}$$

In (3a) and (3b) we have introduced the Riccati–Hankel functions $h_1(kx)$ and $h_2(kx)$ with their asymptotic behavior. In the half-plane, where they are defined, one has

$$f_i(k,x) = h_i(kx) + \int_x^\infty K(x,y) h_i(ky)\, dy, \tag{4}$$

$i = 1$ and 2.

The existence of the representation (4) imposes the condition

$$\int_x^\infty s^\alpha |U(s)|\, ds < \infty \quad \text{for } \alpha = 0,1. \tag{5a}$$

If one adds the requirment that $K(x,y)$ be absolutely integrable, one needs also

$$\int_x^\infty s^\alpha |U(s)|\, ds < \infty \quad \text{for } \alpha = 0,1,\ldots,\ell+1. \tag{5b}$$

When (5a) is satisfied by the potential $U(s)$, the kernel $K(x,y)$ of (4) called hereafter "translation kernel" is *unique*. It obeys the upper bound

$$|K(x,y)| \leqslant \tfrac{1}{2}\left(\frac{y}{x}\right)^\ell \sigma_0\left(\frac{x+y}{2}\right)\exp[\sigma_1(x)] \tag{6a}$$

when only (5a) is verified. The upper bound reads

$$|K(x,y)| \leqslant \tfrac{1}{2}\left(\frac{1}{x}\right)^\ell \sigma_0\left(\frac{x+y}{2}\right)\exp[\sigma_1(x)] \tag{6b}$$

with (5b). In (6a) and (6b) the σ_i are the absolute moments of the interaction $U(x)$ defined as follows:

$$\sigma_i = \int_x^\infty s^i |U(s)|\, ds.$$

A condition similar to (5a) and (5b) is also used; it is denoted (5c)

$$\int_0^\infty s^\alpha |U(s)|\, ds < \infty; \quad \alpha = 1,2. \tag{5c}$$

When the potential $U(x)$ is real, the condition (5c) guarantees that the set of the normalizable states is finite and located in $\mathrm{Im} k \geqslant 0$. No spectral singularities can be present for real $k \neq 0$. The normalizable states happen for $k = ik_n$ (k_n real), when the Jost function $f_1(k)$ vanishes

$$f_1(k) = \lim_{r\to 0}\left[(kr)^\ell \frac{1}{(2\ell-1)!!} f_1(k,r)\right].$$

The vanishing of $f_1(k)$ for $k=0$ may mean a normalizable or a scattering state: a special discussion is therefore needed if $f_1(0)$ vanishes. With (5b) and conditions guaranteeing that the set of the zeros of $f_1(k)$ in $\mathrm{Im} k \geqslant 0$ remains finite, and that $\lim_{k=0}[1 - S(k)] = Ak^{2\ell}$ one proves[5] the existence of a fundamental equation valid for all $(x, y; y \geqslant x)$:

$$K(x,y) + F(x,y) + \int_x^\infty K(x,z) F(z,y)\, dz = 0. \tag{7}$$

The function $F(x,y)$ of Eq. (7) will be called "spectral kernel" for the reason that it is constructed from the set of scattering data which in Marchenko's theory replace the spectral data. The spectral kernel can be divided into two parts, a continuum and a discrete part denoted respectively $F_C(x,y)$ and $F_D(x,y)$.

To obtain Eq. (7) two identities defining the "regular" solution $G(k,x)$

$$-2iG(k,x)k^{\ell+1} f_1(k)^{-1} = f_2(k,x) - S(k) f_1(k,x); \quad k \text{ real} \neq 0,$$

$$-2iG(k,x)k^{\ell+1} f_1(k)^{-1} = \tilde{f}_2(k,x) - \tilde{S}(k) f_1(k,x); \quad k \geqslant 0, |k| \neq 0,$$

are used. Notations were the following:

$$S(k) = f_2(k)/f_1(k),$$

$$\tilde{S}(k) = \tilde{f}_2(k)/f_1(k),$$

with

$$\tilde{f}_2(k) = \lim_{x\to 0} (kx)^\ell \tilde{f}_2(k,x) \frac{1}{(2\ell-1)!!}$$

and

$$\tilde{f}_2(k,x) = h_2(kx) + \frac{1}{2ik} \times \int_a^x h_1(kx) h_2(kt) V(t) \tilde{f}_2(k,t)\, dt + \frac{1}{2ik}\int_x^\infty h_2(kx) h_1(kt) V(t) \tilde{f}_2(k,t)\, dt.$$

The solution $\tilde{f}_2(k,x)$ where the constant a is chosen so as to make the series solution absolutely convergent has been introduced for $\ell = 0$ by Naimark.[6]

An integration from $-\infty$ to $+\infty$ of the r.h.s. of the first identity with a contour integration in the complex upper half plane of the r.h.s. of the second identity are used. The Marchenko equation results if $\lim_{k\to 0} T(k) \equiv [1 - S(k)] = Ak^{2\ell}$. Then one has

$$F_C(x,y) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} h_1(kx)[1 - S(k)] h_1(ky)\, dk. \tag{8a}$$

If $f_1(k)^{-1}$ has only poles of finite order, these poles furnish the definition of $F_D(x,y)$ through their residues.

To be more specific, we use the asymptotic behavior of the physical solution $\psi(k,x)$

$$\psi(k,x) \simeq h_2(kx) + S(k) h_1(kx);$$

$S(k)$ and $T(k) = 1 - S(k)$ are respectively the S and T matrix obtained from Eq. (2).

If the potential is real, the normalizable states are "simple"; they occur for $f_1(ik_n) = 0$, $k = ik_n$ (k_n real); in addition their number p is finite. If, $\lim_{k\to 0} T(k) = Ak^{2\ell}$, we have

$$F_D(x,y) = \sum_{n=1}^{p} h_1(ik_n x) M_n^2 h_1(ik_n y), \tag{8b}$$

with

$$M_n^2 = \int_0^\infty dx\, f_1(ik_n,x) f_1(ik_n,x).$$

The constant M_n^2 is related to the residue of $\tilde{S}(k)$, to be precise:

$$M_n^2 = -(-1)^\ell 2\pi i \tilde{f}_2(ik_n)/f_1'(ik_n).$$

These M_n's defined in (8b) are normalization constants and are therefore positive. In the matrix case not considered here, they are Hermitian matrices associated with semipositive definite quadratic forms. An equation similar to Eq. (7) where F is replaced by M,

$$M(x,y) = m_c(x,y) + M_D(x,y),$$

$$M_C(x,y)=\frac{1}{2\pi}\int_{-\infty}^{\infty}[1-S(k)]\exp ik(x+y)\,dk,$$

$$M_D(x,y)=\sum_{n=1}^{p} M_n^2\exp(-k_m(x+y)),$$

is introduced. This equation is called the Marchenko associated equation.

We assume that $f_1(0)\neq 0$. If the potential satisfies (sufficient conditions):

$$\int_0^\infty |U(s)|s^{2\ell+2}\,dr<\infty;\quad U(s)=O\left(\frac{1}{s^{2l+3+\epsilon}}\right),$$

$T(k)$ is $O(k^{2\ell+1})$ and (8a) is the usual integral. The Landau symbol O just used means that the ratio $U(s)/s^{-2\ell-3-\epsilon}$ is bounded when s approaches infinity and that $T(k)/k^{2\ell+1}$ is bounded when k goes to zero. However if such a high moment of the interactions does not exist and if we have only

$$\int_0^1 |U(s)|s^{2\ell+2}\,dr<\infty \text{ and } U(s)=O(s^{-\alpha-1-\epsilon}),$$

n integer, $n<2\ell+2$, one may have simply $T(k)=O(k^{n-1})$. For instance for $n=2\ell+1$, $T(k)=O(k^{2\ell})$ then (8a) is still the usual integral. But for $\alpha=2l$, $T(k)$ may be only $O(k^{2\ell-1})$ then a pole of order 1 intervenes. The integral (8a) has to be subjected to a regularization process; it becomes a principal value integral.

Extensions of the domain of validity of Eq. (7) require some kind of regularization. The following mode of regularization will be considered in this paper. Some $\epsilon>0$ being chosen, the r.h.s. of the first equation defining the regular solution $G(k,x)$ is integrated on the segments $[-R,-\epsilon][\epsilon,R]$. To complete the integration of radius ϵ, located in the half plane $\operatorname{Im}k\geqslant 0$. Limits are then taken with R going to infinity and ϵ going to zero. Since the integral along the semicircle vanishes, we are left with

$$\lim_{\substack{R\to\infty\\ \epsilon\to 0}}\left[\int_{-R}^{-\epsilon}+\int_{\epsilon}^{R}\right].$$

In the case of a pole of order one, we have a principal value integral.

If $T(k)$ in the vicinity of $k=0$ has an expansion

$$T(k)=\sum_{n=\alpha}A_nk^n,\quad T(k)=O(k^\alpha),$$

which contains only odd powers of k for $n<2\ell$, the method just described applies. If $f_1(0)=0$ the regularization method fails; then, for $\ell\neq 0$, the semicircle surrounds a pole of order two, bringing a limit which is infinite. We keep, therefore, throughout the paper the assumption $f_1(0)\neq 0$.

3. THE FUNDAMENTAL EQUATION

The fundamental equation (7) can be regarded as an equation for the spectral kernel when the translation kernel is given: as such it is a Volterra equation which can be solved by iterations. With the definitions

$$\eta(x)=\tfrac{1}{2}\sigma_0(x)\exp[\sigma_1(x)], \tag{9}$$

$$\eta(x)=\int_x^\infty |Q(x)|\,dx=\tfrac{1}{2}\eta_0(x), \tag{10}$$

$$\int_x^\infty x|Q(x)|\,dx=\eta_1(x), \tag{11}$$

one has

$$|F(x,y)|\leqslant\tfrac{1}{2}\left(\frac{y}{x}\right)^{\ell}\eta_0(x)\exp[\eta_1(x)]. \tag{12}$$

The inequality (12) guarantees the continuity of $F(x,y)$ for all $y\leqslant x$ and that of $F(x,x)$ for all x including $x=0$.

It is tempting at this point to try to solve Eq. (7), when $F(x,y)$ is given, by the method of successive approximations. The translation kernel is then the solution one wants to be obtained.

In this perspective the fundamental equation becomes a Fredholm equation. The same method as before gives

$$|K(x,y)|\leqslant\tfrac{1}{2}\eta_0(x)\exp[\eta_1(x)]\times\sum_{n=1}^{\infty}\left[\int_x^\infty \tfrac{1}{2}\eta_0(z)\exp[\eta_1(z)]\,dz\right]^n; \tag{13}$$

Eq. (13) does not sum up into an exponential formula but rather into a $1/(1-x)$ form. Trouble will arise if there exists an eigenvalue ν, $|\nu|=1$; the method becomes unsuitable. However, a general result can be proved. If $F(x,y)$ is constructed from scattering data generated by a Schrödinger equation (2), where $U(x)$ satisfies (5a,b and c), then the fundamental equation has a unique solution [5]. Such data will be called "authentic".

In order for Eq. (7) to have a unique solution the homogeneous equation

$$K(x,y)+\int_x^\infty K(x,z)F(z,y)\,dz=0 \tag{7'}$$

must have only the trivial solution

$$K(x,y)\equiv 0.$$

Section 4 discusses, therefore, the solutions of the homogeneous equation (7′) related to the fundamental equation (7).

4. DISCUSSION OF THE SOLUTIONS OF EQ. (7′)

A subspace $L'^2(\epsilon,\infty)$, $(\epsilon>0)$ which is dense in $L^2(\epsilon,\infty)$ is used to show that Eq. (7) has no nontrivial solution in $L^2(\epsilon,\infty)$. Afterwards one uses the result that any solution of (7′) in $L^1(\epsilon,\infty)$ belongs also to $L^2(\epsilon,\infty)$ to extend the previous statement for L^2 to L^1; the case where $\epsilon=0$ is considered separately. The study is therefore divided into three parts: (A) solutions in $L^2(\epsilon,\infty)$ with $\epsilon>0$, (B) solutions in $L^1(\epsilon,\infty)$ with $\epsilon>0$ and (C) solutions when ϵ is equal to zero.

A. Solutions in $L^2(\epsilon,\infty)$

Two differential operators D'_+ and D'_- are introduced. Riccati–Hankel functions $h_1^\ell(kx)$ are also used. The order of the operators and that of the Riccati–Hankel functions is indicated by the superscript ℓ

$$h_1^\ell(kx)\simeq\exp[i(kx-\ell\pi/2)],$$

$$D^\ell_+=x^\ell\frac{d}{dx}\frac{1}{x^\ell};\quad D^\ell_-=\frac{1}{x^\ell}\frac{d}{dx}x^\ell. \tag{14}$$

When the two differential operators D^ℓ_+ and D^ℓ_- are applied to the Riccati–Hankel functions $h_1^\ell(kx)$ they give

$$D^\ell_+h_1^{\ell-1}(kx)=-kh_1^\ell(kx), \tag{15}$$

$$D_-^{\ell} h_1^{\ell}(kx) = kh_1^{\ell-1}(kx). \tag{16}$$

We consider now the integral transformation

$$T_1[\phi] = \int_{-\infty}^{\infty} h_1^{\ell}(k\xi)\phi(\xi)\, d\xi, \tag{17}$$

when $\phi(\xi)$ is well behaved at infinity and at the origin with

$$\lim_{\xi\to\infty} \phi(\xi) = 0.$$

After one integration by parts we obtain

$$T_1[\phi] = \int_{-\infty}^{\infty} h_1^{\ell-1}(k\xi)\frac{1}{k} D_1^{\ell}[\phi(\xi)]\, d\xi. \tag{18}$$

Defining now

$$\mathscr{D}_- = D_-^1 \times D_-^2 \times \cdots \times D_-^{\ell}, \tag{19}$$

from the assumptions that one can perform ℓ integrations by parts and still have the limit at infinity equal to zero and no problem at the origin,

$$T_1[\phi] = \int_{-\infty}^{\infty} \exp(ik\xi)\frac{1}{(k)^{\ell}} \mathscr{D}_-[\phi(\xi)]\, d\xi. \tag{20}$$

We move to a second kind of transformation. In the case where the integrand in Eq. (8a) has no pole on the real axis, we have the usual integral running from $-\infty$ to $+\infty$

$$F_C(\xi,t) = \tfrac{1}{2}\int_{-\infty}^{\infty} h_1^{\ell}(k\xi)T(k)h_1^{\ell}(kt)\, dk,$$

$$T(k) = 1 - S(k).$$

We define:

$$T_2[\phi] = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\xi \int_{-\infty}^{\infty} dk\, \phi(\xi)h_1^{\ell}(k\xi)T(k)h_1^{\ell}(kt), \tag{21}$$

$$T_2[\phi] \equiv y(t).$$

The existence of the fundamental equation (7) was obtained from the existence of a bounded translation kernel. Let us assume

$$\lim_{k\to 0} T(k) = Ak^{2\ell}. \tag{22}$$

Condition (22) will be partially removed at the end of the conclusion. Sufficient conditions for (22), are

$$\int_0^1 s^{2\ell+2}|U(s)|\, ds < \infty$$

and (23)

$$\int_1^{\infty} s^{2l+1}|U(s)|\, ds < \infty.$$

Let us assume that $\phi(\xi)$ and its ℓ first derivatives vanish when ξ approaches infinity and are bounded at $\xi = 0$. From Eq. (21) one obtains

$$\mathscr{D}_-[y(t)] = \int_{-\infty}^{\infty} d\xi \int_{-\infty}^{\infty} dk\, \phi(\xi) \times h_1^{\ell}(k\xi)T(k)k^{\ell}\exp[ikt],$$

after integrations by parts.

Equation (20), which expresses $T_1[\phi]$, is used together with the definition $T \equiv 1 - S$:

$$\mathscr{D}_-[y(t)] = \mathscr{D}_-[\phi(-t)] - \frac{1}{2\pi} \times \int_{-\infty}^{\infty} d\xi\, \mathscr{D}_-[\phi(\xi)] \int_{-\infty}^{\infty} dk \exp ik\xi \times S(k)\exp ikt. \tag{24}$$

If we introduce $\tilde{\phi}$ by

$$\tilde{\phi}(-k) = \int_{-\infty}^{\infty} \mathscr{D}_-[\phi(\xi)]\exp ik\xi\, d\xi = \int_{-\infty}^{\infty} \phi(\xi)h_1^{\ell}(k\xi)k^{\ell}\, d\xi, \tag{25}$$

we can write:

$$\mathscr{D}_-[y(t)] = \mathscr{D}_-[\phi(-t)] - \frac{1}{2\pi} \times \int_{-\infty}^{\infty} \tilde{\phi}(-k)S(k)\exp[ikt]\, dk. \tag{26}$$

We retain the condition (22) and define the spaces $L'^2(-\infty,\infty), L'^2(\epsilon,\infty)$. The interest in introducing the L'^2 is its property of being dense in L^2.

A function $\phi(t)$ is an element of $L'^2(-\infty,\infty)$ if

$$\mathscr{D}_-[\phi(t)] \in L^2(-\infty,\infty).$$

A scalar product between two elements of $L'^2(-\infty,\infty)$ is defined as follows:

$$\langle\psi(t),\phi(t)\rangle = \int_{-\infty}^{\infty} \mathscr{D}_-[\psi^*(t)]\mathscr{D}_-[\phi(t)]\, dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{\psi}^*(k)\tilde{\phi}(k)\, dk.$$

Now we introduce

$L^2(\epsilon,\infty)$ and $L'^2(\epsilon,\infty)$ for $\epsilon \geqslant 0$.

A function $\phi(t)$ belongs to $L^2(\epsilon,\infty)$ if

$$\phi(t) \in L^2(-\infty,\infty),$$

$$\phi(t) = 0 \quad \text{for } t < \epsilon.$$

By definition

$$\int_{-\infty}^{\infty} \phi^*(t)\phi(t)\, dt \equiv \int_{\epsilon}^{\infty} \phi^*(t)\phi(t)\, dt.$$

A function $\phi(t)$ belongs to $L'^2(\epsilon,\infty)$ if

$$\mathscr{D}_-[\phi(t)] \in L^2(\epsilon,\infty).$$

Let us reconsider Eq. (21) when $\phi(t)$ belongs to $L'^2(\epsilon,\infty)$:

$$y(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} d\xi\, \phi(\xi)F_C(\xi,t) = \frac{1}{2\pi}\int_{\epsilon}^{\infty} d\xi\, \phi(\xi)F_C(\xi,t). \tag{21}$$

The function $y(t)$ belongs to $L^2(-\infty,+\infty)$. Together with $y(t)$ one may define a function $\bar{y}(t)$ in $L'^2(\epsilon,\infty)$ by

$$\bar{y}(t) = y(t), \quad t > \epsilon,$$

$$\mathscr{D}_-[\bar{y}(t)] = 0, \quad t \leqslant \epsilon.$$

The discussion of T_2 starts with expressions related to its norm. Obviously:

$$\int_\epsilon^\infty \mathscr{D}_-[\bar{y}^*(t)]\mathscr{D}_-[\bar{y}(t)]\,dt$$

$$<\int_{-\infty}^\infty \mathscr{D}_-[y^*(t)]\mathscr{D}_-[y(t)]\,dt.$$

Using Eq. (26) the l.h.s. of this inequality becomes for $\epsilon\geqslant 0$,

$$\int_\epsilon^\infty \mathscr{D}_-[\bar{y}^*(t)]\mathscr{D}_-[\bar{y}(t)]\,dt$$

$$<\int_{-\infty}^\infty dt\,\mathscr{D}_-[\bar{\phi}^*(t)]\mathscr{D}_-[\bar{\phi}(t)]$$

$$=\frac{1}{2\pi}\int_{-\infty}^\infty \tilde{\phi}(-k)S(k)S^*(k)\tilde{\phi}^*(-k)\,dk. \tag{27}$$

Equation (27) suggests that one consider the Fourier transform of $\mathscr{D}_-[y(t)]$. One writes Eq. (27) as

$$\frac{1}{2\pi}\int_{-\infty}^\infty \tilde{y}^*(k)\tilde{y}(k)\,dk$$

$$<\frac{1}{2\pi}\int_{-\infty}^\infty \tilde{\phi}^*(-k)S^*(k)S(k)\tilde{\phi}(-k)\,dk. \tag{27'}$$

Assume now an eigenvalue equation

$$\mathscr{D}_-[\bar{y}(t)]=\lambda\mathscr{D}_-[\phi(t)]\equiv\lambda x(t),$$

where $x(t)\in L^2(\epsilon,\infty)$. The eigenvalue equation reads

$$\lambda x(t)=x(-t)-\frac{1}{2\pi}\int_{-\infty}^\infty \tilde{x}(-k)S(k)\exp ikt\,dk.$$

Since $x(t)\in L^2(\epsilon,\infty)$

$$\int_{-\infty}^\infty x(t)\exp[\pm ikt]\,dt=\int_\epsilon^\infty x(t)\exp ikt\,dt.$$

Applying the Fourier transform to the eigenvalue equation ($\epsilon\geqslant 0$) gives either

$$\lambda\tilde{x}(k)=-\tilde{x}(-k)S(k)$$

or

$$\lambda\tilde{x}(-k)=-\tilde{x}(k)S(-k).$$

The first of these two equations is multiplied by $S(-k)$ and to simplify the result the equality $S(k)S(-k)=1$, valid for non-Hermitian as well as for Hermitian interactions[7] is used. One obtains

$$\lambda\tilde{x}(k)S(-k)=\tilde{x}(-k)=(1/\lambda)\tilde{x}(k)S(-k).$$

Therefore $\lambda^2=1$ or $\lambda=\pm 1$ for all the eigensolutions in $L^2(\epsilon,\infty)$ of the equation

$$\lambda x(t)-x(-t)=\frac{1}{2\pi}\int_{-\infty}^\infty \tilde{x}(-k)S(k)\exp ikt\,dt. \tag{28}$$

Lemma 1: The eigenvalues of Eq. (28) considered as an equation in $L'^2(\epsilon,\infty)$ ($\epsilon\geqslant 0$) are equal to plus or minus one.

Lemma 1 applies to Hermitian and non-Hermitian interactions as well.

To pursue and to shorten the notations we denote

$$\mathscr{D}_-[\phi(t)]=x(t),$$

$$\tilde{\phi}(k)=\tilde{x}(k),$$

$$\phi(t)\in L'^2(\epsilon,\infty),\quad x(t)\in L^2(\epsilon,\infty).$$

The continuous Marchenko associated $M_C(\xi,t)$ kernel is introduced

$$M_C(\xi,t)=\frac{1}{2\pi}\int_{-\infty}^\infty \exp ik\xi\,[1-S(k)]\exp ikt\,dt,$$

$$M_C(\xi,t)\equiv M_C(\xi+t).$$

Equation (28) is equivalent to the Marchenko associated equation

$$\lambda x(t)+\int_\epsilon^\infty x(\xi)M_C(\xi+t)\,d\xi=0. \tag{29}$$

Let us consider the scalar product

$\langle T_2\phi,\phi\rangle$, when ϕ is in $L'^2(\epsilon,\infty)$. Let

$$A\equiv\int_\epsilon^\infty dt\int_\epsilon^\infty x^*(\xi)M_C^*(\xi+t)x(t)\,d\xi.$$

One has

$$A\equiv\int_\epsilon^\infty d\xi\,x^*(\xi)\int_\epsilon^\infty dt\,x(t)\delta(t+\xi)$$

$$-\frac{1}{2\pi}\int_{-\infty}^\infty S(k)\,dk\int_\epsilon^\infty x^*(\xi)\exp ik\xi\,d\xi$$

$$\times\int_\epsilon^\infty x(t)\exp ikt\,dt.$$

When $x(t)\in L^2(\epsilon,\infty)$,

$$A=-\frac{1}{2\pi}\int_{-\infty}^\infty \tilde{x}^*(-k)S^*(k)\tilde{x}(k)\,dk.$$

From the inequality

$$\int_{-\infty}^\infty (f^*+g^*)(f+g)\,dk\geqslant 0$$

one can derive a Schwarz inequality in the form

$$\left|\int_{-\infty}^\infty f^*(k)g(k)\,dk\right|$$

$$\leqslant\frac{1}{2}\left[\int_{-\infty}^\infty f^*(k)f(k)\,dk+\int_{-\infty}^\infty g^*(k)g(k)\,dk\right],$$

$$|A|\leqslant\frac{1}{2}\left[\int_{-\infty}^\infty \tilde{x}^*(-k)S^*(k)\tilde{x}(-k)\,dk\right.$$

$$\left.+\int_{-\infty}^\infty \tilde{x}^*(k)\tilde{x}(k)\,dk\right]. \tag{30}$$

We note now the identities

$$\frac{1}{2\pi}\int_{-\infty}^\infty \tilde{x}^*(k)\tilde{x}(k)\,dk=\int_\epsilon^\infty x^*(t)x(t)\,dt;$$

$$\frac{1}{2\pi}\int_{-\infty}^\infty \tilde{x}(-k)^*S(k)^*S(k)x(k)\,dk$$

$$=\int_\epsilon^\infty dt\int_\epsilon^\infty \tilde{x}^*(\xi)M_C^*(\xi+t)\,d\xi$$

$$\times\int_\epsilon^\infty M_C(\xi'+t)x(\xi')\,d\xi'.$$

Since the inequality

$$|a^*b|\leqslant\tfrac{1}{2}[aa^*+bb^*],$$

$\mu=\pm 1$ implies $\mathrm{Re}[\tfrac{1}{2}aa^*+\tfrac{1}{2}bb^*+\mu a^*b]\geqslant 0$, one has:

$$\mathrm{Re}\left\{\frac{1}{2}\int_\epsilon^\infty x^*(t)x(t)\,dt+\frac{1}{2}\int_\epsilon^\infty dt\int_\epsilon^\infty x^*(\xi)\times M_C{}^*(\xi+t)\,d\xi\int_\epsilon^\infty M_C(\xi'+t)\,x(\xi')\,d\xi' + \mu\int_\epsilon^\infty x(t)\,dt\int_\epsilon^\infty x^*(\xi)M_C{}^*(\xi+t)\right\}\geqslant 0. \quad (31)$$

With $\lambda = 1/\mu$, the l.h.s. of (31) is equal to zero if and only if

$$\int_{-\infty}^{\infty}\tilde{x}^*(-k)\,S^*(k)\,x(k)\,dk = \frac{\lambda}{2}\left\{\int_{-\infty}^{\infty}\tilde{x}^*(-k)S^*(k)\times S(k)\tilde{x}(-k)\,dk+\int_{-\infty}^{\infty}x^*(k)\tilde{x}(k)\,dk\right\}. \quad (32)$$

We look for a solution of (32) of the form

$$\tilde{x}^*(-k)S^*(k)=\alpha^*\tilde{x}^*(k). \quad (33)$$

From (32) one obtains:

$$\alpha^*\int_{-\infty}^{\infty}\tilde{x}^*(k)x(k)\,dk = \tfrac{1}{2}\lambda\left\{\alpha^*\alpha\int_{-\infty}^{\infty}\tilde{x}^*(k)\times x(k)\,dk+\int_{-\infty}^{\infty}\tilde{x}^*(k)x(k)\,dk\right\}.$$

Since λ is equal to ± 1, α is real and equal to λ. The computation of both sides of (32) with $\alpha = \pm 1$ yields an identity. In other words, (33) implies (32).

Conversely, (32) implies (33) with $\alpha = \lambda$. If it does not, there exists a nonzero function $\tilde{z}(k)$ such that

$$\tilde{x}^*(-k)S^*(k)=\lambda\tilde{x}^*(k)+\tilde{z}^*(k). \quad (34)$$

The two sides of (34) are squared ($\lambda\lambda^* = 1$)

$$\int_{-\infty}^{\infty}\tilde{x}^*(-k)S^*(k)S(k)\tilde{x}(-k)\,dk = \int_{-\infty}^{\infty}\tilde{x}^*(k)\tilde{x}(k)\,dk+\lambda\int_{-\infty}^{\infty}\tilde{x}^*(k)z(k)\,dk + \lambda^*\int_{-\infty}^{\infty}z^*(k)\tilde{x}(k)\,dk+\int_{-\infty}^{\infty}\tilde{z}^*(k)\tilde{z}(k)\,dk. \quad (35)$$

We now insert (34) into the l.h.s. of (32) and obtain

$$\lambda\int_{-\infty}^{\infty}\tilde{x}^*(k)\tilde{x}(k)\,dk+\int_{-\infty}^{\infty}\tilde{z}^*(k)\tilde{x}(k)\,dk = \frac{\lambda}{2}\left[\int_{-\infty}^{\infty}\tilde{x}^*(-k)S^*(k)S(k)\tilde{x}(k)\,dk + \int_{-\infty}^{\infty}\tilde{x}^*(k)x(k)\,dk\right].$$

Equation (32b) says that

$$\lambda^*\int_{-\infty}^{\infty}\tilde{z}^*(k)\tilde{x}(k)\,dk=\lambda\int_{-\infty}^{\infty}\tilde{z}(k)\tilde{x}^*(k)\,dk$$

is real. In addition we have

$$2\lambda^*\int_{-\infty}^{\infty}\tilde{z}^*(k)\tilde{x}(k)\,dk = \int_{-\infty}^{\infty}\tilde{x}(-k)^*S(k)^*S(k)\tilde{x}(k)\,dk - \int_{-\infty}^{\infty}\tilde{x}(k)^*\tilde{x}(k)\,dk. \quad (36)$$

With (36) we return to (35) and obtain

$$\int_{-\infty}^{\infty}z^*(k)z(k)\,dk=0,$$

or $z(k)=0$. Equation (32) therefore implies Eq. (33).

At no point of the argument has use been made of a possible unitarity of $S(k)$; therefore the reasoning extends to complex interactions.

The conclusion of this discussion of the scalar product $\langle T_2[y],y\rangle$ results. From the uniqueness of the Fourier transform, the equation

$$\tilde{x}(-k)^*S(k)^*=\lambda\tilde{x}(k)^*$$

is equivalent to the Marchenko associated equation

$$\lambda x(t)+\int_\epsilon^\infty x(\xi)M_C(\xi+t)\,d\xi=0, \quad (37)$$

$$0\leqslant\epsilon\leqslant t<\infty;\quad \lambda=\pm 1;\quad x(t)\in L^2(\epsilon,\infty)$$

or to our initial equation

$$\lambda\mathscr{D}_-[\phi(t)]+\int_\epsilon^\infty\mathscr{D}_-[\phi(\xi)]F_C(\xi,t)\,d\xi=0,$$

$$\epsilon\leqslant t<\infty;\quad \phi(t)\in L'^2(\epsilon,\infty).$$

No use is made here of (31), except in the particular case where $S(k)$ is unitary. This form is derived later.

We move now toward the proof that $\lambda = \pm 1$ cannot be an eigenvalue for (37). Still assuming $\phi(t)$ solution of (37) in $L'^2(\epsilon,\infty)$ one has

$$\tilde{x}^*(-k)S^*(k)=\lambda\tilde{x}^*(k);\quad \lambda=\pm 1. \quad (33')$$

Thus we write

$$K(t)\equiv\lambda x(t)+\int_\epsilon^\infty x(\xi)M_C(\xi,t)\,dt$$

or

$$K(t)\equiv\lambda x(t)+x(-t)-\frac{1}{2\pi}\int_{-\infty}^{\infty}\tilde{x}(-k)S(k)\times\exp[ikt]\,dt. \quad (38)$$

Use of (33′) in the last term of (38) gives

$$K(t)\equiv\lambda x(t)+x(-t)-x(t).$$

Therefore we may write

$$\lambda x(t)+\int_\epsilon^\infty\phi(\xi)M_C(\xi,t)\,d\xi=x(-t),\quad \lambda=\pm 1,$$

which is valid for $-\infty<t<\infty$.

Lemma 2: Let $\epsilon>0$, then the equation

$$\lambda\phi(t)+\int_\epsilon^\infty\phi(\xi)F_C(\xi,t)\,d\xi=0,\quad \epsilon\leqslant t<\infty$$

has no eigenvalue λ with $\lambda=\pm 1$.

The Lemma assumes ϵ strictly positive and its proof is found in Ref. 1; it is not repeated here.

When $[1 - S(k)] = O[k^{2\ell}]$ the transformation generated by F_C is completely continuous in $L'^2(\epsilon, \infty)$. Following Ref. 1 one says that if an eigenvalue exists with $\lambda = \pm 1$, one can construct an infinite set (not countable) of distinct eigensolutions with the same eigenvalue. This construction is excluded by the complete continuity of F_C and therefore no eigenvalue $\lambda = \pm 1$ may exist.

Lemma 3: Let $\epsilon \geqslant 0$, let $S(k)$ be unitary; in order for $\phi(t)$ to be a solution in $L'^2(\epsilon, \infty)$ of the equation

$$\phi(t) + \int_\epsilon^\infty \phi(\xi) F(\xi, t)\, d\xi = 0$$

it must also be a solution of the equation

$$\phi(t) + \int_\epsilon^\infty \phi(\xi) F_C(\xi, t)\, d\xi = 0;$$

in addition each $\tilde{\phi}(ik_n)$ must vanish.

Before proving Lemma 3 we return to Eqs. (28)–(32) assuming this time the unitarity of $S(k)$. The new equations will retain the same number, followed by U.

Instead of (27′) we have

$$\frac{1}{2\pi}\int_{-\infty}^\infty \tilde{y}^*(k)\tilde{y}(k)\, dk \leqslant \frac{1}{2\pi}\int_{-\infty}^\infty \tilde{\phi}^*(-k)\tilde{\phi}(-k)\, dk$$

$$= \int_\epsilon^\infty x^*(t)x(t)\, dt.$$

Equation (27′U) expresses that the norm T_2 is less than one and consequently that the eigenvalues of Eq. (28U) are smaller or equal to one in modulus.

In the study of the scalar product $\langle T_2[\phi], \phi\rangle$ the inequality (30) becomes

$$\left| \frac{1}{2\pi}\int_{-\infty}^\infty \tilde{x}^*(-k)S^*(k)\tilde{x}(k)\, dk \right|$$

$$\leqslant \frac{1}{2\pi}\int_{-\infty}^\infty \tilde{x}^*(-k)\tilde{x}(k)\, dk$$

$$= \int_\epsilon^\infty x^*(t)x(t)\, dt. \qquad (30\text{U})$$

As a consequence inequality (31U) reads ($\mu = \pm 1$):

$$\text{Re}\left\{ \int_\epsilon^\infty x^*(t)x(t)\, dt + \mu\int_\epsilon^\infty dt\, x(t) \times \int_\epsilon^\infty M_C^*(\xi + t)x^*(\xi)\, d\xi \right\} \geqslant 0. \qquad (31\text{U})$$

The necessary and sufficient condition for (31U) to hold is given in two equivalent forms: (32U) and (33U)

$$\int_{-\infty}^\infty \tilde{x}^*(-k)S^*(k)x(k)\, dk = \int_{-\infty}^\infty \tilde{x}^*(k)x(k)\, dk, \qquad (32\text{U})$$

or

$$\tilde{x}^*(-k)S^*(k) = \pm\tilde{x}(k). \qquad (33\text{U})$$

Lemma 3 is now ready for proof. Let us assume that $\phi(t)$ is a solution of the equation

$$\phi(t) + \int_\epsilon^\infty \phi(\xi)F(\xi, t)\, d\xi = 0, \qquad (39\text{a})$$

with $F \equiv F_C + F_D$ and $\phi(t) \in L'^2(\epsilon, \infty)$. Equation (39a) is multiplied by $\phi^*(t)$ and integrated. One gets

$$\int_\epsilon^\infty \phi(t)\phi(t)^*\, dt + \int_\epsilon^\infty \phi^*(t)\, dt \int_\epsilon^\infty \phi(\xi)F_C(\xi, t)\, d\xi + \int_\epsilon^\infty \phi^*(t)\, dt \int_\epsilon^\infty \phi(\xi, t)F_D(\xi, t)\, d\xi = 0. \qquad (39\text{b})$$

The last term of Eq. (40), if $1 - S(k) = O(k^2)$, is

$$D \equiv \sum_{n=1}^{p} \tilde{\phi}(ik_n)M_n^2\tilde{\phi}^*(ik_n)$$

and is semipositive definite.

The real part of Eq. (39b) yields

$$\text{Re}\left\{ \int_\epsilon^\infty \phi^*(t)\phi(t)\, dt + \int_\epsilon^\infty \phi^*(t)\, dt \times \int_\epsilon^\infty \phi(\xi)F_C(\xi, t)\, d\xi + D \right\} \equiv 0.$$

Since D is positive definite, Eq. (39b) is obtained if

(a) $\tilde{\phi}(ik_n)M_n \equiv 0$ for all n,

(b) $$\text{Re}\left[\int_\epsilon^\infty \phi^*(t)\phi(t)\, dt + \int_\epsilon^\infty \phi^*(t)\, dt \times \int_\epsilon^\infty \phi(\xi)F_C(\xi, t)\, d\xi \right] = 0.$$

By (31U) we know that

$$\text{Re}\left\{ \int_\epsilon^\infty \phi^*(t)\phi(t)\, dt + \int_\epsilon^\infty \phi^*(t) \times \int_\epsilon^\infty \phi(\xi)F_C(\xi, t)\, d\xi \right\} \geqslant 0,$$

with equality if and only if for $t > \epsilon$, one has

$$\pm\phi(t) = \int_\epsilon^\infty \phi(\xi)F_C(\xi, t)\, d\xi. \qquad (39\text{c})$$

It results, therefore, that for $\phi(t) \in L'^2(\epsilon, \infty)$ to be an eigensolution

$$\phi(t) + \int_\epsilon^\infty \phi(\xi)F(\xi, t)\, d\xi = 0, \quad \epsilon \leqslant t < \infty,$$

$\phi(t)$ must also be a solution of the equation

$$\phi(t) \pm \int_\epsilon^\infty \phi(\xi)F_C(\xi, t)\, d\xi = 0.$$

Let us now consider a possible reciprocal statement and assume $\phi(t)$ is a solution of

$$\phi(t) \pm \int_\epsilon^\infty \phi(\xi)F_C(\xi, t)\, d\xi = 0$$

in $L'^2(\epsilon, \infty)$. As before one gets

$$\int_\epsilon^\infty \phi^*(t)\phi(t)\, dt + \int_\epsilon^\infty \phi^*(t)\, dt \int_\epsilon^\infty \phi(\xi)F_C(\xi, t)\, d\xi = 0.$$

To obtain (39b) one must add D, therefore one must require $D \equiv 0$ or equivalently require that each

$\tilde{\phi}(ik_n)M_n = 0$

in order to construct a solution of

$$\phi(t) + \int_\epsilon^\infty \phi(\xi)F(\xi, t)\, d\xi = 0.$$

Lemma 3 has in general no equivalent in a non-Hermitian case. When one considers a non-Hermitian problem, one has to require that Eq. (7') have no nontrivial solution. This requirement can be an *a priori* one as in Ref. 6 where Ljance sets up an essential condition 2 for his F-spectral function. It can also be realized by a proof that the set of scattering data are authentic, in the sense defined in Sec. 3. There one finds the essential difference between Hermitian and non-Hermitian inverse problems.

All the results obtained so far have been obtained with the assumption that $\phi(t)$ and its $(\ell-1)$ first derivatives vanish at $t\leqslant\epsilon$. Let $\{R_n(t)\}$ be a denumerable basis for $L'^2(0,\infty)$, the set $\{\bar{R}_n(u)\}\equiv\{R_n(t-\epsilon)\}$ is a denumerable basis for $L'^2(\epsilon,\infty)$. With its help, given any $\phi(t)\in L^2(-\infty,\infty)$ defined for $t\geqslant\epsilon$, one can construct a sequence

$$\{\phi_n(t)\}$$

such that

$$\int_\epsilon^\infty [\phi_n(t)-\phi(t)][\phi_n^*(t)-\phi^*(t)]\,dt$$

is smaller than any prescribed positive quantity. In other words $L'^2(\epsilon,\infty)$ is dense in $L^2(\epsilon,\infty)$. This extends the study to any element $\phi(t)$ of $L^2(\epsilon,\infty)$.

In Appendix B we prove directly that if $\phi(t)$ is a solution of Eq. (32') with plus sign then $\mu(t)\equiv\mathscr{D}_-[\phi(t)]$ is also a solution of Marchenko's associated equation:

$$\mu(t)+\int_\epsilon^\infty \mu(t)\,M_C(\xi+t)\,d\xi=0, \tag{40}$$

and we know from [1] that $\mu(t)$ must then be identically zero.

B. Solutions in $L^1(\epsilon,\infty)$

In Subsec. 4 A concerned with $L^2(\epsilon,\infty)$ solutions, the importance of the study of Marchenko associated equations was made clear. Here in Subsec. 4 B concerned with $L^1(\epsilon,\infty)$ the same need will become apparent. In order to be explicit we restrict ourselves to a complete treatment of the $\ell=1$ case. Then dropping the superscript 1 we have

$$F(\xi,t)=\frac{1}{2\pi}\int_{-\infty}^\infty h_1(k\xi)[1-S(k)]h_1(kt)\,dk + \sum_{n=1}^{p} h_1(ik_n\xi)M_n^2 h_1(ik_n t);$$

$$\mathscr{D}_-(\xi)=\frac{1}{\xi}\frac{d}{d\xi}\xi,$$

$$h_1(k\xi)=i(1+i/k\xi)\exp ik\xi.$$

Assumption is made that $T(k)=O(k^2)$. To shorten the notations, and to mean that in the discrete part $[1-S(k)]$ must be replaced by M_n^2 we write,

$$F(\xi,t)=\sum\!\!\!\!\!\!\int h_1(k\xi)[1-S(k)]h_1(kt)\,dk.$$

The Marchenko associated kernel $M_0(\xi,t)$ reads

$$M_0(\xi,t)=\sum\!\!\!\!\!\!\int \exp ik\xi\,[1-S(k)]\exp ikt\,dk. \tag{41}$$

By definition a function $\phi(t)$ will be said to belong to $L'^1(\epsilon,\infty)$ if

$$\int_\epsilon^\infty |\mathscr{D}\{\phi(t)\}|dt\equiv||\phi||<\infty \tag{42}$$

and $|\phi(\epsilon)|=|\phi|<\infty$.

We may use a norm, with subscript 1, in this space

$$||\phi||_1\equiv||\phi||+|\phi|.$$

By its definition $L'^1(\epsilon,\infty)$ is contained in $L^1(\epsilon,\infty)$. Let us consider

$$y(t)=\int_\epsilon^\infty \phi(\xi)F(\xi,t)\,dt.$$

Application of $\mathscr{D}_-(t)$ to both sides of Eq. (42) gives

$$\mathscr{D}_-[y(t)]=\phi(\epsilon)M_0(t+\epsilon)+\int_\epsilon^\infty \mathscr{D}_-[\phi(\xi)]M_0(\xi,t)\,d\xi.$$

Therefore

$$||y||<|\phi|\int_\epsilon^\infty |M_0(s)|\,ds+||\phi||\int_\epsilon^\infty |M_0(s)|\,ds. \tag{43}$$

We assume $\epsilon\neq0$ and define $M_1(t)$, $M_2(t)$

$$M_1(t)=\sum\!\!\!\!\!\!\int \frac{1}{k}[1-S(k)]\exp ikt\,dk,$$
$$M_2(t)=\sum\!\!\!\!\!\!\int \frac{1}{k^2}[1-S(k)]\exp ikt\,dk. \tag{44}$$

If $T(k)=0(k^2)$, $M_1(t)$ and $M_2(t)$ exist together with $M_0(t)$; any element transformed by M and belonging to $L'^1(\epsilon,\infty)$ belongs to $L^1(\epsilon,\infty)$. This remains true even if $\epsilon=0$ as we shall see later in Subsec. 4 C. Therefore, from now on, we will omit the superscript prime when dealing with L^1:

$$y(\epsilon)=\phi(\epsilon)\left[iM_1(2\epsilon)-\frac{1}{\epsilon}M_2(2\epsilon)\right]+\int_\epsilon^\infty i\mathscr{D}_-[\phi(\epsilon)]M_1(\xi+\epsilon)\,d\xi-\frac{1}{\epsilon}\int_\epsilon^\infty \mathscr{D}_-[\phi(\xi)]M_2(\xi+\epsilon)\,d\xi.$$

If one assumes $|M_1(s)|$ and $|M_2(s)|$ to exist, or if

$$\int_{-\infty}^\infty \frac{1}{k}[1-S(k)]\exp iks\,ds$$

and

$$\int_{-\infty}^\infty \frac{1}{k^2}[1-S(k)]\exp iks\,ds$$

exist, that is if $T(k)$ is $O(k^2)$, then $y(\epsilon)$ is bounded for $\epsilon>0$. Now let $||M||=\sup\{||M_1||,||M_2||\}$;

$$|y|\leqslant[|\phi|+||\phi||](1+\epsilon)||M||$$
$$|y|\leqslant||\phi||_1(1+\epsilon)||M||. \tag{45}$$

Putting together (43) and (44) one has

$$||y||_1<||\phi||_1[(1+\epsilon)||M||+||M_0||]. \tag{46}$$

We have, therefore, a norm for the F-transformation in $L^1(\epsilon,\infty)$.

We follow now Marchenko; for the same reasons as in Ref. 1 the F-transformation is completely continuous in $L^1(\epsilon,\infty)$.

As we defined $L^{1}(\epsilon,\infty)$, we define $L^{2}(\epsilon,\infty)$; we omit the superscript prime since L'^{2} is dense in L^{2}. A function $\phi(t)$ will be said to belong to $L^{2}(\epsilon,\infty)$ if

$$\int_{\epsilon}^{\infty}\mathscr{D}_{-}[\phi(t)^{*}]\mathscr{D}_{-}[\phi(t)]\,dt<\infty$$

and if the product $\phi(\epsilon)\phi^{*}(\epsilon)$ is bounded. The norm in $L^{2}(\epsilon,\infty)$ having a subscript 2 notations follow: $|\phi|$ for L^{2} is identical to $|\phi|$ for L^{1},

$$\|\phi\|=\int_{\epsilon}^{\infty}\mathscr{D}_{-}[\phi(t)^{*}]\mathscr{D}_{-}[\phi(t)]\,dt$$

$$\|\phi\|_{2}=|\phi|+\|\phi\|.$$

Now we consider the equation

$$\phi(t)+\int_{\epsilon}^{\infty}\phi(\xi)F(\xi,t)\,d\xi=0. \tag{47}$$

Applying $\mathscr{D}_{-}(t)$ and integrating by parts gives

$$0=\mathscr{D}_{-}[\phi(t)]+\phi(\epsilon)M_{0}(t+\epsilon)+\int_{\epsilon}^{\infty}\mathscr{D}_{-}[\phi(\xi)]M_{0}(t+\xi)\,d\xi. \tag{48}$$

Following Marchenko (see Appendix B for a summary of his method), we obtain the following result.

Lemma 4: If F has a norm defined as (46) then any solution $\phi(t)$ of Eq. (47) in $L^{1}(\epsilon,\infty)$ is also a solution of (47) in $L^{2}(\epsilon,\infty)$.

Now we know from Sec. 4A that such a solution is necessarily related to a solution of the Marchenko associated equation

$$\mathscr{D}_{-}[\phi(t)]+\int_{\epsilon}^{\infty}\mathscr{D}_{-}[\phi(\xi)]M_{0}(\xi+t)\,d\xi=0.$$

Notations being as in (41) and (44); $M_{0,1,2;\mathrm{C}}$ denotes elements of the continuum part of the kernel M. Lemma 5 is obtained.

Lemma 5: Let $\epsilon>0$. Let $M_{0\mathrm{C}}\in L(-\infty,\infty)$ be the Fourier transform of $T(k)=O(k^{2})$ where $S(k)$ is unitary. Let F be

$$F\equiv F_{\mathrm{C}}+F_{\mathrm{D}}\quad\text{with }F_{\mathrm{D}}=\sum_{n=1}^{p}h_{1}(ik_{n}\xi)M_{n}^{2}h_{1}(ik_{n}t),$$

The M_{n}^{2}'s are positive in finite number p no requirement p). Then the equation

$$f(t)+x(t)+\int_{\epsilon}^{\infty}x(\xi)F(\xi,t)\,d\xi=0$$

for any $f(t)\in L^{1}(\epsilon,\infty)$, where $\epsilon>0$, has a unique solution in $L^{1}(\epsilon,\infty)$.

The proof of this lemma uses Lemmas 1–4; the requirement for a strictly positive ϵ is therefore included.

C. Case $\epsilon=0$

We consider now the case where ϵ is equal to zero. Since

$$F_{\mathrm{C}}\equiv\frac{1}{2\pi}\int_{-\infty}^{+\infty}h_{1}'(k\xi)T(k)h_{1}'(kt)\,dk,$$

one has to consider a set of functions $\phi(\xi)$ such that for some finite α the integral

$$\int_{0}^{\alpha}\phi(\xi)h_{1}'(h\xi)\,d\xi\ \text{ converges.}$$

In other words, one requires

$$\lim_{\xi\to0}\phi(\xi)=A\xi^{\alpha},$$

with $\alpha>\ell-1$.

The set of functions $\phi(t)$ belongs to $L^{1}(0,\infty)$ and in addition satisfies the previous limit at α equal to zero. In other words,

$$\phi(\xi)\in L^{1}(0,\infty);\ \mathscr{D}_{-}[\phi(\xi)]\in L^{1}(0,\infty).$$

Returning to parts A and B, one sees that while Lemmas 1,3, and 4 hold for vanishing ϵ, Lemmas 2 and 5 do not. A study is therefore needed when ϵ goes to zero; this is achieved by discussing the equations which follow.

These are the two equations

$$\phi(t)+\int_{0}^{\infty}\phi(\xi)F(\xi,t)\quad dt=0,$$

$$\phi(t)\pm\int_{0}^{\infty}\phi(\xi)F_{\mathrm{C}}(\xi,t)\quad dt=0.$$

To these two equations, we must add, for reasons which become apparent later, the Hilbert associated equation

$$\phi(t)+\int_{0}^{\infty}\phi(\xi)F_{\mathrm{C}}(-\xi,-t)=0.$$

Together with these three equations we consider the three Marchenko associated equations where F is replaced by M and F_{C} by M_{C}, see Eqs. (40), (41), and (44).

Proposition:

(I) The equation

$$\phi(t)+\int_{0}^{\infty}\phi(\xi)F_{\mathrm{C}}(\xi,t)\,d\xi=0$$

has n linearly independent solutions in $\hat{L}(0,\infty)$ (see Appendix A for the definition of $\hat{L}$); n is the number of the bound states.

(II) The equation

$$\pm\phi(t)+\int_{0}^{\infty}\phi(\xi)F_{\mathrm{C}}(-\xi,-t)\,d\xi=0$$

has no nontrivial solution in $\hat{L}(0,\infty)$.

(III) The equation

$$\phi(t)+\int_{0}^{\infty}\phi(\xi)F(\xi,t)\,d\xi=0$$

has no nontrivial solution in $\hat{L}(0,\infty)$.

Proof: Application of the $\mathscr{D}$ operator and use of integrations by parts transform (I), (II), and (III) into Marchenko associated equations. The proof of the three propositions is therefore found in Ref. 1. It is sufficient that a summary of the method be given here.

(I) The equation is replaced by

$$\tilde{x}(k)=\tilde{x}(-k)S(k).$$

A function $\tilde{\phi}(k)$ is constructed:

$$\tilde{\phi}_{-}(k)=\tilde{x}(k)f_{1}^{-1}(-k),\quad \mathrm{Im}k\leqslant0,$$

$$\tilde{\phi}_{+}(k)=\tilde{x}(-k)f_{1}^{-1}(k),\quad \mathrm{Im}k\geqslant0,$$

$$\tilde{\phi}=\tilde{\phi}_{+}=\tilde{\phi}_{-},\quad \mathrm{Im}k=0.$$

where

$$f_{1}^{-1}(-k)=\lim_{r\to0}(-kr)^{\ell}f_{1}(-k,r).$$

Since $\phi(t)$ of the original equation belongs to $\hat{L}(0,\infty)$ the $\tilde{x}(k)$ vanishes when k goes to infinity. The function ϕ is meromorphic in the k plane. It has poles at every bound state location. The proposition follows a simple computation.

(II) The equation is replaced by

$$\pm \tilde{x}(k) = \tilde{x}(k) S(-k)$$

A function $\tilde{\phi}(k)$ is constructed

$$\tilde{\phi}_-(k) = \pm \tilde{x}(k) f_1^*(-k^*), \quad \mathrm{Im}k<0,$$
$$\tilde{\phi}_+(-k) = \tilde{x}(-k) f_1^*(k^*), \quad \mathrm{Im}k>0,$$
$$\tilde{\phi} = \tilde{\phi}_- = \tilde{\phi}_+, \quad \mathrm{Im}k=0.$$

By construction $\tilde{\phi}(k)$ is a bounded function of k which vanishes at $k=\infty$. One has therefore

$$\tilde{\phi}(k) \equiv 0$$

and the proposition results.

(III) (Valid for Hermitian potentials)

From Lemma 3 one knows that in order for $\phi(t)$ to be a solution of (III) one must have

$$\begin{cases} \tilde{x}(k)=0, & |k|\to\infty \\ \tilde{x}(k)=\tilde{x}(-k)S(k) \end{cases}$$

$\tilde{x}(ik_p x)=0$ for n values ik_p, $p=1,\dots n$. Since (I) has n independent solutions, a solution of (III) must satisfy $(n+1)$ conditions.

$$\tilde{x}(ik_p x)=0$$
$$\tilde{x}(k)=0, \quad |k|=\infty.$$

Only the trivial solutions satisfies these requirements; the proposition results.

Since n intervening in (I) and (III) is the number of bound states, the number of terms in F_D must also be n. In other words the number of terms of F_D is specified by Levinson's theorem.[8]

Converse Proposition: This converse proposition depends on the special relationship between the solutions of (I), (II), and (III) and consequently does not apply to non-Hermitian systems where the relationship fails to exist.

Let n be the number of bound states as given by Levinson's formula.[8] Let p be the number of the terms belonging to F_D. Assume $p<n$, then III has $(n-p)$ linearly independent solutions. Assume $p>n$, then III has no independent solutions. However there will be $(n-p)$ values ik_n for which

$$\tilde{\phi}(ik_n)=0 \text{ and } f_1(ik_n)\neq 0.$$

The equation

$$\tilde{\phi}(k)=\tilde{\phi}(-k)S(k)$$

implies that $f_1(-k)$ has a pole for $k=ik_n$. Consequently (I) will have $(n-p)$ independent solutions. A statement that (I) and (III) have no independent solutions is therefore equivalent to the statement that the number of terms of the discrete part F_D is identical to the number of bound states as given by Levinson's formula.

CONCLUSIONS

This paper has been undertaken for two purposes: (a) to show that a direct study of the $\ell\neq 0$ inverse problem was natural and purpose (b) to specify where differences between Hermitian and non-Hermitian systems occur. We arrived at some conclusions. They will be expressed in terms of conditions A, B, C, D and E. The conditions are now enumerated.

A) A matrix $S(k)$ is given: It is unitary and $T(k)=0(k^{2\ell})$ for $k=0$. We will show later cases where the condition $T(k)=0(k^{2\ell})$ for $k=0$ can be removed.

B) The Fourier transform $M_C(t)$ of $[1-S(k)]$ exists.

C) An arbitrary but finite number of real numbers M_n is given so that a kernel

$$F(\xi,t)=\frac{1}{2\pi}\int_{-\infty}^{\infty} h_1^{\ell}(k\xi)[1-S(k)]h_1^{\ell}(kt)\,dk + \sum_{n=1}^{p} h_1^{\ell}(ik_n\xi)[M_h^2]h_1^{\ell}(ik_n t)$$

for real k_n can be constructed.

Conclusion 1: If A, B, and C are verified, the equation

$$K(x,y)+F(x,y)+\int_x^{\infty} K(x,z)F(z,y)\,dz=0$$

has always a unique solution for $x>0$. If, on the one hand, conditions, A and B concerning S or, on the other hand, condition C concerning the M_n's and the k_n's are not satisfied, no conclusion can be obtained. There lies the answer for purpose (b).

D) The requirements on $F(\xi,t)$ are strong enough for $K(x,x)$ to be differentiable and for $K(x,y)$ to be y-absolutely integrable.

Conclusion 2: If the requirement D is not satisfied no potential can be obtained from the solution of the fundamental equation and the Marchenko representation of the irregular solution is not valid.

E) On the one hand, the equation

$$x(t)+\int_0^{\infty} x(\xi)M_C(\xi+t)\,d\xi=0$$

has n-independent solutions where n is the index of S (for the definition of the index, see Gakhov[9]). On the other hand, the equation

$$x(t)+\int_0^{\infty} x(\xi)M_C(-\xi,-t)\,d\xi=0$$

has no nontrivial solutions.

Conclusion 3: (Sec. C; $\epsilon>0$). The two conditions of E state that $S(k)\equiv f_1(-k)/f_1(k)$ is factorizable. These conditions are equivalent to the Levinson theorem.[8] If they are not verified

$$S(k)\neq \lim_{x\to 0}\frac{f_1(-k,x)}{f_1(k,x)}(-1)^{\ell}.$$

Conclusion 4: If all the conditions A, B, C, D, E are satisfied then the equation

$$K(x,y)+F(x,y)+\int_x^{\infty} K(x,z)F(z,y)\,dz=0$$

has a unique solution $K(x,y)$ when $S(k)$, M_n^2, k_n are given to construct $F(x,y)$. One obtains $V(x)$ by

$$V(x)=-\tfrac{1}{2}\frac{d}{dx}K(x,x).$$

Let now a $V(x)$ be given which satisfies (5a)–(5c) and (23); it is possible to construct a unique translation kernel $K(x,y)$ by the Riemann method. This unique $K(x,y)$ generates a unique spectral matrix $\bar{F}(x,y)$ by the fundamental equation. Since $\bar{F}(x,y)$ is unique, by identification we are assured that the spectral elements $\bar{S}(k)$, $\bar{M}_n^2$, $\bar{k}_n$, of $\bar{F}(x,y)$ are identical to the spectral elements

$$\bar{S}(k)\equiv S(k),$$
$$\bar{M}_n^2\equiv M_n,$$
$$\bar{k}_n\equiv k_n,$$

of $F(x,y)$.

In other words, the exact scattering data are identical to the scattering data used for the construction of the fundamental equation; the circle is closed. Let us now remove the restriction we set in A, namely that $T(k)$ be $O(k^{2\ell})$, i.e., condition (23), while the sufficient condition for the existence of an integrable $K(x,y)$ is only 5(b). The gap between (23) and (5b) can be partially removed. If $T(k)$ is only $0(k^{2\ell-2n-1})$, $n=0,1,2,\cdots$ a residual term appears in Eq. (8b); it is due to the poles of the Hankel functions for $k=0$. Nonetheless from Marchenko[1] and Faddeev[3] we know that if conditions A, B, C, D, and E are satisfied there exists a potential $V(x)$ which satisfies (5a) and has $T(k)$ as its T-function. Since $V(x)$ satisfies (5a) a translation kernel $K(x,y)$ which may be constructed by the Riemann method exists; in addition

$$|K(x,y)|\leqslant\tfrac{1}{2}\sigma_0(x)\exp\sigma_1(x)\times(y/x)^\ell.$$

Assume now that a) $V(x)$ exists which satisfies (5b), and b) a fundamental equation exists; the corresponding spectral kernel $F(x,y)$ satisfies inequality (12). The situation is the following: a fundamental equation exists which has a solution $K(x,y)$ generating a potential $V(x)$ satisfying (5a). But is this solution unique? According to the result reported in Sec. 3, if one deals with authentic scattering data the solution of the fundamental equation is unique. Now Faddeev's result[3] reported in the introduction states that any set of scattering data suitable for $\ell=0$ is authentic. Therefore *when* one possesses such data (their characterization being as in Ref. 1 or 3 and *when* a fundamental equation exists, this fundamental equation has necessarily a unique solution.

APPENDIX A:

Marchenko's lemmas on integral equations. The following function spaces are introduced:

$L^1(\epsilon,\infty)$: The function $x(t)$ belongs to $L^1(\epsilon,\infty)$

$$\text{if} \int_\epsilon^\infty |x(t)|\,dt<\infty.$$

$L^2(\epsilon,\infty)$: The function $x(t)$ belongs to $L^2(\epsilon,\infty)$

$$\text{if} \int_\epsilon^\infty x^*(t)x(t)\,dt<\infty.$$

$L^\infty(\epsilon,\infty)$: The function $x(t)$ belongs to $L^\infty(\epsilon,\infty)$ if for $t\in[\epsilon,\infty]$

$$|x(t)|\leqslant M.$$

$\hat{L}(\epsilon,\infty)$: The function $x(t)=x_1(t)+x_2(t)$ belongs to $\hat{L}(\epsilon,\infty)$

if $x_1(t)\in L^1(\epsilon,\infty)$
and $x_2(t)\in L^2(\epsilon,\infty)$ with $x_2(t)\in L^\infty(\epsilon,\infty)$.

Let us recall Marchenko's results concerning kernels of integral equations $L(\epsilon,t)$ which are dominated by an additive term

$$|L(\xi,t)|<A(\xi+t).$$

If $A(t)$ is $L^1(\epsilon,\infty)$ for some $\epsilon\geqslant0$ then, $L(\xi,t)$ is a completely continuous operator in $L^1(\epsilon,\infty)$.

Let $A=A_1+A_2$ with $A_1\in L^1(\epsilon,\infty)$ and $A_2\in L^2(\epsilon,\infty)$ and $A_2\in L^\infty(\epsilon,\infty)$; then any solution $f(t)$ of the equation

$$f(t)+\int_\epsilon^\infty f(\xi)L(\xi,t)\,d\xi=0$$

in $\hat{L}(\epsilon,\infty)$ belongs to both $L^2(\epsilon,\infty)$ and $L^\infty(\epsilon,\infty)$.

Corollary: If A belongs only to $L^1(\epsilon,\infty)$, any solution of the equation

$$f(t)+\int_\epsilon^\infty f(\xi)L(\xi,t)\,d\xi=0$$

in $L^1(\epsilon,\infty)$ belongs to both $L^\infty(\epsilon,\infty)$ and $L^2(\epsilon,\infty)$. Also one can prove that any solution in $L^\infty(\epsilon,\infty)$ belongs to both $L^1(\epsilon,\infty)$ and $L^2(\epsilon,\infty)$.

Suppose now that $A(t)$ belongs to $L^1(-\infty,\infty)$, suppose also that its Fourier transform is continuous over the real axis and vanishes when $|k|\to\infty$; then for any $\epsilon>-\infty$ the operator

$$L[x]=\int_\epsilon^\infty x(\xi)L(\xi,t)\,d\xi$$

is completely continuous in $L^2(\epsilon,\infty)$.

Also if $\epsilon\geqslant0$, and if $A(t)$ belongs to $L^1(\epsilon,\infty)$, the operator $L[x]$ is completely continuous in $L^2(\epsilon,\infty)$.

APPENDIX B: Discussion of Eq. (39c)

We write as in (39c)

$$\pm\phi(t)=\int_\epsilon^\infty \phi(\xi)F_C(\xi,t)\,dt. \tag{B1}$$

$$\pm\mu(t)=\int_\epsilon^\infty \mu(\xi)M_C(\xi,t)\,dt. \tag{B2}$$

Together with (B1) we consider the Marchenko equation where F_C is replaced by M_C and which is denoted (B2). For simplicity only the case $\ell=1$ is fully discussed, then

$$F_C(\xi,t)=\frac{1}{2\pi}\int_{-\infty}^\infty h_1(k\xi)[1-S(k)]h_1(kt)\,dk,$$

$$M_C(\xi,t)=\frac{1}{2\pi}\int_{-\infty}^\infty [1-S(k))\exp ik(\xi+t)\,dk]$$

$$h_1(k\xi)=\frac{1}{k}\xi\frac{d}{d\xi}\frac{1}{\xi}\exp ik\xi.$$

Let us apply the operator $D\equiv(1/t)(d/dt)t$, to both sides of equation (B1). Define $\mu(t)$ by

$$\pm\dot{\phi}(t)\pm\phi(t)/t=\mu(t).$$

The function $\mu(t)$ satisfies Eq. (B3):

$$\pm\mu(t)=\int_\epsilon^\infty \phi(\xi)\,d\xi\frac{1}{2\pi}\times\int_{-\infty}^\infty h_1(k\xi)[1-S(k)]k\exp ikt\,dk. \tag{B3}$$

Notice that when $\mu(t)$ is given, a specific constant $\phi(\epsilon)$ is needed to reconstruct $\phi(t)$:

$$\phi(t) = \frac{1}{t}\int_\epsilon^t s\mu(s)\,ds + \phi(s).$$

We write

$$\phi(t) = \bar{\phi}(t) + \phi(\epsilon).$$

By its definition

$$D\phi(t) = D\bar{\phi}(t).$$

We note also

$$\int_\epsilon^\infty \phi(\xi)d\xi\, h_1(k\xi) = \tilde{\phi}(-k).$$

Equation (B3) is integrated by parts; one gets

$$\pm\mu(t) = \phi(\epsilon)\frac{1}{2\pi}\int_{-\infty}^{\infty}[\exp ik\epsilon][1-S(k)]\exp ikt\,dk + \int_\epsilon^\infty d\xi\frac{1}{2\pi}\int_{-\infty}^{\infty}\mu(\xi)\exp ik\xi\,[1-S(k)]\exp ikt\,dk. \tag{B4}$$

Let us denote

$$\int_\epsilon^\infty \mu(\xi)\exp ik\xi = \tilde{\tilde{\phi}}(-k),$$

and assume $\mu(t) = 0$ for $t<\epsilon$, nothing being said for $t=\epsilon$ of $\mu(\epsilon)$.

Equation (B4) reads

$$\pm\mu(t) = \phi(\epsilon)\frac{1}{2\pi}\int_{-\infty}^{\infty}\exp ik\epsilon[1-S(k)]\exp ikt\,dk - \frac{1}{2\pi}\int_{-\infty}^{\infty}\tilde{\tilde{\phi}}(-k)S(k)\exp ikt\,dk. \tag{B5}$$

We obtain at $t=\epsilon$

$$\pm\mu(\epsilon) = \phi(\epsilon)M_0(\epsilon,\epsilon) + \frac{1}{2\pi}\int_{-\infty}^{\infty}\tilde{\tilde{\phi}}(-k)S(k)\exp ik\epsilon\,dk,$$

or

$$\phi(\epsilon) = \left[\pm\mu(\epsilon) - \frac{1}{2\pi}\int_{-\infty}^{\infty}\tilde{\tilde{\phi}}(-k)S(k)\exp ik\,dk\right]\times[M(\epsilon,\epsilon)]^{-1}. \tag{B6}$$

If $\mu(t)$ is a solution of the Marchenko equation

$$\pm\mu(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\mu(\xi)\,d\xi \times\int_{-\infty}^{\infty}\exp ik\xi\,[1-S(k)]\exp ikt\,dk$$

the bracket in Eq. (B6) vanishes.

A Fourier transform of Eq. (B4) yields

$$\pm\tilde{\tilde{\phi}}(k) = \phi(\epsilon)\exp ik\epsilon[1-S(k)] - \tilde{\tilde{\phi}}(-k)S(k), \tag{B7}$$

$$\pm\tilde{\tilde{\phi}}(-k) = \phi(\epsilon)\exp - ik\epsilon[1-S(k)] - \tilde{\tilde{\phi}}(k)S(-k).$$

Since $S(k)$ is factorizable

$$S(k) = f_1(-k)/f_1(k),$$

the two equations (B7) read

$$[\pm\tilde{\tilde{\phi}}(k)f_1(k) + \tilde{\tilde{\phi}}(-k)f_1(-k)][f_1(k)-f_1(-k)]^{-1} = \phi(\epsilon)\exp ik\epsilon, \tag{B8}$$

$$[\pm\tilde{\tilde{\phi}}(-k)f_1(-k) + \tilde{\tilde{\phi}}(k)f_1(k)][f_1(k)-f_1(-k)]^{-1} = \phi(\epsilon)\exp(-ik\epsilon). \tag{B9}$$

In (B8) and (B9), k is real. A condition for the compatibility of (B8) and (B9) is

either $\phi(\epsilon)\exp ik\epsilon + \phi(\epsilon)\exp(-ik\epsilon) = 0$

or $S(k) = S(-k) = \pm 1.$

The first equality happens only if $\phi(\epsilon)$ vanishes. Returning to (B4) and(B6) one sees that $\phi(t)$ is the solution of the Marchenko associated equation. For the second condition if one has $S(k) = +1$, one considers a transport potential V which does not possess any bound state since the Levinson formula is assumed to be valid; therefore

$$V(x)\equiv 0.$$

If $S(k) = -1$, $M_C(\xi+t)$ did not exist since then we would have

$$M_C(\xi+t) \equiv \frac{2}{2\pi}\int_{-\infty}^{\infty}\exp ik(\xi+t)\,d\xi,$$

which is not bounded. The second condition does not need to be considered.

Proposition: In order for a function $\phi(t)$ to be solution of the $\ell=1$ equation (B1) it must be related to a solution $\mu(t)$ of the $\ell=0$ associated Marchenko equation (B2) by

$$\phi(t) = \frac{1}{t}\int_\epsilon^t s\mu(s)\,ds.$$

Extension to higher ℓ requires simply a recursion method. The result is first proved for $\ell = m-1$, the operators

$$\frac{1}{\xi^m}\frac{d}{d\xi}\xi^m,$$

$$\xi^m\frac{d}{d\xi}\frac{1}{\xi^m}$$

are introduced, and Fourier transforms are replaced by Hankel transforms of order $m-1$. Using arguments similar to the one we used in Appendix A, the results extend to $\ell = m$.

[1] L. S. Agranovitch and V. A. Marchenko, *The Inverse Problem of Scattering Theory* (Gordon and Breach, New York, 1963).
[2] M. M. Crum, Quart. J. Math. **6** (2), 121 (1955).
[3] L. D. Faddeev, J. Math. Phys. **4**, 72 (1963).
[4] M. Coz and P. Rochus, J. Math. Phys. **17**, 894 (1976); **18**, 2223 (1977); **18**, 2232 (1977).
[5] M. Coz and P. Rochus, Ann. Phys. **126**, 460 (1980).
[6] M. A. Naimark, *Ordinary Differential Operators*, Vol. 2, edited by F. Hungar W. T. (1968); M. A. Naimark, Am. Math. Soc. Transl. **16**, 103 (1960); V. E. Ljance, USSR Math. Sbornik, **72**, 114 (1967); Am. Mat. Transl. **60**, 485 (1967).

[7] The equality $S(k)S(-k) = 1$ follows the factorization $S(k) = f_1(-k)/f_1(k)$.

[8] N. Levinson, Mat.-Fys. Medd. Danske Vid. Selsk. (Copenhagen), **25**, 9 (1963).
[9] F. D. Gakhov, *Boundary Value Problems* (Addison-Wesley, Reading, Mass., 1966).

Chain of the Bäcklund transformation for the KdV equation

Akira Nakamura
Department of Mathematics, Clarkson College of Technology, Potsdam, New York 13676

(Received 18 November 1980; accepted for publication 13 February 1981)

We study the chain of the Bäcklund transformation ($\equiv$BT) by the example of the KdV equation. The previously obtained chain of the BT, KdV$\rightarrow$modified KdV ($\equiv$mKdV)$\rightarrow$second mKdV, has been extended one step further to the third mKdV case. From this lowest order example, the structure of "infinity" of the chain process has been foreseen.

PACS numbers: 02.30.Jr

I. INTRODUCTION

In the study of solitons, Bäcklund transformation (hereafter abbreviated as BT) plays a very important role in the analytical investigation. The BT enables us to obtain the so-called multiple soliton solution of a given nonlinear evolution equation. At the same time the BT produces a new nonlinear evolution equation which usually again has soliton solutions. In the previous work,[1] we have shown, by explicit example of the KdV equation, that this process actually goes to the higher order as KdV$\rightarrow$modified KdV (abbreviated hereafter as mKdV)$\rightarrow$second mKdV, where the arrows denote the BT. In this paper, we show that this process can be extended one step further as KdV$\rightarrow$mKdV$\rightarrow$second mKdV$\rightarrow$third mKdV.

Significant here is the fact that, from the present study, very naturally appears the possibility that this process actually continues "infinitely." In fact, this lowest-order process of "chain" of the BT sufficiently clearly reveals that they have actually a very simple and regular structure.

In this paper, we explicitly calculate the BT of the second mKdV equation and derive the third mKdV equation. In the last section, we give certain analysis on the structure of the general BT connecting the "nth" and "$(n+1)$th" mKdV equations.

As has been done previously, we calculate the BT using the language of Hirota's bilinear formalism. One definite advantage of this method is that if we once transform from the original variable to the bilinear variable, BT becomes nothing but an almost trivial simple exchange between bilinear variables.[2,3] This simplicity in the transformed variable has already been shown by many examples of the physically interesting nonlinear evolution equations such as KdV, sine–Gordon equations,[4] the integrodifferential type equation of Benjamin and Ono,[5] and the cylindrical KdV equation.[6] We believe that the present work adds another interesting example to these.

II. BASIC SCHEME OF THE BILINEAR BT

Before going into specific problems of the chain of the BT in the next section, it is worthwhile to consider bilinear BT theory in general. In 1974, Hirota first proposed the idea of the bilinear BT together with the explicit method of derivation. Although he has discovered the most important prescription of how to perform the calculation fully explicitly, the underlying reasons why certain procedures are taken have not been explained much in detail. Thus we give here an explicit explanation of Hirota's prescription, especially about his starting expression [which is Eq. (2.8) of this section].

For this purpose, we take the simplest example of the KdV equation

$$u_t + 6uu_x + u_{xxx} = 0. \tag{2.1}$$

Throughout this paper, subscripts x and t denote partial differentiation. We consider the different solution u' of the same equation as

$$u'_t + 6u'u'_x + u'_{xxx} = 0. \tag{2.2}$$

The BT is the relation relating one solution u to another solution u' (which may be called as relating the old solution u to the new solution u'). Namely, the BT is the relation which connects Eq. (2.1) to Eq. (2.2). We find that the simplest and the most natural way of combining Eqs. (2.1) and (2.2) is direct subtraction,

$$(u_t + 6uu_x + u_{xxx}) - (u'_t + 6u'u'_x + u'_{xxx}) = 0. \tag{2.3}$$

Clearly, if Eq. (2.3) holds and u satisfies Eq. (2.1), then u' satisfies Eq. (2.2). Thus Eq. (2.3) is the most primitive form of BT.

From here, we consider, the dependent variable transformation

$$u = (2\log f)_{xx}, \quad u' = (2\log g)_{xx}. \tag{2.4}$$

By the direct insertion of Eq. (2.4) into (2.3), we have

$$\left\{\frac{F(D_x,D_t)f\cdot f}{f\cdot f} - \frac{F(D_x,D_t)g\cdot g}{g\cdot g}\right\}_x = 0, \tag{2.5}$$

where

$$F(D_x,D_t) \equiv D_xD_t + D_x^4, \tag{2.6}$$

and bilinear differential operators D_x and D_t are defined by[2,3]

$$D_x^mD_t^n a(x,t)\cdot b(x,t)$$
$$\equiv (\partial_x - \partial_{x'})^m(\partial_t - \partial_{t'})^n a(x,t)b(x',t')|_{\substack{x'=x\\ t'=t}}, \tag{2.7}$$

for arbitrary functions $a(x,t)$ and $b(x,t)$. By integrating Eq. (2.5) once with x, taking the integration constant to be zero, and multiplying by f^2g^2, we have

$$P \equiv \{F(D_x,D_t)f\cdot f\}gg - ff\{F(D_x,D_t)g\cdot g\} = 0. \tag{2.8}$$

This is the primitive form of BT in transformed variables f and g.

Now how should we deal with Eq. (2.8)? In the theory of

 0022-2488/81/081608-06$01.00 © 1981 American Institute of Physics

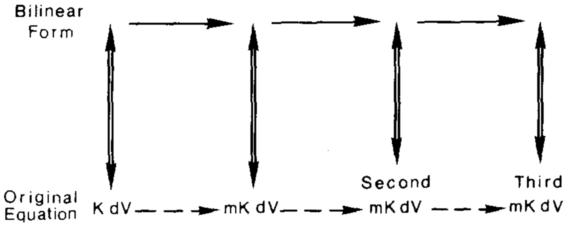


FIG. 1. Schematic picture of the chain of the Bäcklund transform and bilinear formalism for the KdV equation. Symbols $\Leftrightarrow, \rightarrow, -\,-\,\rightarrow$ represent the dependent variable transform, the BT in bilinear form, and the BT in original variables respectively.

bilinear formalism, every quantity is to be reduced or decomposed into bilinear form. This comes from the implicit but basic postulate of the bilinear theory that the most fundamental nonlinearity is the lowest order nonlinearity which is second order (because first order is by definition linear) or bilinear. We notice Eq. (2.8) is fourth order nonlinear in variables f and g. Thus we try to decompose Eq. (2.8) into (coupled) bilinear relations between old and new solutions, i.e., between f and g. For the purpose, in each term of Eq. (2.8) we try to interchange one f with one g under appropriate rules of D_x and D_t operators corresponding to the given form of $F(D_x, D_t)$. As the results, we have several decoupled bilinear equations instead of Eq. (2.8), which is nothing but the BT in bilinear variables. Above is the explanation of the basic strategy of the bilinear BT method.

III. SECOND mKdV EQUATION AND ITS BILINEAR BT

In this section, we apply bilinear BT technique explained just above to the second mKdV equation. In the previous paper, we showed that the BT of the mKdV equation generates the second mKdV equation.[1] For simplicity, we consider a simple case of the second mKdV equation with only one arbitrary constant,

$$u_t + u_{xxx} + \tfrac{1}{2}(u_x)^3 + 6\alpha^2 u_x \sin^2 u = 0, \tag{3.1}$$

where α is an arbitrary constant. When $\alpha = 0$, we consider u as a "potential" function of the variable v defined by $v = -u_x$. Then by differentiating Eq. (3.1) once with x, we have the modified KdV equation[7]

$$v_t + v_{xxx} + \tfrac{3}{2} v^2 v_x = 0. \tag{3.2}$$

Thus the second mKdV equation, Eq. (3.1), is not only the equation connected to the mKdV equation by BT, but is actually a generalization of the mKdV equation with the additional parameter α, which includes the mKdV equation as its special case $\alpha = 0$.

Now we transfer from the original variable u to the bilinear variable. By the dependent variable transformation

$$u = i \log(f_1' f_2' / f_1 f_2), \tag{3.3}$$

Eq. (3.1) can be transformed to the coupled bilinear form

$$(D_t + D_x^3) f_i' \cdot f_i = 0 \quad (i = 1,2), \tag{3.4a}$$

$$D_x^2 f_i' \cdot f_i = 0 \quad (i = 1,2), \tag{3.4b}$$

$$(D_x + \alpha) f_1 \cdot f_2 - \alpha f_1' f_2' = 0, \tag{3.4c}$$

$$(D_x + \alpha) f_1' \cdot f_2' - \alpha f_1 f_2 = 0. \tag{3.4d}$$

Note that for $\alpha = 0$ Eqs. (3.4c,d) become $D_x f_1 \cdot f_2 = D_x f_1' \cdot f_2' = 0$, which is equivalent to $f_2 \propto f_1$, $f_2' \propto f_1'$. Then Eqs. (3.4a,b) reduce to only the $i = 1$ case ($i = 2$ being redundant) and Eq. (3.3) to $u = 2i \log(f_1'/f_1)$, which precisely agrees with the bilinear form of the mKdV equation given by Hirota,[2,3] as it should.

We consider another solution of Eq. (3.1) and the same dependent variable transformation as

$$u' = i \log(g_1' g_2' / g_1 g_2), \tag{3.3'}$$

which leads to the same bilinear equations for g_1, g_2, g_1', g_2' as

$$(D_t + D_x^3) g_i' \cdot g_i = 0 \quad (i = 1,2), \tag{3.4a'}$$

$$D_x^2 g_i' \cdot g_i = 0 \quad (i = 1,2), \tag{3.4b'}$$

$$(D_x + \alpha) g_1 \cdot g_2 - \alpha g_1' g_2' = 0, \tag{3.4c'}$$

$$(D_x + \alpha) g_1' \cdot g_2' - \alpha g_1 g_2 = 0. \tag{3.4d'}$$

The BT, which connects two solutions $u \longleftrightarrow u'$, becomes, in the language of the bilinear formalism, the relation between transformed bilinear variables, $f_1 f_2, f_1', f_2' \longleftrightarrow g_1, g_2, g_1', g_2'$.

Having equations rewritten in bilinear form, we follow the standard procedure of Hirota explained in the previous section.

Following the prescription given in detail in Appendix B, we obtain the explicit form of the BT as

$$(D_t + D_x^3 + \tfrac{3}{4} k^2 D_x) f_i^{(\prime)} \cdot g_i^{(\prime)} = 0 \quad (i = 1,2), \tag{3.5a}$$

$$(D_x^2 - k^2/4) f_i^{(\prime)} \cdot g_i^{(\prime)} = 0 \quad (i = 1,2), \tag{3.5b}$$

$$D_x f_i' \cdot g_i = \tfrac{1}{2}\sqrt{-1}\, k f_i g_i' (-1)^i \quad (i = 1,2), \tag{3.5c}$$

$$D_x f_i \cdot g_i' = \tfrac{1}{2}\sqrt{-1}\, k f_i' g_i (-1)^{i+1} \quad (i = 1,2), \tag{3.5d}$$

$$\tfrac{1}{2} ik f_2 g_1 - \alpha f_2' g_1' = \gamma f_1 g_2, \tag{3.5e}$$

$$-\tfrac{1}{2} ik f_2' g_1' - \alpha f_2 g_1 = \gamma f_1' g_2', \tag{3.5f}$$

with γ being an arbitrary constant. Here and in the following, our notation $\cdots f^{(\prime)} \cdot g^{(\prime)} = 0$ denotes two independent equations $\cdots f \cdot g = 0$ and $\cdots f' \cdot g' = 0$. The structure of the present BT is very regular in the following sense. We notice that the form of Eqs. (3.5a,b) is precisely the same as the BT of the KdV equations,[2,3] and Eqs. (3.5c,d) the same as the space part BT of the mKdV,[2,4] both repeated twice for subscripts $i = 1,2$. On the other hand, Eqs. (3.5e,f) are the first example of BT which contains *no derivative* at all. We can summarize this as follows. As we go to higher order in the BT chain, KdV→mKdV→second mKdV→, the D-operator functional form of time part BT *does not change at all*, while, in the space part BT, each time, we have the addition of equations containing derivatives of decreasing order *one by one* as $D_x^2 \rightarrow D_x^1 \rightarrow D_x^0 = 1$. Thus, in the language of the bilin-

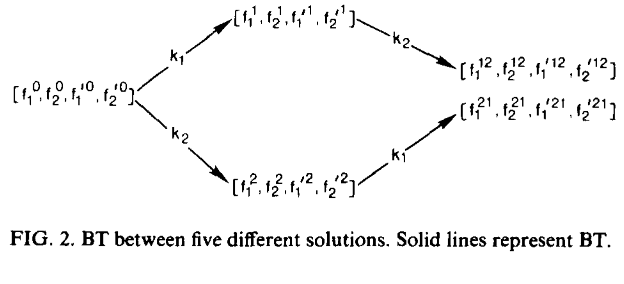


FIG. 2. BT between five different solutions. Solid lines represent BT.

ear formalism, we see the BT chain has a very simple and regular structure.

We can check that the present BT (3.5) actually generates arbitrary N-soliton solutions as shown in the following. The generation of one-soliton solutions is seen as follows. For $f_1 = f_2 = f'_1 = f'_2 = 1$ (which corresponds to the trivial vacuum solution $u = 0$) Eqs. (3.5) become pure linear equations (with each D_x, D_t replaced respectively by $-\partial_x, -\partial_t$) whose solution can be easily obtained as

$$\begin{aligned}
g_1 &= \exp[\tfrac{1}{2}(\theta + \phi + \phi')] + \exp[-\tfrac{1}{2}(\theta + \phi + \phi')],\\
g_2 &= \exp[\tfrac{1}{2}(\theta - \phi - \phi')] + \exp[-\tfrac{1}{2}(\theta - \phi - \phi')],\\
g'_1 &= \exp[\tfrac{1}{2}(\theta - \phi + \phi')] + \exp[-\tfrac{1}{2}(\theta - \phi + \phi')],\\
g'_2 &= \exp[\tfrac{1}{2}(\theta + \phi - \phi')] + \exp[-\tfrac{1}{2}(\theta + \phi - \phi')],\\
\theta &\equiv kx + \omega t + \theta_0, \quad \omega \equiv -k^3, \qquad (3.6)\\
e^{\phi} &\equiv i, \quad e^{2\phi'} \equiv (-k + 2\alpha)/(k + 2\alpha), \qquad (3.7)
\end{aligned}$$

with the proper choice of $\gamma = ie^{\phi'}(k/2 + \alpha) = [(ik/2)^2 + \alpha^2]^{1/2}$. This corresponds to the one-soliton solution

$$\begin{aligned}
u' &= i \log\left(\frac{g'_1 g'_2}{g_1 g_2}\right)\\
&= i \log\left[\frac{\cosh\frac{1}{2}(\theta - \phi + \phi')\cosh\frac{1}{2}(\theta + \phi - \phi')}{\cosh\frac{1}{2}(\theta + \phi + \phi')\cosh\frac{1}{2}(\theta - \phi - \phi')}\right]. \qquad (3.8)
\end{aligned}$$

Next, we consider multisoliton solutons. We denote the BT (3.5) by an arrow $(f_1 f_2 f'_1 f'_2) \xrightarrow{k} (g_1 g_2 g'_1 g'_2)$. Then we consider the four BT as depicted in Fig. 2. If we assume commutability of BT, i.e., $f_i^{(\prime)12} = f_i^{(\prime)21}$, using relation (3.5b), we can obtain the superposition formula[4]

$$f_i^{(\prime)0} f_i^{(\prime)12} \propto D_x f_i^{(\prime)1} \cdot f_i^{(\prime)2} \quad (i = 1,2). \qquad (3.9)$$

Next, we verify that this superposition formula holds even without assuming $f_i^{(\prime)12} = f_i^{(\prime)21}$. For that purpose, we define $\tilde{f}_i^{(\prime)12}$ by

$$\tilde{f}_i^{(\prime)12} \equiv (1/f_i^{(\prime)0}) D_x f_i^{(\prime)1} \cdot f_i^{(\prime)2} \quad (i = 1,2) \qquad (3.10)$$

and assume the BT relation depicted by solid line in Fig. 3. We can prove without the *a priori* assumption of $f_i^{(\prime)12} = f_i^{(\prime)21}$, that broken lines in Fig. 3 are actually the BT relation defined by Eqs. (3.5), and therefore the function $\tilde{f}_i^{(\prime)12}$ constructed from three old solutions $f_i^{(\prime)0}, f_i^{(\prime)1}, f_i^{(\prime)2}$ by (3.10) is actually a new solution. We leave the details of the proof to Appendix C. Superposition formula (3.10) is known to generate higher order multisolitons from lower order solitons.[4]

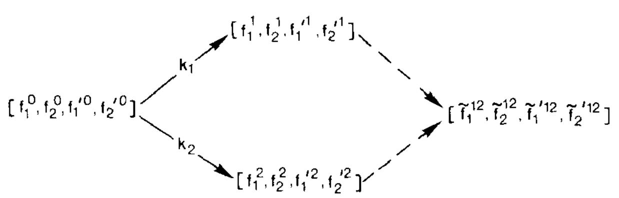


FIG. 3. Solid lines represent assumed BT and broken lines represent BT to be proved.

IV. BT IN ORIGINAL VARIABLE AND THE THIRD mKdV EQUATION

In the previous sections, we have obtained the BT in bilinear variables. Now we transform the results back to the original variables $u + u'$ and $u - u'$. For the purpose, it is convenient to introduce the following symbols:

$$\begin{aligned}
&\Phi_i \equiv \log(f_i/g_i), \quad \Phi'_i \equiv \log(f'_i/g'_i),\\
&\rho_i \equiv \log f_i g_i, \quad \rho'_i \equiv \log f'_i g'_i \quad (i = 1,2).
\end{aligned} \qquad (4.1)$$

From Eqs. (3.3) and (3.3′), we have

$$\begin{aligned}
i(u - u') &= \Phi_1 + \Phi_2 - \Phi'_1 - \Phi'_2, \qquad (4.2)\\
i(u + u') &= \rho_1 + \rho_2 - \rho'_1 - \rho'_2.
\end{aligned}$$

By these variables $\Phi_i^{(\prime)}, \rho_i^{(\prime)}$, Eqs. (3.5) can be rewritten as

$$\Phi_{it}^{(\prime)} + \Phi_{ixxx}^{(\prime)} + 3\Phi_{ix}^{(\prime)}\rho_{ixx}^{(\prime)} + \Phi_{ix}^{(\prime)3} + \tfrac{3}{4}k^2\Phi_{ix}^{(\prime)} = 0 \quad (i = 1,2), \qquad (4.3a)$$

$$\rho_{ixx}^{(\prime)} + \Phi_{ix}^{(\prime)2} - k^2/4 = 0 \quad (i = 1,2), \qquad (4.3b)$$

$$\tfrac{1}{2}(\Phi'_i + \rho'_i + \Phi_i - \rho_i)_x = (-1)^i \tfrac{1}{2}\sqrt{-1}k \exp(\Phi_i - \Phi'_i) \quad (i = 1,2), \qquad (4.3c)$$

$$\tfrac{1}{2}(\Phi_i + \rho_i + \Phi'_i - \rho'_i)_x = (-1)^{i+1} \tfrac{1}{2}\sqrt{-1}k \exp(\Phi'_i - \Phi_i) \quad (i = 1,2), \qquad (4.3d)$$

$$\begin{aligned}
&\tfrac{1}{2}ik - \alpha \exp[\tfrac{1}{2}(\Phi_1 - \Phi_2 - \Phi'_1 + \Phi'_2) - \tfrac{1}{2}i(u + u')]\\
&\quad = \gamma \exp(\Phi_1 - \Phi_2), \qquad (4.3e)
\end{aligned}$$

$$\begin{aligned}
&-\tfrac{1}{2}ik - \alpha \exp[-\tfrac{1}{2}(\Phi_1 - \Phi_2 - \Phi'_1 + \Phi'_2) + \tfrac{1}{2}i(u + u')]\\
&\quad = \gamma \exp(\Phi'_1 - \Phi'_2). \qquad (4.3f)
\end{aligned}$$

Subtracting Eq. (4.3d) from (4.3c) and adding for $i = 1$ and 2, we have

$$\begin{aligned}
(u + u')_x &= i(\rho'_1 + \rho'_2 - \rho_1 - \rho_2)_x\\
&= ik\sqrt{-1}[\cosh(\Phi_2 - \Phi'_2) - \cosh(\Phi_1 - \Phi'_1)]\\
&= 2ik\sqrt{-1}\sinh\tfrac{1}{2}(\Phi_1 + \Phi_2 - \Phi'_1 - \Phi'_2)\\
&\quad \times \sinh\tfrac{1}{2}(\Phi_2 - \Phi_1 - \Phi'_2 + \Phi'_1)\\
&= -2k\sqrt{-1}\sin\tfrac{1}{2}(u - u')\\
&\quad \times \sinh\tfrac{1}{2}(\Phi_2 - \Phi_1 - \Phi'_2 + \Phi'_1). \qquad (4.4)
\end{aligned}$$

Dividing Eq. (4.3e) by Eq. (4.3f), we have

$$k\cosh\tfrac{1}{2}(\Phi_1 - \Phi_2 - \Phi'_1 + \Phi'_2) = -2\alpha\sin[(u + u')/2]. \qquad (4.5)$$

Eliminating $\Phi_1 - \Phi_2 - \Phi'_1 + \Phi'_2$ from Eqs. (4.4) and (4.5), we have

$$\begin{aligned}
&(u + u')_x\left[1 - \left(\frac{2\alpha}{k}\right)^2 \sin^2\left(\frac{u + u'}{2}\right)\right]^{-1/2}\\
&\quad = -2k\sin\left(\frac{u - u'}{2}\right). \qquad (4.6)
\end{aligned}$$

This is the space part of the BT in the original variables. When $\alpha = 0$, Eq. (4.6) reduces to the BT of the mKdV equation given by Wadati.[8] As shown in Appendix D, the time part of the BT can be written as

$$(u-u')_t + (u-u')_{xxx} + \tfrac{3}{2}k^2(u-u')_x - 2N = 0,$$

$$\begin{aligned} N \equiv \tfrac{3}{16}(u-u')_x &\left\{(u+u')_x \cot\left(\frac{u-u'}{2}\right)\right\}^2 \\ &+ 3\alpha^2(u-u')_x \sin^2\left(\frac{u-u'}{2}\right)\sin^2\left(\frac{u+u'}{2}\right) \\ &- 6\alpha^2(u+u')_x \cos\left(\frac{u+u'}{2}\right)\cos\left(\frac{u-u'}{2}\right) \\ &\times \sin\left(\frac{u+u'}{2}\right)\sin\left(\frac{u-u'}{2}\right) \\ &- 3\alpha^2(u-u')_x \cos^2\left(\frac{u+u'}{2}\right)\sin^2\left(\frac{u-u'}{2}\right) \\ &- \frac{1}{16}[(u-u')_x]^3. \end{aligned} \tag{4.7}$$

Equations (4.6) and (4.7) are the BT of the second mKdV in the original variables. If we consider it as a coupled equation for $(u+u')/2 \equiv V$ and $(u-u')/2 \equiv U$, then the equation for U is nothing but the "third mKdV" equation. The third mKdV equation can be written in single equation form by eliminating $(u+u')/2\ (=V)$ from Eqs. (4.6) and (4.7). This can be done by solving Eq. (4.6) as

$$\sin\tfrac{1}{2}(u+u') = \mathrm{sn}\left(-k\int_0^x dx\,\sin\left(\frac{u-u'}{2}\right); \frac{2\alpha}{k}\right) \tag{4.8}$$

and putting this into every place for $\frac{1}{2}(u+u')$ in Eq. (4.7):

$$\begin{aligned} U_t &+ U_{xxx} + \tfrac{1}{2}(U_x)^3 + (\tfrac{3}{2}k^2 + 6\alpha^2)U_x\sin^2 U \\ &- 6\alpha^2 U_x(3\sin^2 U - 1)\mathrm{sn}^2\left(-k\int_0^x dx\,\sin U; \frac{2\alpha}{k}\right) \\ &+ 6\alpha^2\left\{\mathrm{sn}^2\left(-k\int_0^x dx\,\sin U; \frac{2\alpha}{k}\right)\right\}_x \sin U\cos U = 0. \end{aligned} \tag{4.9}$$

V. CHARACTERISTICS OF THE CHAIN OF THE BT IN BILINEAR VARIABLES

We have seen that the BT has a very regular structure when written in bilinear variables. From the lowest order examples of the KdV BT series, one can foresee the following features as a natural extention. In the bilinear form, in each step of the BT we have the addition of the space part with decreasing order of x derivative. Since we have reached the lowest order of x derivative $D_x^2 \to D_x^1 \to 1$ at the BT of the second mKdV, from here on we expect no addition of new D_x-operator functional forms, but simply an exchange of bilinear variables under the same D_x operators. We expect that the nth mKdV equation [which is essentially the same as the BT of the $(n-1)$th mKdV equation] can be reduced to bilinear form by the dependent variable transform

$$u = \sqrt{-1}\log\frac{f_1'\cdots f_N'}{f_1\cdots f_N}, \quad N \equiv 2^{n-1} \quad (n \geqslant 1) \tag{5.1}$$

and has the one-soliton solution of the form

$$f_i^{(\prime)} = \cosh(\theta \pm \phi_1 \pm \cdots \pm \phi_n), \tag{5.2}$$

where total 2^n different $f_i^{(\prime)}$'s (for $i = 1,\ldots,2^{n-1}$ and with and without the prime) correspond to 2^n different combinations of plus-or-minus signs $\pm\phi_i$ $(i = 1,\ldots,n)$. N-soliton solutions follow the KdV case with the formal replacement of $\theta \to \theta \pm \phi_1 \pm \cdots \pm \phi_n$.

ACKNOWLEDGMENTS

I thank Dr. Ryogo Hirota for discussions, including the original notion of the "consecutive chain" of the BT. I also enjoyed discussions with Dr. Junkichi Satsuma and Dr. Mark Ablowitz. This work is supported by the Air Force Office of Scientific Research #78-3674B and the Office of Naval Research #N00014-76-C-0867.

APPENDIX A: IDENTITY PROPERTY OF THE BILINEAR OPERATOR

The following identities are used in the derivation of the BT:

$$D_x c_0 a(x)\cdot a(x) = D_x a(x)\cdot c_0 a(x) = 0 \quad (c_0 = \text{const.}), \tag{A1}$$

$$\begin{aligned}(D_x a\cdot b)cd - ab(D_x c\cdot d) &= (D_x a\cdot c)bd - ac(D_x b\cdot d) \\ &= D_x(ad)\cdot(bc),\end{aligned} \tag{A2}$$

$$\begin{aligned}(D_x^2 a\cdot b)cd - ab(D_x^2 c\cdot d) &= (D_x^2 a\cdot c)bd - ac(D_x^2 b\cdot d) \\ &- 2D_x ad\cdot(D_x b\cdot c),\end{aligned} \tag{A3}$$

$$\begin{aligned}(D_x^3 a\cdot b)cd - ab(D_x^3 c\cdot d) &= (D_x^3 a\cdot c)bd - ac(D_x^3 b\cdot d) \\ &- 3D_x(D_x a\cdot d)\cdot(D_x b\cdot c).\end{aligned} \tag{A4}$$

APPENDIX B: BT

From Eqs. (3.4a),(3.4a'),(A2),(A4), we have

$$\begin{aligned} P_1 &\equiv \{(D_t + D_x^3)f_i'\cdot f_i\}g_i'g_i - f_i'f_i\{(D_t + D_x^3)g_i'\cdot g_i\} \\ &= \{(D_t + D_x^3 + \tfrac{3}{4}k^2 D_x)f_i'\cdot g_i'\}f_i g_i \\ &\quad - f_i'g_i'\{(D_t + D_x^3 + \tfrac{3}{4}k^2 D_x)f_i\cdot g_i\} \\ &\quad - 3D_x(D_x f_i'\cdot g_i)\cdot(D_x f_i\cdot g_i') - \tfrac{3}{4}k^2 D_x f_i'g_i\cdot f_i g_i'. \end{aligned}$$

From the BT equations (3.5a–d), we see that $P_1 = 0$. Similarly, from Eqs. (3.4b),(3.4b'),(A3), we have

$$\begin{aligned} P_2 &\equiv (D_x^2 f_i'\cdot f_i)g_i'g_i - f_i'f_i(D_x^2 g_i'\cdot g_i) \\ &= \{(D_x^2 - k^2/4)f_i'\cdot g_i'\}f_i g_i - f_i'g_i'\{(D_x^2 - k^2/4)f_i\cdot g_i\} \\ &\quad - 2D_x(f_i'g_i)\cdot(D_x f_i\cdot g_i'), \end{aligned}$$

which also vanishes provided the BT equations (3.5b–d) are satisfied. From Eqs. (3.4c),(3.4d'),(A2), we have

$$\begin{aligned} P_3 &\equiv \{(D_x + \alpha)f_1\cdot f_2 - \alpha f_1'f_2'\}g_1'g_2' - f_1 f_2\{(D_x + \alpha)g_1'\cdot g_2' - \alpha g_1 g_2\} \\ &= (D_x f_1\cdot g_1')f_2 g_2' - f_1 g_1'(D_x f_2\cdot g_2') - \alpha f_1'f_2'g_1'g_2' + \alpha f_1 f_2 g_1 g_2 \\ &= \tfrac{1}{2}ik f_1' g_1 f_2 g_2' + \tfrac{1}{2}ik f_1 g_1' f_2' g_2 - \alpha f_1' f_2' g_1' g_2' + \alpha f_1 f_2 g_1 g_2 \end{aligned}$$

$$=f_1'g_2'(\tfrac{1}{2}ikf_2g_1-\alpha f_2'g_1')+f_1g_2(\tfrac{1}{2}ikg_1'f_2'+\alpha f_2g_1),$$

which is seen to vanish provided the BT equations (3.5e,f) hold. In a similar way, we see

$$P_4\equiv\{(D_x+\alpha)f_1'f_2'-\alpha f_1f_2\}g_1g_2-f_1'f_2'\{(D_x+\alpha)g_1\cdot g_2-\alpha g_1'g_2'\}$$
$$=\cdots=-P_3=0.$$

APPENDIX C

We prove that in Fig. 3 if two solid line BT are satisfied, then a broken line is actually a BT. We check the upper broken line in Fig. 3. As is mentioned after Eq. (3.5), since the essential new part in the present BT is (3.5e,f), we confine our attention to this part of the BT only. Proof for (3.5a,b) can be obtained in a manner similar to the recent work on cylindrical KdV BT[6] and for (3.5c,d) to the work on the Benjamin–Ono BT.[5]

We will prove that relations (3.5e) hold for variables $f_i^{(\prime)1}\longleftrightarrow\tilde{f}_i^{(\prime)12}$. Namely Q, defined by

$$Q\equiv\tfrac{1}{2}ik_2f_2^1\tilde{f}_1^{12}-\alpha f_2'^1\tilde{f}_1'^{12}-\gamma_2f_1^1\tilde{f}_2^{12},\tag{C1}$$

vanishes provided the following are satisfied:

$$D_xf_i'^0\cdot f_i^P=(-1)^i\tfrac{1}{2}\sqrt{-1}k_Pf_i^0f_i'^P\quad(i=1,2,\,P=1,2),\tag{C2a}$$

$$D_xf_i^0\cdot f_i'^P=(-1)^{i+1}\tfrac{1}{2}\sqrt{-1}k_Pf_i'^0f_i^P\quad(i=1,2,\,P=1,2),\tag{C2b}$$

$$\tfrac{1}{2}ik_Pf_2^0f_1^P-\alpha f_2'^0f_1'^P=\gamma_Pf_1^0f_2^P\quad(i=1,2,\,P=1,2),\tag{C2c}$$

$$-\tfrac{1}{2}ik_Pf_2'^0f_1'^P-\alpha f_2^0f_1^P=\gamma_Pf_1'^0f_2'^P\quad(i=1,2,\,P=1,2),\tag{C2d}$$

$$f_i^{(\prime)0}\tilde{f}_i^{(\prime)12}=D_xf_i^{(\prime)1}\cdot f_i^{(\prime)2}\quad(i=1,2).\tag{C2e}$$

From (C2e) we have

$$\begin{aligned}f_1^0f_2^0f_1'^0f_2'^0Q&=\tfrac{1}{2}ik_2f_2^0f_1'^0f_2'^0f_2^1(D_xf_1^1\cdot f_1^2)-\alpha f_1^0f_2^0f_2'^0f_2'^1(D_xf_1'^1\cdot f_1'^2)-\gamma_2f_1^0f_1'^0f_2'^0f_1^1(D_xf_2^1\cdot f_2^2),\\&=\tfrac{1}{2}ik_2[-f_2^0f_2'^0f_2^1f_1^2(D_xf_1'^0\cdot f_1^1)+f_2^0f_2'^0f_2^1f_1^1(D_xf_1'^0\cdot f_1^2)]\\&\quad-\alpha[-f_2^0f_2'^0f_2'^1f_1'^2(D_xf_1^0\cdot f_1'^1)+f_2^0f_2'^0f_2'^1f_1'^1(D_xf_1^0\cdot f_1'^2)]\\&\quad-\gamma_2[-f_1^0f_1'^0f_1^1f_2^2(D_xf_2'^0\cdot f_2^1)+f_1^0f_1'^0f_1^1f_2^1(D_xf_2'^0\cdot f_2^2)],\end{aligned}$$

which becomes by (C2a,b), and then by (C2c,d)

$$\begin{aligned}&=\tfrac{1}{2}ik_2(+\tfrac{1}{2}ik_1f_2^0f_2'^0f_2^1f_1^2f_1^0f_1'^1-\tfrac{1}{2}ik_2f_2^0f_2'^0f_2^1f_1^1f_1^0f_1'^2)\\&\quad-\alpha(-\tfrac{1}{2}ik_1f_2^0f_2'^0f_2'^1f_1'^2f_1'^0f_1^1+\tfrac{1}{2}ik_2f_2^0f_2'^0f_2'^1f_1'^1f_1'^0f_1^2)\\&\quad-\gamma_2(-\tfrac{1}{2}ik_1f_1^0f_1'^0f_1^1f_2^2f_2^0f_2'^1+\tfrac{1}{2}ik_2f_1^0f_1'^0f_1^1f_2^1f_2^0f_2'^2)\\&=\tfrac{1}{2}ik_2(-\alpha f_2^0f_1^1-\gamma_1f_1'^0f_2'^1)f_2^0f_2^1f_1^2f_1^0+\tfrac{1}{4}k_2^2f_2^0f_2'^0f_2^1f_1^1f_1^0f_1'^2\\&\quad+\tfrac{1}{2}i\alpha k_1f_2^0f_2'^0f_2'^1f_1'^2f_1'^0f_1^1+\tfrac{1}{2}ik_2f_2^0f_2'^1f_1'^0f_1^2(\gamma_1f_1^0f_2^1-\tfrac{1}{2}ik_1f_2^0f_1^1)+\tfrac{1}{2}ik_1(\tfrac{1}{2}ik_2f_2^0f_1^2-\alpha f_2'^0f_1'^2)f_1'^0f_1^1f_2^0f_2'^1\\&\quad+\tfrac{1}{2}ik_2(\tfrac{1}{2}ik_2f_2'^0f_1'^2+\alpha f_2^0f_1^2)f_1^0f_1^1f_2^1f_2^0=0.\end{aligned}$$

Similarly, we can prove

$$Q'\equiv-\tfrac{1}{2}ik_2f_2'^1\tilde{f}_1'^{12}-\alpha f_2^1\tilde{f}_1^{12}-\gamma_2f_1'^1\tilde{f}_2'^{12}=0.$$

Therefore we have proved that the upper broken line in Fig. 3 is actually a BT. Similarly, we can also prove that the lower broken line in Fig. 3 is a BT.

APPENDIX D: TIME PART OF BT IN ORIGINAL VARIABLES

Elimination of ρ_{ixx} from Eqs. (4.3a,b) with use of (4.2) gives

$$i(u-u')_t+i(u-u')_{xxx}-2iN+\tfrac{3}{2}ik^2(u-u')_x=0,\tag{D1}$$

$$\begin{aligned}iN&\equiv(\Phi_{1x})^3-(\Phi_{1x}')^3+(\Phi_{2x})^3-(\Phi_{2x}')^3\\&=\tfrac{3}{16}(\Phi_1+\Phi_2-\Phi_1'-\Phi_2')_x[(\Phi_1+\Phi_2+\Phi_1'+\Phi_2')_x]^2+\tfrac{3}{16}(\Phi_1+\Phi_2-\Phi_1'-\Phi_2')_x[(\Phi_1-\Phi_2+\Phi_1'-\Phi_2')_x]^2\\&\quad+\tfrac{3}{8}(\Phi_1-\Phi_2-\Phi_1'+\Phi_2')_x(\Phi_1+\Phi_2+\Phi_1'+\Phi_2')_x(\Phi_1-\Phi_2+\Phi_1'-\Phi_2')_x\\&\quad+\tfrac{3}{16}(\Phi_1+\Phi_2-\Phi_1'-\Phi_2')_x[(\Phi_1-\Phi_2-\Phi_1'+\Phi_2')_x]^2+\tfrac{1}{16}[(\Phi_1+\Phi_2-\Phi_1'-\Phi_2')_x]^3.\end{aligned}\tag{D2}$$

From Eqs. (4.3c,d) we have

$$(\Phi_1+\Phi_2+\Phi_1'+\Phi_2')_x = ik\,[-\sinh(\Phi_1-\Phi_1')+\sinh(\Phi_2-\Phi_2')]$$
$$= 2ik\cosh\frac{\Phi_1+\Phi_2-\Phi_1'-\Phi_2'}{2}\sinh\frac{\Phi_2-\Phi_1+\Phi_1'-\Phi_2'}{2}$$
$$= -(u+u')_x\cot\Big(\frac{u-u'}{2}\Big), \tag{D3}$$
$$(\Phi_1-\Phi_2+\Phi_1'-\Phi_2')_x = -4\alpha\sin\Big(\frac{u-u'}{2}\Big)\sin\Big(\frac{u+u'}{2}\Big). \tag{D4}$$

From Eqs. (4.4) and (4.5), we have

$$(\Phi_1-\Phi_2-\Phi_1'+\Phi_2')_x = \partial_x 2\,\mathrm{arccosh}\Big[-\frac{2\alpha}{k}\sin\Big(\frac{u+u'}{2}\Big)\Big] = -4\alpha i\cos\Big(\frac{u+u'}{2}\Big)\sin\Big(\frac{u-u'}{2}\Big). \tag{D5}$$

Inserting Eqs. (D3)–(D5) into Eq. (D2), we have Eq. (4.7).

[1]A. Nakamura and R. Hirota, J. Phys. Soc. Jpn. **48**, 1755 (1980).
[2]R. Hirota, Prog. Theor. Phys. **52**, 1498 (1974).
[3]R. Hirota, "Recent Developments of Direct Method in Soliton Theory," Tech. Rep. No. A-1 (Dept. of Appl. Math., Faculty of Engineering, Hiroshima Univ., Hiroshima, Japan, 1979).
[4]R. Hirota and J. Satsuma, J. Phys. Soc. Jpn. **45**, 1741 (1978).
[5]A. Nakamura, J. Phys. Soc. Jpn. **47**, 1335 (1979).
[6]A. Nakamura, J. Phys. Soc. Jpn. **49**, 2380 (1980).
[7]M. Wadati, J. Phys. Soc. Jpn. **32**, 1681 (1972).
[8]M. Wadati, J. Phys. Soc. Jpn. **36**, 1498 (1974).

A limit on the variation of bounded positive operators

Gordon N. Fleming
Department of Physics, The Pennsylvania State University, University Park, Pennsylvania 16802

(Received 13 March 1980; accepted for publication 14 November 1980)

An upper limit is obtained for the rate at which the expectation value, of a bounded positive operator in an arbitrary state, can change with the parameters of a unitary transformation of the state.

PACS numbers: 02.30.Tb

1. STATEMENT OF THE LIMIT

Let ρ be a positive quantum mechanical density operator with unit trace. Let $U(\boldsymbol{\lambda})$ be a member of a family of unitary operators parametrized by the real ordered n-tuple, $\boldsymbol{\lambda} = (\lambda_1,...,\lambda_n)$. Let A be a positive semidefinite bounded operator with norm $\|A\|$, and let ρ, $U(\boldsymbol{\lambda})$, and A all be defined in a complex Hilbert space $\mathscr{H}$. Finally, let $G_a(\boldsymbol{\lambda})$ be defined by

$$G_a(\boldsymbol{\lambda}) \equiv -i\frac{\partial U(\boldsymbol{\lambda})}{\partial \lambda_a} U^{-1}(\boldsymbol{\lambda}) \quad (a = 1,...,n). \tag{1.1}$$

Now consider the expectation value of A in the mixed state $\rho_{\boldsymbol{\lambda}} \equiv U(\boldsymbol{\lambda})\rho U^{-1}(\boldsymbol{\lambda})$, transformed from the "original" mixed state ρ by $U(\boldsymbol{\lambda})$. This expectation value is

$$E(A;\boldsymbol{\lambda}) \equiv \mathrm{Tr}\{\rho_{\boldsymbol{\lambda}} A\}. \tag{1.2}$$

Writing

$$A \equiv \|A\|\bar{A}, \tag{1.3}$$

we have

$$E(A;\boldsymbol{\lambda}) = \|A\|E(\bar{A};\boldsymbol{\lambda}) \equiv \|A\|\bar{E}, \tag{1.4}$$

where

$$0 \leqslant \bar{E} \leqslant 1. \tag{1.5}$$

Consequently, there exists a real angle, $\theta(\bar{A};\boldsymbol{\lambda})$, such that

$$\bar{E} \equiv E(\bar{A};\boldsymbol{\lambda}) = \sin^2\theta(A;\boldsymbol{\lambda}). \tag{1.6}$$

If $E(A;\boldsymbol{\lambda})$ is differentiable with respect to λ_a, then $\theta(\bar{A};\boldsymbol{\lambda})$ can be chosen so that it also is differentiable with respect to λ_a.

From these definitions we will show that the derivative of θ is bounded by the positive square root of some expectation values of the operator $-\Delta_{\bar{A}}G_a^2$, where

$$\Delta_{\bar{A}}G_a(\boldsymbol{\lambda}) \equiv \bar{A}^{1/2}G_a(\boldsymbol{\lambda})(I-\bar{A})^{1/2} - (I-\bar{A})^{1/2}G_a(\boldsymbol{\lambda})\bar{A}^{1/2}, \tag{1.7}$$

with all indicated square roots taken positive. The expectation values in question are calculated using the modified normalized density operators

$$\rho_{\bar{A},\boldsymbol{\lambda}} \equiv \bar{A}^{1/2}\rho_{\boldsymbol{\lambda}}\bar{A}^{1/2}/\mathrm{Tr}\{\rho_{\boldsymbol{\lambda}}\bar{A}\} \tag{1.8a}$$

or

$$\rho_{(I-\bar{A}),\boldsymbol{\lambda}} \equiv (I-\bar{A})^{1/2}\rho_{\boldsymbol{\lambda}}(I-\bar{A})^{1/2}/\mathrm{Tr}\{\rho_{\boldsymbol{\lambda}}(I-\bar{A})\}. \tag{1.8b}$$

In short, suppressing the $\boldsymbol{\lambda}$ dependence, we show that

$$\left|\frac{\partial\theta}{\partial\lambda_a}\right| \leqslant |\,\mathrm{Tr}\{\rho_{\bar{A}}\Delta_{\bar{A}}G_a^2\}\,|^{1/2}, \tag{1.9}$$

$$|\,\mathrm{Tr}\{\rho_{I-\bar{A}}\Delta_{\bar{A}}G_a^2\}\,|^{1/2} \leqslant \|\Delta_{\bar{A}}G_a\|.$$

The last inequality is obvious and is to be employed when the expectation values are too complicated to be useful.

There is, of course, no priori guarantee that the expectation values or norm on the right-hand side of (1.9) are finite and we learn something only when they are. Nevertheless, as we will see, there are circumstances in which they are usefully finite. To begin with we note that if A commutes with U they all vanish, forcing the derivative to vanish as we know it must. Secondly, we note that if $\bar{E}$ is "very close" to unity or zero, then since

$$\frac{\partial\bar{E}}{\partial\lambda_a} = 2\sin\theta\cos\theta\frac{\partial\theta}{\partial\lambda_a}, \tag{1.10}$$

any finite expectation value in (1.9) yields

$$\left|\frac{\partial\bar{E}}{\partial\lambda_a}\right| \ll \left|\frac{\partial\theta}{\partial\lambda_a}\right|. \tag{1.11}$$

2. DERIVATION OF THE LIMIT

The derivation is based primarily on the following property of the trace:

$$|\mathrm{Tr}(AB)|^2 \leqslant \mathrm{Tr}(A^\dagger A)\,\mathrm{Tr}(B^\dagger B), \tag{2.1}$$

which holds for any operators A,B, for which the indicated quantities exist.

We have

$$\frac{\partial\bar{E}}{\partial\lambda_a} = \frac{\partial}{\partial\lambda_a}\mathrm{Tr}\{\rho_{\boldsymbol{\lambda}}\bar{A}\} = \frac{\partial}{\partial\lambda_a}\mathrm{Tr}\{\rho U^{-1}\bar{A}U\}$$
$$= i\,\mathrm{Tr}\{\rho_{\boldsymbol{\lambda}}[\bar{A},G_a]\}. \tag{2.2}$$

But, referring back to (1.7)

$$[\bar{A},G_a] = \bar{A}^{1/2}\Delta_{\bar{A}}G_a(I-\bar{A})^{1/2} + (I-\bar{A})^{1/2}\Delta_{\bar{A}}G_a\bar{A}^{1/2}. \tag{2.3}$$

Substituting (2.3) into (2.2) we have

$$\frac{\partial\bar{E}}{\partial\lambda_a} = -2\,\mathrm{Im}[\,\mathrm{Tr}\{\rho_{\boldsymbol{\lambda}}\bar{A}^{1/2}\Delta_{\bar{A}}G_a(I-\bar{A})^{1/2}\}\,], \tag{2.4}$$

where "Im" denotes imaginary part. Consequently

$$\left|\frac{\partial\bar{E}}{\partial\lambda_a}\right| \leqslant |2\,\mathrm{Tr}\{\rho_{\boldsymbol{\lambda}}\bar{A}^{1/2}\Delta_{\bar{A}}G_a(I-\bar{A})^{1/2}\}|. \tag{2.5}$$

If we now write the absolute value on the right side of (2.5) in the form

$$|\,\mathrm{Tr}\{\rho_{\boldsymbol{\lambda}}^{1/2}\bar{A}^{1/2}\Delta_{\bar{A}}G_a(I-\bar{A})^{1/2}\rho_{\boldsymbol{\lambda}}^{1/2}\}|,$$

then we can apply (2.1) to obtain

 0022-2488/81/081614-05$01.00 © 1981 American Institute of Physics

$$\left|\frac{\partial \bar{E}}{\partial \lambda_a}\right| \leqslant 2\,|\,\mathrm{Tr}\{\rho_\lambda \bar{A}\}\,|^{1/2} \times |\,\mathrm{Tr}\{(I-\bar{A})^{1/2}\rho_\lambda (I-\bar{A})^{1/2}\Delta_{\bar{A}} G_a^2\}\,|^{1/2},$$

or (2.6a)

$$\left|\frac{\partial \bar{E}}{\partial \lambda_a}\right| \leqslant 2\,|\,\mathrm{Tr}\{\rho_\lambda (I-\bar{A})\}\,|^{1/2} \times |\,\mathrm{Tr}\{\bar{A}^{1/2}\rho_\lambda \bar{A}^{1/2}\Delta_{\bar{A}} G_a^2\}\,|^{1/2}. \tag{2.6b}$$

In the first case (2.6a) the second factor on the right side is

$$|\,\mathrm{Tr}\{\rho_\lambda (I-\bar{A})\}\,|^{1/2}|\,\mathrm{Tr}\{\rho_{I-\bar{A},\lambda}\Delta_{\bar{A}} G_a^2\}\,|^{1/2}, \tag{2.7a}$$

while in the second case (2.6b) the second factor is

$$|\,\mathrm{Tr}\{\rho_\lambda \bar{A}\}\,|^{1/2}|\,\mathrm{Tr}\{\rho_{\bar{A},\lambda}\Delta_{\bar{A}} G_a^2\}\,|^{1/2}, \tag{2.7b}$$

where the definitions (1.8a,b) have been used. Substituting these expressions back into (2.6a,b) and introducing the compact notation,

$$\langle \Delta_{\bar{A}} G_a^2\rangle_{\bar{A}} \equiv \mathrm{Tr}\{\rho_{\bar{A},\lambda}\Delta_{\bar{A}} G_a(\lambda)^2\}, \tag{2.8a}$$

$$\langle \Delta_{\bar{A}} G_a^2\rangle_{I-\bar{A}} \equiv \mathrm{Tr}\{\rho_{I-\bar{A},\lambda}\Delta_{\bar{A}} G_a(\lambda)^2\}, \tag{2.8a}$$

we have

$$\left|\frac{\partial \bar{E}}{\partial \lambda_a}\right| \leqslant 2|\bar{E}|^{1/2}|1-\bar{E}|^{1/2}\begin{bmatrix} |\langle \Delta_{\bar{A}} G_a^2\rangle_{\bar{A}}|^{1/2} \\ |\langle \Delta_{\bar{A}} G_a^2\rangle_{I-\bar{A}}|^{1/2}\end{bmatrix}. \quad \begin{matrix}(2.9a)\\(2.9b)\end{matrix}$$

We now invoke (1.6) which yields

$$\left|\frac{\partial \bar{E}}{\partial \lambda_a}\right| = \left|2\sin\theta\cos\theta\frac{\partial\theta}{\partial\lambda_a}\right| = 2|\bar{E}|^{1/2}|1-\bar{E}|^{1/2}\left|\frac{\partial\theta}{\partial\lambda_a}\right|. \tag{2.10}$$

Comparison with (2.9) immediately gives

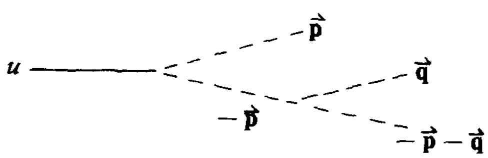
$$\left|\frac{\partial\theta}{\partial\lambda_a}\right| \leqslant \begin{cases} |\langle \Delta_{\bar{A}} G_a^2\rangle_{\bar{A}}|^{1/2} & (2.11a)\\ |\langle \Delta_{\bar{A}} G_a^2\rangle_{I-\bar{A}}|^{1/2} & (2.11b)\end{cases}$$

as claimed.

3. ILLUSTRATION OF THE LIMIT

A. Sequential decay

We will first illustrate the limit (1.9) by application to a simple model of sequential decay. Working in the center of mass frame we denote the parent particle state by $|u)$ and the two particle daughter state of relative momentum $\mathbf{p}$ by $|\mathbf{p}\rangle$. The three particle grandaughter state resulting from subsequent decay of one of the daughters is denoted by $|\mathbf{p},\mathbf{q}\rangle$ where the following diagram explains the notation:



Our normalization conventions are given by

$$(u|u) = 1, \tag{3.1a}$$

$$\langle \mathbf{p}'|\mathbf{p}\rangle = \delta^3(\mathbf{p}'-\mathbf{p}), \tag{3.1b}$$

$$\langle \mathbf{p}',\mathbf{q}'|\mathbf{p},\mathbf{q}\rangle = \delta^3(\mathbf{p}'-\mathbf{p})\delta^3(\mathbf{q}'-\mathbf{q}). \tag{3.1c}$$

Letting H denote the Hamiltonian of the system, we consider the time dependence of

$$|(u|\exp[(i/\hbar)Ht]|u)|^2$$

and

$$\int d^3p|\langle \mathbf{p}|\exp[(i/\hbar)Ht]|u)|^2.$$

The former is the probability that the parent particle has not yet decayed after the time interval t and the latter is the probability for finding the daughter state with any momentum after the time interval t. The first probability is of the general form (1.2) if we put $\rho = A = |u)(u|$ and $G = \hbar^{-1}H$. The second probability requires the choices $\rho = |u)(u|, G = \hbar^{-1}H$, and $A = \int d^3p|\mathbf{p}\rangle\langle\mathbf{p}|\equiv\pi$.

In the first case we have

$$\Delta_{\bar{A}} G = [\,|u)(u|H - H|u)(u|\,]\hbar^{-1} \tag{3.2}$$

and

$$|\langle \Delta_{\bar{A}} G^2\rangle_{\bar{A}}| = \hbar^{-2}(u|H-\bar{\epsilon})^2|u) \equiv \Delta\epsilon^2/\hbar^2, \tag{3.3}$$

wnere

$$\bar{\epsilon} \equiv (u|H|u). \tag{3.4}$$

Thus

$$\left|\frac{\partial\theta}{\partial t}\right| \leqslant \frac{\Delta\epsilon}{\hbar}, \tag{3.5}$$

where

$$|(u|\exp[(i/\hbar)Ht]|u)|^2 \equiv \sin^2\theta(t) \tag{3.6}$$

This limit, the rms deviation of the energy spectrum in the state $|u)$, is the same as that obtained from the norm $\|\Delta_{\bar{A}} G\|$, and has been reported before.[1] For the alternative expectation value we get

$$|\langle \Delta_{\bar{A}} G^2\rangle_{I-\bar{A}}| = \hbar^{-2}|(u|(H-\bar{\epsilon})\exp[(i/\hbar)Ht]|u)|^2 \times [1-|(u|\exp[(i/\hbar)Ht]|u)|^2]^{-1}. \tag{3.7}$$

We make the model more specific by putting

$$H|u) = |u)\bar{\epsilon} + \int d^3p|\mathbf{p}\rangle v(\mathbf{p}); \tag{3.8}$$

then

$$|\langle \Delta_{\bar{A}} G^2\rangle_{I-\bar{A}}| = \hbar^{-2}\left|\int d^3p \;\; v^*(\mathbf{p})\langle \mathbf{p}|u_t)\right|^2 \times\left[\int d^3p|\langle \mathbf{p}|u_t)|^2 + \int d^3p\,d^3q|\langle \mathbf{p},\mathbf{q}|u_t)|^2 + ,...,\right]^{-1}, \tag{3.9}$$

where $|u_t)\equiv\exp[(i/\hbar)Ht]|u)$ has been introduced. This limit is less than (3.3) (except for $t=0$) since

$$\left|\int d^3p\, v^*(\mathbf{p})\langle \mathbf{p}|u_t)\right|^2 \leqslant \int d^3p|v(\mathbf{p})|^2\int d^3p|\langle \mathbf{p}|u_t)|^2 \tag{3.10}$$

and

$$\int d^3p|v(\mathbf{p})|^2 = \Delta\epsilon^2. \tag{3.11}$$

In particular this limit drops to zero for large times with the depletion of the $|\mathbf{p}\rangle$ states.[2]

For the second probability $\int d^3p|\langle \mathbf{p}|u_t)|^2$, we will retain (3.8), change (3.2) to

$$\Delta_{\bar{A}} G = \hbar^{-1}[\pi H - H\pi], \tag{3.12}$$

and add, for definiteness,

$$H|\mathbf{p}\rangle = |\mathbf{p}\rangle 2\epsilon(\mathbf{p}) + |u)v^*(\mathbf{p}) + \int d^3q|\mathbf{p},\mathbf{q}\rangle v_\mathbf{p}(q), \tag{3.13a}$$

$$H|\mathbf{p},\mathbf{q}\rangle = |\mathbf{p},\mathbf{q}\rangle[\epsilon(\mathbf{p}) + \bar{\epsilon}(\mathbf{q}) + \bar{\epsilon}(\mathbf{p}+\mathbf{q})] + |\mathbf{p}\rangle v_\mathbf{p}^*(\mathbf{q}). \tag{3.13b}$$

With these specifications a little calculation yields

$$|\langle \Delta_{\bar{A}} G^2\rangle_{\bar{A}}| = \hbar^{-2}\Big[\Big|\int d^3p \quad v^*(\mathbf{p})\langle \mathbf{p}|u_t)\Big|^2 + \int d^3p\, d^3q |v_\mathbf{p}(\mathbf{q})|^2 |\langle \mathbf{p}|u_t)|^2\Big] \times \Big[\int d^3p |\langle \mathbf{p}|\langle \mathbf{p}|u_t|^2\Big]^{-1}, \tag{3.14}$$

and

$$|\langle \Delta_{\bar{A}} G^2\rangle_{I-\bar{A}}| = \hbar^{-2}\Big[\Delta\epsilon^2|(u|u_t)|^2 + \int d^3p\Big|\int d^3q\, v_\mathbf{p}^*(\mathbf{q})\langle \mathbf{p},\mathbf{q}|u_t)\Big|^2\Big] \times\Big[|(u|u_t)|^2 + \int d^3p\, d^3q|\langle \mathbf{p},\mathbf{q}|u_t)|^2\Big]^{-1}. \tag{3.15}$$

Both (3.14) and (3.15) yield $\Delta\epsilon^2/\hbar^2$ for very small times t, as can be shown from a power series expansion of $\exp[(i/\hbar)Ht]$ in each case. For later times (3.14) is difficult to assess while (3.15) approaches

$$\int d^3p\Big|\int d^3q \quad v_\mathbf{p}^*(\mathbf{q})\langle \mathbf{p},\mathbf{q}(+)|u) \times \exp\{(i/\hbar)(\epsilon(\mathbf{p}) + \bar{\epsilon}(\mathbf{q}) + \bar{\epsilon}(\mathbf{p}+\mathbf{q}))t\}\Big|^2 \Big/ \int d^3p\, d^3q|\langle \mathbf{p},\mathbf{q}(+)|u)|^2, \tag{3.16}$$

where $\langle \mathbf{p},\mathbf{q}(+)|$ is the outgoing scattering eigenbra of H for the stable grandaughter particles. For "sufficiently smooth" potential $v_\mathbf{p}(\mathbf{q})$, and decay amplitude $\langle \mathbf{p},\mathbf{q}(+)|u_t)$, this expression (3.16), vanishes by the Reimann–Lesbesque Lemma as $t \to \infty$. We speculate that detailed examination of the threshold behavior of the potential and the decay amplitude in (3.16) would yield inverse power law bounds on $\dot{\theta}$ for large times, where

$$\sin^2\theta(t) = \int d^2p|\langle \mathbf{p}|u_t)|^2. \tag{3.17}$$

Finally we note, for completeness, that the norm of (3.12) in the presence of (3.13) and

$$I = |u)(u| + \int d^3p|\mathbf{p}\rangle\langle \mathbf{p}| + \int d^3p\, d^3q|\mathbf{p},\mathbf{q}\rangle\langle \mathbf{p},\mathbf{q}| \tag{3.18}$$

is $|\lambda|$, where

$$1 = \int d^3p \frac{|v(\mathbf{p})|^2}{\lambda^2 - \int d^3q|v_\mathbf{p}(\mathbf{q})|^2}. \tag{3.19}$$

B. Field theory correlation

As an indication of possibly extensive applications of the limit to quantum field theory we consider the spatially smeared two point function for a neutral scalar field

$$\langle 0|\phi(f,x_0)\phi(f',x_0')|0\rangle, \tag{3.20}$$

where

$$\phi(f,x_0) \equiv \int d^3x\, f(\mathbf{x})\phi(\mathbf{x},x_0), \tag{3.21}$$

with

$$\int d^3x|f(\mathbf{x})|^2 < \infty. \tag{3.22}$$

Writing

$$\tilde{f}(\mathbf{q}) = (2\pi)^{-3}\int d^3x\, r^{i\mathbf{q}\cdot\mathbf{x}} f(\mathbf{x}), \tag{3.23}$$

we have

$$\langle 0|\phi(f^*,x_0)\phi(f',x_0')|0\rangle = (2\pi)^6\int_0^\infty d\sigma^2\eta(\sigma^2)\frac{d^3q}{2(\mathbf{q}^2+\sigma^2)^{1/2}}\tilde{f}^*(\mathbf{q})\tilde{f}'(\mathbf{q}) \times \exp i[(\mathbf{q}^2+\sigma^2)^{1/2}(x_0'-x_0)], \tag{3.24}$$

where $\eta(\sigma^2)$ is the Källen–Lehmann spectral function[3]

$$\eta(q^2) = (\textstyle\int\Sigma)dn\delta^4(q-p_n)|\langle n|\phi(0)|0\rangle|^2. \tag{3.25}$$

Since[4]

$$\int_0^\infty d\sigma^2\,\eta(\sigma^2) = 1, \tag{3.26}$$

we have

$$\langle 0|\phi(f^*,0)\phi(f,0)|0\rangle < 1 \tag{3.27}$$

and we can put

$$\rho = \bar{A} = \phi(f,0)|0\rangle\langle|\phi(f^*,0)\phi(f,0)|0\rangle^{-1} \times\langle 0|\phi(f^*,0) \tag{3.28}$$

and

$$U = e^{(i/\hbar)P_0(x_0'-x_0)}. \tag{3.29}$$

The statement of the result of applying our limit, with the indicated definitions, to this quantity (3.24), will be greatly facilitated by introducing the notation

$$\tilde{f}_N(\mathbf{q}) \equiv \tilde{f}(\mathbf{q}) \Big/ \int d\sigma^2\,\eta(\sigma^2)\frac{d^3q'}{2(\mathbf{q}'^2+\sigma^2)^{1/2}}|\tilde{f}(\mathbf{q}')|^2. \tag{3.30}$$

We then have, putting

$$\sin^2\theta(x_0'-x_0) = \langle 0|\phi(f,x_0)\phi(f,x_0')|0\rangle/\langle 0|\phi(f,0)\phi(f,0)|0\rangle, \tag{3.31}$$

that

$$\left|\frac{\partial\theta(x_0)}{\partial x_0}\right| \leqslant \frac{\Delta p_0}{\hbar}, \tag{3.32}$$

where

$$\Delta p_0^2 = \int d\sigma^2\,\eta(\sigma^2)\frac{d^3q}{2(\mathbf{q}^2+\sigma^2)^{1/2}}|\tilde{f}_N(\mathbf{q})|^2[(\mathbf{q}^2+\sigma^2) - \mathbf{p}_0^2] \tag{3.33}$$

and

$$\bar{p}_0 = \int d\sigma^2\,\eta(\sigma^2)\frac{d^3q}{2(\mathbf{q}^2+\sigma^2)^{1/2}}|\,\tilde{f}_N(\mathbf{q})\,|^2(\mathbf{q}^2+\sigma^2)^{1/2}$$

$$= \frac{1}{2}\int d^3q|\tilde{f}_N(\mathbf{q})|^2. \tag{3.34}$$

The finiteness of these quantities depends on how fast $\eta(\sigma^2)$ vanishes at infinity.

C. Collisions from a mixed state

If $\rho = \rho_{\text{in}}$ is the density operator for the incident state of a collision process then the corresponding density operator for the final state is[5]

$$\rho_{\text{out}} = T\rho_{\text{in}}T^+, \tag{3.35}$$

where $S = I - 2iT$ is the unitary scattering operator. If the projection operator π projects out the eigenvectors of the observed final state parameters, then the probability of the observed transition is proportional to

$$\text{Tr}\{\pi T\rho_{\text{in}}T^+\pi\} = \text{Tr}\{\rho_{\text{in}}T^+\pi T\}, \tag{3.36}$$

where $T^+\pi T$ is bounded by unity. A matter of frequent interest is the change induced in this transition probability by a change in the relative spin orientations of the participating particles. Such a change would be induced by the unitary rotation operator $U(\hat{n}\phi)$ generated by the spin operator $\mathbf{S}$ for some subset of the initial particles. Thus we consider

$$\rho_{\text{in},(\hat{n}\phi)} = U(\hat{n}\phi)\rho_{\text{in}}U^{-1}(\hat{n}\phi), \tag{3.37}$$

where

$$U(\hat{n},\phi) = e^{-(i/\hbar)\mathbf{S}\cdot\hat{n}\phi}. \tag{3.38}$$

We define, in this case,

$$\boldsymbol{\lambda} = \hat{n}\phi, \quad \delta\boldsymbol{\lambda} = \hat{n}\delta\phi + \phi\delta\hat{n} \tag{3.39}$$

and

$$\frac{\partial}{\partial\boldsymbol{\lambda}} = \hat{n}\frac{\partial}{\partial\phi} + \frac{1}{\phi}\frac{\partial}{\partial\hat{n}}, \tag{3.40}$$

with the understanding that $\hat{n}\cdot(\partial/\partial\hat{n}) = 0$ since $\hat{n}^2 = 1$. Then noticing that [from (1)]

$$U(\hat{n}\phi + \delta\hat{n}\phi + \hat{n}\delta\phi)U^{-1}(\hat{n}\phi) - I$$
$$\simeq i\mathbf{G}(\hat{n}\phi)\cdot(\hat{n}\delta\phi + \delta\hat{n}\phi) \tag{3.41}$$

and

$$R(\hat{n}\phi + \hat{n}\delta\phi + \delta\hat{n}\phi)R^{-1}(\hat{n}\phi)\mathbf{x}$$
$$\approx \mathbf{x} + \{(1 - \cos\phi)\hat{n}\times\delta\hat{n} + \sin\phi\delta\hat{n} + \delta\phi\hat{n}\}\times\mathbf{x}, \tag{3.42}$$

where

$$R_{ab}(\hat{n}\phi) = n_a n_b + (\delta_{ab} - n_a n_b)\cos\phi - \epsilon_{abc}n_c\sin\phi, \tag{3.43}$$

we conclude

$$\mathbf{G}(\hat{n}\phi)\cdot(\hat{n}\delta\phi) + \delta n\phi)$$
$$= -(i/\hbar)\mathbf{S}\cdot\{(1 - \cos\phi)\hat{n}\times\delta\hat{n} + \sin\phi\delta\hat{n} + \delta\phi\hat{n}\}, \tag{3.44}$$

which yields

$$\mathbf{G}(\hat{n}\phi)$$
$$= -(1/\hbar)\{[(1 - \cos\phi)/\phi]\mathbf{S}\times\hat{n} + (\sin\phi/\phi)[\mathbf{S} - \hat{n}(\hat{n}\cdot\mathbf{S})]$$
$$+ \hat{n}(\hat{n}\cdot S)\}. \tag{3.45}$$

Having presented this example of a calculation of $G_a(\boldsymbol{\lambda})$ for a nonabelian group we now return to (3.36) and consider the simplest case of the *total* transition probability, i.e., $\pi = I$. Then defining $\theta(\hat{n}\phi)$ by,

$$\text{Tr}\{\rho_{\text{in}_{(\hat{n}\phi)}}T^+T\} \equiv \sin^2\theta(\hat{n}\phi), \tag{3.46}$$

we have

$$\left|n_a\frac{\partial\theta}{\partial\phi} + \frac{1}{\phi}\frac{\partial\theta}{\partial n_a}\right| \leqslant |\langle\Delta_{T^+T}G_a(\hat{n}\phi)^2\rangle|^{1/2}, \tag{3.47}$$

where the expectation value on the right employs either $\rho_{\text{in}_{(\hat{n}\phi)},T^+T}$ or $\rho_{\text{in}_{(\hat{n}\phi)},I-T^+T}$. Let us examine $\Delta_{T^+T}\mathbf{G}(\hat{n}\phi)$. Introducing the reaction operator K via

$$S = (I - iK)/(I + iK), \tag{3.48}$$

we have

$$T = K/(I + iK), \quad T^+T = K^2/(I + K^2) \tag{3.49}$$

and, consequently,

$$\Delta_{T^+T}\mathbf{G} = (I + K^2)^{-1/2}[K,\mathbf{G}](I + K^2)^{-1/2} \tag{3.50}$$

and

$$(\Delta_{T^+T}G_a)^2 = (I + K^2)^{-1/2}[K,G_a](I + K^2)^{-1}$$
$$\times [K,G_a](I + K^2)^{-1/2}. \tag{3.51}$$

Thus,

$$|\langle(\Delta_{T^+T}G_a)^2\rangle| \leqslant |\langle[K,G_a]^2\rangle|$$
$$= |\langle[K,G_a]\rangle|^2 + (\Delta_{\text{rms}}i[K,G_a])^2$$
$$\leqslant (\Delta_{\text{rms}}K)(\Delta_{\text{rms}}iG_a)^2 + (\Delta_{\text{rms}}i[K,G_a])^2, \tag{3.52}$$

where Δ_{rms} denotes the rms deviation of the (Hermitian) operator following Δ_{rms} in the same (mixed) state used to calculate the original expectation value.

We recognize that the limit (3.52) is probably not very good, and indeed, none of the examples are offered for their stunning power. They are offered as suggestive of the generality of the basic result (2.11) and in the hope that they will stimulate more substantive application of that basic result.

4. EXAMINATION OF THE LIMIT

In this final section a few observations are made on the general structure of the limit (2.11) for the purpose of gaining insight into future applications. To this end we denote the eigenvalue of $\bar{A}$ by $\sin^2\alpha$ and introduce the spectral decomposition[6] of $\bar{A}$ via

$$\bar{A} = \int_0^{\pi/2} \sin^2\alpha \, d\Pi(\alpha), \tag{4.1}$$

where $\Pi(\alpha)$ is a projective resolution of unity satisfying

$$\Pi(0) = 0. \quad \Pi(\pi/2) = I,$$
$$\Pi(\alpha)\Pi(\alpha') = \Pi(\min(\alpha,\alpha')), \tag{4.2}$$
$$\bar{A}\,d\Pi(\alpha) = d\Pi(\alpha)\sin^2\alpha.$$

From (4.1) we have

$$\bar{A}^{1/2} = \int_0^{\pi/2} \sin\alpha \, d\Pi(\alpha), \quad (I - \bar{A})^{1/2} = \int_0^{\pi/2} \cos\alpha \, d\Pi(\alpha) \tag{4.3}$$

and, from (1.7)

$$\Delta_{\bar{A}}G_a = \int_0^{\pi/2} \sin(\alpha - \alpha')\,d\Pi(\alpha)G_a\,d\Pi(\alpha'). \tag{4.4}$$

The density operators $\rho_{\lambda,A}$ and $\rho_{\lambda,I-\bar{A}}$ become,

$$\rho_{\lambda,\bar{A}} = \int_0^{\pi/2} \sin\alpha \sin\alpha' \, d\Pi(\alpha)\rho_\lambda d\Pi(\alpha') \Big/ \int_0^{\pi/2} \sin^2\alpha \mathrm{T_r}\{\rho_\lambda d\Pi(\alpha)\}, \tag{4.5a}$$

$$\rho_{\lambda,I-\bar{A}} = \int_0^{\pi/2} \cos\alpha \cos\alpha' \, d\Pi(\alpha)\rho_\lambda d\Pi(\alpha') \Big/ \int_0^{\pi/2} \cos^2\alpha \mathrm{T_r}\{\rho_\lambda d\Pi(\alpha)\}, \tag{4.5b}$$

and $\Delta_{\bar{A}} G_a^2$ becomes

$$\Delta_{\bar{A}} G_a^2 = \int_0^{\pi/2} \sin(\alpha - \alpha')\sin(\alpha' - \alpha'') \times d\Pi(\alpha) G_a d\Pi(\alpha') G_a d\Pi(\alpha''). \tag{4.6}$$

But

$$2\sin(\alpha - \alpha')\sin(\alpha' - \alpha'') = \cos(\alpha + \alpha'' - 2\alpha') - \cos(\alpha - \alpha''). \tag{4.7}$$

Hence

$$\Delta_{\bar{A}} G_a^2 = -\tfrac{1}{2}\int_0^{\pi/2} [\cos(\alpha - \alpha'')\, d\Pi(\alpha) G_a^2 d\Pi(\alpha'') - \cos(\alpha + \alpha'' - 2\alpha') d\Pi(\alpha) G_a d\Pi(\alpha') G_a d\Pi(\alpha'')]. \tag{4.8}$$

This last equation displays a kinship of $\Delta_{\bar{A}} G_a^2$ to the square of the rms deviation of G_a with respect to the eigenspaces of $\bar{A}$.

A simple limit, lying somewhere between (2.11a) and (2.11b) is obtained by writing

$$\left|\frac{\partial\theta}{\partial\lambda_a}\right|^2 = \mathrm{Tr}\{\rho_\lambda \bar{A}\}\left|\frac{\partial\theta}{\partial\lambda_a}\right|^2 + \mathrm{Tr}\{\rho_\lambda(I-\bar{A})\}\left|\frac{\partial\theta}{\partial\lambda_a}\right|^2 \leqslant \mathrm{Tr}\{\rho_\lambda \bar{A}^{1/2}\Delta_{\bar{A}} G_a^2 \bar{A}^{1/2}\} + \mathrm{Tr}\{\rho_\lambda (I-\bar{A})^{1/2}\Delta_{\bar{A}} G_a^2 (I-\bar{A})^{1/2}\}. \tag{4.9}$$

In the presence of (4.3), (4.5), and (4.8) this becomes

$$\left|\frac{\partial\theta}{\partial\lambda_a}\right|^2 \leqslant \tfrac{1}{2}\int_0^{\pi/2} \cos(\alpha - \alpha'')[\mathrm{Tr}\{\rho_\lambda d\Pi(\alpha) G_a^2 d\Pi(\alpha'')\}\cos(\alpha - \alpha'') - \mathrm{Tr}\{\rho_\lambda d\Pi(\alpha) G_a d\Pi(\alpha'')\}\cos(\alpha + \alpha'' - 2\alpha')]. \tag{4.10}$$

ACKNOWLEDGMENTS

This work was begun while the author was visiting the Center for Theoretical Physics at the Massachusetts Institute of Technology. It was completed during a visit to the International Centre for Theoretical Physics in Trieste. Accordingly I wish to thank Professor A. Kerman for the hospitality of the CTP at MIT and Professor A. Salam for the hospitality of the ICTP in Trieste. Special thanks for stimulating conversations are extended to Professors K. Johnson of MIT, L. Fonda, and G. C. Ghirardi at Trieste, and H. Ekstein from Marseille.

[1]G. N. Fleming, Nuovo Cimento **16**, 232 (1973); **46**, 579 (1978). See also A. Peres, "Non Exponential Decay Law", Technion-Israel Institute of Technology preprint (1980).

[2]L. Fonda, G. C. Ghirardi, C. Omero, A. Rimini, and T. Weber, Phys. Rev. D **18**, 4757 (1978).

[3]H. Lehmann, Nuovo Cimento **11**, 342 (1954).

[4]J. D. Bjorken and S. D. Drell, *Relativistic Quantum Fields* (McGraw-Hill, New York, 1965), p. 140-141.

[5]U. Fano, Rev. Mod. Phys. **29**, 74 (1975). See also R. G. Newton, Fund. Phys. **9**, 929 (1979).

[6]F. W. Byron, Jr. and R. W. Fuller, *Mathematics of Classical and Quantum Physics* (Addison-Wesley, Reading, Mass., 1970), Vol. 2, p. 529.

1619,1622

Adjoints of nondensely defined Hilbert space operators

Peter Pfeifer
Laboratory of Physical Chemistry, ETH Zürich, CH-8092 Zürich, Switzerland[a)]

(Received 27 January 1981; accepted for publication 20 March 1981)

To each linear operator T acting in a Hilbert space $\mathscr{H}$, an adjoint operator T^* is assigned which coincides with the usual adjoint whenever the domain of T is dense in $\mathscr{H}$. General properties of T^* are: If $\mathscr{H}$ is countably infinite-dimensional, then the set of all closed operators equals the set of all adjoints; if the domain (range) of T is closed, then so is the domain (range) of T^*; T is closable (bounded) if and only if T^* is densely (everywhere) defined. Noteworthy corollaries are the closed graph and the closed range theorems, as well as basic crosslinks between adjoints and inverses. Applications to the problem of extending formally self-adjoint operators are given.

PACS numbers: 02.30.Tb

1. INTRODUCTION

In quantum theory, one is accustomed to the fact that the basic objects like observables, symmetries, or propagators are always represented by operators which are (at least) densely defined in the relevant Hilbert space $\mathscr{H}$ (complex, and with inner product $\langle . | . \rangle$). In many cases, however, the mathematical development leads naturally to operators whose domain is no longer dense in $\mathscr{H}$. Typical examples are scattering theory, spectral analysis of Hamiltonians perturbed by finite-rank operators (e.g., in pseudo-potential theory, or for construction of lower energy bounds), Trotter-type approximations of semigroups,[1] approximation of Schrödinger operators,[2] "coordinate notation"[3] for direct integrals of operators in Hilbert spaces of nonconstant dimension.

For such a linear operator T with domain $\mathscr{D}(T)$ and range $\mathscr{R}(T)$ in $\mathscr{H}$ [the set of all these will be denoted by $\mathscr{L}(\mathscr{H})$] where $\mathscr{D}(T)$ is not dense in $\mathscr{H}$, the conventional wisdom is to regard T as a densely defined operator $\widetilde{T}$ from the closed subspace $\overline{\mathscr{D}(T)}$ to $\mathscr{H}$, say, so that the fundamental notion of the adjoint operator is available as $\widetilde{T}^*$ (see e.g., Ref. 4, p. 50).[5] Clearly, this approach requires careful bookkeeping of domain and range spaces and yields a rather unwieldy calculus when more than a single operator is involved. It is the purpose of this paper to show that the usual definition of the adjoint T^* for densely defined $T\in\mathscr{L}(\mathscr{H})$ can be naturally extended to a map $*:\mathscr{L}(\mathscr{H})\to\mathscr{L}(\mathscr{H})$, which not only preserves the former's manipulative simplicity, but also sheds light on many standard results [i.e., for $\overline{\mathscr{D}(T)} = \mathscr{H}$] by eliminating all nonintrinsic, "local-reference-frame" $\overline{\mathscr{D}(T)}$-dependencies. Applications additional to those in Sec. 5 will be presented elsewhere.[6]

2. DEFINITION AND PROPERTIES OF THE ADJOINT

For $T\in\mathscr{L}(\mathscr{H})$, let (in addition to the foregoing) $\mathscr{G}(T)$, $\mathscr{N}(T)$, P_T be the graph, null space, and orthogonal projector onto $\overline{\mathscr{D}(T)}$, respectively; for closable T, $\overline{T}$ is the closure.

Of the following notions of adjointness, the first one is standard.

Definition 1: Let $T\in\mathscr{L}(\mathscr{H})$. $T'\in\mathscr{L}(\mathscr{H})$ is said to be a formal adjoint of T if $\langle\varphi|T\psi\rangle = \langle T'\varphi|\psi\rangle$ for all $\psi\in\mathscr{D}(T)$, $\varphi\in\mathscr{D}(T')$. The set of all formal adjoints of T is denoted by $\mathscr{A}(T)$. T is called formally self-adjoint (f.s.a.) if $T\in\mathscr{A}(T)$.

Definition 2: Let $T\in\mathscr{L}(\mathscr{H})$ and put $\mathscr{D}(T^*) = \{\varphi|\varphi\in\mathscr{H}$; the linear functional $\langle\varphi|T.\rangle$ with domain $\mathscr{D}(T)$ is bounded$\} = \{\varphi|\varphi\in\mathscr{H}$; there is a $\varphi'\in\mathscr{H}$ [unique if chosen from $\overline{\mathscr{D}(T)}$] such that $\langle\varphi|T\psi\rangle = \langle\varphi'|\psi\rangle$ for all $\psi\in\mathscr{D}(T)\}$. The adjoint $T^*\in\mathscr{L}(\mathscr{H})$ of T then is defined by $T^*\varphi = \varphi'$, $\varphi'\in\overline{\mathscr{D}(T)}$ being as in the preceding equation, for every $\varphi\in\mathscr{D}(T^*)$. T is called self-adjoint if $T = T^*$.

Remark 1: The few assertions underlying Def. 2 are elementary. It follows from Theorems 2 and 5 that this definition of self-adjointness coincides with the customary one.

Remark 2: T^* can equivalently be defined as the (usual) adjoint of the densely defined operator TP_T, or as the single-valued part of the adjoint subspace[7] of $\mathscr{G}(T)$, whereas the relation

$$\mathscr{A}(T) = \{T^* + X\,|X\in\mathscr{L}(\mathscr{H}); R(X)\subset\mathscr{D}(T)^\perp\} \tag{1}$$

shows that T^* is singled out from $\mathscr{A}(T)$ by having maximal domain and minimal range.

If $S,T\in\mathscr{L}(\mathscr{H})$, then

$$\mathscr{A}(S) + \mathscr{A}(T)\subset\mathscr{A}(S+T),$$
$$P_{S+T}(S^* + T^*)\subset(S+T)^*;$$
$$\mathscr{A}(S).\mathscr{A}(T)\subset\mathscr{A}(TS),$$
$$P_{TS}S^*T^*\subset(TS)^*;$$
$$\mathscr{A}(S)\subset\mathscr{A}(T) \text{ and } P_TS^*\subset T^* \text{ for } T\subset S.$$

Less trivial results on composition and extension of adjoints are collected in the following theorem:

Theorem 1: Let $S,T\in\mathscr{L}(\mathscr{H})$. Then

$$(S+T)^* = S^* + P_ST^* \quad \text{if } \mathscr{D}(S)\subset\mathscr{D}(T) \text{ and } T \text{ is bounded;} \tag{2}$$
$$(TS)^* = S^*T^* \quad \text{if } \mathscr{R}(S)\subset\mathscr{D}(T) \text{ and } T \text{ is bounded;} \tag{3a}$$
$$(TS)^* = S^*T^* \quad \text{if } \mathscr{D}(T)\subset\mathscr{R}(S),\ S \text{ has a bounded inverse, and } \overline{S^{-1}P_TS} \text{ is bounded}^{8}; \tag{3b}$$

[a)]Present address: Department of Physical Chemistry, Hebrew University of Jerusalem, 91904 Jerusalem, Israel.

 0022-2488/81/081619-04$01.00 © 1981 American Institute of Physics

$(TS)^* = S^*T^*$ if S is closed and there is a closed subspace $\mathscr{D} \subset \mathscr{D}(T)$ with $\mathscr{R}(S)\cap\mathscr{D} = \{0\}$,
$\mathscr{R}(S) + \mathscr{D} = \overline{\mathscr{D}(T)}$, and the restriction of T to $\mathscr{D}$ being bounded[9]; (3c)

$T^* = \overline{P_T S^*}$ if $T\subset S$, T is bounded, and S is closable; (4a)

$T^* = S^*$ if $T\subset S$, $\overline{\mathscr{R}(T)} = \overline{\mathscr{R}(S)}$, and S has a bounded inverse. (4b)

Proof: (2), (3a): Adapt the well-known proof for densely defined S and T. (3b): The hypotheses imply $T = TSS^{-1}$, $T^* \supset P_T(S^{-1})^*(TS)^*$, and by (3a) $S^*T^* \supset (\overline{S^{-1}P_TS})^*(TS)^*$. So boundedness of $\overline{S^{-1}P_TS}$ and Theorem 5 yield $\mathscr{D}(S^*T^*) \supset \mathscr{D}((TS)^*)$, whence $(TS)^* = P_{TS}S^*T^* = S^*T^*$. (3c): Since $\mathscr{R}(S)$ must be closed (Theorem IV. 1.12 in Ref. 4) the (bounded) projector from $\overline{\mathscr{D}(T)}$ onto $\mathscr{D}$ along $\mathscr{R}(S)$ is well-defined. Let T_0 and S_0 be the restriction of T to $\mathscr{D}$ and of S to $\mathscr{D}(S)\cap\mathscr{N}(S)^\perp$. Boundedness of T_0 and S_0^{-1} gives $T = T_0P + TSS_0^{-1}(1-P)$, so that $T^* \supset (T_0P)^* + (S_0^{-1}(1-P))^*(TS)^*$ and [by $S^* = S_0^*$ and (3a)]

$$S^*T^* \supset S^*(T_0P)^* + S_0^*(S_0^{-1}(1-P))^*(TS)^*$$
$$= (T_0PS)^* + (S_0^{-1}(1-P)S_0)^*(TS)^* = (TS)^*.$$

(4a): Follows from Theorems 2 and 5. (4b): Take T in (3b) to be an appropriate restriction of the unit operator. Q.E.D.

Remark 3: Equations (3b,c) generalize Problem 4.18 in Ref. 10 and Theorem 6 in Ref. 11. Further results in Ref. 11, and Theorem 5.27 in Ref. 10 admit similar extensions.

Remark 4: None of the conditions $\mathscr{R}(S)\subset\mathscr{D}(T)$, boundedness of $\overline{S^{-1}P_TS}$, $\overline{\mathscr{R}(T)} = \overline{\mathscr{R}(S)}$ in (3a), (3b), (4b) can be dropped.

To see that only T^* and T^{**} are relevant (because $T^{***} = T^*P_{T^*}$ and $T^{****} = T^{**}$) we have:

Theorem 2: Let $T\in\mathscr{L}(\mathscr{H})$. Then T^* is closed, $P_{T^*}T$ is closable, and $T^{**} = \overline{P_{T^*}T}\,P_T$.

Proof: Closedness is clear from Remark 2. Since $TP_T \in \mathscr{A}(T^*)$ and by (1) $P_{T^*}TP_T \subset T^{**}$, the densely defined operator $P_{T^*}TP_T$ is closable and hence satisfies $\overline{P_{T^*}TP_T} = (P_{T^*}TP_T)^{**} = ((TP_T)^*P_{T^*})^* = (T^*P_{T^*})^* = T^{**}$.Q.E.D.

Remark 5: The operator $\overline{P_{T^*}T}$ equals the single-valued part[7] of $\overline{\mathscr{G}(T)}$. $P_{T^*}T$ is distinguished in that it effects a unique decomposition of T into a maximal closable part $P_{T^*}T$ and a minimal nonclosable part $(1-P_{T^*})T$ [indeed, the adjoint of the latter is the restriction of the zero operator to $\overline{\mathscr{D}(T^*)}$, which is minimal in the sense that, for $S\in\mathscr{L}(\mathscr{H})$, $S^*\subset 0$ if and only if $\mathscr{D}(S^*)$ consists just of the trival part $\mathscr{R}(S)^\perp$]. A similar decomposition is known for positive quadratic forms.[2] For further characteristics of $P_{T^*}T$, see Remarks 6 and 7, and Theorem 8.

There is a partial converse of Theorem 2:

Theorem 3: For separable $\mathscr{H}$, the equality

$$\{T^*\,|\,T\in\mathscr{L}(\mathscr{H})\} = \{S\,|\,S\in\mathscr{L}(\mathscr{H});\ S \text{ is closed}\}$$

holds if and only if $\dim(\mathscr{H})$ is either 0 or ∞.

Proof: A little thought shows that we only need to prove that, for closed $S\in\mathscr{L}(\mathscr{H})$ with $\overline{\mathscr{D}(S)} \neq \mathscr{H}$ and $\mathscr{H}$ countably infinite-dimensional, there is a $T\in\mathscr{L}(\mathscr{H})$ such that $S = T^*$. Pick an orthonormal basis $(\varphi_n)_{n=1}^\infty$ in $\mathscr{D}(S^*)$ and a dense sequence $(\xi_n)_{n=1}^\infty$ in $\mathscr{D}(S)^\perp$, define $X\in\mathscr{L}(\mathscr{H})$ by $\mathscr{D}(X) = \operatorname{span}\{\varphi_1, \varphi_2 \ldots\}$ and $X\varphi_n = (1 + \|S^*\varphi_n\|)\xi_n$ $(n = 1,2,\ldots)$, and put $T = S^* + X$. If follows that

$$\mathscr{D}(T^*) = \{\psi + \xi\,|\,\psi\in\overline{\mathscr{D}(S)}; \xi\in\mathscr{D}(S)^\perp;$$
$$\langle\psi|S^*.\rangle + \langle\xi|X.\rangle \text{ is bounded}\}. \quad (5)$$

But for every $\psi\in\overline{\mathscr{D}(S)}$ and $\xi\in\mathscr{D}(S)^\perp$ we have

$$|\langle\psi|S^*\varphi_n\rangle + \langle\xi|X\varphi_n\rangle|$$
$$\geqslant |\langle\xi|X\varphi_n\rangle| - |\langle\psi|S^*\varphi_n\rangle|$$
$$\geqslant |\langle\xi|\xi_n\rangle| + (|\langle\xi|\xi_n\rangle| - \|\psi\|)\|S^*\varphi_n\|$$

$(n = 1,2,\ldots)$ where, by denseness of $(\xi_n)_{n=1}^\infty$ in $\mathscr{D}(S)^\perp$, the right hand side remains bounded for $n\to\infty$ only if $\xi = 0$. Thus Eq. (5) reduces to $\mathscr{D}(T^*) = \mathscr{D}(S^{**})\cap\overline{\mathscr{D}(S)}$. Since $S^{**} = SP_S$ (Theorem 5), one actually finds $T^* = S$. Q.E.D.

The next result collects the fundamental domain and range properties of T^*.

Theorem 4: Let $T\in\overline{\mathscr{L}(\mathscr{H})}$. Then

$$\mathscr{N}(T^*) = \mathscr{R}(T)^\perp, \quad \overline{\mathscr{R}(T^*)} = \mathscr{N}(\overline{P_{T^*}T})^\perp \cap \overline{\mathscr{D}(T)}. \quad (6)$$

$$\overline{\mathscr{D}(T^*)} = \{\lim_{n\to\infty} T\psi_n\,|\,(\psi_n)_{n=1}^\infty \text{ is a sequence in } \overline{\mathscr{D}(T)}$$
$$\text{with } \lim_{n\to\infty}\psi_n = 0 \text{ and } (T\psi_n)_{n=1}^\infty \text{ convergent}\}^\perp, \quad (7)$$

$$\overline{\mathscr{D}(T^*)} \supset \overline{\mathscr{D}(T)} \text{ if } \{\langle\psi|T\psi\rangle\,|\,\psi\in\mathscr{D}(T);\ \|\psi\| = 1\} \neq \mathbb{C}, \quad (8)$$

$\mathscr{D}(T^*)$ is closed if $\mathscr{D}(T)$ is so, (9a)

$\mathscr{R}(T^*)$ is closed if $\mathscr{R}(T)$ is so. (9b)

Proof: (6): Clear from definition 2 and Theorem 2. (7) is well known in view of Remark 2. (8): Cf. proof of Theorem V.3.4 in Ref. 12. (9a): let $\varphi\in\overline{\mathscr{D}(T^*)}$, and let $(\varphi_n)_{n=1}^\infty$ be an approximating sequence in $\mathscr{D}(T^*)$. Then $\langle\varphi|T\psi\rangle = \lim_{n\to\infty}\langle\varphi_n|T\psi\rangle$ for all $\psi\in\mathscr{D}(T)$, where $\langle\varphi_n|T.\rangle$ is bounded $(n = 1,2,\ldots)$. For closed $\mathscr{D}(T)$, the uniform boundedness principle implies that the limit functional $\langle\varphi|T.\rangle$ is also bounded. (9b): Let $\varphi'\in\overline{\mathscr{R}(T^*)}$. By (6), there is a sequence $(\varphi_n)_{n=1}^\infty$ in $\mathscr{D}(T^*)\cap\overline{\mathscr{R}(T)}$ with $\varphi' = \lim_{n\to\infty}T^*\varphi_n$. Then $\langle\varphi'|\psi\rangle = \lim_{n\to\infty}\langle\varphi_n|T\psi\rangle$ for all $\psi\in\mathscr{D}(T)$. Since $\varphi_n\in\overline{\mathscr{R}(T)}$ $(n = 1,2,\ldots)$ it follows by twofold application of the uniform boundedness principle that, for closed $\mathscr{R}(T)$, $(\varphi_n)_{n=1}^\infty$ is bounded and coverges weakly to some $\varphi\in\mathscr{R}(T)$. But weak convergence of $(\varphi_n)_{n=1}^\infty$ and $(T^*\varphi_n)_{n=1}^\infty$ to φ and φ' implies, as T^* is closed, that $\varphi\in\mathscr{D}(T^*)$ and $\varphi' = T^*\varphi$. Q.E.D

Remark 6: The conditions in (9) can be weakened to that the respective $\mathscr{M}\in\{\mathscr{D}(T), \mathscr{R}(T)\}$ are of second category in

$\overline{\mathscr{M}}$.[13] Also, (9) and Theorem 2 imply that $\mathscr{D}(T^*)$ [$\mathscr{R}(T^*)$] is closed if and only if $\mathscr{D}(\overline{P_T \cdot T})$ [$\mathscr{R}(\overline{P_T \cdot T})$] is so.

3. CLOSABILITY AND BOUNDEDNESS

The results of Sec. 2 allow very quick generalization of some classical tools to operators which are not densely defined. Particularly simple proofs emerge for the closed graph and the closed range theorems (Theorem 6).

Theorem 5: For $T\in\mathscr{L}(\mathscr{H})$, the following unprimed (primed) statements are equivalent:

(a)	T is closable.	(a')	T is bounded.
(b)	$\overline{\mathscr{D}(T^*)} = \mathscr{H}$.	(b')	$\mathscr{D}(T^*) = \mathscr{H}$.
(c)	The functional[14] $\|T.\|$ is lower semicontinuous.	(c')	The functional $\|T.\|$ is continuous.

If T is closable, then $\overline{T}^* = T^*$ and $T^{**} = \overline{T}P_T$. If T is bounded, then T^* is also bounded and $\|T^*\| = \|T\|$.

Proof: For (a)⇔(c) and (a')⇔(c'), see Theorem 6.1.3 and Sec. 4.1.2 in Ref. 13. Everything else follows, e.g., from Remark 2, Theorem 2, and standard arguments.

Theorem 6. For closed $T\in\mathscr{L}(\mathscr{H})$, the following unprimed (primed) statements are equivalent:

(a) $\mathscr{D}(T)$ is closed.	(a') $\mathscr{R}(T)$ is closed.
(b) $\mathscr{D}(T^*)$ is closed.	(b') $\mathscr{R}(T^*)$ is closed.
(c) T is bounded.	(c') $\mathscr{R}(T) = \mathscr{N}(T^*)^\perp$
(d) $\overline{T^*}$ is bounded.	(d') $\mathscr{R}(T^*) = \mathscr{N}(T)^\perp \cap \mathscr{D}(T)$

Proof: Elementary combination of (6), (9), Remark 6, and Theorem 5.

Remark 7: Most parts of Theorems 5 and 6 admit numerous variants and corollaries in view of (1), (6), Theorem 2, and Remark 6. Some of these are well known (see, e.g., Theorem III.5.28 in Ref. 12, or Theorem 2.12.3 in Ref. 15); others read as follows: For every $T\in\mathscr{L}(\mathscr{H})$, T^* is bounded if and only if $P_T \cdot T$ is so. For closed $T\in\mathscr{L}(\mathscr{H})$ with $\mathscr{R}(T)\subset \mathscr{D}(T^*)$, every subspace dense in $\mathscr{D}(T)$ is a core of T.

4. INVERSES

First we show how to obtain $(T^{-1})^*$ from T^* (and vice versa) for invertible $T\in\mathscr{L}(\mathscr{H})$. Theorem 8 then prepares for spectral analysis.

Theorem 7: For $T\in\mathscr{L}(\mathscr{H})$, the restriction of T^* to $\mathscr{D}(T^*)\cap \mathscr{R}(T)$, denoted by T^+, is closed and invertible. If T is invertible, then

$$(T^{-1})^* = (T^+)^{-1}P_T, \tag{10}$$

$$\mathscr{D}((T^{-1})^*) = \mathscr{R}(T^*) + \mathscr{D}(T)^\perp, \tag{11}$$

$$\overline{\mathscr{R}((T^{-1})^*)} = \mathscr{D}(T^*)\cap \overline{\mathscr{R}(T)}. \tag{12}$$

Proof: The first part is clear from Theorem 2 and (6). Since $\langle (T^{-1})^*\varphi \,|\, T\psi\rangle = \langle \varphi \,|\, \psi\rangle$ for all $\varphi\in\mathscr{D}((T^{-1})^*)$ and $\psi\in\mathscr{D}(T)$, we have $\mathscr{R}((T^{-1})^*)\subset\mathscr{D}(T^*)$ and hence $\mathscr{R}((T^{-1})^*)\subset\mathscr{D}(T^+)$. So relation $T^*(T^{-1})^*\subset P_T$ may be rewritten as $(T^{-1})^*\subset(T^+)^{-1}P_T$ and we obtain (10) by noting that $\mathscr{D}((T^+)^{-1}P_T) = \mathscr{R}(T^*) + \mathscr{D}(T)^\perp$, $\mathscr{R}(T^*)\subset\mathscr{D}((T^{-1})^*)$, and $\mathscr{D}(T)^\perp = \mathscr{N}((T^{-1})^*)$. Q.E.D.

Theorem 8: For $T\in\mathscr{L}(\mathscr{H})$, the following unprimed (primed) statements are equivalent:

(a) T has a closable inverse.	(a') T has a bounded inverse.
(b) $\overline{\mathscr{R}(T^*)} = \overline{\mathscr{D}(T)}$.	(b') $\mathscr{R}(T^*) = \overline{\mathscr{D}(T)}$.
(c) $\overline{P_T \cdot T}$ has an inverse.	(c') $\overline{P_T \cdot T}$ has a bounded inverse.

If T has a closable inverse, then $\overline{T^{-1}} = (\overline{P_T \cdot T})^{-1}P_{T\cdot}$. If T has a bounded inverse, then $(T^+)^{-1}$ is also bounded and $\|(T^+)^{-1}\| = \|(P_T \cdot T)^{-1}\|$.

Proof: (a)⇔(b), (a')⇔(b'): Clear from (11), (6), and Theorem 5. (b)⇔(c): Use (6). (b')⇔(c'): Use (b)⇔(c) and Theorem 6. For closable T^{-1}, compute $\overline{T^{-1}}$ by taking the inverse of $\overline{\mathscr{G}(T)}$, where

$\overline{\mathscr{G}(T)} = \{(\psi, \overline{P_T \cdot T}\,\psi + \varphi) \,|\, \psi\in\mathscr{D}(\overline{P_T \cdot T});\ \varphi\in\mathscr{D}(T^*)^\perp\}$

by Remark 5 and Eq. (7). For bounded T^{-1}, use (b'), (11), and the closed graph theorem; and combine Eq. (10) with the formula for T^{-1}. Q.E.D

Remark 8: Again, most known results on inverses can be recovered from Theorems 5–8. Whereas the fact that T^* (respectively T^+) and $(T^{-1})^*$ are Tseng generalized inverses[16] of each other suggests extensions involving generalized inverses in place of T^{-1}.

5. FORMALLY SELF-ADJOINT OPERATORS

To see in what sense f.s.a. operators generalize the notion of a symmetric (i.e., densely defined f.s.a.) operator, we have:

Theorem 9: For $T\in\mathscr{L}(\mathscr{H})$, the following are equivalent:
(a) T is f.s.a (b) $P_T T\subset T^*$. (c) $T = T^* + (1-P_T)T$.
(d) There is an $S\in\mathscr{L}(\mathscr{H})$ with $T = S + S^*$.

Proof: (a)⇒(b), (c)⇒(a): Use (1). (b)⇒(c), (d)⇒(a): Trivial. (a)⇒(d): Put $S = (1-\frac{1}{2}P_T)T$, verify that $S + S^*$ $= (1-\frac{1}{2}P_T)T + T^*(1-\frac{1}{2}P_T) = (1-\frac{1}{2}P_T)T + \frac{1}{2}T^*$, and apply (a)⇒(b). Q.E.D.

As indicated in Sec. 1, in physical applications a typical problem is to find self-adjoint extensions (e.g., as for a Hamiltonian), if any, of a given f.s.a. operator $T\in\mathscr{L}(\mathscr{H})$. This is well understood if $\overline{\mathscr{D}(T)} = \mathscr{H}$, so the problem reduces to finding symmetric extensions of T. Necessary (and presumably also sufficient) for this to be possible, is that T be closable. {Note that nonclosable f.s.a. T's exist [e.g., with $\mathscr{R}(T)\subset\mathscr{D}(T)^\perp = \mathscr{N}(T)^\perp$ and $\mathscr{D}(T)\neq\mathscr{N}(T)$] already for $\dim(\mathscr{D}(T)^\perp) = 1$. On the other hand, the operators $P_T T$ obtained from f.s.a. T's do not exhaust the set of all closable f.s.a. operators.}

The next result bounds the set of possible symmetric extensions and shows how to construct f.s.a. extensions from arbitrary ones.

Theorem 10: Let $T\in\mathscr{L}(\mathscr{H})$ be f.s.a. and closable. Then

$\{S \,|\, S\in\mathscr{L}(\mathscr{H});\ T\subset S;\ S \text{ is symmetric}\}$

$\subset \{\frac{1}{2}[T^* + (1-P_T)S + (T^* + (1-P_T)S)^*] \,|\, S\in\mathscr{L}(\mathscr{H});\ T\subset S;\ \overline{\mathscr{D}(T^*)\cap\mathscr{D}(S)} = \mathscr{H}\}$

$\subset \{S \,|\, S\in\mathscr{L}(\mathscr{H});\ T\subset S;\ S \text{ is f.s.a}\}$

$\supset \{\frac{1}{2}[T^* + (1-P_T)S + T^{**} + S^*(1-P_T)] \,|\, S\in\mathscr{L}(\mathscr{H});\ T\subset S\}$.

Proof: Let $S \in \mathscr{L}(\mathscr{H})$ be f.s.a. with $T \subset S$, so that $\mathscr{D}(T) \subset \mathscr{D}(S) \subset \mathscr{D}(S^*) \subset \mathscr{D}(T^*)$. Hence by Theorem 9, (b) and (c), $P_T S = P_T S^* + P_T(1 - P_S)S = P_T S^* \subset T^*$, $T^* + (1 - P_T)S = S$, and $\frac{1}{2}[T^* + (1 - P_T)S + (T^* + (1 - P_T)S)^*] = \frac{1}{2}(1 - P_S)S$. On considering the special case $\overline{\mathscr{D}(S)} = \mathscr{H}$, this proves the first inclusion. For the others, let $S \in \mathscr{L}(\mathscr{H})$ with $T \subset S$ only, so $T \subset T^* + (1 - P_T)S$ by Theorem 9(c). Since $T \in \mathscr{A}(T^* + (1 - P_T)S)$, Eq. (1) and condition $\overline{\mathscr{D}(T^*) \cap \mathscr{D}(S)} = \mathscr{H}$ give also $T \subset (T^* + (1 - P_T)S)^*$, whence the second inclusion. The last one is clear from Theorem 5. Q.E.D.

If applied to extensions $S = \overline{T} P_T + B(1 - P_T)$ of T, where $B \in \mathscr{L}(\mathscr{H})$ is bounded and everywhere defined, the last-mentioned set of operators in Theorem 10 is characterized as follows and, for bounded T, gives all self-adjoint extensions of T:

Theorem 11: Let $T \in \mathscr{L}(\mathscr{H})$ be f.s.a. and closable, let $B \in \mathscr{L}(\mathscr{H})$ bounded and self-adjoint, and put

$$\begin{aligned} S &= T^* + (1 - P_T)T^{**} + (1 - P_T)B(1 - P_T) \\ &= T^{**} + T^*(1 - P_T) + (1 - P_T)B(1 - P_T). \end{aligned}$$

Then S is a symmetric (bounded self-adjoint) extension of T if and only if $(1 - P_T)T$ is closable (T is bounded, in which case every self-adjoint extension of T has this form).

Proof: Since $\mathscr{D}(S) = \mathscr{D}(\overline{T}) + (\mathscr{D}(T^*) \cap \mathscr{D}(T)^\perp)$ by $T^{**} = \overline{T} P_T$ and Theorem 9(b), the stated equality for S and the "if and only if" part are clear from Theorems 10 (last inclusion) and 5. If T is bounded and $T \subset B$, then the previous relation $T^* + (1 - P_T)B = B$ (proof of Theorem 10) shows that $S = B$. Q.E.D.

Remark 9: If T is f.s.a. and $(1 - P_T)T$ is closable, then T is closable. But even self-adjoint operators may have restrictions T such that $(1 - P_T)T$ is not closable.

Remark 10: The operator T^* coincides with the extension of $P_T T$ as carried out in Krein's classical construction to prove that every bounded f.s.a. T admits a norm-preserving self-adjoint extension.

ACKNOWLEDGMENTS

The author would like to thank E. Gutknecht for some valuable discussions, and T. Kato for encouragement of an earlier version of this paper.

[1] T. Kato, in *Topics in Functional Analysis*, edited by I. Gohberg and M. Kac (Academic, New York, 1978), pp. 185–195.
[2] B. Simon, J. Funct. Anal. **28**, 377–385 (1978).
[3] M. A. Naimark, *Normed Algebras* (Wolters-Noordhoff, Groningen, 1972), paragraph 41.
[4] S. Goldberg, *Unbounded Linear Operators* (McGraw-Hill, New York, 1966).
[5] The alternative method of adjoint subspaces (see, e.g., Ref. 7) suffers from the drawback that concepts like closability and boundedness cease to be meaningful for the ensuing multivalued "operators".
[6] P. Pfeifer, "Tensor-product integration and abstract adiabatic decoupling" (in preparation).
[7] E. A. Coddington, Mem. Am. Math. Soc. **134** (1973).
[8] $\overline{S^{-1} P_T S}$ is bounded, e.g., if $\mathscr{R}(S) \subset \overline{\mathscr{D}(T)}$; if $\mathscr{D}(S^*) \subset \overline{\mathscr{D}(T)}$ (Theorem 8); or if $\overline{\mathscr{D}(T)} \subset \mathscr{D}(S^*)$ (Theorem 5).
[9] These conditions hold, e.g., if S is closed and $\mathscr{R}(S)$ has finite codimension with respect to $\overline{\mathscr{D}(T)}$ (Ref. 11).
[10] J. Weidmann, *Linear Operators in Hilbert Spaces* (Springer, New York, 1980).
[11] J.A.W. Van Casteren and S. Goldberg, Stud. Math. **38**, 125–130 (1970).
[12] T. Kato, *Perturbation Theory for Linear Operators*, 2nd ed. (Springer, Berlin, 1976).
[13] A. I. Plesner, *Spectral Theory of Linear Operators* (Ungar, New York, 1969), par. 4.5.1.
[14] Subadditive and with domain $\mathscr{D}(T)$.
[15] E. Hille and R. S. Phillips, *Functional Analysis and Semigroups* (Amer. Math. Soc., Providence, RI, 1957).
[16] M. Z. Nashed and G. F. Votruba, in *Generalized Inverses and Applications*, edited by M. Z. Nashed (Academic, New York, 1976), pp. 1–109.

The criticality problem for an exponential atmosphere

G. C. Pomraning
School of Engineering and Applied Science, University of California, Los Angeles, Los Angeles, California 90024

(Received 25 November 1980; accepted for publication 6 March 1981)

The criticality problem for a halfspace with an exponential single scatter albedo is analyzed. Analytic results are presented in the limits of very weak and very strong exponential behavior, and numerical results are given for general exponential behavior.

PACS numbers: 02.60.Nm, 42.10.Fa

I. INTRODUCTION

In 1966 Chamberlain and McElroy[1] gave an approximate solution to the problem of reflection of light from a semi-infinite atmosphere with a single scatter albedo c which approaches zero expontially with depth into the half-space, i.e.,

$$c(z) = c_0 e^{-z/s}, \tag{1}$$

where c_0 and s are positive constants, with z representing the spatial variable. This problem was reconsidered by Martin[2] who derived a singular integral equation which the specific intensity of radiation exiting the halfspace must satisfy. Mullikin and Siewert[3] have recently rederived this integral equation by constructing a set of singular eigenfunctions of the equation of transfer and using the associated full-range orthogonality relationship. For sufficiently rapid exponential behavior of the single scatter albedo (i.e., for small s) they obtained excellent numerical results from this equation using a particular collocation method, the so-called F–N method. In a subsequent paper by Larsen, Pomraning, and Badham[4] the F–N method was shown to break down for weak exponential behavior ($s \gg 1$). It was conjectured in that paper that these singular eigenfunctions may not be complete for s sufficiently large, thus accounting for the poor numerical results from the F–N method for large s. It was proved in that paper that these eigenfunctions are in fact complete for s sufficiently small. More precisely, it was shown that completeness is extant if c_0 and s satisfy the inequality

$$\frac{c_0}{2}\left\{\left(\frac{s}{s+1}\right)\ln(1+2s) + \min\left[\frac{\pi s}{s+1} + \left(\frac{s}{s+1}\right)^{1/2}, \pi\right]\right\} < 1. \tag{2}$$

In a recent paper by Garcia and Siewert[5] a modification to the F–N method extended the range of parameters c_0 and s for which good numerical results are obtained. This rather confusing situation was partially clarified by Larsen and Mullikin[6] who very recently proved that the continuum eigenfunctions are complete, and the corresponding singular integral equation possesses a unique solution, for all subcritical atmospheres. That is, if the parameters c_0 and s are such that a solution exists to the diffuse reflection problem first considered by Chamberlain and McElroy,[1] then the singular eigenfunction technique and the associated singular integral equation provide the solution. A separate, still not entirely understood, question is the applicability of the F–N method as a solution technique for solving this singular integral equation.

The purpose of this short note is to give explicit results for the region in the c_0–s plane for which a solution to the diffuse reflection problem exists (i.e., for which the halfspace is subcritical). For small and large s, we obtain analytic relationships between c_0 and s defining the criticality condition, and for intermediate s we give numerical results based upon the use of the Rayleigh quotient applied to non-self-adjoint operators.

2. ANALYSIS

The eigenvalue equation to be analyzed is

$$\mu \frac{\partial \psi(z,\mu)}{\partial z} + \psi(z,\mu) = \frac{c_0 e^{-z/s}}{2} \int_{-1}^{1} d\mu' \psi(z,\mu'), \tag{3}$$

with boundary conditions

$$\psi(0,\mu) = 0, \quad 0 \leqslant \mu \leqslant 1, \tag{4}$$

$$\psi(\infty,\mu) = 0, \quad -1 \leqslant \mu \leqslant 0. \tag{5}$$

Given s, we seek the smallest value of c_0 such that a non-zero solution exists to Eqs. (3)–(5). Physically these equations describe a monoenergetic critical halfspace with isotropic reemission according to a single scatter albedo given by Eq. (1), and $\psi(z,\mu)$ is the specific intensity of radiation at optical depth z streaming at an angle $\cos^{-1}(\mu)$ with respect to the z axis. The smallest value of c_0 corresponds to the fundamental mode [the physically meaningful mode in which the eigenfunction $\psi(z,\mu)$ is non-negative].

For $s \gg 1$ the albedo is a slowly varying function of space and Eq. (3) then describes transport in a source-free, essentially homogeneous (constant properties) medium which is well known to be equivalently described by asymptotic diffusion theory[7–9]

$$D \frac{\partial^2 \phi(z)}{\partial z^2} + (c_0 e^{-z/s} - 1)\phi(z) = 0, \tag{6}$$

where $\phi(z)$ is given by

$$\phi(z) = 2\pi \int_{-1}^{1} d\mu\, \psi(z,\mu). \tag{7}$$

The diffusion coefficient D follows from

$$D = \frac{1-c}{\kappa^2}; \frac{2\kappa}{c} = \ln\left(\frac{1+\kappa}{1-\kappa}\right). \tag{8}$$

Clearly D depends upon position since $c = c(z)$ in Eq. (8). However, for $s \gg 1$ this dependence is very weak and can be neglected in the limit $s \to \infty$. The boundary conditions on

 0022-2488/81/081623-03$01.00 © 1981 American Institute of Physics

Eq. (6) can be taken as

$$\phi(0) = \phi(\infty) = 0, \tag{9}$$

i.e., we can neglect the extrapolation distance at $z = 0$ for large s. To solve Eq. (6) we change variables according to

$$x = 2s(c_0/D)^{1/2} e^{-z/2s}; \quad \theta(x) = \phi(z), \tag{10}$$

which gives

$$x^2 \frac{\partial^2 \theta(x)}{\partial x^2} + x \frac{\partial \theta(x)}{\partial x} + (x^2 - 4s^2/D)\theta(x) = 0. \tag{11}$$

Equation (11) is Bessel's equation of order $2s/\sqrt{D}$, and the solution for $\phi(z)$ which vanishes at infinity is[10]

$$\phi(z) = (\text{const}) J_{2s/D^{1/2}}(2s(c_0/D)^{1/2} e^{-z/2s}). \tag{12}$$

Applying the boundary condition at $z = 0$, setting $D = 1/3$ since $c_0 \approx 1$ for $s \gg 1$, we find the criticality condition

$$J_{2(3s)^{1/2}}(2(3c_0)^{1/2} s) = 0. \tag{13}$$

Using the properties of Bessel functions with simultaneous large order and large argument[10] we obtain the final result

$$(c_0 - 1)^{3/2} s \underset{s\to\infty}{\longrightarrow} x_0^{3/2} 3^{-1/2} = 2.064\,125\,31, \tag{14}$$

where $-x_0$ is the first zero of the Airy function $Ai(x)$ given by[10]

$$x_0 = 2.338\,107\,41. \tag{15}$$

For other values of s we find it convenient to rewrite Eqs. (3)–(5) in the equivalent integral form (Peierls' equation). A formal integration of Eq. (3) yields this equation:

$$\phi(z) = \frac{c_0}{2} \int_0^\infty dz'\, E_1(|z - z'|) e^{-z'/s} \phi(z'), \tag{16}$$

where $E_1(z)$ is the first-order exponential integral.[10] The equation adjoint to Eq. (16) is

$$\phi^*(z) = \frac{c_0}{2} e^{-z/s} \int_0^\infty dz'\, E_1(|z - z'|) \phi^*(z'), \tag{17}$$

and it is clear that $\phi^*(z)$, the adjoint function, is related to $\phi(z)$ by

$$\phi^*(z) = e^{-z/s} \phi(z). \tag{18}$$

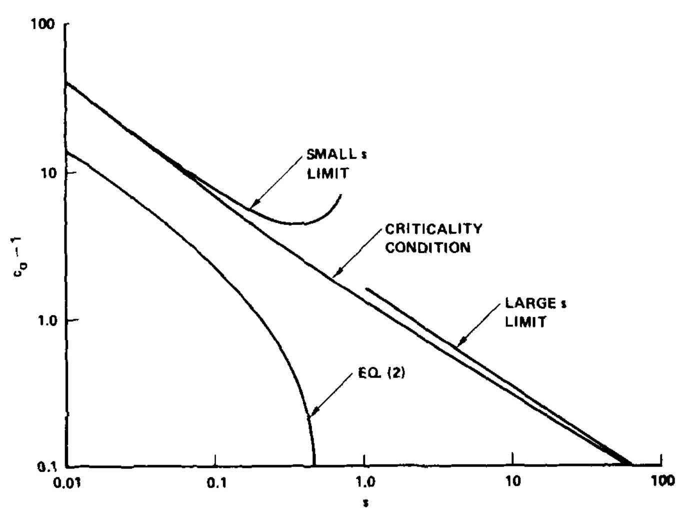


FIG. 1. c_0 vs s.

Given s, we consider c_0 to be the eigenvalue and the standard Rayleigh quotient for estimating this eigenvalue is

$$\bar{c}_0 = \frac{2\int_0^\infty dz\, \bar{\phi}^*(z)\bar{\phi}(z)}{\int_0^\infty dz\, \bar{\phi}^*(z) \int_0^\infty dz'\, E_1(|z - z'|) e^{-z'/s} \bar{\phi}(z')}, \tag{19}$$

where $\bar{c}_0$ is the variational estimate of c_0, and $\bar{\phi}(z)$ and $\bar{\phi}^*(z)$, the trial functions, are envisioned as first-order approximations to $\phi(z)$ and $\phi^*(z)$. The value $\bar{c}_0$ given by Eq. (19) differs from the eigenvalue c_0 by terms quadratic in the first-order errors in $\bar{\phi}(z)$ and $\bar{\phi}^*(z)$. In view of Eq. (18) it is reasonable to write the adjoint trial function as

$$\bar{\phi}^*(z) = \bar{\phi}(z) e^{-z/s}, \tag{20}$$

and Eq. (19) then becomes

$$\bar{c}_0 = \frac{2\int_0^\infty dz\, \bar{\phi}^2(z) e^{-z/s}}{\int_0^\infty dz\, \bar{\phi}(z) e^{-z/s} \int_0^\infty dz'\, E_1(|z - z'|) e^{-z'/s} \bar{\phi}(z')}. \tag{21}$$

Equation (21) can be used to obtain the analytic dependence of c_0 upon s for small s. For $s \ll 1$, the integrands only contribute to the integrals in the vicinity of $z = z' = 0$, and hence

$$c_0 \underset{s\to 0}{\longrightarrow} \bar{c}_0 \underset{s\to 0}{\longrightarrow} \frac{2\int_0^\infty dz\, e^{-z/s}}{\int_0^\infty dz\, e^{-z/s} \int_0^\infty dz'\, E_1(|z - z'|) e^{-z'/s}}, \tag{22}$$

which gives

$$c_0 \underset{s\to 0}{\longrightarrow} 2[s \ln(1/s)]^{-1}. \tag{23}$$

For intermediate values of s, we use Eq. (21) with a trial function

$$\bar{\phi}(z) = \sum_{n=0}^{N} a_n z^n e^{-z}, \tag{24}$$

where the a_n are determined by demanding that the first variation of $\bar{c}_0$ vanish within this restricted function space. This trial function is chosen because it falls off exponentially for large z, and this is the proper behavior for $\phi(z)$ since the halfspace becomes a pure absorber ($c = 0$) for large z. This leads to the matrix eigenvalue problem

$$\sum_{n=0}^{N} A_{mn} a_n = \bar{c}_0 \sum_{n=0}^{N} B_{mn} a_n, \tag{25}$$

where the matrix elements are given by

$$A_{mn} = 2(m+n)!\left(\frac{s}{2s+1}\right)^{m+n+1}, \tag{26}$$

$$B_{mn} = M_{mn}\left(\frac{s+1}{s}, \frac{s+1}{s}\right). \tag{27}$$

Here the function $M_{mn}(x,y)$ is defined as

$$M_{mn}(x,y) = \int_0^\infty dz \int_0^\infty dz'\, z^m z'^n\, E_1(|z - z'|) e^{-xz} e^{-yz'}. \tag{28}$$

TABLE I. c_0 vs s as determined from Eq.(25).

s	c_0
0.01	42.740
0.02	24.956
0.04	14.937
0.07	10.096
0.1	7.9661
0.2	5.1964
0.4	3.5672
0.7	2.7450
1.	2.37054
2.	1.86437
4.	1.54979
7.	1.38307
10.	1.30461
20.	1.195328
40.	1.125244
70.	1.08741
100.	1.06947

For $m = n = 0$ we have

$$M_{00}(x,y) = \frac{1}{x+y}\left[\frac{\ln(1+x)}{x} + \frac{\ln(1+y)}{y}\right], \tag{29}$$

and the higher-order indices results follow from recurrence relationships which can be developed from Eq. (29) in conjunction with the observations

$$M_{m+1,n}(x,y) = -\frac{\partial}{\partial x} M_{mn}(x,y), \tag{30}$$

$$M_{m,n+1}(x,y) = -\frac{\partial}{\partial y} M_{mn}(x,y). \tag{31}$$

Equation (25) was solved by a standard matrix eigenenvalue routine, and Table I gives results for various values of s. The values tabulated are believed to be accurate to the number of digits given. We were restricted as to the value of N which could be used ($N < 20$) due to numerical roundoff. Figure 1 displays these same results, together with the large and small s limiting forms, Eqs. (14) and (23), and the earlier estimate of the criticality (completeness) condition given by Eq. (2).

ACKNOWLEDGMENTS

This work was partially supported by the National Science Foundation. The author is grateful to Mr. I. R. Shokair who performed the numerical calculations.

[1]J. B. Chamberlain and M. B. McElroy, Astrophys. J. **144**, 1148 (1966).
[2]B. J. Martin, SIAM J. Appl. Math. **20**, 703 (1971).
[3]T. W. Mullikin and C. E. Siewert, Ann. Nucl. Energy 7, 205 (1980).
[4]E. W. Larsen, G. C. Pomraning, and V. C. Badham, J. Math. Phys. **21**, 2448 (1980).
[5]R. D. M. Garcia and C. E. Siewert, "Radiative Transfer in Inhomogeneous Atmospheres-Numerical Results," J. Quant. Spectros. Rad. Transfer (in press).
[6]E. W. Larsen and T. W. Mullikin, "Linear Transport in an Exponential Atmosphere," J. Math. Phys. **22**, 856 (1981).
[7]B. Davidson, *Neutron Transport Theory* (Oxford U. P., London, 1957).
[8]A. M. Winslow, Nucl. Sci. Eng, **32**, 101 (1968).
[9]G. C. Pomraning, *The Equations of Radiation Hydrodynamics* (Pergamon, Oxford, 1973).
[10]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (U. S. Goverment Printing Office, Washington, D. C. 1964).

Conservation laws and discrete symmetries in classical mechanics

M. Lutzky
10111 Quinby Street, Silver Spring, Maryland 20901

(Received 3 September 1980; accepted for publication 17 April 1981)

A method is given for deriving conserved quantities from discrete symmetries in classical mechanics.

PACS numbers: 03.20. + i

It seems to be widely believed that conservation laws in classical mechanics can be deduced from symmetries of the system only when these symmetries are continuous; this is in contrast to the quantum mechanical situation, where conserved quantities can sometimes be associated with discrete symmetries (for example, parity and coordinate inversion). We will show that, in fact, conserved quantities can be found from discrete symmetries even in classical mechanics; our method utilizes a recent result concerning nonequivalent Lagrangians.

Consider the second order dynamical system

$$\ddot{q}_l = \alpha_l(q,\dot{q},t), \quad (l = 1,2,...,N) \tag{1}$$

and suppose that two distinct Lagrangians $\tilde{L}(q,\dot{q},t)$ and $L(q,\dot{q},t)$ both lead to (1). It has been shown in Ref. 1 that a constant of the motion is given by

$$\Phi = \tilde{D}/D, \tag{2}$$

where

$$\tilde{D} = \det\{\partial^2\tilde{L}/\partial\dot{q}_i\partial\dot{q}_j\}, \quad D = \det\{\partial^2 L/\partial\dot{q}_i\partial\dot{q}_j\}.$$

If the Lagrangians are equivalent [that is, if the Lagrangians differ only by the total time derivative of some function $f(q,t)$] then the result is trivial, since, in this case, we have $\tilde{D}/D = 1$. Of interest to us are situations where the Lagrangians are nonequivalent. We note further that the validity of the theorem does not require that $\tilde{L}$ be derivable from L by means of a symmetry transformation. The case in which $\tilde{L}$ does arise from L through the action of a continuous symmetry group has been treated in Ref. 1, where it has been shown that knowledge of the group generators allows the determination of conserved quantities. Here we will be particularly concerned with the case in which $\tilde{L}$ is related to L through a *discrete* symmetry.

Let the q_l, t variables be related to the Q_l, T variables through the transformation[2]

$$q_l = q_l(Q,T), \quad t = t(Q,T), \; l = 1,2,...N. \tag{3}$$

Then the time derivatives along trajectories transform according to

$$\dot{q}_l = \left\{\frac{dq_l}{dT}\right\}\bigg/\left\{\frac{dt}{dT}\right\}$$

$$= \left\{\frac{\partial q_l}{\partial T} + \frac{\partial q_l}{\partial Q_k}Q'_k\right\}\bigg/\left\{\frac{\partial t}{\partial T} + \frac{\partial t}{\partial Q_k}Q'_k\right\},$$

$$\ddot{q}_l = \left\{\frac{\partial \dot{q}_l}{\partial T} + \frac{\partial \dot{q}_l}{\partial Q_k}Q'_k + \frac{\partial \dot{q}_l}{\partial Q'_k}Q''_k\right\}\bigg/\left\{\frac{\partial t}{\partial T} + \frac{\partial t}{\partial Q_k}Q'_k\right\},$$

where $Q'_k = dQ_k/dT$. Let us assume that (1) is invariant under (3); this means that (1) takes the form

$$Q''_l = \alpha_l(Q,Q',T), \quad l = 1,2...,N, \tag{4}$$

when expressed in terms of the Q_l, T variables. If the equations (1) are the Euler equations associated with the action integral $\int_{t_1}^{t_2} L(q,\dot{q},t)\,dt$, then by expressing this integral in terms of Q_l, T we can show that Eqs. (4) are the Euler equations arising from

$$\int_{T_1}^{T_2} \tilde{L}(Q,Q',T)\,dT,$$

where

$$\tilde{L}(Q,Q',T) = L(q,\dot{q},t)\left(\frac{\partial t}{\partial T} + \frac{\partial t}{\partial Q_k}Q'_k\right). \tag{5}$$

Here the $(q,\dot{q},t)$ are considered to be expressed in terms of the (Q,Q',T), using the relations (3) given above.[3] We conclude that $\tilde{L}(q,\dot{q},t)$ and $L(q,\dot{q},t)$ both lead to (1), so that a conserved quantity for (1) can be obtained from (2). It is natural to associate this conserved quantity with the discrete symmetry (3).

We illustrate this technique with the simplest possible example, namely the one-dimensional free particle, with Lagrangian $L = \frac{1}{2}\dot{q}^2$ and equation of motion $\ddot{q} = 0$. Let Q and T be defined by the discrete transformation

$$q = Q/(1+Q),$$
$$t = T/(1+Q).$$

Then

$$\dot{q} = Q'/(1 + Q - TQ'),$$
$$\ddot{q} = Q''\{1+Q\}^3/\{1+Q-TQ'\}^3.$$

Under this transformation $\ddot{q} = 0$ goes to $Q'' = 0$, so that the free particle equation is invariant. The new Lagrangian is obtained from (5), and assumes the form

$$\tilde{L}(Q,Q',T) = \tfrac{1}{2}\{Q'\}^2/\{1+Q-TQ'\}\{1+Q\}^2.$$

It is easy to verify that $\tilde{L}(Q,Q',T)$ does indeed yield $Q'' = 0$. We have therefore demonstrated that $L = \frac{1}{2}\dot{q}^2$ and

$$\tilde{L} = \tfrac{1}{2}\dot{q}^2/\{1+q-t\dot{q}\}\{1+q\}^2$$

both yield $\ddot{q} = 0$; then (2) gives $\Phi = \{1+q-t\dot{q}\}^{-3}$, which is indeed a conserved quantity for the system.

This procedure may be applied to any system posessing a discrete symmetry, although the conserved quantity so obtained can be nontrivial only if the Lagrangians L and $\tilde{L}$ turn out to be nonequivalent. It is thus clear that the applicability of the method described here depends crucially on the existence of inequivalent Lagrangians; in this connection we

 0022-2488/81/081626-02$01.00 © 1981 American Institute of Physics

note that it has occasionally been implied in the literature that inequivalent Lagrangians cannot exist for (nontrivial) N-dimensional systems (see, e.g., Ref. 4). That this view is in error is amply demonstrated in an unjustly neglected paper of Jesse Douglas on the inverse problem of the variational calculus.[5] The question considered is that of finding Lagrangians which yield specified Euler equations, and a complete analysis is given of the various cases which may arise for the general two-dimensional system $\ddot{x} = f(t,x, y,\dot{x}, \dot{y})$, $\ddot{y} = g(t,x, y,\dot{x}, \dot{y})$. Of particular interest here is the conclusion that there exist classes of two-dimensional systems for which appropriate Lagrangians can be found as solutions of a set of completely integrable linear partial differential equations. The existence of these solutions can be proven, along with the determination of their generality, that is, the number and nature of the arbitrary functions or constants which are involved. For example, it is proved that the set of possible Lagrangians yielding $y\ddot{y} = 1 + \dot{y}^2 + \dot{x}^2$, $\ddot{x} = 0$ is doubly infinite, each member being determined by the specification of two arbitrary functions. In this classification, Lagrangians which differ by total time derivatives are not considered to be distinct, so that inequivalent Lagrangians certainly exist in this case. Since Douglas does not explicitly solve any of the systems of partial differential equations which define the Lagrangians, his demonstration is in the nature of an existence proof; however, for the purpose of showing that the method of this paper has potential for application to N-dimensional systems, existence of inequivalent Lagrangians is sufficient. We conclude by remarking that is is of considerable theoretical and conceptual interest that procedures exist, even in classical mechanics, for determining conserved quantities from discrete symmetries.

Note added in proof: It has recently been demonstrated by Marmo, Saletan, Simoni, and Zaccaria [J. Math. Phys. **22**, 835 (1981)] that conserved quantities may be associated with discrete symmetries in both Hamiltonian and Liouville mechanics.

[1]M. Lutzky, Phys. Lett. A **75**, 8 (1979).

[2]The transformation (3) is to be regarded as a transformation of trajectories; thus, given a trajectory $Q_l = Q_l(T)$, $l = 1,2,\ldots,N$ in (Q,T) space, (3) associates with each point of this trajectory a point in (q,t) space, thereby defining a trajectory $q_l = q_l(t)$ in (q,t) space. Since the total derivatives along these trajectories transform in a well-defined way [see the equations immediately following (3)], the procedure can be used to investigate the manner in which differential equations such as (1) transform under operations such as (3).

[3]A special case $N = 1$ of Eq. (5) is given in Gel'fand and Fomin, *Calculus of Variations* (Prentice-Hall, Englewood Cliffs, N.J., 1963), p. 30; see also M. Lutzky, J. Phys. A **11**, 249 (1978).

[4]G. Rosen, *Formulations of Classical and Quantum Dynamical Theory* (Academic, New York, 1969).

[5]J. Douglas, Trans. Amer. Math. Soc. **50**, 71 (1941).

On reduction of the four-dimensional harmonic oscillator

Toshihiro Iwai
Department of Applied Mathematics, Gifu University, Kagamigahara, Gifu 504, Japan

(Received 4 November 1980; accepted for publication 6 March 1981)

This paper deals with reduction of the four-dimensional harmonic oscillator by use of a one-parameter subgroup U(1) of the symmetry group SU(4), U(1) being the symmetry subgroup generated by an "angular momentum." The angular momentum determines in the energy surface S^7 an "energy-momentum" manifold $S^3 \times S^3$ on which a subgroup $SU(2) \times SU(2)$ of SU(4) acts. The reduction process yields a manifold $S^3 \times S^2 = S^3 \times S^3/U(1)$ on which SO(4) acts effectively.

PACS numbers: 03.20.+ i, 02.20. − a

1. INTRODUCTION

The harmonic oscillator, one of the well-known dynamical systems whose symmetries are studied to the full extent, is no longer interesting by itself. However, new use for it can be found when it is associated with other dynamical systems. For example, the four-dimensional harmonic oscillator is employed in Ref. 1 for analyzing the three-dimensional Kepler problem. There is an analogy in quantum mechanics. Ikeda and Miyachi[2] studied the symmetry of the three-dimensional hydrogen atom, the correspondent to the Kepler problem, by using the symmetry of the four-dimensional harmonic oscillator.

The purpose of this paper is to study reduction of the symmetry of the four-dimensional harmonic oscillator. The results to be obtained will be utilized in the next paper[3] for getting insight into the symmetry of the three-dimensional Kepler problem in the large.

As regards reduction of dynamical systems, Marsden and Weinstein presented a theory of reduction of symplectic manifolds with symmetry.[4,5] A general setting for reduction is discussed in Ref. 6.

The material of this paper is organized as follows. Section 2 contains a review of the symmetry of harmonic oscillators which are concisely described in the complex vector space $\mathbb{C}^n$. The symmetry Lie algebra formed by constants of the motion is identified with anti-Hermitian matrices with vanishing traces. The linear group SU(n) acts on $\mathbb{C}^n$ as a group of symmetry transformations. Particular interest will center on the kinematical symmetry for the four-dimensional harmonic oscillator. The set of 4×4 antisymmetric matrices, isomorhpic with the kinematical symmetry Lie algebra, is the Lie algebra of $SU(2) \times SU(2)$ as well as of SO(4). A question as to which of $SU(2) \times SU(2)$ and SO(4) is the kinematical symmetry group will be cleared up. Though the answer is well known, the question deserves mention for comparison with the case of the symmetry group which will appear in Sec. 4. In the succeeding sections only four-dimensional harmonic oscillators are treated.

Section 3 is concerned with the subgroup that commutes with a one-parameter group U(1), where U(1) is generated by the antisymmetric matrix corresponding to an angular momentum, a constant of the motion. A subgroup $SU(2) \times SU(2) \times U(1)$ will be obtained.

Section 4 deals with invariant manifolds in $\mathbb{C}^4$ under the action of the subgroup $SU(2) \times SU(2) \times U(1)$ obtained in Sec. 4; those manifolds are determined by assigning values of the angular momentum mentioned above. Of particular interest is a hypersurface $S^3 \times S^3$ of the energy surface S^7, which is obtained by setting the angular momentum equal to zero. A subgroup $SU(2) \times SU(2)$ of SU(4) acts on $S^3 \times S^3 \subset \mathbb{C}^4$ effectively.

Finally, Sec. 5 is devoted to obtaining a reduced manifold $S^3 \times S^2$ defined by $S^3 \times S^3/U(1)$. Accompanying the reduction to the manifold $S^3 \times S^2$, the symmetry subgroup $SU(2) \times SU(2)$ reduces to SO(4), acting on $S^3 \times S^2$.

2. REVIEW OF THE SYMMETRY OF HARMONIC OSCILLATORS

Suppose that we have an n-dimensional harmonic oscillator. Let[7] (x_j, p_j) be the Cartesian coordinates of $\mathbb{R}^n \times \mathbb{R}^n = T^*\mathbb{R}^n$, the cotangent bundle of $\mathbb{R}^n$. By introducing the coordinates

$$z_j = \lambda x_j + i p_j \quad (i = \sqrt{-1}), \tag{2.1}$$

λ being a positive constant, we endow $T^*\mathbb{R}^n$ with the structure of an n-dimensional complex vector space $\mathbb{C}^n$. Then the Hamiltonian H and the canonical two-form ω are written in the form

$$H = \frac{1}{2}\sum(p_j)^2 + (\lambda^2/2)\sum(x_j)^2 = \frac{1}{2}\sum|z_j|^2, \tag{2.2}$$

$$\omega = dp_j \wedge dx_j = (1/2\lambda i) dz_j \wedge d\bar{z}_j. \tag{2.3}$$

The time evolution of the dynamical system is generated by the vector field X_H determined by

$$i(X_H)\omega = -dH, \tag{2.4}$$

where $i(X_H)$ denotes the interior product by X_H. The vector field X_H is easily integrated to give $z_j(t) = e^{i\lambda t} z_j(0)$, from which $z_j \bar{z}_k$ proves to be a constant of the motion.

Let (C_{kj}) be a constant matrix. Define a function F by

$$F = (1/2i) C_{kj} z_j \bar{z}_k. \tag{2.5}$$

Of course, F is a constant of the motion. We impose the condition that F is real-valued. Then (C_{kj}) becomes an anti-Hermitian matrix, so that it is expressed in the form

$$C_{kj} = A_{kj} + iB_{kj}, \tag{2.6}$$

where (A_{kj}) and (B_{kj}) are antisymmetric and symmetric real matrices, respectively. According to (2.6), F reads

$$F = \tfrac{1}{2}A_{kj}L_{jk} + \tfrac{1}{2}B_{kj}Q_{jk}, \tag{2.7}$$

where L_{jk} and Q_{jk} are the imaginary and real parts of $z_j\bar{z}_k$, respectively

$$L_{jk} = \operatorname{Im} z_j\bar{z}_k = -\lambda x_j p_k + \lambda x_k p_j, \tag{2.8a}$$

$$Q_{jk} = \operatorname{Re} z_j\bar{z}_k = p_j p_k + \lambda^2 x_j x_k. \tag{2.8b}$$

As is well known, the constants of the motion (2.7) with $\operatorname{tr}(B_{kj}) = 0$ form the symmetry Lie algebra for the harmonic oscillator with respect to the Poisson bracket.

The infinitesimal canonical transformation X_F determined by (2.4) with F substituted for H takes the form

$$X_F = -\lambda C_{jk} z_k \frac{\partial}{\partial z_j} - \lambda\, \overline{C_{jk} z_k}\, \frac{\partial}{\partial \bar{z}_j}. \tag{2.9}$$

Let K be a constant of the motion with a coefficient matrix (D_{jk}). Then one has commutation relations

$$\begin{aligned}(1/\lambda)X_{\{F,K\}} &= -[(1/\lambda)X_F,(1/\lambda)X_K] \\ &= (C_{jh}D_{hk} - D_{jh}C_{hk})z_k \frac{\partial}{\partial z_j} \\ &\quad + \overline{(C_{jh}D_{hk} - D_{jh}C_{hk})z_k}\, \frac{\partial}{\partial \bar{z}_j},\end{aligned} \tag{2.10}$$

where $\{F, K\}$ denotes the Poisson bracket of F and K. Equations (2.9) and (2.10) show that the mappings

$$F \to -(1/\lambda)X_F, \tag{2.11a}$$

$$-(1/\lambda)X_F \to (C_{jk}) \tag{2.11b}$$

are Lie algebra homomorphisms. Therefore, we can take anti-Hermitian matrices (C_{jk}) with vanishing traces to be the symmetry Lie algebra. This Lie algebra, the Lie algebra of $\mathrm{SU}(n)$, will be integrated to give a symmetry group of transformations.

Given an anti-Hermitian matrix $C = (C_{jk})$ with $\operatorname{tr} C = 0$, one can construct a one-parameter subgroup $\exp tC$ of $\mathrm{SU}(n)$. The subgroup acts on $\mathbb{C}^n$, as a symmetry group preserving ω and H, in the form

$$z \to \exp(-tC)z, \quad \bar{z} \to \overline{\exp(-tC)z}, \tag{2.12}$$

where $z = (z_j)$ and $\bar{z} = (\bar{z}_j)$ are column vectors. Conversely, (2.12) gives rise to the infinitesimal canonical transformation

$$\frac{1}{\lambda}X_F f = \frac{d}{dt} f(\exp(-tC)z, \overline{\exp(-tC)z})\Big|_{t=0}. \tag{2.13}$$

Here f is any function of z and $\bar{z}$, and F is the constant of the motion with the coefficient matrix C. Since $\mathrm{SU}(n)$ is connected and simply connected, it is generated by all the matrices $\exp tC$ with $\operatorname{tr} C = 0$. Therefore, it acts on $\mathbb{C}^n$ as the symmetry group in the same manner as (2.12).

Now we concentrate on four-dimensional harmonic oscillators. The kinematical symmetry is expressed by the constants of the motion $L = \frac{1}{2}A_{kj}L_{jk} = -\lambda A_{kj}x_j p_k$ [see (2.7)]. According to (2.11), we treat the symmetry in terms of antisymmetric matrices (A_{kj}). As is well known, the 4×4 antisymmetric matrices are the Lie algebra of SO(4) as well as of $\mathrm{SU}(2)\times\mathrm{SU}(2)$. The basis and the commutation relations are given in the following:[8]

$$2M_1 = \begin{pmatrix} & & & -1 \\ & & 1 & \\ & -1 & & \\ 1 & & & \end{pmatrix}, \quad 2M_2 = \begin{pmatrix} & & -1 & \\ & & & -1 \\ 1 & & & \\ & 1 & & \end{pmatrix}, \quad 2M_3 = \begin{pmatrix} & -1 & & \\ 1 & & & \\ & & & 1 \\ & & -1 & \end{pmatrix}, \tag{2.14}$$

$$2N_1 = \begin{pmatrix} & & & 1 \\ & & 1 & \\ & -1 & & \\ -1 & & & \end{pmatrix}, \quad 2N_2 = \begin{pmatrix} & & -1 & \\ & & & 1 \\ 1 & & & \\ & -1 & & \end{pmatrix}, \quad 2N_3 = \begin{pmatrix} & -1 & & \\ 1 & & & \\ & & & -1 \\ & & 1 & \end{pmatrix}, \tag{2.15}$$

$$[M_j, M_k] = \epsilon_{jkh}M_h, \quad [N_j, N_k] = \epsilon_{jkh}N_h, \quad [M_j, N_k] = 0. \tag{2.16}$$

Here ϵ_{jkh} are Eddington's epsilons. The set of $\exp tM_j$, $j = 1,2,3$ and $\exp tN_j$, $j = 1,2,3$ generate, respectively, the matrices

$$\begin{pmatrix} a_1 & -a_2 & -a_3 & -a_4 \\ a_2 & a_1 & a_4 & -a_3 \\ a_3 & -a_4 & a_1 & a_2 \\ a_4 & a_3 & -a_2 & a_1 \end{pmatrix}, \tag{2.17a}$$

$$\begin{pmatrix} b_1 & b_2 & b_3 & b_4 \\ -b_2 & b_1 & b_4 & -b_3 \\ -b_3 & -b_4 & b_1 & b_2 \\ -b_4 & b_3 & -b_2 & b_1 \end{pmatrix}, \tag{2.17b}$$

with $\sum_{k=1}^{4}(a_k)^2 = \sum_{k=1}^{4}(b_k)^2 = 1$.

This fact may be proved by straightforward calculation. Conversely, the matrix (2.17a) [respectively, (2.17b)] is broken up into a product of $\exp tM_j$, $j = 2,3$ (respectively, $\exp tN_j$, $j = 2,3$). In fact, denoting (2.17a) and (2.17b) by $M(a_k)$ and $N(b_k)$, respectively and letting

$$\begin{aligned} a_1(= b_1) &= \cos\frac{\theta}{2}\cos\frac{\psi+\varphi}{2}, \\ a_2(= b_2) &= \cos\frac{\theta}{2}\sin\frac{\psi+\varphi}{2}, \\ a_3(= b_3) &= \sin\frac{\theta}{2}\cos\frac{\psi-\varphi}{2}, \end{aligned} \tag{2.18}$$

$$a_4(=b_4)=\sin\frac{\theta}{2}\sin\frac{\psi-\varphi}{2},$$

we obtain the decomposition

$$M(a_k)=(\exp\varphi M_3)(\exp\theta M_2)(\exp\psi M_3), \tag{2.19}$$

$$N(b_k)=(\exp(-\varphi N_3))(\exp(-\theta N_2))(\exp(-\psi N_3)). \tag{2.20}$$

Of course, $\exp tM_1$ and $\exp tN_1$ are special cases of (2.19) and (2.20), respectively. Because $\Sigma(a_k)^2=\Sigma(b_k)^2=1$, and because of the commutativity of $M(a_k)$ and $N(b_k)$, the matrices $M(a_k)$ and $N(b_k)$ form the group SU(2)×SU(2).

Now, SU(2)×SU(2) acts on $\mathbb{C}^4$ in the same manner as in (2.12), leaving ω and H invariant. We now consider whether the action of SU(2)×SU(2) is effective or not. Since (2.17) are real matrices, SU(2)×SU(2) induces the same action on each $\mathbb{R}^4$ of the real and imaginary parts of $\mathbb{C}^4$. Hence we choose to treat the action restricted to $\mathbb{R}^4$, the real part of $\mathbb{C}^4$.

Introducing the complex variables by

$$u=x_1+ix_2,\quad v=x_3+ix_4, \tag{2.21}$$

$$\begin{cases}\xi_1=a_1+ia_2,\\ \eta_1=a_3+ia_4,\end{cases}\quad \begin{cases}\xi_2=b_1+ib_2,\\ \eta_2=b_3+ib_4,\end{cases} \tag{2.22}$$

we define the complex matrices by

$$X=\begin{pmatrix}u & -\bar{v}\\ v & \bar{u}\end{pmatrix}, \tag{2.23}$$

$$g_1=\begin{pmatrix}\xi_1 & -\bar{\eta}_1\\ \eta_1 & \bar{\xi}_1\end{pmatrix}, \tag{2.24a}$$

$$g_2=\begin{pmatrix}\xi_2 & -\bar{\eta}_2\\ \eta_2 & \bar{\xi}_2\end{pmatrix}. \tag{2.24b}$$

It is obvious that $|\xi_j|^2+|\eta_j|^2=1, j=1,2$. Then $\mathbb{R}^4$ is identified with the vector space consisting of all matrices of the form (2.23), and a pair (g_1,g_2) belongs to SU(2)×SU(2). The action of (2.17) on $\mathbb{R}^4$ can now be represented in the form

$$X\rightarrow g_1Xg_2^{-1}. \tag{2.25}$$

The transformation (2.25) makes it easy to show that the subgroup $\Gamma=\{(I,I),(-I,-I)\}$ is the one and only subgroup that fixes all of X, where I stands for the identity of SU(2). Thus the action of SU(2)×SU(2) proves to be not effective. We then form the group SO(4) = SU(2)×SU(2)/Γ to establish the conclusion that the kinematical symmetry group is SO(4).

3. THE SUBGROUP OF SU(4) WHICH COMMUTES WITH A ONE-PARAMETER SUBGROUP U(1)

Let U(1) be the one-parameter subgroup generated by N_3 given in (2.15). From (2.15) U(1) has the form

$$\exp tN_3=\begin{pmatrix}T(t) & \\ & T(t)\end{pmatrix},\ T(t)=\begin{pmatrix}\cos(t/2) & -\sin(t/2)\\ \sin(t/2) & \cos(t/2)\end{pmatrix}. \tag{3.1}$$

We proceed to obtain subgroups of SU(4) which commute with the group U(1). In the first place, we determine the Lie subalgebra that commutes with N_3. To do this, instead of dealing with commutation relations of anti-Hermitian matrices, we employ the Poisson brackets of the constants of the motion L_{jk} and Q_{jk} for the sake of later use. We note here that N_3 is the matrix corresponding to an "angular momentum" $L_{12}+L_{34}$ [see (2.8a), (2.11), and (2.15)]. The Poisson brackets of L_{jk} and Q_{jk} show that the constants of the motion that commute with $L_{12}+L_{34}$ are linear combinations of $\frac{1}{2}A_{kj}L_{jk}$ and $\frac{1}{2}B_{kj}Q_{jk}$, each of which commutes with $L_{12}+L_{34}$ independently. The constants of the motion $\frac{1}{2}A_{kj}L_{jk}$ that commute with $L_{12}+L_{34}$ are immediately obtained from the commutation relations (2.16). They correspond to $M_j, j=1,2,3$. As regards $\frac{1}{2}B_{kj}Q_{jk}$ commuting with $L_{12}+L_{34}$, calculating the Poisson brackets in full yields the linearly independent constants of the motion: $Q_{11}+Q_{22}$, $Q_{13}+Q_{24}$, $-Q_{14}+Q_{23}$, and $Q_{33}+Q_{44}$. Imposing the condition $\mathrm{tr}(B_{kj})=0$, we obtain $Q_{13}+Q_{24}$, $-Q_{14}+Q_{23}$, and $Q_{11}+Q_{22}-Q_{33}-Q_{44}$.

The coefficient matrices of $Q_{13}+Q_{24}$, $-Q_{14}+Q_{23}$, and $\frac{1}{2}(Q_{11}+Q_{22}-Q_{33}-Q_{44})$ turn out to be

$$2B_1=\begin{pmatrix} & & i & \\ & & & i\\ i & & & \\ & i & & \end{pmatrix},\quad 2B_2=\begin{pmatrix} & & & -i\\ & & i & \\ & i & & \\ -i & & & \end{pmatrix},\quad 2B_3=\begin{pmatrix} i & & & \\ & i & & \\ & & -i & \\ & & & -i\end{pmatrix}, \tag{3.2}$$

respectively. Thus we have obtained a basis of the Lie algebra that commutes with N_3: M_j, B_j $(j=1,2,3)$, where M_j are given in (2.14). The commutation relations are

$$[N_3,M_j]=[N_3,B_j]=0, \tag{3.3a}$$

$$[M_j,M_k]=\epsilon_{jkh}M_h,$$

$$[B_j,B_k]=\epsilon_{jkh}M_h,\ [M_j,B_k]=\epsilon_{jkh}B_h. \tag{3.3b}$$

Let

$$H_j=\tfrac{1}{2}(M_j+B_j),\quad K_j=\tfrac{1}{2}(M_j-B_j). \tag{3.4}$$

Then from (3.3) we get

$$[N_3,H_j]=[N_3,K_j]=0, \tag{3.5a}$$

$$[H_j,H_k]=\epsilon_{jkh}H_h,\quad [K_j,K_k]=\epsilon_{jkh}K_h,$$

$$[H_j,K_k]=0. \tag{3.5b}$$

Thus we get the following:

Theorem 3.1: The symmetry Lie algebra that commutes with the constant of the motion $L_{12}+L_{34}$ is isomorphic with the Lie algebra of SU(2)×SU(2)×U(1), and is represented in terms of the matrices H_j, K_j, and N_3 $(j=1,2,3)$ with the structure equation (3.5).

We now proceed to Lie subgroups which commute with U(1). To get the subgroups we need to compute $\exp tH_j$, $\exp tK_j$, and $\exp tN_3$, and to form products of them. (Of course, there is no need to fix t for all the matrices.) We here give the following:

Theorem 3.2: The set of $\exp tH_j, j=1,2,3$, generates SU(2) in the form

$$U(z,w) = \begin{pmatrix} \frac{1}{2}(z+1) & (i/2)(z-1) & -\frac{1}{2}\bar{w} & -(i/2)\bar{w} \\ -(i/2)(z-1) & \frac{1}{2}(z+1) & (i/2)\bar{w} & -\frac{1}{2}\bar{w} \\ \frac{1}{2}w & (i/2)w & \frac{1}{2}(\bar{z}+1) & (i/2)(\bar{z}-1) \\ -(i/2)w & \frac{1}{2}w & -(i/2)(\bar{z}-1) & \frac{1}{2}(\bar{z}+1) \end{pmatrix}, \tag{3.6}$$

with

$|z|^2 + |w|^2 = 1,$

and the set of $\exp tK_j, j = 1,2,3$, also generates SU(2) in the form $\overline{U(z,w)}$, the complex conjugate of (3.6).

Proof: The second part of the theorem is obvious from the fact that $\overline{K_j} = H_j$. We prove the first part. It can be shown that each of $\exp tH_j$ takes the form (3.6). In fact, we have

$$\exp tH_1 = U(\cos(t/2), i\sin(t/2)), \tag{3.7a}$$

$$\exp tH_2 = U(\cos(t/2), \sin(t/2)), \tag{3.7b}$$

$$\exp tH_3 = U(e^{it/2}, 0). \tag{3.7c}$$

Multiplying $U(z,w)$ by $\exp tH_2$ and $\exp tH_3$ results in

$$\begin{cases} (\exp tH_2)U(z,w) \\ = U\left(z\cos\frac{t}{2} - w\sin\frac{t}{2}, z\sin\frac{t}{2} + w\cos\frac{t}{2}\right), \\ U(z,w)(\exp tH_2) \\ = U\left(z\cos\frac{t}{2} - \bar{w}\sin\frac{t}{2}, \bar{z}\cos\frac{t}{2} + w\sin\frac{t}{2}\right), \end{cases} \tag{3.8a}$$

$$\begin{cases} (\exp tH_3)U(z,w) = U(ze^{it/2}, we^{-it/2}), \\ U(z,w)(\exp tH_3) = U(ze^{it/2}, we^{it/2}). \end{cases} \tag{3.8b}$$

Note that matrices in the right-hand side of Eqs. (3.8) actually belong to (3.6). From (3.7) and (3.8b) we obtain

$$\exp tH_1 = (\exp(-(\pi/2)H_3))(\exp tH_2)(\exp(\pi/2)H_3). \tag{3.9}$$

It follows from (3.7), (3.8), and (3.9) that any product of $\exp tH_j, j = 1,2,3$, takes the form (3.6). We next show that the converse is true. Given a matrix $U(z,w)$, we let

$$z = e^{i(\psi+\varphi)/2}\cos(\theta/2), \quad w = e^{i(\psi-\varphi)/2}\sin(\theta/2). \tag{3.10}$$

Then by using (3.8) and (3.7b) we obtain

$$(\exp(-\varphi H_3))U(z,w)(\exp(-\psi H_3)) = U(\cos(\theta/2), \sin(\theta/2)) = \exp\theta H_2. \tag{3.11}$$

Equation (3.11) implies that

$$U(z,w) = (\exp\varphi H_3)(\exp\theta H_2)(\exp\psi H_3). \tag{3.12}$$

Thus we conclude that the matrices of the form (3.6) actually form the group SU(2), since the space of the matrices (3.6) is homeomorphic to $S^3(|z|^2 + |w|^2 = 1)$. This completes the proof.

The following is clear:

Theorem 3.3: $SU(2)\times SU(2)\times U(1)$ is a subgroup of SU(4), which commute with U(1) generated by N_3.

4. AN INVARIANT MANIFOLD $S^3\times S^3$ OF $SU(2)\times SU(2)$

At the beginning of Sec. 3 we obtained the constants of the motion that commute with $L_{12} + L_{34}$. Let $F = (1/2i)C_{kj}z_j\bar{z}_k$ be one of those constants. Then the coefficient matrix of F is a linear combination of M_j, B_j, and N_3 ($j = 1,2,3$). By the definition of the Poisson bracket, X_F acts on the hypersurface $L_{12} + L_{34} = \text{const}$. In fact we have $X_F(L_{12} + L_{34}) = -\{F, L_{12} + L_{34}\} = 0$. The set of F's is the symmetry Lie algebra which was defined in Theorem 3.1. We now show that the group $SU(2)\times SU(2)\times U(1)$ defined in Theorem 3.3 indeed acts on the hypersurface $L_{12} + L_{34} = \text{const}$.

For this purpose we introduce the coordinates (w_j) by

$$\begin{aligned} w_1 &= z_1 + iz_2, \quad w_2 = z_3 + iz_4, \\ w_3 &= z_1 - iz_2, \quad w_4 = z_3 - iz_4. \end{aligned} \tag{4.1}$$

Then in these coordinates the matrices $U(z,w)$ and $\overline{U(z,w)}$ given in Theorem 3.2 are transformed into

$$\begin{pmatrix} z & -\bar{w} & & \\ w & \bar{z} & & \\ & & 1 & \\ & & & 1 \end{pmatrix}, \tag{4.2a}$$

$$\begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \bar{z} & -w \\ & & \bar{w} & z \end{pmatrix}, \tag{4.2b}$$

respectively. The matrix $\exp tN_3$ given in (3.1) is diagonalized: $D(e^{it/2}, e^{it/2}, e^{-it/2}, e^{-it/2})$. Thus we have the elements of $SU(2)\times SU(2)$ and of U(1), respectively, in the form

$$\begin{pmatrix} u_1 & -\bar{v}_1 & & \\ v_1 & \bar{u}_1 & & \\ & & u_2 & -\bar{v}_2 \\ & & v_2 & \bar{u}_2 \end{pmatrix}, \tag{4.3a}$$

$$\begin{pmatrix} e^{it/2} & & & \\ & e^{it/2} & & \\ & & e^{-it/2} & \\ & & & e^{-it/2} \end{pmatrix}, \tag{4.3b}$$

with

$|u_j|^2 + |v_j|^2 = 1 \quad (j = 1,2).$

Theorem 4.1: Real hypersurfaces $L_{12} + L_{34} = \text{const}$ of $\mathbb{C}^4$ are invariant manifolds for the group $SU(2)\times SU(2)\times U(1)$ defined in Theorem 3.3.

Proof: First we recall that $L_{12} + L_{34} = \text{Im}(z_1\bar{z}_2 + z_3\bar{z}_4)$. From (4.1) we have the following:

$$|w_1|^2 + |w_2|^2 = \Sigma|z_j|^2 + 2\text{Im}(z_1\bar{z}_2 + z_3\bar{z}_4), \tag{4.4a}$$

$$|w_3|^2 + |w_4|^2 = \Sigma|z_j|^2 - 2\text{Im}(z_1\bar{z}_2 + z_3\bar{z}_4). \tag{4.4b}$$

The matrices (4.3) show that $SU(2)\times SU(2)\times U(1)$ leaves $|w_1|^2 + |w_2|^2$ and $|w_3|^2 + |w_4|^2$ invariant. Moreover, $SU(2)\times SU(2)\times U(1)$ preserves $\Sigma|z_j|^2$, because it is a subgroup of SU(4). Therefore Eqs. (4.4) show that hypersurfaces

$\text{Im}(z_1\bar{z}_2 + z_3\bar{z}_4) = \text{const.}$ are invariant under $SU(2)\times SU(2)\times U(1)$. This completes the proof.

Remark 1: The action of $SU(2)\times SU(2)\times U(1)$ on $\mathbb{C}^4$ is not effective. This is because the elements of $(SU(2)\times SU(2), U(1))$ that fix all the points w_j are (I,I) and $(-I,-I)$, where I denotes the 4×4 unit matrix. This can be proved by means of (4.3).

Remark 2: On the other hand, the subgroup $SU(2)\times SU(2)$ acts effectively on $\mathbb{C}^4$ as a symmetry Lie group. This should be compared with the case of the kinematical symmetry group discussed in Sec.2.

The hypersurface $L_{12} + L_{34} = 0$ is of particular interest. Let S^7 be the energy surface $\Sigma|z_j|^2 = 1$. Then from (4.4) the intersection of S^7 with the above hypersurface turns out to be a product manifold $S^3\times S^3 \subset S^7$. The following is then immediate.

Theorem 4.2: The condition $L_{12} + L_{34} = 0$ defines a hypersurface $S^3\times S^3$ of the energy surface S^7, on which $SU(2)\times SU(2)$ acts effectively.

We conclude this section with the remark that the Hamiltonian flows are given by $w_j(t) = e^{-i\lambda t}w_j(0), j = 1,2,3,4$. Of course, the flows run on $S^3\times S^3$.

5. A REDUCED MANIFOLD $S^3\times S^2$

First we show the following.

Theorem 5.1: Let $S^3\times S^3$ and $U(1)$ be the manifold defined in Theorem 4.2 and the one-parameter group (4.3b) acting on $S^3\times S^3$, respectively. Then one has

$$S^3\times S^3/U(1) = S^3\times S^2. \tag{5.1}$$

Proof: The manifold $S^3\times S^3$, given by the conditions

$$|w_1|^2 + |w_2|^2 = 1, \quad |w_3|^2 + |w_4|^2 = 1, \tag{5.2}$$

has the local coordinates

$$w_1 = e^{i(\psi_1 + \varphi_1)/2}\cos(\theta_1/2), \quad w_2 = e^{i(\psi_1 - \varphi_1)/2}\sin(\theta_1/2), \tag{5.3a}$$

$$w_3 = e^{i(\psi_2 + \varphi_2)/2}\cos(\theta_2/2), \quad w_4 = e^{i(\psi_2 - \varphi_2)/2}\sin(\theta_2/2). \tag{5.3b}$$

To obtain the orbit space $S^3\times S^3/U(1)$ in a manifest way, we introduce the variables

$$\chi = \tfrac{1}{2}(\psi_2 - \psi_1), \quad \psi = \tfrac{1}{2}(\psi_2 + \psi_1). \tag{5.4}$$

Then the action of (4.3b) is expressed in the form

$$\chi \to \chi - t \text{ and the others fixed.} \tag{5.5}$$

From (5.5) we see that the orbit space is independent of χ. In order to get an idea of the topology of the orbit space, we make (4.3b) with $t = \psi + \chi\ (= \psi_2)$ act on (5.3). Then we obtain the equalities of equivalence classes:

$$[(w_1,w_2)] = \left[(e^{i(2\psi + \varphi_1)/2}\cos\frac{\theta_1}{2}, e^{i(2\psi - \varphi_1)/2}\sin\frac{\theta_1}{2})\right], \tag{5.6a}$$

$$[(w_3,w_4)] = \left[(e^{i\varphi_2/2}\cos\frac{\theta_2}{2}, e^{-i\varphi_2/2}\sin\frac{\theta_2}{2})\right]. \tag{5.6b}$$

It follows from (5.6) that $[(w_1,w_2)]$ and $[(w_3,w_4)]$ represent a point of S^3 and S^2, respectively. Thus we get (5.1). It is worth mentioning that (5.1) is an extension of the Hopf mapping: $S^3 \to S^2$. This ends the proof.

We point out incidentally that the Hamiltonian flows, $w_j(t) = e^{-i\lambda t}w_j(0)$, are projected on $S^3\times S^2$ to be expressed in the coordinates appearing in (5.6) as $\psi \to \psi - z\lambda t$, with the other coordinates fixed.

Now that we have obtained a reduced manifold $S^3\times S^2$ from the energy surface S^7, we turn the crank to consider what group acts on $S^3\times S^2$. Let $p\in S^3\times S^3$ and $g\in SU(2)\times SU(2)$. Denote the action of g by the same letter, that is, $g: p \to gp$. Then the induced action on $S^3\times S^3/U(1) = S^3\times S^2$ is defined by

$$g[p] = [gp] \quad \text{for } [p]\in S^3\times S^3/U(1). \tag{5.7}$$

Since $SU(2)\times SU(2)$ and $U(1)$ commute, the action is well defined. Now suppose that g fixes all the points $[p]$. Then for g there is an element $r\in U(1)$ such that $rgp = p$ for all the points p of $S^3\times S^3$. Further, one has $rgp = p$ for all $p\in\mathbb{C}^4$. Incidentally, the elements of $SU(2)\times SU(2)\times U(1)$ that fix all the points of $\mathbb{C}^4$ have been obtained in Remark 1, which followed Theorem 4.1. Accordingly, the subgroup of $SU(2)\times SU(2)$ that fixes all the points $[p]$ turns out to be $S^0 = \{I, -I\}$, where I is 4×4 unit matrix. Therefore, we have the transformation group $SO(4) = SU(2)\times SU(2)/S^0$ acting on $S^3\times S^2$ effectively.

Theorem 5.2: Accompanying the reduction of the energy surface S^7 to $S^3\times S^2$ stated in Theorem 5.1, the symmetry group SU(4) reduces to a transformation group SO(4) acting on $S^3\times S^2$ effectively.

[1]E. L. Stiefel and G. Scheifele, *Linear and Regular Celestial Mechanics* (Springer, Berlin, 1971).

[2]M. Ikeda and Y. Miyachi, Math. Japon. **15**, 127 (1970).

[3]T. Iwai, J. Math. Phys. **22** xxx (1981).

[4]J. Marsden and A. Weinstein, Rep. Math. Phys. **5**, 121 (1974).

[5]R. Abraham and J. Marsden, *Foundations of Mechanics*, 2nd ed. (Addison-Wesley, Reading, MA., 1978).

[6]G. Marmo, E. J. Saletan, and S. Simoni, J. Math. Phys. **20**,856 (1979).

[7]Until the four-dimensional harmonic oscillator is discussed, Latin indices take values from 1 to n. The summation convention will be adopted.

[8]Hereafter, missing matrix entries are all zero.

On a "conformal" Kepler problem and its reduction

Toshihiro Iwai
Department of Applied Mathematics, Gifu University, Kagamigahara, Gifu 504, Japan

(Received 4 November 1980; accepted for publication 6 March 1981)

A "conformal" Kepler problem is defined in order to associate the Kepler problem with the harmonic oscillator. The four-dimensional conformal Kepler problem which shares an energy surface with the four-dimensional harmonic oscillator reduces to the ordinary three-dimensional Kepler problem. By use of the reduction the symmetry group SO(4) of the Kepler problem is brought out from a symmetry subgroup SU(2)×SU(2) of the conformal Kepler problem; the subgroup is the same as a subgroup of the symmetry group SU(4) of the harmonic oscillator.

PACS numbers: 03.20. + i, 02.20. − a

1. INTRODUCTION

The Kepler problem has received much attention because of its marked symmetry. It is widely known that the symmetry Lie algebra consisting of the constants of motion is isomorphic with the Lie algebra of SO(4). As regards global theory, Bacry, Ruegg, and Souriau[1] showed that the Kepler problem of a negative energy has the symmetry group SO(4) indeed (see also Ref. 2). The regularized (or compactified) energy surface of the n-dimensional Kepler problem is known to be the unit tangent bundle of the n-sphere S^n.[3,4] For $n = 3$ the unit tangent bundle becomes a trivial bundle $S^3 \times S^2$. The method for getting the topology of the energy surface is ultimately due to Fock,[5] and is developed for studying symmetry of the Kepler problem in both classical and quantum mechanics.[6–9] The point of this method is projecting the momentum space stereographically upon a unit sphere.

This paper aims to show that the regularized energy surface $S^3 \times S^2$ for the three-dimensional Kepler problem of a negative energy is obtained together with the symmetry group SO(4) from the energy surface S^7 of the four-dimensional harmonic oscillator with the symmetry group SU(4). A dynamical system called a "conformal" Kepler problem is defined and analyzed in order that the harmonic oscillator may be associated with the Kepler problem. The point of this article is that defining the conformal Kepler problem makes it feasible to treat the ordinary Kepler problem in the large. This article will make full use of the results obtained in a previous paper.[10]

Here brief mention should be made of the relation between the Kepler problem and the harmonic oscillator. Introducing the time coordinate and its conjugate momentum (the negative total energy) as dependent variables, Baumgarte[11,12] brought the three-dimensional Kepler problem into the four-dimensional harmonic oscillator to discuss the noninvariance algebra of the Kepler problem. Beside this, by means of the so-called KS transformation[13] the Kepler problem is put into the four-dimensional harmonic oscillator. In Ref. 13 Stiefel and Scheifele intensively utilized the KS transformation to study perturbed two-body motions. This transformation has its use in quantum mechanics.[14,15]

On the other hand, the present article is founded on a principle similar to one which Ikeda and Miyachi used in Ref. 16 for associating the four-dimensional harmonic oscillator with the three-dimensional hydrogen atom in quantum mechanics. The leading idea is as follows.

Consider the four-dimensional harmonic oscillator. The Hamiltonian is given on $\mathbb{R}^4 \times \mathbb{R}^4$, the phase space, by

$$H = \frac{1}{2}\sum p_j^2 + \frac{\lambda^2}{2}\sum x_j^2, \tag{1.1}$$

where $(x_j, p_j) \in \mathbb{R}^4 \times \mathbb{R}^4$ and λ is a positive constant. Introduce in $\mathbb{R}^4$ the coordinates

$$x_1 = R\cos\frac{\psi+\varphi}{2}\cos\frac{\theta}{2}, \quad x_2 = R\sin\frac{\psi+\varphi}{2}\cos\frac{\theta}{2},$$
$$x_3 = R\cos\frac{\psi-\varphi}{2}\sin\frac{\theta}{2}, \quad x_4 = R\sin\frac{\psi-\varphi}{2}\sin\frac{\theta}{2}, \tag{1.2}$$

where $R^2 = \Sigma x_j^2$. Furthermore, let

$$R^2 = r. \tag{1.3}$$

Then, the Hamiltonian (1.1) is expressed in terms of the new coordinates and their conjugate momentums as

$$H = 2rp_r^2 + \frac{2}{r}\left(\frac{p_\psi^2 + p_\varphi^2 - 2\cos\theta p_\psi p_\varphi}{\sin^2\theta} + p_\theta^2\right) + \frac{\lambda^2}{2}r. \tag{1.4}$$

Since the Hamiltonian is independent of ψ, the conjugate momentum p_ψ is conserved. When p_ψ is equal to zero, the energy conservation $H = \text{const}$ is put into the form

$$-\lambda^2/8 = \tfrac{1}{2}(p_r^2 + (1/r^2)(p_\theta^2 + p_\varphi^2/\sin^2\theta)) - H/4r. \tag{1.5}$$

This may be interpreted as the conservation of negative energy for the three-dimensional Kepler problem, if the coordinates r, θ, and φ are to be thought of as the spherical polar coordinates. However, it should be pointed out here that ψ, φ, and θ have the range

$$0 \leqslant (\psi+\varphi)/2 \leqslant 2\pi, \quad -\pi \leqslant (\psi-\varphi)/2 \leqslant \pi, \quad 0 \leqslant \theta/2 \leqslant \pi/2. \tag{1.6}$$

Questions then arise as to whether the range of φ reduces to $0 \leqslant \varphi \leqslant 2\pi$ or not, and as to how the coordinate ψ is gotten rid of. Succeeding sections will give answers to these questions, which could be ignored when one is interested only in formal correspondence between the harmonic oscillator and the Kepler problem. As far as symmetry Lie algebras for these dynamical systems are concerned, this formal correspondence could work well. The present paper, however aims to look into symmetry groups.

In Sec.2 a conformal Kepler problem is defined and

 0022-2488/81/081633-07$01.00 © 1981 American Institute of Physics

analyzed in connection with the harmonic oscillator. Here the conformal Kepler problem has the Hamiltonian having the kinetic energy defined through a conformally flat metric ds_c^2 and the central potential $-k/r$, where r is the radial distance with respect to ds_c^2 and k is a positive constant. In the course of analysis, the singularity at the collision, i.e., $r=0$, in the conformal Kepler problem is regularized.

Section 3 is concerned with reduction of the four-dimensional conformal Kepler problem to the three-dimensional ordinary Kepler problem.

Section 4 shows that the regularized energy surface S^7 of the conformal Kepler problem reduces to the regularized energy surface $S^3 \times S^2$ of the ordinary Kepler problem.

Section 5 shows how the symmetry group SU(4) of the four-dimensional harmonic oscillator reduces to the symmetry group SO(4) of the three-dimensional Kepler problem.

2. THE CONFORMAL KEPLER PROBLEM

Let (x_j) be the Cartesian coordinates of $\mathbb{R}^n$.[17] By ds_0^2 and $d\tau^2$ we mean the standard flat metric on $\mathbb{R}^n$ and the canonical metric on the unit sphere S^{n-1}, respectively. Let

$$R^2 = \sum x_j^2 = r. \tag{2.1}$$

Then the metric ds_0^2 takes the form

$$ds_0^2 = dR^2 + R^2 d\tau^2 = (1/4r)(dr^2 + 4r^2 d\tau^2). \tag{2.2}$$

We define a conformally flat metric ds_c^2 on $\mathbb{R}^n - \{0\}$ by

$$ds_c^2 = dr^2 + 4r^2 d\tau^2. \tag{2.3}$$

Let T_0 and T_c be the kinetic energies associated with ds_0^2 and ds_c^2, respectively. Then from (2.2) and (2.3) we have on $(\mathbb{R}^n - \{0\}) \times \mathbb{R}^n$

$$T_c = 4R^2 T_0. \tag{2.4}$$

By (x_j, p_j) we mean the standard coordinates of $\mathbb{R}^n \times \mathbb{R}^n$, the cotangent bundle of $\mathbb{R}^n$. Then $T_0 = \frac{1}{2}\Sigma p_j^2$. Equation (2.4) therefore means that

$$T_c = 4R^2\Big(\frac{1}{2}\sum p_j^2\Big). \tag{2.5}$$

A one-form θ_c relevant to ds_c^2 is defined by

$$\theta_c = 4R^2 \sum p_j dx_j. \tag{2.6}$$

Definition 2.1: We define the conformal Kepler problem to be a dynamical system on $(\mathbb{R}^n - \{0\}) \times \mathbb{R}^n$, with the Hamiltonian

$$H_c = 4R^2(\tfrac{1}{2}\sum p_j^2) - \frac{k}{R^2} \tag{2.7}$$

and the symplectic form $\omega_c = d\theta_c$, where θ_c is given by (2.6).

Remark: It is to be noted that $r = R^2$ is the geodesic distance in the radial direction with respect to ds_c^2.

Now we multiply (2.7) by $4R^2$ to get, after a change of form,

$$4k = (4R^2)^2\Big(\frac{1}{2}\sum p_j^2\Big) - 4H_c R^2. \tag{2.8}$$

In view of this, we define the following Hamiltonian K, and the relevant one- and two-forms θ_K and ω_K on $(\mathbb{R}^n - \{0\}) \times \mathbb{R}^n$

$$K = (4R^2)^2\Big(\frac{1}{2}\sum p_j^2\Big) + \frac{1}{2}\lambda^2 R^2, \tag{2.9}$$

$$\theta_K = 4R^2 \sum p_j dx_j, \tag{2.10a}$$

$$\omega_K = d\theta_K, \tag{2.10b}$$

where λ is a positive constant. The Hamiltonian K and the symplectic form ω_K determine a dynamical system on $(\mathbb{R}^n - \{0\}) \times \mathbb{R}^n$, which is closely connected with the conformal Kepler problem.

Proposition 2.2: Both dynamical systems $((\mathbb{R}^n - \{0\}) \times \mathbb{R}^n, \omega_c, H_c)$ and $((\mathbb{R}^n - \{0\}) \times \mathbb{R}^n, \omega_K, K)$ have the same energy surfaces when $H_c = -\lambda^2/8$ and $K = 4k$. On the energy surface prescribed the Hamiltonian flows of respective dynamical systems coincide within a change of parameters.

Proof: From (2.7) and (2.9), and from (2.6) and (2.10) we obtain

$$K = 4R^2(H_c + \lambda^2/8) + 4k, \tag{2.11}$$

$$\omega_K = \omega_c, \tag{2.12}$$

respectively. The first part of the proposition is an immediate consequence of (2.11). We prove the second part. The Hamiltonian flows are generated by the vector fields X_{H_c} and X_K determined, respectively, by

$$\mathrm{i}(X_{H_c})\omega_c = -dH_c, \tag{2.13a}$$

$$\mathrm{i}(X_K)\omega_K = -dK, \tag{2.13b}$$

where i( ) denotes the interior product. Writing out (2.13) by use of (2.11) and (2.12), we obtain

$$\mathrm{i}(X_K - 4R^2 X_{H_c})\omega_K = -8(H_c + \lambda^2/8)R\,dR. \tag{2.14}$$

Since ω_K is nondegenerate on $(\mathbb{R}^n - \{0\}) \times \mathbb{R}^n$, it follows from (2.14) that if $R \neq 0$

$$X_K = 4R^2 X_{H_c} \quad \text{on } K = 4k$$

or

$$H_c = -\lambda^2/8. \tag{2.15}$$

Equation (2.15) shows that both X_K and X_{H_c} generate the same Hamiltonian flows as sets on the energy surface $K = 4k$ or $H_c = -\lambda^2/8$. Flows of X_K and X_{H_c} are transformed to each other by a change of parameters. This completes the proof.

We now define the mapping of $(\mathbb{R}^n - \{0\}) \times \mathbb{R}^n$ into $\mathbb{R}^n \times \mathbb{R}^n$

$$(x_j, p_j) \mapsto (x_j, p_j') = (x_j, 4R^2 p_j). \tag{2.16}$$

The mapping is singular outside of $(\mathbb{R}^n - \{0\}) \times \mathbb{R}^n$. By (2.16) we put K, θ_K, and ω_K into

$$K' = \tfrac{1}{2}\sum p_j'^2 + (\lambda^2/2)\sum x_j^2, \tag{2.17}$$

$$\theta' = \sum p_j' dx_j, \tag{2.18a}$$

$$\omega' = d\theta', \tag{2.18b}$$

respectively. A pair K' and ω' determines the harmonic oscillator on $\mathbb{R}^n\times\mathbb{R}^n$. The following is obvious.

Proposition 2.3: Within a region $(\mathbb{R}^n-\{0\})\times\mathbb{R}^n$, the dynamical system $(\mathbb{R}^n\times\mathbb{R}^n, \omega', K')$ is isomorphic with the system $((\mathbb{R}^n-\{0\})\times\mathbb{R}^n, \omega_K, K)$:

$$K=K', \tag{2.19a}$$

$$\omega_K=\omega'. \text{ (Ref. 18)} \tag{2.19b}$$

So far we have obtained a series of dynamical systems $((\mathbb{R}^n-\{0\})\times\mathbb{R}^n, \omega_c, H_c)$, $((\mathbb{R}^n-\{0\})\times\mathbb{R}^n, \omega_K, K)$, and $(\mathbb{R}^n\times\mathbb{R}^n, \omega', K')$. By Propositions 2.2 and 2.3 the energy surface $H_c=-\lambda^2/8$ of the conformal Kepler problem is mapped into the energy surface $K'=4k$ of the harmonic oscillator. This makes it possible to extend the noncompact energy surface $H_c=-\lambda^2/8$ to be compact. We begin by considering what occurs outside of the domain $(\mathbb{R}^n-\{0\})\times\mathbb{R}^n$ of the conformal Kepler problem.

Flows of X_{H_c} which tend to $\{0\}\times\mathbb{R}^n$ will reach within a finite time an ideal point $(0,\infty)$ which represents the collision of two bodies at the origin $R=0$ with the momentum ∞. This may be seen from (2.7). From Proposition 2.2, the flows which approach along the energy surface $H_c=-\lambda^2/8$ to $\{0\}\times\mathbb{R}^n$ have corresponding flows on the energy surface $K=4k$ of the dynamical system $((\mathbb{R}^n-\{0\})\times\mathbb{R}^n, \omega_K, K)$. These flows go out of the domain $(\mathbb{R}^n-\{0\})\times\mathbb{R}^n$ to an ideal region $\{0\}\times S^{n-1}(\infty)$, where $S^{n-1}(\infty)$ is an ideal sphere of infinite radius: $\Sigma p_j^2=\infty$. This may be observed from (2.9). The reason why we should imagine the ideal sphere is that the corresponding flows on the energy surface $K'=4k$ of the dynamical system $(\mathbb{R}^n\times\mathbb{R}^n, \omega', K')$ tend to the subset $\{0\}\times S^{n-1}$ of the energy surface $K'=4k$, where S^{n-1} is a sphere determined by $\Sigma p_j'^2=8k$. We refer to the space enlarged by gluing the ideal region $\{0\}\times S^{n-1}(\infty)$ to the energy surface $H_c=-\lambda^2/8$ as the regularized energy surface $\overline{H}_c=-\lambda^2/8$.

Conversely, let $(0,b_j')$ be any point of $\{0\}\times S^{n-1}$, the subset of the energy surface $K'=4k$, and let $(x_j(t), p_j'(t))$ be the flow passing $(0,b_j')$. Then the curve $(x_j(t))$ comes to zero when and only when the curve $(p_j'(t))$ arrives at $(\pm b_j')$. Hence Propositions 2.2 and 2.3 show that the flow $(x_j(t), p_j'(t))$ without points $(0,\pm b_j')$ has the corresponding flow $(x_j(t), p_j(t))$, which represents a collision orbit of $((\mathbb{R}^n-\{0\})\times\mathbb{R}^n, \omega_c, H_c)$. Thus the subset $\{0\}\times S^{n-1}$ of the energy surface $K'=4k$ is understood to be what corresponds to the ideal region $\{0\}\times S^{n-1}(\infty)$ for the conformal Kepler problem. Therefore we have proved

Theorem 2.4: The regularized energy surface $\overline{H}_c=-\lambda^2/8$ for the conformal Kepler problem is mapped onto the compact energy surface $K'=4k$ of the harmonic oscillator. The Hamiltonian flows of respective dynamical systems coincide within a change of parameters.

3. REDUCTION OF THE FOUR-DIMENSIONAL CONFORMAL KEPLER PROBLEM

Consider the four-dimensional conformal Kepler problem. We employ the local coordinates (r,θ,φ,ψ) introduced in (1.2) and (1.3). Then the metric ds_c^2 defined by (2.3) takes form

$$ds_c^2=dr^2+r^2(d\theta^2+\sin^2\theta d\varphi^2+(d\psi+\cos\theta d\varphi)^2). \tag{3.1}$$

By p_r, p_θ, p_φ, and p_ψ we denote the conjugate momentums with respect to ds_c^2.[19] The Hamiltonian (2.7) and the one-form (2.6) then take the form

$$H_c=\tfrac{1}{2}\left(p_r^2+\frac{1}{r^2}\left(p_\theta^2+\frac{p_\varphi^2-2p_\varphi p_\psi\cos\theta+p_\psi^2}{\sin^2\theta}\right)\right)-\frac{k}{r}, \tag{3.2}$$

$$\theta_c=p_r dr+p_\theta d\theta+p_\varphi d\varphi+p_\psi d\psi, \tag{3.3}$$

respectively. Since H_c is independent of ψ, the momentum p_ψ is conserved. Therefore, if ψ is got rid of, and if p_ψ is set equal to zero, a pair of (3.2) and (3.3) may determine the ordinary Kepler problem. We are to show that this is the case.

To gain an insight into the reduced manifold determined by $p_\psi=0$, we ought to deal with p_ψ in the large. To this end, we consider the one-parameter group of transformations which is relevant to the momentum p_ψ. Let [20]

$$N=\begin{pmatrix}A & \\ & A\end{pmatrix} \text{ with } A=\begin{pmatrix} & -1/2\\ 1/2 & \end{pmatrix}. \tag{3.4}$$

Then one has

$$\exp tN=\begin{pmatrix}T(t) & \\ & T(t)\end{pmatrix} \text{ with } T(t)=\begin{pmatrix}\cos(t/2) & -\sin(t/2)\\ \sin(t/2) & \cos(t/2)\end{pmatrix}. \tag{3.5}$$

Taking account of (1.2), we can express the action of $\exp tN$ on $\mathbb{R}^4$ in the coordinates (r,θ,φ,ψ) as

$$\psi\to\psi+t, \text{ and the others fixed.} \tag{3.6}$$

Equating the infinitesimal generator of (3.5) with that of (3.6), we obtain

$$\frac{\partial}{\partial\psi}=\frac{1}{2}\left(x_1\frac{\partial}{\partial x_2}-x_2\frac{\partial}{\partial x_1}+x_3\frac{\partial}{\partial x_4}-x_4\frac{\partial}{\partial x_3}\right). \tag{3.7}$$

The transformation $\exp tN$ of $\mathbb{R}^4$ lifts to that of $\mathbb{R}^4\times\mathbb{R}^4=T^*\mathbb{R}^4$, the cotangent bundle of $\mathbb{R}^4$, therefore so does the infinitesimal generator. By $\exp tN^*$ and X^* we denote the lifts of $\exp tN$ and its infinitesimal generator, respectively. Then we have

$$\exp tN^*: (x,p)\to((\exp tN)x, (\exp tN)p), \tag{3.8}$$

$$X^*=\frac{1}{2}\left(x_1\frac{\partial}{\partial x_2}-x_2\frac{\partial}{\partial x_1}+x_3\frac{\partial}{\partial x_4}-x_4\frac{\partial}{\partial x_3}\right)+\frac{1}{2}\left(p_1\frac{\partial}{\partial p_2}-p_2\frac{\partial}{\partial p_1}+p_3\frac{\partial}{\partial p_4}-p_4\frac{\partial}{\partial p_3}\right), \tag{3.9}$$

where $x=(x_j)$ and $p=(p_j)$ are column vectors, $j=1,2,3,4$. Since $\exp tN$ is an orthogonal matrix, $\exp tN^*$ given by (3.8) leaves the canonical one-form $\Sigma p_j dx_j$ invariant. (This is, however, a special case of the well-known fact that the lift of a base space transformation leaves invariant the canonical one-form on the cotangent bundle.[21]) Therefore, from (2.6) it follows that θ_c is invariant under $\exp tN^*$. As a consequence of the invariance, we see from (3.3) and (3.6) that the conjugate momentums are all invariant under $\exp tN^*$. Thus we obtain in (an open subset of) $\mathbb{R}^4\times\mathbb{R}^4$

$$X^* = \frac{\partial}{\partial\psi}. \tag{3.10}$$

We are now ready to express the momentum p_ψ without reference to the local coordinates. From (3.3) and (3.10), and from (2.6) and (3.9), we obtain

$$p_\psi = \theta_c(X^*) = 2R^2(x_1p_2 - x_2p_1 + x_3p_4 - x_4p_3). \tag{3.11}$$

Of course, $\theta_c(X^*)$ is a conserved quantity, as H_c is invariant under $\exp tN^*$. (This is a special case of the conservation theorem which is established under fairly general conditions.[21]) We are now in a position to show

Proposition 3.1: For the four-dimensional conformal Kepler problem, the condition $p_\psi = 0$ determines a "momentum" manifold $(\mathbb{R}^4 - \{0\})\times\mathbb{R}^3$.

Proof: Recall that our conformal Kepler problem has the domain $(\mathbb{R}^4 - \{0\})\times\mathbb{R}^4$, so that $R\neq 0$. From (3.11) the condition $p_\psi = 0$ then becomes

$$\tfrac{1}{2}(x_1p_2 - x_2p_1 + x_3p_4 - x_4p_3) = \langle Nx, p\rangle = 0, \tag{3.12}$$

where $\langle\ ,\ \rangle$ denotes the standard inner product. We here understand that for each $x\neq 0$ Eq. (3.12) gives a linear equation in the cotangent space $T^*_x\mathbb{R}^4$ at x. Then, to Eq. (3.12) there exists linearly independent solutions $s_k(x)$, $k = 1,2,3$, which depend continuously on x:

$$s_1(x) = \begin{pmatrix} -x_3 \\ x_4 \\ x_1 \\ -x_2 \end{pmatrix},\quad s_2(x) = \begin{pmatrix} -x_4 \\ -x_3 \\ x_2 \\ x_1 \end{pmatrix},\quad s_3(x) = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}. \tag{3.13}$$

The basis $s_k(x)$, $k = 1,2,3$, spans the hyperplane $\mathbb{R}^3$ in each cotangent space

$$\mathbb{R}^3 = \left\{(\sigma_k);\ \left\langle Nx, \sum_{k=1}^{3}\sigma_k s_k(x)\right\rangle = 0\right\}. \tag{3.14}$$

Thus Eq. (3.12) determines the product space $(\mathbb{R}^4 - \{0\})\times\mathbb{R}^3$. This completes the proof.

So far we have decreased the degree of freedom by one. We now proceed to get rid of ψ by means of the group $\exp tN^*$, under which all the coordinates other than ψ are invariant. Obviously, the group $\exp tN^*$ acts on $(\mathbb{R}^4 - \{0\})\times\mathbb{R}^3$. Indeed the form $\langle Nx, p\rangle$ is invariant under $\exp tN^*$

$$\langle N(\exp tN)x, (\exp tN)p\rangle = \langle Nx, p\rangle. \tag{3.15}$$

The action of $\exp tN^*$ is given by the following.

Lemma 3.2: Let $S(t)$ be a 3×3 matrix given by

$$S(t) = \begin{pmatrix} T(2t) & \\ & 1 \end{pmatrix}, \tag{3.16}$$

where $T(2t)$ comes from (3.5). Then the action of $\exp tN^*$ on $(\mathbb{R}^4 - \{0\})\times\mathbb{R}^3$ takes the form

$$(x,\sigma)\to((\exp tN)x,\ S(t)\sigma), \tag{3.17}$$

where $\sigma = (\sigma_k)$ is an element of the second factor space $\mathbb{R}^3$.

Proof: The action on the first factor space $\mathbb{R}^4 - \{0\}$ is already obtained in (3.8): $x\to(\exp tN)x$. We turn to the action on the second factor space $\mathbb{R}^3$. In the first place, we have to determine how the basis $s_k(x)$, $k = 1,2,3$, given by the (3.13) transforms under the transformation $x\to x' = (\exp tN)x$. From (3.5) it follows that

$$\begin{pmatrix} -x_3' \\ x_4' \end{pmatrix} = T(-t)\begin{pmatrix} -x_3 \\ x_4 \end{pmatrix},\quad \begin{pmatrix} x_1' \\ -x_2' \end{pmatrix} = T(-t)\begin{pmatrix} x_1 \\ -x_2 \end{pmatrix}, \tag{3.18a}$$

$$\begin{pmatrix} x_4' \\ x_3' \end{pmatrix} = T(-t)\begin{pmatrix} x_4 \\ x_3 \end{pmatrix},\quad \begin{pmatrix} x_2' \\ x_1' \end{pmatrix} = T(-t)\begin{pmatrix} x_2 \\ x_1 \end{pmatrix}. \tag{3.18b}$$

From (3.13) and (3.18), we obtain the action of $\exp tN$ in the form

$$s_1((\exp tN)x) = \exp(-tN)s_1(x), \tag{3.19a}$$

$$s_2((\exp tN)x) = \exp(-tN)s_2(x), \tag{3.19b}$$

$$s_3((\exp tN)x) = (\exp tN)s_3(x). \tag{3.19c}$$

Assume that a point (x_j,σ_k) of $(\mathbb{R}^4 - \{0\})\times\mathbb{R}^3$ is carried to (x_j',σ_k') by the action of $\exp tN^*$. Then one has from (3.19) and (3.15)

$$\left\langle Nx', \sum_{k=1}^{3}\sigma_k' s_k(x')\right\rangle = \left\langle Nx, \sum_{k=1}^{2}\sigma_k'\exp(-2tN)s_k(x) + \sigma_3's_3(x)\right\rangle. \tag{3.20}$$

With reference to the last equation, simple calculation yields

$$\sum_{k=1}^{2}\sigma_k'\exp(-2tN)s_k(x) = (\sigma_1'\cos t + \sigma_2'\sin t)s_1(x) + (-\sigma_1'\sin t + \sigma_2'\cos t)s_2(x). \tag{3.21}$$

Substituting (3.21) in (3.20) and using the invariance of the form $\langle Nx, p\rangle$, we obtain

$$\begin{aligned}\left\langle Nx, \sum_{k=1}^{3}\sigma_k s_k(x)\right\rangle &= \left\langle Nx', \sum_{k=1}^{3}\sigma_k' s_k(x')\right\rangle \\ &= \langle Nx,(\sigma_1'\cos t + \sigma_2'\sin t)s_1(x) + (-\sigma_1'\sin t \\ &\quad + \sigma_2'\cos t)s_2(x) + \sigma_3's_3(x)\rangle.\end{aligned} \tag{3.22}$$

From (3.22) we deduce that $\sigma = S(-t)\sigma'$, where $S(t)$ is given by (3.16). This ends the proof.

Using Lemma 3.2, we prove

Proposition 3.3: Let $(\mathbb{R}^4 - \{0\})\times\mathbb{R}^3$ be the momentum manifold stated in Proposition 3.1. Then one has the orbit space

$$(\mathbb{R}^4 - \{0\})\times\mathbb{R}^3/U(1) = (\mathbb{R}^3 - \{0\})\times\mathbb{R}^3, \tag{3.23}$$

where $U(1)$ denotes the group $\exp tN^*$.

Proof: To make effective use of (3.17), we employ (1.2) and (1.3), setting

$$z = x_1 + ix_2,\quad w = x_3 + ix_4, \tag{3.24}$$

and denote $\sigma = (\sigma_k)$ in the form

$$\zeta = \sigma_1 + i\sigma_2 = \rho e^{i\chi}\sin\gamma,\quad \sigma_3 = \rho\cos\gamma. \tag{3.25}$$

We write x and σ as $x = (z,w)$ and $\sigma = (\zeta,\sigma_3)$, respectively. Then the action of $\exp tN^*$ is expressed as

$$((z,w),\ (\zeta,\sigma_3))\to((e^{it/2}z, e^{it/2}w),\ (e^{it}\zeta,\ \sigma_3)). \tag{3.26}$$

Introducing new coordinates α and β by

$$\alpha = \tfrac{1}{2}(\chi - \psi), \quad \beta = \tfrac{1}{2}(\chi + \psi), \tag{3.27}$$

we make use of (3.26) with $t = -(\beta - \alpha) = -\psi$ to obtain a representative (x_0, σ_0) of the equivalence class $[(x, \sigma)]$ with

$$x_0 = (\sqrt{(r)}e^{i\varphi/2}\cos(\theta/2), \sqrt{(r)}e^{-i\varphi/2}\sin(\theta/2)), \tag{3.28a}$$

$$\sigma_0 = (\rho e^{i2\alpha}\sin\gamma, \rho\cos\gamma). \tag{3.28b}$$

Equation (3.28a) means that (r, θ, φ) are the spherical polar coordinates in $\mathbb{R}^3$. In fact, one has for $[x] = [x_0]$ the point of $\mathbb{R}^3$

$$(2\mathrm{Re}z\overline{w}, 2\mathrm{Im}z\overline{w}, |z|^2 - |w|^2)$$
$$= (r\sin\theta\cos\varphi, r\sin\theta\sin\varphi, r\cos\theta), \tag{3.29}$$

where Re and Im indicates the real and imaginary parts, respectively. Equation (3.29) is a slight extension of the Hopf mapping $S^3 \to S^2$.[22] Thus Eq. (3.28) implies that $[(x, \sigma)]$ determines a point of $(\mathbb{R}^3 - \{0\}) \times \mathbb{R}^3$. Conversely, given any point of $(\mathbb{R}^3 - \{0\}) \times \mathbb{R}^3$, we can invert the above reasoning to get the inverse image in $(\mathbb{R}^4 - \{0\}) \times \mathbb{R}^3$. This completes the proof.

So far we have reduced the cotangent bundle $(\mathbb{R}^4 - \{0\}) \times \mathbb{R}^4$ to the cotangent bundle $(\mathbb{R}^3 - \{0\}) \times \mathbb{R}^3$ of $\mathbb{R}^3 - \{0\}$. Furthermore, it is evident that $(\mathbb{R}^3 - \{0\}) \times \mathbb{R}^3$ has the local coordinates $(r, \theta, \varphi, p_r, p_\theta, p_\varphi)$ because of the fact that they are invariant under $\exp tN^*$. Here a question arises as to whether ds_c^2 on $\mathbb{R}^4 - \{0\}$ reduces to the standard flat metric on $\mathbb{R}^3 - \{0\}$ or not. To work out the question, we first note that Eq. (3.12) also defines a three-dimensional subspace $\mathbb{R}_x^3$ of the tangent space $T_x\mathbb{R}^4$ at $x \neq 0$. The basis are $s_k(x)$, $k = 1,2,3$. Then they read

$$s_1(x) = -x_3\frac{\partial}{\partial x_1} + x_4\frac{\partial}{\partial x_2} + x_1\frac{\partial}{\partial x_3} - x_2\frac{\partial}{\partial x_4}, \tag{3.30a}$$

$$s_2(x) = -x_4\frac{\partial}{\partial x_1} - x_3\frac{\partial}{\partial x_2} + x_2\frac{\partial}{\partial x_3} + x_1\frac{\partial}{\partial x_4}, \tag{3.30b}$$

$$s_3(x) = x_1\frac{\partial}{\partial x_1} + x_2\frac{\partial}{\partial x_2} + x_3\frac{\partial}{\partial x_3} + x_4\frac{\partial}{\partial x_4}. \tag{3.30c}$$

In the coordinates $(r, \theta, \varphi, \psi)$ the above vectors have the form

$$\tfrac{1}{2}s_1(x) = \cos\psi\frac{\partial}{\partial\theta} + \sin\psi\csc\theta\frac{\partial}{\partial\psi} - \sin\psi\cot\theta\frac{\partial}{\partial\psi}, \tag{3.31a}$$

$$\tfrac{1}{2}s_2(x) = \sin\psi\frac{\partial}{\partial\theta} - \cos\psi\csc\theta\frac{\partial}{\partial\varphi} + \cos\psi\cot\theta\frac{\partial}{\partial\psi}, \tag{3.31b}$$

$$\tfrac{1}{2}s_3(x) = r\frac{\partial}{\partial r}. \tag{3.31c}$$

We now treat ds_c^2 in terms of one-forms defined by

$$\pi_1 = \cos\psi d\theta + \sin\psi\sin\theta d\varphi, \tag{3.32a}$$

$$\pi_2 = \sin\psi d\theta - \cos\psi\sin\theta d\varphi, \tag{3.32b}$$

$$\pi_3 = d\psi + \cos\theta d\varphi. \tag{3.32c}$$

Then ds_c^2 takes the form

$$ds_c^2 = dr^2 + r^2(\pi_1^2 + \pi_2^2 + \pi_3^2). \tag{3.33}$$

That ds_c^2 is invariant under $\exp tN^*$ can be verified from (3.32) and (3.33) by using (3.6).

We are now in a position to reduce ds_c^2. By S^1 we mean the group $\exp tN$. Let u and v be tangent vectors in $\mathbb{R}_x^3$, the subspace of $T_x\mathbb{R}^4$. Let $[u]$ and $[v]$ denote the equivalence classes in $\mathbb{R}_x^3/T_x(S^1\cdot x)$, which are thought of as the tangent vectors to $\mathbb{R}^3 - \{0\} = (\mathbb{R}^4 - \{0\})/S^1$, where $T_x(S^1\cdot x)$ is the tangent space at x to the S^1 orbit of x. Now the reduced metric is uniquely determined by the condition

$$ds_c^2(u,v) = (ds_c^2)^{rd}([u], [v]). \tag{3.34}$$

Proposition 3.4: The quotient space $\mathbb{R}^3 - \{0\} = \mathbb{R}^4 - \{0\}/S^1$ is equipped with the standard flat metric which is reduced from ds_c^2 defined on $\mathbb{R}^4 - \{0\}$:

$$(ds_c^2)^{rd} = dr^2 + r^2(\pi_1^2 + \pi_2^2) = dr^2 + r^2(d\theta^2 + \sin^2\theta d\varphi^2). \tag{3.35}$$

Proof: With reference to (3.31) and (3.32), one has

$$\pi_k(\tfrac{1}{2}s_j(x)) = \delta_{kj}, \quad \pi_3(\tfrac{1}{2}s_j(x)) = 0, \quad k = 1,2. \tag{3.36}$$

The proof is then accomplished by using (3.36) and the fact that $T_x(S^1\cdot x)$ is spanned by $(\partial/\partial\psi)$. This ends the proof.

We return to (3.2) and (3.3). Through the reduction process stated above, the Hamiltonian (3.2) and the one-form (3.3) reduce, respectively, to

$$H_c^{rd} = \tfrac{1}{2}(p_r^2 + (1/r^2)(p_\theta^2 + p_\varphi^2/\sin^2\theta)) - k/r, \tag{3.37}$$

$$\theta_c^{rd} = p_r dr + p_\theta d\theta + p_\varphi d\varphi. \tag{3.38}$$

From Proposition 3.4 we see that $(p_r, p_\theta, p_\varphi)$ can be thought of as the conjugate momentums with respect to the standard flat metric on $\mathbb{R}^3 - \{0\}$. Thus we have obtained

Theorem 3.5: The conformal Kepler problem on $(\mathbb{R}^4 - \{0\}) \times \mathbb{R}^4$ with the Hamiltonian (3.2) and the one-form (3.3) reduces to the ordinary Kepler problem on $(\mathbb{R}^3 - \{0\}) \times \mathbb{R}^3$ with the Hamiltonian (3.37) and the one-form (3.38).

To make things precise, we point out how H_c^{rd} and $\omega_c^{rd} = d\theta_c^{rd}$ are characterized. Let $M = (\mathbb{R}^4 - \{0\}) \times \mathbb{R}^3$ be the momentum manifold stated in Proposition 3.1. Then H_c^{rd} and ω_c^{rd} are characterized as follows[21]:

$$H_c^{rd}([m]) = H_c(m) \quad \text{for } m \in M, \tag{3.39}$$

$$\omega_c^{rd}([u], [v]) = \omega_c(u, v) \quad \text{for } u,v \in T_m(M), \tag{3.40}$$

where $[m] \in M/U(1)$ and $[u], [v] \in T_{[m]}(M/U(1)) = T_m(M)/T_m(U(1)\cdot m)$. Here $U(1)\cdot m$ denotes the $U(1)$ orbit of m.

4. REDUCTION OF THE ENERGY SURFACE

From Theorem 3.5 we may expect that the regularized energy surface $\overline{H}_c = -\lambda^2/8$ (which is by Theorem 2.4 with $n = 4$ the same as the energy surface $K' = 4k$ of the harmonic oscillator) will reduce to the regularized energy surface of the ordinary Kepler problem. With this in mind, we treat the harmonic oscillator in parallel with the conformal Kepler problem.

We shall begin by considering what momentum for the harmonic oscillator corresponds to p_ψ for the conformal Kepler problem. Recall that p_ψ is associated with $\exp tN^*$, the lift of $\exp tN$. In the case of the harmonic oscillator, $\exp tN$ and its infinitesimal generator lift to[23]

$$\exp tN^*: (x, p') \to ((\exp tN)x, (\exp tN)p'), \tag{4.1}$$

$$X^* = \frac{1}{2}\left(x_1 \frac{\partial}{\partial x_2} - x_2 \frac{\partial}{\partial x_1} + x_3 \frac{\partial}{\partial x_4} - x_4 \frac{\partial}{\partial x_3}\right) + \frac{1}{2}\left(p_1' \frac{\partial}{\partial p_2'} - p_2' \frac{\partial}{\partial p_1'} + p_3' \frac{\partial}{\partial p_4'} - p_4' \frac{\partial}{\partial p_3'}\right), \tag{4.2}$$

respectively. For θ' and θ_c, one has $\theta' = \theta_c$ in $(\mathbb{R}^4 - \{0\})\times\mathbb{R}^4$ from (2.6), (2.16), and (2.18).[24] We should note here that θ_c is degenerate on $\{0\}\times\mathbb{R}$ but θ' is not. Now, applying $\theta' = \theta_c$ to X^*, we obtain

$$2\theta'(X^*) = x_1p_2' - x_2p_1' + x_3p_4' - x_4p_3' = 2\theta_c(X^*). \tag{4.3}$$

Thus we see that the conditions $\theta_c(X^*) = p_\psi = 0$ and $\theta'(X^*) = 0$ coincide at least in $(\mathbb{R}^4 - \{0\})\times\mathbb{R}^4$. We now show that the condition $\theta_c(X^*) = 0$ may work outside of $(\mathbb{R}^4 - \{0\})\times\mathbb{R}^4$. Consider flows of the harmonic oscillator which pass $\{0\}\times\mathbb{R}^4$. Then $\theta'(X^*)$ equals zero along the flows. This is because $\theta'(X^*)$ is a constant of the motion and $\theta'(X^*)$ equals zero when $(x_j) = 0$. We turn to the corresponding flows of the conformal Kepler problem which represent flows going beyond the domain $(\mathbb{R}^4 - \{0\})\times\mathbb{R}^4$. Along the flows, $\theta_c(X^*)$ also equals zero by means of (4.3). Thus from the conservation law of the momentum we may understand that the condition $\theta_c(X^*) = 0$ makes sense outside of $(\mathbb{R}^4 - \{0\})\times\mathbb{R}^4$. We indicate this understanding by $\bar{p}_\psi = \bar{\theta}_c(X^*) = 0$.

Now that we have seen that the conditions $\bar{H}_c = -\lambda^2/8$ and $\bar{\theta}_c(X^*) = 0$ are equivalent to $K' = 4k$ and $\theta'(X^*) = 0$, we are ready to apply Theorem 4.2 in Ref. 10, which is a theorem about the submanifold determined by $K' = 4k$ and $\theta'(X^*) = 0$ for the harmonic oscillator (but the notations are slightly different). According to that theorem, we obtain

Proposition 4.1: For the four-dimensional conformal Kepler problem, the condition $\bar{p}_\psi = 0$ determines a regularized "energy-momentum" manifold $S^3\times S^3$ in the regularized energy surface S^7 given by $\bar{H}_c = -\lambda^2/8$.

We take the next step to get the regularized energy surface $\bar{H}_c^{rd} = -\lambda^2/8$ of the ordinary Kepler problem. On account of the reduction process which went on in Proposition 3.3 and Theorem 3.5, we have to construct an orbit space of the regularized energy-momentum manifold $S^3\times S^3$.

We consider below the conformal Kepler problem and the harmonic oscillator simultaneously. By $U(1)$ we mean the group $\exp tN^*$ whether it is defined by (3.8) or (4.1). For the harmonic oscillator the action of $U(1)$ on $S^3\times S^3$ was studied in Ref. 10, where $S^3\times S^3$ is defined by $K' = 4k$ and $\theta'(X^*) = 0$. On the other hand, for the conformal Kepler problem we still have to study the action of $U(1)$ on the regularized energy-momentum manifold determined by $\bar{H}_c = -\lambda^2/8$ and $\bar{\theta}_c(X^*) = 0$. It is clear that the energy-momentum manifold before regularization admits the action of $U(1)$, since H_c and $\theta_c(X^*)$ are invariant under $U(1)$. Thus the remaining problem to work out is to show that the action of $U(1)$ makes sense beyond the domain $(\mathbb{R}^4 - \{0\})\times\mathbb{R}^4$.

From the definition (3.8), we may conceive that the ideal region $\{0\}\times S^3(\infty)$, described already in Sec. 2, is invariant under $U(1)$. Furthermore, the ideal region is mapped onto $\{0\}\times S^3$, the subset of the energy surface of the harmonic oscillator. From (4.1) it appears that the range $\{0\}\times S^3$ admits the action of $U(1)$. Thus we can interpret that $U(1)$ acts outside of the domain of the conformal Kepler problem. With this in mind, we apply Theorem 5.1 of Ref. 10, which refers to orbit space $S^3\times S^3/U(1) = S^3\times S^2$. From Theorem 3.5 and Proposition 4.1 we may conclude that the $S^3\times S^2$ obtained is the regularized energy surface $\bar{H}_c^{rd} = -\lambda^2/8$. Before stating this conclusion as a theorem, we verify that $S^3\times S^2$ is indeed constructed by gluing an ideal region $\{0\}\times S^2(\infty)$ to the energy surface $H_c^{rd} = -\lambda^2/8$ in $(\mathbb{R}^3 - \{0\})\times\mathbb{R}^3$.

Consider Hamiltonian flows on the energy surface $H_c^{rd} = -\lambda^2/8$, which tend to $\{0\}\times\mathbb{R}^3$. Then the flows will arrive in an ideal region $\{0\}\times S^2(\infty)$ within a finite time. This region may be considered as the orbit space $(\{0\}\times S^3(\infty))/U(1)$ of the ideal region for the conformal Kepler problem. To see this, we turn to the corresponding region for the harmonic oscillator. As was stated in Sec. 2, $\{0\}\times S^3(\infty)$ is mapped onto $\{0\}\times S^3$, the subset of the energy surface for the harmonic oscillator. The $U(1)$-action on $\{0\}\times S^3$ is given by (4.1). Then we can obtain the orbit space $(\{0\}\times S^3)/U(1) = \{0\}\times S^2$ by the same method as in the proof of Proposition 3.3. The space thus obtained corresponds to the ideal region $\{0\}\times S^2(\infty)$ for the Kepler problem. We now state

Theorem 4.2: The regularized energy surface S^7 of the four-dimensional conformal Kepler problem reduces to $S^3\times S^2$ to give the regularized energy surface of the ordinary Kepler problem.

5. REDUCTION OF SYMMETRY

In the preceding sections we have shown that the conformal Kepler problem reduces to the ordinary problem. In this section we first point out that along with the reduction a certain symmetry subgroup for the conformal Kepler problem also reduces to a symmetry group of the ordinary Kepler problem. After doing so, we proceed to obtain the symmetry group $SO(4)$ of the ordinary Kepler problem.

Proposition 5.1: Assume that one has a symmetry group G of the four-dimensional conformal Kepler problem such that the momentum manifold $(\mathbb{R}^4 - \{0\})\times\mathbb{R}^3$ stated in Proposition 3.1 is an invariant manifold of G, and G and $U(1)$ commute, where $U(1)$ is the group $\exp tN^*$ given by (3.8). Then G becomes a symmetry group of the ordinary Kepler problem.

Proof: We denote $(\mathbb{R}^4 - \{0\})\times\mathbb{R}^3$ by M. By F_g we mean the action of $g\in G$. Then G induces the action F_g^{rd} on $M/U(1)$

$$F_g^{rd}([m]) = [F_g(m)] \quad \text{for } m\in M. \tag{5.1}$$

It is easy to see that F_g^{rd} is well defined. In order to prove that G leaves H_c^{rd} and ω_c^{rd} invariant, we require (3.39) and (3.40). By using them together with the invariance of H_c and ω_c under F_g, we can show, after a calculation, that

$$H_c^{rd}(F_c^{rd}([m])) = H_c^{rd}([m]), \tag{5.2}$$

$$((F_g^{rd})^*\omega_c^{rd})_{[m]}([u],[v]) = (\omega_c^{rd})_{[m]}([u],[v]), \tag{5.3}$$

where the superscript asterisk indicates the pullback. Thus from Theorem 3.5, G is looked on as a symmetry group of the ordinary Kepler problem, as was expected.

Now we recall that conditions $\bar{H}_c = -\lambda^2/8$ and $\bar{p}_\psi = \bar{\theta}_c(X^*) = 0$ are equivalent to $K' = 4k$ and $\theta'(X^*) = 0$, from which Proposition 4.1 was deduced. Incidentally, the previous paper[10] contains the results on groups acting on the manifold defined by $K' = 4k$ and $\theta'(X^*) = 0$. According to Theorem 4.2 in Ref. 10, a subgroup $G = SU(2)\times SU(2)$ of the symmetry group SU(4) acts on the manifold mentioned just above. Moreover, G and $U(1)$ commute by Theorem 3.3 of Ref. 10. We then see from Proposition 4.1 that G acts on the regularized energy-momentum manifold $S^3\times S^3$ of the conformal Kepler problem.

We continue to consider the symplectic forms ω_c and ω'. So long as ω_c is not degenerate, one has $\omega_c = \omega'$ from (2.12) and (2.19b). Recall that ω_c is degenerate on $\{0\}\times\mathbb{R}^4$. However, on account of Theorem 2.4 we may conceive that $\omega_c = \omega'$ will be true throughout $\bar{H}_c = -\lambda^2/8$ or $K' = 4k$. Indeed, the Hamiltonian flows of respective dynamical systems, which by Theorem 2.4 coincide within a change of parameters, leave the respective symplectic forms ω_c and ω' invariant, so that the equality $\omega_c = \omega'$ may be preserved even if the flows of X_{H_c} reach the ideal region $\{0\}\times S^3(\infty)$. Since G leaves ω' invariant, we understand that ω_c is also invariant under the action of G. Therefore G may be thought of as a symmetry group of the conformal Kepler problem.

Theorem 5.2: A symmetry subgroup $SU(2)\times SU(2)$ for the four-dimensional conformal Kepler problem reduces to the symmetry group SO(4) for the ordinary Kepler problem which acts effectively on the regularized energy surface $S^3\times S^2$.

Proof: Since $G = SU(2)\times SU(2)$ and $U(1)$ commute, G induces the action on $S^3\times S^3/U(1) = S^3\times S^2$. By the same method as in the proof of Proposition 5.1, we can see that ω_c^{rd} is invariant under G. Incidentally, as was shown in Ref. 10, the action of G on $S^3\times S^2$ is not effective. According to Theorem 5.2 in Ref. 10, $SO(4) = G/S^0$ acts on $S^3\times S^2$ effectively. Of course, ω_c^{rd} is also left invariant under G/S^0, so that SO(4) turns out to be a symmetry group of the ordinary Kepler problem. This completes the proof.

[1]H. Bacry, H. Ruegg, and J. M. Souriau, Commun. Math. Phys. **3**, 323 (1966).
[2]J. M. Souriau, *Symposia Matematica*, Vol. XIV (Academic, London, 1974), pp. 343–360.
[3]J. Moser, Commun. Pure Appl. Math. **23**, 609 (1970).
[4]V. Guillemin and S. Sternberg, *Geometric Asymptotics* (American Mathematical Society, Providence, RI, 1977), Chap.4.
[5]V. Fock, Z. Phys. **98**, 145 (1935).
[6]G. Györgyi, Nuovo Cimento A **53**, 717 (1968).
[7]H. H. Rogers, J. Math. Phys. **14**, 1125 (1973).
[8]M. Bander and H. V. Itzykson, Rev. Mod. Phys. **38**, 330 (1966).
[9]Y. Akyildiz, J. Math. Phys. **21**, 665 (1980).
[10]T. Iwai, "On reduction of the four-dimensional harmonic oscillator," preceding paper, J. Math. Phys. **22**, xxx (1981).
[11]J. Baumgarte, J. Reine Angew. Math. **301**, 59 (1978).
[12]J. Baumgarte, Z. Angew. Math. Mech. **59**, 177 (1979).
[13]E. L. Steifel and G. Scheifele, *Linear and Regular Celestial Mechanics* (Springer-Verlag, Berlin, 1971), Chaps. 2 and 3.
[14]A. O. Barut, C. K. E. Schneider, and R. Wilson, J. Math. Phys. **20**, 2244 (1979).
[15]G. A. Ringwood and J. T. Devreese, J. Math. Phys. **21**, 1390 (1980).
[16]M. Ikeda and Y. Miyachi, Math. Japon. **15**, 127 (1970).

[17]In this section, Latin indices range over 1 to n.
[18]To be precise, we have to use the pullback of the mapping (2.16), but we omit doing it.
[19]Since (3.1) is not the standard flat metric, the conjugate momentums are not the same as in (1.4).
[20]This matrix was used in the previous paper(Ref. 10). Missing matrix entries are all zero.

[21]R. Abraham and J. E. Marsden, *Foundations of Mechanics*, 2nd ed. (Benjamin, London, 1978), Chap.4.
[22]J. Dieudonné, *Éléments d'analyses*, Tome III, 2nd ed. (Gauthier-Villars, Paris, 1974), Chap. 16.

[23]No confusion will arise if use is made of the same notations as in (3.8) and (3.9).
[24]By the transformation (2.16), the domain $(\mathbb{R}^4 - \{0\})\times\mathbb{R}^4$ of the conformal Kepler problem is identified with its range in the domain of the harmonic oscillator.

Classification of orbits of Fokker's time-asymmetric relativistic two-body problem

Donald E. Fahnline
The Pennsylvania State University, Altoona, Pennsylvania 16603

(Received 24 November 1980; accepted for publication 27 March 1981)

The requirements that the proper velocities of the particles in Fokker's time-asymmetric relativistic two-body problem be timelike and future pointing restrict the variables ρ_1 and ρ_2 associated with the distances between the particles as measured in their rest frames. Employing these restrictions in an algebraic equation relating ρ_1 and ρ_2 to the total angular momentum and the total four-momentum, assumed timelike and future pointing, classifies the physical orbits of the system. The results include orbits similar to those of the nonrelativistic Kepler problem and several new types in which the angular velocity is opposite to the total angular momentum. This information is required for the integration of the equations of motion to determine the orbits in four-space.

PACS numbers: 03.30. + p

I. INTRODUCTION

In Fokker's time-asymmetric relativistic two-body problem,[1] one spinless electrically charged particle responds without self-action to the retarded Liénard–Wiechert field of a second, while the second responds similarly to the advanced field of the first. The purposes of this paper are to simplify a previously given reduction to quadratures[2] and to identify and classify all possible types of motion of the system consistent with the assumption that the total four-momentum is timelike and future pointing.

Several authors have presented special-case solutions of this problem. Rudd and Hill,[3] Staruszkiewicz,[4] Bruhns,[5] and Künzle[6] obtained solutions for one-dimensional motion. Bruhns[5] also exhibited particular circular motion solutions. Künzle[6] reduced the problem to quadratures and gave numerical results for the bounded motion of equal rest masses resulting from an attractive force and for circular motion with arbitrary mass ratios. The present author gave a reduction to quadratures[2] and a general solution for circular motion.[7]

The second section reviews previous results,[2] presents a slightly simpler reduction to quadratures, and expresses the important quantities of the problem in terms of just two variables. The third section shows that a single algebraic equation relating these two variables to the total energy and angular momentum is a convenient substitute for the nonrelativistic equivalent one-dimensional radial energy equation; its real and positive solutions for one of the variables in terms of the other characterize the possible motions of the system. The results of the analysis of this equation include motions corresponding to the usual nonrelativisitc orbits and several new types of motion in which the angular velocity is opposite to the total angular momentum. The discussion of these orbits in the fourth section includes a proof that they guarantee timelike and future pointing velocities for the two particles. The analysis also shows that the zero angular momentum limit and the limit in which one of the masses becomes infinite are singular.

II. REDUCTION TO QUADRATURES

This section reviews the notation and results of a previous paper,[2] and presents some slight extensions. The space–time positions of the particles x_n^{μ} always have null separation $r^{\mu}r_{\mu} = 0$, where $r^{\mu} \equiv x_1^{\mu} - x_2^{\mu}$ and the particle leading in time is labeled 1 so that $r^0 > 0$. The subscripts n, $f = 1, 2$ always refer to the particles and are never equal when they appear in the same equation. The metric tensor is $g^{ii} = -g^{00} = 1$, $g^{\mu\nu} = 0$ for $\mu \neq \nu$.

The particle momenta are

$$p_n^{\mu} = m_n v_n^{\mu} + g v_f^{\mu}/c\rho_f - g\psi r^{\mu}/2, \tag{1}$$

where c is the speed of light, $g \equiv e_1 e_2/c$ is the coupling constant in Gaussian units, m_n and e_n are the constant rest mass and electric charge of particle n, $v_n^{\mu} \equiv dx_n^{\mu}/d\tau_n$ is the proper velocity of n obeying $v_n^{\mu} v_{n\mu} = -c^2$, $\rho_n \equiv -v_n^{\mu} r_{\mu}/c$, and $\psi \equiv -v_1^{\mu} v_{2\mu}/c^2 \rho_1 \rho_2$. The proper velocities are timelike and future pointing; hence, the ρ_n and ψ are greater than zero.

The total momentum $P^{\mu} \equiv p_1^{\mu} + p_2^{\mu}$, assumed timelike and future pointing, and the total angular momentum $J^{\mu\nu} \equiv \Sigma_n (x_n^{\mu} p_n^{\nu} - x_n^{\nu} p_n^{\mu})$ are conserved. The scalar $m > 0$ defined by

$$m^2c^2 \equiv -P^{\mu}P_{\mu} = \sum_n (m_n c + g/\rho_n)^2 + 2\eta\psi, \tag{2}$$

where $\eta \equiv m_1 m_2 c^2 \rho_1 \rho_2 - g^2$, is conserved and represents the rest mass of the system considered as a composite particle.

The system possesses a "center of motion" x^{μ} with constant proper velocity $v_x^{\mu} \equiv dx^{\mu}/d\tau_x = P^{\mu}/m$. The motion of the system is simplest in the center of motion frame shown in Fig. 1, where x^{μ}, P^{μ}, and $J^{\mu\nu}$ have zero components except for $x^0 = c\tau_x$, $P^0 = mc > 0$, and $J^{12} = -J^{21} \equiv J_3 \equiv J \geqslant 0$. In this frame the center of mass z^{μ} moves in a circle of radius $z \equiv |\mathbf{z}| = J/mc$ about the origin:

$$z^0 = x^0 = c\tau_x, \quad \mathbf{z} = \mathbf{J} \times \mathbf{r}/mcr, \tag{3}$$

where $r \equiv |\mathbf{r}| = \rho_x > 0$, and $J_i = \frac{1}{2}\epsilon_{ijk} J_{jk}$.

The particle positions are

 0022-2488/81/081640-07$01.00 © 1981 American Institute of Physics

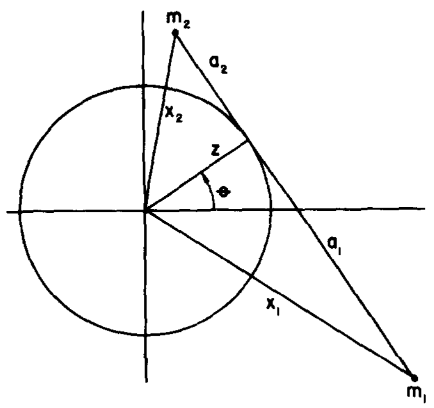


FIG. 1. Geometry of the time-asymmetric problem in the center of motion frame: $z = J/mc$, $a_n = (m_f c\rho_f + g)/mc$.

$$x_n{}^\mu = z^\mu - (-1)^n \Gamma_f r^\mu/\Gamma, \tag{4}$$

where $\Gamma \equiv -P^\mu r_\mu = mc\rho_x$, $\rho_x \equiv -v_x{}^\mu r_\mu/c$, and $\Gamma_f \equiv -p_f{}^\mu r_\mu = m_f c\rho_f + g$. Hence, the particles move in the plane perpendicular to **J**, and **r** always passes through the center of mass perpendicular to **z**. If $J = 0$, the center of mass coincides with the center of motion.

The definition of P^μ provides one relation between the various ρ's

$$\Gamma = mc\rho_x = \Gamma_1 + \Gamma_2 = m_1 c\rho_1 + m_2 c\rho_2 + 2g. \tag{5}$$

A second relation

$$m^2c^2J^2/\eta = m^2c^2 - mc\rho_x(m_1c\rho_1{}^{-1} + m_2c\rho_2{}^{-1}) \tag{6}$$

and an expression for ψ,

$$2\psi = m^2c^2J^2/\eta^2 + \rho_1{}^{-2} + \rho_2{}^{-2}, \tag{7}$$

both evolve from the definition of $J^{\mu\nu}$.

Fokker's problem now reduces to solving the equations

$$m_n c\dot{\rho}_n = (-1)^n\sigma(m_n c\rho_n\psi - m_n c\rho_n{}^{-1} + g\rho_f{}^{-2}) \tag{8}$$

and

$$\mathbf{\Omega} \equiv \dot{\theta}\hat{\mathbf{e}} = mc\sigma\mathbf{J}/\rho_x\eta, \tag{9}$$

where a dot above the variable indicates differentiation with respect to τ_x, $\sigma \equiv mc^2/(m_1c\rho_1{}^{-1} + m_2c\rho_2{}^{-1})$, $\hat{\mathbf{e}}$ is the unit vector along **J**, and θ is the angle between **z** and the x axis in the center of motion frame.

Using (7) to eliminate ψ in (8) yields a pair of coupled ordinary differential equations

$$m_n c\dot{\rho}_n/\sigma = \tfrac{1}{2}m_n c\rho_n[\rho_1{}^{-2} - \rho_2{}^{-2} + (-1)^n(mcJ/\eta)^2] + (-1)^n g\rho_f{}^{-2}. \tag{10}$$

Equation (6), with ρ_x eliminated by (5), is already an integral of (10) and is a cubic equation for ρ_f in terms of ρ_n. Using its algebraic solution to eliminate ρ_f in (10) yields an ordinary differential equation for the single variable ρ_n. Equation (9) for the angular velocity becomes integrable once the solution to (10) is known. Alternatively, dividing (10) by (9) yields the orbital differential equation for ρ_n.

It is also possible to add the two equations in (10) to obtain an expression for $\dot{\rho}_x$. This approach is more difficult because using (5) and (6) to eliminate ρ_n in favor of ρ_x involves solving a biquadratic algebraic equation.[2]

Once the equations of motion are solved for the ρ_n and θ, (4) provides the positions of the particles. All the physically significant variables are determined by the ρ_n in combination with m and J. For example, the scalar product of (1) with P^μ yields the center of motion frame values for $\gamma_n \equiv v_n{}^0/c$

$$mc\gamma_n = m_n c + m_f c\rho_1\rho_2\psi + g/\rho_n; \tag{11}$$

eliminating ψ via (2), (6), and (7) yields

$$2mm_n\gamma_n = m^2 - m_1{}^2 - m_2{}^2 + g^2m^2J^2/\eta^2 + 2(m_n{}^2 - gm_f/c\rho_f)$$

III. GRAPHS OF THE ρ_2,ρ_1 EQUATION

The last section has shown that the solution of (6) for ρ_f in terms of ρ_n, m, and J is essential for integrating the equations of motion and for determining the other variables in Fokker's problem. A third reason for studying (6) is that the requirement that ρ_2 be real and positive imposes limits on the ranges of ρ_1 and m for given values of J, and vice versa. In fact, analysis of (6) yields a complete classification of the types of motion of the system. The following substitutions conveniently abbreviate the notation in dimensionless quantities; reversing them returns the equations to Gaussian units:

$$\begin{aligned} &m_1c\rho_n/|g| \to \rho_n > 0, && m_1c\rho_x/|g| \to \rho_x > 0,\\ &m_n/m_1 \to m_n, && m/m_1 \to m > 0,\\ &g/|g| \to g = \pm 1, && \eta/g^2 \to \eta,\\ &J/|g| \to J \geqslant 0, && m_1{}^2c^2\sigma/g^2 \to \sigma > 0,\\ &m_1c^2\tau_n/|g| \to \tau_n, && m_1{}^2c^2\tau_x/|g| \to \tau_x. \end{aligned}$$

Substituting (5) into (6) and using the above abbreviations yield an equation relating ρ_1, ρ_2, m, and J

$$f \equiv \eta m\rho_x(\rho_2 + m_2\rho_1) - m^2(\eta - J^2)\rho_1\rho_2 = 0, \tag{12}$$

where both $\eta \equiv m_2\rho_1\rho_2 - 1$ and $m\rho_x \equiv \rho_1 + m_2\rho_2 + 2g > 0$ are now considered abbreviations for combinations of ρ_1 and ρ_2. (If $g = 0$, the abbreviations are still convenient, but division by $|g|$ must be replaced by division by 1 times the units of g, and η is $m_2\rho_1\rho_2$. For now it is assumed that $g \neq 0$.)

The inequalities $\rho_n > 0$ result from the assumptions that $r^0 > 0$ and that the proper velocities are timelike and future pointing. However, reversing the signs of both ρ_1 and ρ_2 in (12) is equivalent to reversing the sign of g; hence, it is convenient to fix $g = -1$ and consider (12) for both positive and negative values of ρ_1 and ρ_2. The first quadrant of the ρ_1,ρ_2 plane then corresponds to attraction, and the third quadrant to repulsion. The second and fourth quadrants are not physical in the present context, but they are included because they contain connecting links between the physical branches of the graph and because they aid the analysis.

Equation (12) is cubic for ρ_f in terms of ρ_n

$$\rho_f{}^3 + 3a_n\rho_f{}^2 + b_n\rho_f + c_n = 0, \tag{13}$$

where

$$\begin{aligned} c_n &\equiv -(\rho_n + 2g/m_n)/m_2,\\ e_n &\equiv [(1 + m_2{}^2 - m^2)\rho_n + 2gm_n]/m_2,\\ a_n &\equiv (e_n - 1/m_2\rho_n)/3, \end{aligned}$$

and

$$b_n \equiv (mJ/m_2)^2 - m_2\rho_n c_n - e_n/m_2\rho_n.$$

The m_n appearing in the definitions of c_n and e_n is equal to 1 for $n = 1$. The value of g is -1 in accordance with the last paragraph. Let

$$A_n \equiv a_n^{\,2} - b_n/3,$$

$$B_n \equiv a_n b_n/2 - a_n^{\,3} - c_n/2,$$

and

$$D_n \equiv B_n^{\,2} - A_n^{\,3}.$$

If $D_n > 0$, (13) has one real root

$$\rho_f = G_n + H_n - a_n$$

and two conjugate imaginary roots

$$\rho_f = -(G_n + H_n)/2 \pm i(3)^{1/2}(G_n - H_n)/2 - a_n,$$

where $G_n^{\,3} \equiv B_n + \sqrt{D_n}$, and $H_n^{\,3} \equiv B_n - \sqrt{D_n}$. If $D_n \leqslant 0$, (13) has three real roots

$$\rho_f = 2A_n^{\,1/2}\cos(\phi/3 + 120°k) - a_n,$$

where $\cos\phi = B_n A_n^{\,-3/2}$, and $k = 0,1,2$. These solutions are useful for drawing graphs of (12). Figure 2 shows the curves for $m_2 = 2$, $J = 1.25$, and different values of m.

For equal rest masses a simpler procedure is available. Letting $m_2 = 1$ reduces (12) to a quadratic for η in terms of ρ_x

$$\eta^2 - (\rho_x^{\,2} + 2\rho_x/m - 1 + J^2)\eta - J^2 = 0. \tag{14}$$

Where the solution for η is real, the values for ρ_1 and ρ_2 are

$$2\rho_n = m\rho_x + 2 \pm (-1)^n[(m\rho_x + 2)^2 - 4(\eta + 1)]^{1/2}. \tag{15}$$

Neither of the above algebraic solutions is convenient

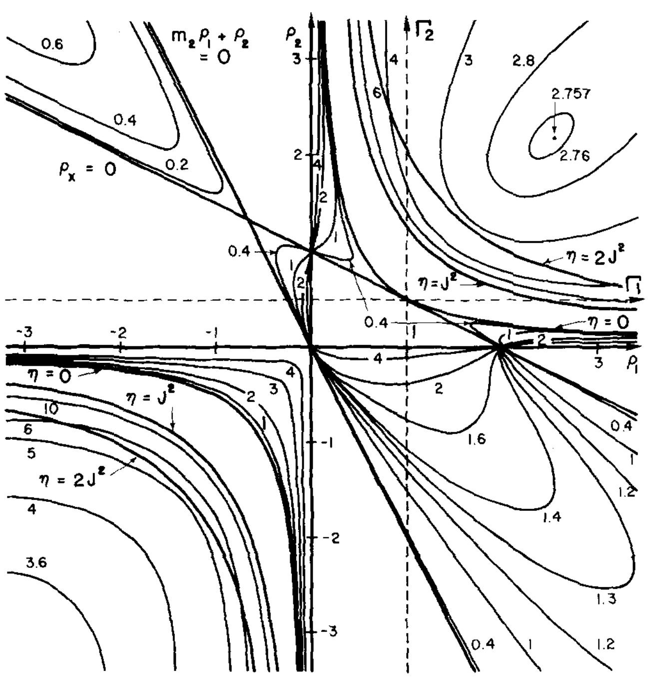


FIG. 2. Examples of the locus of Eq. (12) in the ρ_1,ρ_2 plane for $g = -1$, $m_2 = 2$, $J = 1.25$, and the values of m indicated near each curve. Also shown are the lines $\rho_x = 0$ and $m_2\rho_1 + \rho_2 = 0$; the hyperbolas $\eta = 0$, $\eta = J^2$ and $\eta = 2J^2$; and the Γ_1 and Γ_2 axes. The point labeled $m = 2.757$ corresponds to circular motion.

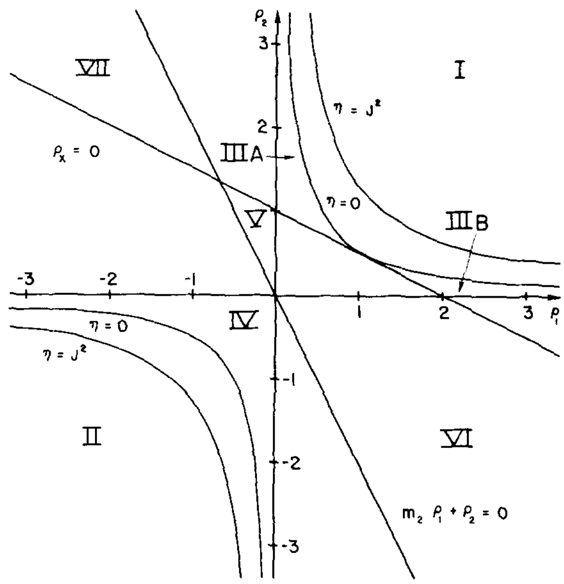


FIG. 3. The regions of the ρ_1,ρ_2 plane for $g - 1$, $m_2 = 2$, $J = 1.25$.

for a general analysis of the types of motion of the system. It is easier to consider (12) directly for different ranges of η and for increasing values of m within each range of η. Figure 3 shows the locations of the various regions of the ρ_1,ρ_2 plane discussed in the following. Until the case $J = 0$ is considered at the end of this section, J has an arbitrary fixed value greater than zero.

a. $\eta < -1$: This range of η corresponds to the second and forth quadrants of the ρ_1,ρ_2 plane. Eq. (12) then requires

$$\rho_x(\rho_2 + m_2\rho_1) = m(\eta - J^2)\rho_1\rho_2/\eta < 0.$$

Therefore, the locus of (12) is further restricted to the regions V, VI, and VII between the lines $\rho_x = 0$ and $\rho_2 + m_2\rho_1 = 0$. If $m_2 > 1$, these lines intersect in the second quadrant at $\rho_1 = -2/(m_2^{\,2} - 1)$, $\rho_2 = 2m_2/(m_2^{\,2} - 1)$ as shown in Figs. 2 and 3 for $m_2 = 2$. If $m_2 < 1$, the intersection lies in the fourth quadrant, and the roles of the second and fourth quadrant are simply interchanged from the $m_2 > 1$ case. If $m_2 = 1$, the lines are parallel, and the graphs are symmetrical with respect to the line $\rho_1 = \rho_2$. It will be assumed for the remainder of the discussion that $m_2 > 1$.

On the lines $\rho_x = 0$ and $\rho_2 + m_2\rho_1 = 0$ the value of the function f defined by (12) is $m^2(J^2 - \eta)\rho_1\rho_2 < 0$. Along the segment $0 < \rho_2 < 2/m_2$ of the ρ_2 axis bounding region V, the value of f is $-(m_2\rho_2 - 2)\rho_2 > 0$. Hence a branch of $f = 0$ always exists in region V for all values of $m > 0$ and $J \geqslant 0$. Similarly, there must be a branch in region VI because the value of f is $-(\rho_1 - 2)m_2\rho_1 > 0$ along the segment $0 < \rho_1 < 2$ of the ρ_1 axis.

The branch in region V is bounded by the region itself; whether the branch in region VI is bounded or unbounded depends on the value of m. In terms of the general quadratic

$$h \equiv m_2^{\,2}\rho_2^{\,2} + m_2[(1 + m_2^{\,2} - m^2)\rho_1 - 2]\rho_2 + m_2^{\,2}\rho_1^{\,2} - 2m_2^{\,2}\rho_1 + K, \tag{16}$$

where K is an arbitrary parameter, the function f is

$$f = (h\eta + K)/m_2 + (m^2J^2 - K)\rho_1\rho_2. \tag{17}$$

The discriminant of h is

$$D = m_2^2[(1 + m_2^2 - m^2)^2 - 4m_2^2]. \tag{18}$$

For $m < |m_2 - 1|$ this discriminant is positive; hence the curves $h = 0$ exist and are hyperbolas lying in regions V, VI, and VII and in the lower left corner of the first quadrant. Their transverse axis lies 45° below the ρ_1 axis. (The inequality cannot occur for $m_2 = 1$.) On the particular curve $h_1 = 0$ with $K = m^2J^2$, Eq. (17) shows that $f = m^2J^2/m_2 > 0$. On the curve $h_2 = 0$ with $K = m^2J^2 - 1$, however, (17) gives $f = \rho_1\rho_2 + (m^2J^2 - 1)/m_2$, which is negative for all points such that $\rho_1\rho_2 < (1 - m^2J^2)/m_2$. Hence, unbounded branches of (12) exist in regions VI and VII for all m such that $0 < m < |m_2 - 1|$.

In region VII rearranging f as

$$f = \eta\{m_2(\rho_1 + \rho_2)^2 - 2(\rho_2 + m_2\rho_1) + [(m_2 - 1)^2 - m^2]\rho_1\rho_2\} + m^2J^2\rho_1\rho_2 \tag{19}$$

shows that $f < 0$ for all $m \geqslant |m_2 - 1|$, since $\rho_2 + m_2\rho_1$ is negative. Therfore, no branches of (12) exist in region VII for $m \geqslant |m_2 - 1|$. The branches which always exist in region VI are still unbounded for $m = |m_2 - 1|$, because $h_1 = 0$ and $h_2 = 0$ are then parabolas in region VI. For $m > |m_2 - 1|$ and all sufficiently negative $\rho_1\rho_2$, the term $[(m_2 - 1)^2 - m^2]\rho_1\rho_2$ in (19) dominates the $2(\rho_2 + m_2\rho_1)$ term and forces $f < 0$; hence, the locus of (12) in region VI is now bounded.

b. $\eta = -1$. The locus of $\eta = -1$ consists of the ρ_1 and ρ_2 axes. Letting $\rho_1 = 0$ and solving (12) for ρ_2 yield $\rho_2 = 0$, $2/m_2$, and $\pm\infty$; setting $\rho_2 = 0$ yields $\rho_1 = 0$, 2, and $\pm\infty$. These solutions are valid for all values of m and J. For points near these solutions, the curves are approximately given by

$$(\rho_1 = 2, \rho_2 = 0), \qquad \rho_2 \simeq m_2(\rho_1 - 2)/(m^2J^2 + m^2 - m_2^2), \tag{20}$$

$$(\rho_1 = 0, \rho_2 = 2/m_2), \qquad \rho_1 \simeq m_2(\rho_2 - 2/m_2)/(m^2J^2 + m^2 - 1), \tag{21}$$

$$(\rho_1 = \pm\infty, \rho_2 = 0), \qquad \rho_2 \simeq (1 - m^2J^2/m_2^2\rho_1^2)/m_2\rho_1, \tag{22}$$

$$(\rho_1 = 0, \rho_2 = \pm\infty), \qquad \rho_1 \simeq (1 - m^2J^2/m_2^2\rho_2^2)/m_2\rho_2, \tag{23}$$

$$(\rho_1 = 0, \rho_2 = 0), \qquad \rho_2 \simeq -m_2\rho_1. \tag{24}$$

c) $-1 < \eta < 0$. The inequality $-1 < \eta < 0$ defines region IV in the third quadrant, and a similar region in the first quadrant. Equation (12), however, restricts ρ_x to positive values in the first quadrant:

$$\rho_x = m(\eta - J^2)(\eta + 1)/m_2\eta(\rho_2 + m_2\rho_1) > 0.$$

Hence, branches of (12) exist in the first quadrant for this range of η only in regions IIIA and IIIB. For region IIIB, f is less than zero on the boundary $\rho_2 = 0, \rho_1 > 2$; greater than zero on the boundary $\eta = 0$; and greater than zero on the boundary $\rho_x = 0$. These and similar inequalities for the other two regions show that branches of (12) exist in regions IIIA, IIIB, and IV for all $m > 0$ and $J > 0$. It is also evident from (20) and (24) that the branch in region IIIA extends through $(\rho_1 = 0, \rho_2 = 2/m_2)$ to the branch in region V and through the origin to region VI. Equation (21) shows that the branch in region VI, which may be unbounded, is joined through $(\rho_1 = 2, \rho_2 = 0)$ to the branch in region IIIB.

d. $0 \leqslant \eta \leqslant J^2$. For any value of $\eta \geqslant 0$, the value of $m\rho_x(\rho_2 + m_2\rho_1)$ is positive (except at $\rho_1 = 1, \rho_2 = 2/m_2$):

$$m\rho_x(\rho_2 + m_2\rho_1) = m_2(\rho_1 - 1)^2 + (m_2\rho_2 - 1)^2/m_2 + (m_2^2 + 1)\eta/m_2 > 0.$$

This and the additional restriction $\eta \leqslant J^2$ imply that $f > 0$. Hence there is no locus of (12) in the regions of the first and third quadrants where $0 \leqslant \eta \leqslant J^2$. Furthermore, no branch of (12) with $\eta < 0$ can join a branch with $\eta > J^2$.

e. $\eta > J^2$. In region I of the first quadrant, where $\eta > J^2$, it is more convenient to consider the function

$$F \equiv f/\rho_1\rho_2(\eta - J^2) = \eta m\rho_x(\rho_1^{-1} + m_2\rho_2^{-1})/(\eta - J^2) - m^2 \tag{25}$$

than f itself; the $F = 0$ and $f = 0$ graphs are identical here. Setting the combinations $\rho_1\,\partial F/\partial\rho_1 \pm \rho_2\partial F/\partial\rho_2$ equal to zero yields the following conditions for the critical points of F:

$$\rho_2(m_2\rho_2 - 1) = m_2\rho_1(\rho_1 - 1) \tag{26}$$

and

$$J^2 = \eta^2/(m_2\rho_1\rho_2 m\rho_x + \eta). \tag{27}$$

Equation (26) requires $\rho_1 > 1$ and

$$\rho_2 = 1/2m_2 + [1/4m_2^2 + \rho_1(\rho_1 - 1)]^{1/2} > 1/m_2, \tag{28}$$

because $\rho_1 \leqslant 1$ would imply $m_2\rho_2 \leqslant 1$ and thus $\rho_x \leqslant 0$ which is impossible in region I. The algebraic equation for ρ_1 in terms of the independent variable J resulting from using (28) to eliminate ρ_2 from (27) is complicated; the existence of a unique solution is most easily inferred indirectly.

First, let $\rho_1 > 1$ be the independent variable. Then (27) and (28) give functions for ρ_2 and J such that (ρ_1, ρ_2) is a critical point of F. It is easily checked that the function for J is a strictly increasing function of ρ_1. Hence an inverse must exist: given any $J > 0$, there exists a unique $\rho_1 > 1$ such that (ρ_1, ρ_2) with ρ_2 given by (28) is a critical point of F.

At the critical point, (12) shows that there is a unique value m_c of m such that $F = 0$. For this value of m the critical point is an absolute minimum of F as determined by the inequalities

$$\begin{aligned} \Delta &\equiv (\partial^2F/\partial\rho_1\partial\rho_2)^2 - (\partial^2F/\partial\rho_1^2)(\partial^2F/\partial\rho_2^2) \\ &= -4(1 - J^2/\eta)^{-1}\rho_1^{-4}\rho_2^{-4}\{m_2(\rho_2 + m_2\rho_1)(\rho_1^2 + \rho_2^2) \\ &\quad + (\eta + 1)[\eta + (m_2^2\rho_1^2 + 1)(\rho_1 - 1) + (\rho_2^2 + 1) \\ &\quad \times (m_2\rho_2 - 1)]\} < 0 \end{aligned} \tag{29}$$

and

$$\partial^2F/\partial\rho_1^2 < 0. \tag{30}$$

Hence, $f = F = 0$ has no locus in region I for $0 < m < m_c$; for $m = m_c$ the locus is the critical point; and for $m > m_c$, the locus is a real curve. The $m = m_c$ critical point locus determined by (12), (26), and (27) corresponds to the circular motion solution of Fokker's problem. This shows that $m_c < 1 + m_2$.[7]

The region I locus for $m > m_c$ may be bounded or unbounded. If $m_c < |m_2 - 1|$, then m can satisfy $m_c < m \leqslant |m_2 - 1|$; in this case rewriting f as

$$\begin{aligned} f = \eta\{&m_2(\rho_1 - 1)^2 + m_2(\rho_2 - 1/m_2)^2 + 2m_2\rho_1\rho_2 \\ &+ [(1 - m_2)^2 - m^2]\rho_1\rho_2 \\ &- (m_2^2 + 1)/m_2\} + m^2J^2\rho_1\rho_2 \end{aligned}$$

shows that $f > 0$ for sufficiently large ρ_1 or ρ_2 and that the locus is bounded. For $|m_2 - 1| \leqslant m_c < m < m_2 + 1$, the discriminant (18) of the general quadratic h defined in (16) is negative. Hence $h = 0$ represents an ellipse, a circle, a point, or no locus. According to (17), the particular quadratic h_1 with $K = m^2J^2$ has the value $h_1 = -m^2J^2/\eta < 0$ on the locus $f = 0$, which exists according to the last paragraph. Since $h_1 > 0$ at the origin, the curve $h_1 = 0$ exists and must be an ellipse surrounding $f = 0$. Outside the ellipse h_1 is greater than zero. Hence (17) shows that

$$f = (h_1\eta + m^2J^2)/m_2 > 0$$

outside the ellipse, and the locus $f = 0$ is bounded.

In the third quadrant rewriting f as

$$f = \eta\{m_2(\rho_1 - \rho_2)^2 - 2(\rho_2 + m_2\rho_1) + [(m_2+1)^2 - m^2]\rho_1\rho_2\} + m^2J^2\rho_1\rho_2$$

shows that $f > 0$ for $\eta > J^2$ and $m \leqslant 1 + m_2$. Therefore, (12) has no locus in region II for $m \leqslant 1 + m_2$.

For $m \geqslant m_2 + 1$ the discriminant (18) implies that the locus $h = 0$ exists in region I as a parabola with its axis lying 45° above the ρ_1 axis or as one branch of a hyperbola with its transverse axis at 45°. In either case (17) implies that $f = m^2J^2/m_2 > 0$ on the curve $h_1 = 0$ with $K = m^2J^2$. For all points such that $\rho_1\rho_2 > (m^2J^2 + 1)/m_2$ on the curve $h_3 = 0$ defined by $K = m^2J^2 + 1$, Eq. (17) yields $f = -\rho_1\rho_2 + (m^2J^2+1)/m_2 > 0$. Hence the locus of (12) is unbounded in region I for $m \geqslant m_2 + 1$. The same analysis shows that the locus of (12) exists and is unbounded in region II for $m > m_2 + 1$.

The case $J = 0$, which has been excluded above, reduces (12) to $\eta = 0$ and

$$f_1 \equiv m\rho_x(\rho_2 + m_2\rho_1) - m^2\rho_1\rho_2 = 0. \quad (31)$$

Fig. 4 shows loci of $\eta = 0$ and (31) for $m_2 = 2$ and various values of m. The conditions for the existence of bounded and unbounded branches in regions I and II remain the same as for $J > 0$. The loci of (12) for fixed m have an interesting

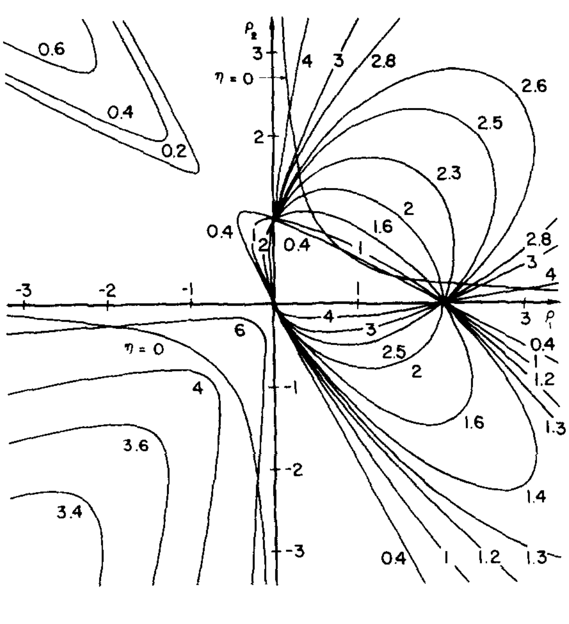


FIG. 4. Examples of the locus of Eq. (31) and $\eta = 0$ for $g = -1$, $m_2 = 2$, $J = 0$, and the values of m indicated near each curve.

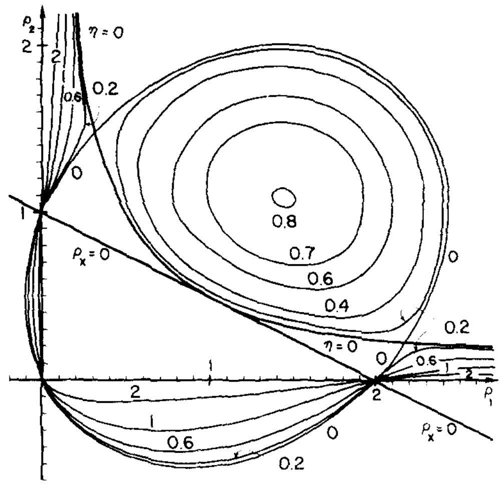


FIG. 5. The $J \to 0$ limit of Eq. (12) illustrated for $g = -1$, $m_2 = 2$, $m = 2.42$, and the values of J indicated near each curve.

singularity in the limit $J \to 0$, as illustrated for the first quadrant by Fig. 5 with $m_2 = 2$ and $m = 2.42$. For any arbitrarily small positive value of J, the loci in regions I, IIIA, and IIIB are separated by the region $0 < \eta < J^2$; at $J = 0$ they merge into two intersecting curves. A similar singularity occurs in the third quadrant for values of m such that (31) intersects $\eta = 0$. Fig. 5 also serves to illustrate the general variation of the locus with J for fixed m.

The effect of increasing m_2 is to move the ρ_2 axis intercept of the line $\rho_x = 0$ closer to the origin and to move the curves $\eta = 0$ and $\eta = J^2$ closer to the ρ_1 and ρ_2 axes. Figure 6 shows the first quadrant loci for $J = 0.8$, $m - 1 - m_2 = -0.6$, and increasing values of m_2. Also included is the locus for $m_2 = \infty$, i.e., the graph for the relativistic one-body Coulomb problem with energy E and rest mass such that $(E - m_0c^2)/m_0c^2 = -0.6$ and with angular momentum $J = 0.8$. It is evident that the form of the locus is singular in the limit $m_2 \to \infty$, because the two-body loci never drop below $\eta = J^2$ in the first quadrant for any arbitrarily large value of m_2. (However, the Coulomb problem locus descends below the ρ_1 axis only for values of $J < 1$.)

For $g = 0$, Eq. (12) reduces to a quadratic relation between ρ_1 and ρ_2. It is easily shown that the loci of (12) exist only in region I and only for $m > 1 + m_2$, and that they are unbounded.

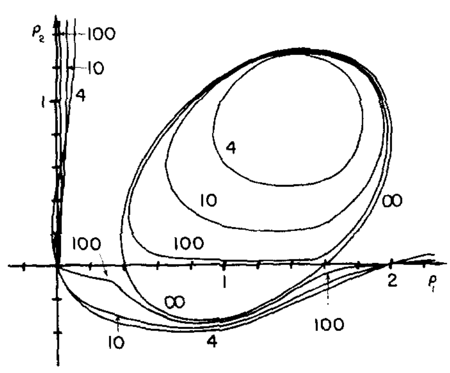


FIG. 6. The $m_2 \to \infty$ limit of Eq. (12) illustrated for $g = -1$, $J = 0.8$, $m - 1 - m_2 = -0.6$, and the values of m_2 indicated near each curve.

IV. PHYSICAL ORBITS

The basic requirement for physical orbits in the context of Fokker's problem is that the $v_n{}^\mu$ be timelike and future pointing. The additional assumptions that P^μ is timelike and future pointing and that (for definiteness) $r^0 > 0$ then imply that the ρ_n are real and positive, and thus restrict the solutions of (12). Conversely, given that P^μ is timelike and future pointing, that $r^0 > 0$, and that the ρ_n are real and positive, it follows that the $v_n{}^\mu$ are timelike and future pointing. Squaring (11) and subtracting (2) yield

$$
\begin{aligned}
(2\psi - \rho_f{}^{-2})m^2c^2(\gamma_n{}^2 - 1) \\
&= [(2\psi - \rho_f{}^{-2})g + m_f c\rho_f(\psi - \rho_f{}^{-2})]^2 \\
&\quad + (m_f c\psi\rho_1\rho_2)^2(2\psi - \rho_1{}^{-2} - \rho_2{}^{-2}).
\end{aligned}
$$

Since $2\psi \geqslant \rho_1{}^{-2} + \rho_2{}^{-2} > \rho_f{}^{-2}$ by (7), this shows that $\gamma_n{}^2 \geqslant 1$, that the $v_n{}^\mu$ are timelike, and that a rest frame exists for each particle. In the rest frame of particle n, the assumptions and the definition of ρ_n imply that $v_n{}^\mu$ is future pointing. Hence, the first quadrant loci found for (12) with $g = -1$ in the last section determine physical orbits for the two particles for an attractive force. Since changing the signs of the ρ_n is equivalent to reversing the sign of g, the third quadrant loci are physical for repulsion. The general direction determined by (10) for the motion of $(\rho_1(\tau_x), \rho_2(\tau_x))$ on the curves as τ_x increases is clockwise around each quadrant. Equation (4) shows that bounded branches of (12) determine bounded orbits, and unbounded branches determine unbounded orbits.

According to the last paragraph, the loci in regions I and II for $J > 0$, $g \neq 0$ determine physical orbits similar in many respects to the orbits of the nonrelativistic two-body Kepler problem

Region I Attraction	Orbit Type	Region II Repulsion
$0 < m < m_c$	No orbits	$0 < m \leqslant 1 + m_2$
$m = m_c < 1 + m_2$	circular orbit	
$m_c < m < 1 + m_2$	bounded orbits	
$1 + n_2 \leqslant m$	unbounded orbits	$1 + m_2 < m$

However, there are several unusual features. For the smallest allowed values of m in both regions I and II, the branches lie entirely outside the curve $\eta = 2J^2$. As m increases, the branches enlarge, move toward the curve $\eta = J^2$, and eventually cross $\eta = 2J^2$ as shown in Fig. 2. Hence, for any J there is a value of m such that for larger m every orbit contains a segment where $\eta < 2J^2$ and

$$\dot{z}^\mu \dot{z}_\mu = (\sigma/\rho_x)^2(2J^2/\eta - 1) > 0; \tag{32}$$

i.e., $\dot{z}^\mu$ is spacelike and the center of mass moves with an ordinary velocity larger than the speed of light.[2] Similarly, Fig. 3 illustrates that in region I the branches of (12) eventually meet and cross one or both of the lines $\Gamma_1 = \rho_1 - 1 = 0$ and $\Gamma_2 = m_2\rho_2 - 1 = 0$ as m increases. Along the segment of such a branch where Γ_f passes through zero and becomes temporarily negative, (4) shows that particle n moves through the center of mass so that both particles are temporarily on the same side of the center of mass.

The $J = 0$ loci in regions I and II correspond to physical one-dimensional motion, and the conditions for bounded and unbounded orbits are identical to those for $J > 0$. However, the physical significance of the discontinuity in the form of the graphs in the limit $J \to 0$ is unclear. The same is true of the discontinuity in the limit $m_2 \to \infty$.

The branches in region IIIA, IIIB, and IV correspond to physical motion for all values of $m > 0$ and $J > 0$, but they have even more unusual features. Since $\eta < 0$ for these branches, (9) shows that the angular velocity of $\mathbf{r}$ is directed opposite to the angular momentum $\mathbf{J}$. This is due to the presence of an interaction angular momentum which is larger than and opposite to the mechanical angular momentum. In both regions IIIA and IIIB, one of the Γ_n is always negative and the other always positive; hence the particles will always be on the same side of the center of mass. Equation (32) shows that $\dot{z}^\mu$ is always timelike in all three of these regions.

V. DISCUSSION

The analysis of (12) presented here specifies the relations between ρ_1 and ρ_2 for all possible orbits for given values of $J \geqslant 0$, $m > 0$, and timelike total four-momentum. This information is required in order to integrate (9) and (10) to find the angular and temporal dependence of ρ_1 and ρ_2. Then (5) determines r^μ, (3) determines z^μ, and (4) determines the orbits of the two particles in the four-space of the center of motion system. This approach using ρ_1 and ρ_2 appears to be easier than using the single variable ρ_x alone. The results of this continuing work will be presented in a future paper.

Complications arise because (12) is fourth-order; for given fixed values of $J \geqslant 0$, $m > 0$, and g there may be different orbits for different initial conditions. These orbits divide into two principal types: those with $\eta > J^2$ corresponding roughly to the orbits of the nonrelativistic two-body problem and those with $\eta < 0$. In essence, the variable η determines the relation between the mechanical and interaction angular momenta, the sum of which gives the conserved total angular momentum. For $\eta > J^2$ Eq. (9) shows that the velocity has the same direction as $\mathbf{J}$; for $\eta < 0$ the interaction angular momentum dominates the mechanical angular momentum, and the angular velocity is in the opposite direction from $\mathbf{J}$. The initial conditions determine the initial value of η, and no branch of (12) with $\eta < 0$ can join a branch with $\eta > J^2$.

The last section describes the various orbits possible for each of the principal types. Equation (12) plays a role similar to that of the equivalent one-dimensional energy equation in Newtonian physics: both determine many important features of the orbits without requiring the full solution of the angular and temporal equations of motion.

The analysis for timelike and future pointing particle velocities and total momentum generalizes in two ways. First, the continuity of the region IIIA and IIIB curves with those in regions V and VI already suggests that the corresponding orbits are physical in both sets of regions despite singularities at the cross-over points. Such an interpretation is easily accommodated within the present formalism and simply requires that the particle four-velocities be allowed to become infinite and past pointing. Second, for any given initial particle positions and velocities, there are values of g

such that the total momentum is not timelike. Even in this case, (6) (with $m^2c^2J^2$ replaced by $W^{\mu}W_{\mu}\geqslant 0$) and much of the formalism are still valid,[2] but the interpretation of the centers of mass and motion must be modified. These extensions and the singularities in the $J\rightarrow 0$ and $m_2\rightarrow\infty$ limits will also be addressed in future papers.

[1]A. D. Fokker, Physica **9**, 33 (1929).
[2]D. E. Fahnline, J. Math. Phys. **18**, 1006 (1977).
[3]R. A. Rudd and R. N. Hill, J. Math. Phys. **11**, 2704 (1970).
[4]A. Staruszkiewicz, Ann. Physik **25**, 362 (1970).
[5]B. Bruhns, Phys. Rev. D 8, 2370 (1973).
[6]H. P. Künzle, Int. J. Theor. Phys. **11**, 395 (1974).
[7]D. E. Fahnline, J. Math. Phys. **20**, 1118 (1979).

Inverse scattering inverse source theory

Norbert N. Bojarski [a), b)]
16 Pine Valley Lane, Newport Beach, California 92660

(Received 23 April 1980; accepted for publication 20 March 1981)

The inverse scattering inverse source problem associated with the inhomogeneous Helmholtz wave equation, the (special case) Sturm–Liouville (acoustic wave) equation, and the time-independent Schrödinger equation is treated . To this end, the concepts of a reference wave velocity and an associated free reference space Green's function spectrum are introduced. A modified Kirchhoff surface integral, containing only the gradient of the real part of this free reference space Green's function spectrum and the fields on a measurement surface is formulated, yielding an integral equation for the unknown fields and sources in the interior of the closed piecewise smooth surface on which the (remotely sensed) fields are known. A well-posed, analytic closed form solution of this integral equation for the unknown fields and their Laplacians is obtained with the aid of a (modified) spatial Fourier transform in which the reference velocity is continually varied in such a fashion that the Ewald sphere shell sweeps to fill the entire transform space. The unknown potential or medium properties and the unknown sources are then determined algebraically for the inverse scattering and inverse source problems respectively. The effects of finite sampling density and incomplete observation domain are discussed briefly.

PACS numbers: 03.40.Kf, 03.65.Nk, 43.20. – f

I. INTRODUCTION

A unified formulation and solution is presented to the inverse scattering inverse source problems for the time-independent Schrodinger equation

$$(\hbar^2/2m)\nabla^2\phi + (E - V)\phi = 0 \tag{1}$$

the (special case) Sturm–Liouville (acoustic wave) equation

$$\nabla^2\phi(\mathbf{x},\omega) + [\omega^2/c^2(\mathbf{x},\omega)]\phi(\mathbf{x},\omega) = 0, \tag{2}$$

and the inhomogeneous Helmholtz wave equation

$$\nabla^2\phi(\mathbf{x},\omega) + (\omega^2/c^2)\phi(\mathbf{x},\omega) = -\rho(\mathbf{x},\omega), \tag{3}$$

subject to the constitutive equation

$$\rho(\mathbf{x},\omega) = V(\mathbf{x},\omega)\,\phi(\mathbf{x},\omega). \tag{4}$$

To this end, the single mixed wave equation

$$\nabla^2\phi(\mathbf{x},\omega) + [\omega^2/c^2(\mathbf{x},\omega)]\phi(\mathbf{x},\omega) = -\rho(\mathbf{x},\omega) \tag{5}$$

is introduced, [still subject to the constitutive equation (4)], which reduces to (1), (2), or (3), depending on the choice of c and ρ in Eq. (5); i.e., (5) reduces to (3) if c is a known constant, (5) reduces to (2) if $\rho = 0$, and (5) reduces to (1) if $\omega^2/c^2 = (2m/\hbar^2)E$ and ρ is given by the constitutive equation (4).

It is argued that the inverse scattering inverse source solution presented is an alternative (to the direct 1859 Kirchhoff) integration of the wave equation. It is thus appropriate to review this direct integration of (5), as well as some of the properties of this direct integration.

The direct integration of (5) is accomplished in the following fashion: let a reference potential V_r, a reference source distribution ρ_r, and a total source distribution ρ_t be defined, respectively, as

$$V_r \equiv \omega^2/c^2(\mathbf{x},\omega) - \omega^2/\upsilon^2, \tag{6}$$

$$\rho_r \equiv V_r\phi, \tag{7}$$

$$\rho_t \equiv \rho + \rho_r, \tag{8}$$

where υ is an arbitrarily chosen constant "free space" reference velocity. With the aid of Eqs (6)–(8) Eqs. (5) can be rewritten as

$$\nabla^2\phi(\mathbf{x},\omega) + (\omega^2/\upsilon^2)\phi(\mathbf{x},\omega) = -\rho_t(\mathbf{x},\omega). \tag{9}$$

If G is chosen as the free space Green's function associated with the constant reference velocity υ, then Kirchhoff's direct integration of Eq. (9) is

$$\int_v dv\, G\rho_t + \oint_s d\mathbf{s}\cdot(G\nabla\phi - \phi\nabla G) = \begin{cases}\phi, & \forall \mathbf{x}\in v\\ 0, & \forall \mathbf{x}\notin v\end{cases}, \tag{10}$$

which is an equivalent *integral representation* of the partial differential equation (9). It is assumed that the closed surface s is piecewise smooth.

If the free space Green's function G satisfies the Sommerfeld radiation condition at infinity, then the Kirchhoff surface integral in (10) can be recognized as the incident field (the field in v due to all the sources not in v);i.e.,

$$\oint_s d\mathbf{s}\cdot(G\,\nabla\phi - \phi\,\nabla G) = \phi_i. \tag{11}$$

If the total source distribution ρ_t is related to the field ϕ by the constitutive equation

$$\rho_t = V_t\phi. \tag{12}$$

then Eqs. (10)–(12) can be combined to yield the direct scattering Lippmann–Schwinger *integral equation*

$$\phi - \int_v dv\, G\, V_t\phi = \phi_i. \tag{13}$$

A brief review of some of the properties of the Kirchhoff surface integral

$$\oint_s d\mathbf{s}\cdot(G\nabla\phi - \phi\nabla G) \tag{14}$$

is now in order. Specifically, this Kirchhoff surface integral is an equivalence statement relating the field at a field point on one side of the closed surface produced by all the sources

a)This research was supported by the Office of Naval Research under contract N00014-76-C-0082.

b)Research contractor and consultant to the Department of Defense.

 © 1981 American Institute of Physics

on the other side of the closed surface, via the fields produced by these sources on this closed surface [an equivalence statement that permitted the identification of the incident field (11) and the formulation of the Lippmann-Schwinger direct scattering integral equation (13)]. The inverse scattering inverse source problem is, however, characterized by both the field point for the unknown fields as well as all the unknown sources (that produce these fields) being on the same side of the closed surface (on which the remote sensing is accomplished), for which situation the Kirchhoff surface integral vanishes, thus rendering this Kirchhoff surface integral useless for the inverse scattering inverse source problem. A modified Kirchhoff surface integral, which does not suffer from this pathology, is introduced next.

II. THE INVERSE SCATTERING INVERSE SOURCE INTEGRAL EQUATION

Let G be the free reference space Green's function satisfying the inhomogeneous Helmholtz wave equation

$$\nabla^2 G + (\omega^2/\mathscr{c}^2)G = -\delta \tag{15}$$

and the Sommerfeld radiation condition at infinity where $\mathscr{c}$ is any arbitrarily chosen reference velocity.

Next, let an effectal field θ be defined as

$$\theta \equiv \oint_s d\mathbf{s}\cdot(G_r \nabla\phi - \phi\nabla G_r), \tag{16}$$

where

$$G_r \equiv \operatorname{Re} G, \tag{17}$$

which, by (15), also satisfies the inhomogeneous Helmholtz wave equation

$$\nabla^2 G_r + (\omega^2/\mathscr{c}^2)G_r = -\delta. \tag{18}$$

It should be noted that, by Eq. (16), the effectal field $\theta(\mathbf{x})$ can be computed for $\forall \mathbf{x}\in v$ from mere knowledge of the field $\phi(\mathbf{x})$ $\forall \mathbf{x}\in s$.

By Green's theorem, Eq. (16) reduces to

$$\theta = \int_v dv\,(G_r \nabla^2\phi - \phi\nabla^2 G_r), \tag{19}$$

which, by Eqs. (5) and (18), further reduces to

$$\theta = \int_v dv\left[G_r\left(-\frac{\omega^2}{c^2} - \rho\right) - \phi\left(-\frac{\omega^2}{\mathscr{c}^2} - \delta\right)\right] \tag{20}$$

$$= \int_v dv\,\delta\phi - \int_v dv\,G_r\left[\left(\frac{\omega^2}{c^2} - \frac{\omega^2}{\mathscr{c}^2}\right)\phi + \rho\right], \tag{21}$$

where it should be noted that $c = c(\mathbf{x},\omega)$.

By the very definition of the Dirac delta function

$$\int_v dv'\,\delta(\mathbf{x}-\mathbf{x}')\phi(\mathbf{x}') = \begin{cases}\phi(\mathbf{x}), & \forall\mathbf{x}\in v\\ 0, & \forall\mathbf{x}\notin v\end{cases}. \tag{22}$$

With the aid of Eqs. (6)–(8)

$$(\omega^2/c^2 - \omega^2/\mathscr{c}^2)\phi + \rho = \rho_{\mathscr{c}}. \tag{23}$$

With the aid of Eqs. (22) and (23), Eq. (21) reduces to

$$\theta = \begin{cases}\phi - \int_v dv\,G_r\rho_{\mathscr{c}}, & \forall\mathbf{x}\in v\\ -\int_v dv\,G_r\rho_{\mathscr{c}}, & \forall\mathbf{x}\notin v\end{cases}, \tag{24}$$

which is a proper (i.e. $\mathbf{x},\mathbf{x}'\in v$) inverse scattering inverse source integral equation.

With the aid of the direct integration Eq. (10) and (11), this inverse scattering inverse source integral Eq. (24) reduces to the less general 1973 inverse scattering inverse source integral equation of this author,[1] which was studied extensively by Bleistein and this author,[2] Bleistein and Cohen,[3] and others.

III. SOLUTION OF THE INVERSE SCATTERING INVERSE SOURCE INTEGRAL EQUATION

For the purpose of solving the inverse scattering inverse source integral equation (24), it becomes convenient to introduce the characteristic function $\gamma(\mathbf{x})$ for the volume of integration v, defined by

$$\gamma(\mathbf{x}) \equiv \begin{cases}1, & \forall\mathbf{x}\in v\\ 0, & \forall\mathbf{x}\notin v\end{cases}. \tag{25}$$

with the aid of Eq. (25), Eq. (24) can be written as

$$\theta(\mathbf{x},\omega,\mathscr{c}) = \gamma(\mathbf{x})\,\phi(\mathbf{x},\omega) - \int_{-\infty}^{\infty} G_r(\mathbf{x}|\mathbf{x}',\omega,\mathscr{c})\,\gamma(\mathbf{x}')\rho_{\mathscr{c}}(\mathbf{x}',\omega,\mathscr{c})\,d^n x'. \tag{26}$$

Introducing the characteristic source distribution $\rho_\gamma(\mathbf{x},\omega,\mathscr{c})$, defined by

$$\rho_\gamma(\mathbf{x},\omega,\mathscr{c}) \equiv \gamma(\mathbf{x})\,\rho_{\mathscr{c}}(\mathbf{x},\omega,\mathscr{c}), \tag{27}$$

permits the rewriting of Eq. (26) as

$$\theta(\mathbf{x},\omega,\mathscr{c}) = \gamma(\mathbf{x})\,\phi(\mathbf{x},\omega) - \int_{-\infty}^{\infty} G_r(\mathbf{x}|\mathbf{x}',\omega,\mathscr{c})\rho_\gamma(\mathbf{x}',\omega,\mathscr{c})\,d^n x', \tag{28}$$

which, in cartesian coordinates, can be rewritten as

$$\theta(\mathbf{x},\omega,\mathscr{c}) = \gamma(\mathbf{x})\,\phi(\mathbf{x},\omega) - \int_{-\infty}^{\infty} G_r(\mathbf{x}-\mathbf{x}',\omega,\mathscr{c})\rho_\gamma(\mathbf{x}',\omega,\mathscr{c})\,d^n x', \tag{29}$$

Since the Green's function G_r in Eq. (29) is a spatial difference kernel in cartesian coordinates Eq. (29) can be further rewritten as the n-dimensional spatial convolution

$$\theta(\mathbf{x},\omega,\mathscr{c}) = \gamma(\mathbf{x})\,\phi(\mathbf{x},\omega) - G_r(\mathbf{x},\omega,\mathscr{c})^*\rho_\gamma(\mathbf{x},\omega,\mathscr{c}). \tag{30}$$

Taking the n-dimensional spatial Fourier transform of Eq. (29) thus yields the algebraic product equation

$$\tilde{\theta}(\mathbf{k},\omega,\mathscr{c}) = \tilde{\gamma}(\mathbf{k})^*\tilde{\phi}(\mathbf{k},\omega) - \tilde{G}_r(\mathbf{k},\omega,\mathscr{c})\tilde{\rho}_\gamma(\mathbf{k},\omega,\mathscr{c}). \tag{31}$$

It should be noted here that the field ϕ, and its spatial Fourier transform $\tilde{\phi}$, does not depend on the arbitrarily chosen reference velocity $\mathscr{c}$. This is mathematically self evident since the wave equation (5) does not contain this arbitrarily chosen reference velocity $\mathscr{c}$, and physically self evident since the physical field and its spatial Fourier transform cannot depend on the arbitrary choice of the reference velocity $\mathscr{c}$.

A digression examining some of the properties of the spatial Fourier transform of the Green's function is now in order. The spatial Fourier transform of the Green's function is

$$\tilde{G}(\mathbf{k},\omega,c) = \mathrm{P}\frac{1}{k^2-\omega^2/c^2} + (i\pi/2k)[\delta(k-\omega/c)-\delta(k+\omega/c)], \quad (32)$$

where P denotes the principal value; i.e.,

$$\mathrm{P}\frac{1}{k^2-\omega^2/c^2} \equiv \begin{cases} \dfrac{1}{k^2-\omega^2/c^2}, & k^2 \neq \omega^2/c^2 \\ 0, & \forall k^2 = \omega^2/c^2 \end{cases}. \quad (33)$$

It should be noted that the functional form of the spatial Fourier transform of the Green's function in k space is invariant to the dimensionality of the space, which is not the case for the Green's function in x space. Furthermore, the spatial Fourier transform of the real and imaginary parts of the Green's function are respectively

$$\tilde{G}_r(\mathbf{k},\omega,c) = \mathrm{P}\frac{1}{k^2-\omega^2/c^2} \quad (34)$$

$$\tilde{G}_i(\mathbf{k},\omega,c) = (\pi/2k)[\delta(k-\omega/c)-\delta(k+\omega/c)]. \quad (35)$$

The notation used here for the imaginary part of the Green's function and its spatial Fourier transform is consistent with the notation used for the real part of the Green's function and its spatial Fourier transform; i.e., $\mathrm{Re}G = G_r \leftrightarrow \tilde{G}_r$ and $\mathrm{Im}G_i = G_i \leftrightarrow \tilde{G}_i$.

Next, the support of the spatial Fourier transform of the real and imaginary parts of the Green's function is examined. By Eqs. (32)–(35)

$$\mathrm{Sup}\tilde{G}_r(\mathbf{k},\omega,c) \in \forall\, k^2 \neq \omega^2/c^2, \quad (36)$$

$$\mathrm{Sup}\tilde{G}_i(\mathbf{k},\omega,c) \in \forall\, k^2 = \omega^2/c^2, \quad (37)$$

i.e., $\tilde{G}_r(\mathbf{k},\omega,c)$ is nonzero everywhere except on the Ewald sphere shell $k^2 = \omega^2/c^2$, and $\tilde{G}_i(\mathbf{k},\omega,c)$ is non-zero only on the Ewald sphere shell $k^2 = \omega^2/c^2$; and conversely, $\tilde{G}_r(\mathbf{k},\omega,c)$ is zero only on the Ewald sphere shell $k^2 = \omega^2/c^2$ and $\tilde{G}_i(\mathbf{k},\omega,c)$ is zero everywhere except on the Ewald sphere shell $k^2 = \omega^2/c^2$. Thus (on the Ewald sphere shell)

$$\tilde{G}_r(\mathbf{k},\omega,c)|_{c=\omega/k} = \tilde{G}_r(\mathbf{k},\omega,\omega/k) = 0, \quad (38)$$

where $k = |\mathbf{k}|$. Equation (31), with the aid of Eq. (38), thus yields (on the Ewald sphere shell)

$$\tilde{\theta}(\mathbf{k},\omega,\omega/k) = \tilde{\gamma}(\mathbf{k})*\tilde{\phi}(\mathbf{k},\omega). \quad (39)$$

It is noteworthy that for far-fields ϕ in Eq. (16), for which the volume v is infinite and $\gamma(\mathbf{k}) = \delta(\mathbf{k})$, on the Ewald sphere shell in the spatial Fourier transform space, the known effectal field is equal to the unknown field.

Earlier attempts at solving the inverse scattering inverse source problems have yielded somewhat similar results; i.e., a solution for the characteristics source distribution on the Ewald sphere shell in the Fourier transform space. The difficulty with such solutions, however, is that the characteristic source distribution depends on the arbitrarily chosen reference velocity, which precluded the determination of the characteristic source distribution off the Ewald sphere shell, whereas the field, as used in this solution, does not depend on this arbitrarily chosen reference velocity. It is thus possible to vary the arbitrary reference velocity in such a fashion that the Ewald sphere shell sweeps to fill the entire Fourier transform space. This is accomplished simply as follows.

Taking the spatial inverse Fourier transformof Eq. (39), and recalling the definition (25), thus yields the desired solution for the field ϕ in volume v; i.e.,

$$\frac{1}{(2\pi)^\eta}\int_{\infty}^{-\infty} e^{-i\mathbf{k}\cdot\mathbf{x}}\tilde{\theta}(\mathbf{k},\omega,\omega/k)\,d^\eta k = \begin{cases}\phi(\mathbf{x},\omega), & \forall\ \mathbf{x}\in v \\ 0, & \forall\ \mathbf{x}\notin v\end{cases}, \quad (40)$$

and, by the Fourier transform differentiation rule, the solution for the Laplacian of the field $\nabla^2\phi$ in the volume v; i.e,

$$-\frac{1}{(2\pi)^\eta}\int_{\infty}^{-\infty} e^{-i\mathbf{k}\cdot\mathbf{x}}\tilde{\theta}(\mathbf{k},\omega,\omega/k)\,k^2 d^\eta k = \begin{cases}\nabla^2\phi(\mathbf{x},\omega), & \forall\mathbf{x}\in v \\ 0, & \forall\mathbf{x}\notin v\end{cases}. \quad (41)$$

This solution [(40)and (41)] thus has the (previously mentioned) desired properties converse to the properties of the direct Kirchhoff integration; i.e., for the case of the (unknown) sources being in the volume v, a solution is yielded for the fields if the field point is also in this volume v, and zero is yielded if the field point is not in this volume v; or, more generally, a solution for the fields is yielded if the sources and the field point are on the same side of the closed surface s, and zero is yielded if the sources and the field point are on different sides of the closed surface s.

The above solution [(40) and (41)] depends only on $\tilde{\theta}(\mathbf{k},\omega,\omega/k)$, i.e., only on the values of the effectal field in the spatial Fourier transform space which are on the Ewald sphere shell $k^2 = \omega^2/c^2$. Examination of Eq. (16) which defines this effectal field ϕ in real space shows it to consist of two terms, the first term depending only on the real part of the Green's function and the second term depending only on the gradient of the real part of the Green's function. Thus, in the spatial Fourier transform space, this effectal field consists also of two terms, the first depending only on the spatial Fourier transform of the real part of the Green's function, and the second term depending only on the spatial Fourier transform of the gradient of the real part of the Green's function. However, on the Ewald sphere shell $k^2 = \omega^2/c^2$ in the spatial Fourier transform space this first term vanishes by (38). The second term, however, does not vanish on the Ewald sphere shell $k^2 = \omega^2/c^2$ in the spatial Fourier transform space, since its support in k space behaves like the support of the spatial Fourier transform of the imaginary part of the Green's function [see Eq. (32) *et seq.*]. It thus follows that this first term contributes nothing to the solution, which depends only on this second term. For the purpose of the solution (40) and (41), Eq. (16) can thus be redefined as consisting only of this second term; i.e.,

$$\theta \equiv -\oint_s d\mathbf{s}\cdot\nabla\tilde{G}_r\phi. \quad (42)$$

This is a particularly gratifying result since it obviates the need to evaluate (or measure) the gradient of the measured field, which in practice is difficult to accomplish accurately. It should be noted that solution Eq. (40) and (41) cannot be implemented for the D.C. case (i.e., $\omega = 0$), since $\omega = 0$ precludes the required sweeping of the Ewald sphere shell $k^2 = \omega^2/c^2$ over all values of k by varying c and locks the Ewald sphere shell to the null sphere shell $k = 0$. It should further be noted that because of the functional form of the reference potential V_r which goes to ∞ as υ goes to zero (see Eqs. (6) and (21) *et seq.*) and the functional form of the argument $\omega\gamma/\upsilon$ of the Green's function G_r, in practice, $\widetilde{\theta}(\mathbf{k}=\omega,\upsilon)$ cannot be evaluated numerically for $\upsilon=0$, and hence not for $k=\infty$. It thus follows (from the spatial Fourier transform relationships involved) that infinite spatial resolution for the fields and their Laplacian cannot be achieved.

The entire solution of Eqs. (16)–(41) contains the spatial Fourier transforms of the field and the Green's function in such a fashion that the effects of finite sampling density [of ϕ in Eq. (16)] and incomplete observation domain [partial, open, not closed, surface of integration in Eq. (16)] results in solution of Eq. (40) and (41) yielding a degraded resolution of the fields and their Laplacian. The degree of this degraded resolution is determined by the spatial Fourier transform uncertainty principle. Once the field ϕ and its Laplacian $\nabla^2\phi$ has been determined by Eqs. (40) and (41), the unknown potential $V(\mathbf{x},\omega)$ or the unknown medium propagation velocity $c(\mathbf{x},\omega)$ can be evaluated algebraically by the appropriate wave equation and constitutive Eqs. (1)–(4) for the inverse scattering case; and similarly, the unknown source distribution $\rho(\mathbf{x},\omega)$ for the inverse source case.

[1]N. N. Bojarski, "Inverse Scattering" Third Quarterly Company Report, Naval Air systems Command Contract N00019-73-C-0312, Sec. II (1973).
[2]N. Bleistein and N. N. Bojarski, "Recently Developed Formulations of the Inverse Problem in Acoustics and Electromagnetics," Denver Research Institute Report MS-R-7501, #AD/A:003 588 (1974).
[3]N. Bleistein and J. K. Cohen, "Non-Uniqueness in the Inverse Source Problems in Acoustics and Electromagnetics" J. Math. Phys. **18**, 194–201 (1977).

A generalized Weyl correspondence: Applications

M. Gadella[a)] and E. Santos
Departamento de Física Teórica, Universidad de Santander, Santander, Spain

(Received 26 February 1980; accepted for publication 20 February 1981)

A so-called generalized Weyl correspondence is defined among random variables on one side and linear operators in a separable Hilbert space $\mathscr{H}$ on the other. Besides such a correspondence, there is a relation among states on $\mathscr{H}$ (considered as positive nuclear operators on $\mathscr{H}$) and the distribution functions of the random variables. By adding some new assumptions, several relations are shown. Later, we study two particularly interesting cases. In the first we connect dichotomic random variables with number operators in a Grassmann algebra $\mathscr{H}$, and nuclear operators on $\mathscr{H}$ with probability measures in the set of all sequences made up of zero and one. In the second case we relate states between stochastic and quantum electrodynamics.

PACS numbers: 03.65. − w

INTRODUCTION

The aim of the present paper is to present a further contribution to the theory which tries to write quantum mechanics in terms of some kind of probability theory. Some attempts have already been made in that direction. The starting point of such attempts is usually to establish some rule of correspondence between quantum and classical observables, together with a further assumption, which allows the establishment of a correspondence between states. (This assumption is, of course, that any observable has the same average value in any state in both theories.) These rules of correspondence were mainly compiled and studied by Agarwal and Wolf.[1] One of the most natural was the Weyl correspondence.[2] It put in correspondence p_j and q_j, which are respectively momentum and position, in classical mechanics, with $\hat{p}_j$ and $\hat{q}_j$ which are the corresponding variables in quantum mechanics by means of the conditions

$$\langle e^{i(\alpha_1 p_1 + \beta_1 q_1 + \cdots + \alpha_n p_n + \beta_n q_n)}\rangle_{\text{av}} = \text{tr}\rho e^{i(\alpha_1 \hat{p}_1 + \beta_1 \hat{q}_1 + \cdots + \alpha_n \hat{p}_n + \beta_n \hat{q}_n)}, \tag{I.1}$$

where the α_j and β_j's are any real numbers, ρ is the state of the quantum mechanical system, and $\langle -- \rangle_{\text{av}}$ means average on some measure (state) in phase space. This implies immediately that the correspondence between observables must follow the "symmetrizer" rule. For instance, the classical observable $p_j^2 q_j$ under this correspondence is written, in quantum mechanics,

$$\tfrac{1}{3}(\hat{p}_j^2\hat{q}_j + \hat{q}_j\hat{p}_j^2 + \hat{p}_j\hat{q}_j\hat{p}_j). \tag{I.2}$$

One of the nicer features of the Weyl correspondence is that, following Krüger and Poffin,[3] it is uniquely defined by the statements of Galilei invariance, unitarity, reality, and normalization. Krüger and Poffyn also show that the requirement that free particles behave classically and the conditions for obtaining the correct mixed distributions also lead to the same result. We have also linearity for the sum and product by scalars. (That means that the classical variable $\alpha p_j + \beta q_j$ corresponds to $\alpha\hat{p}_j + \beta\hat{q}_j$, etc.) But we have the difficulty that the measure μ, defined by

$$\left\langle \exp\left[i\sum_{j=1}^{n}(\alpha_j p_j + \beta_j q_j)\right]\right\rangle_{\text{av}} = \int \exp\left[i\sum_{j=1}^{n}(\alpha_j p_j + \beta_j q_j)\right] d\mu \tag{I.3}$$

is not positive in general, and so with the given correspondence we cannot write quantum states as probability measures in phase space; consequently, taking into account that this correspondence is the only one with nice physical properties, it follows that we cannot look at the ordinary nonrelativistic quantum mechanics like classical statistical theory, at least as long as the statistics remains as is the case in the present theory. On the other hand, looking for the quantum states which give us a positive definite probability measure is a rather hard task.[4]

In the present paper we extend the above definition to a more general set of random variables and linear operators on a Hilbert space. The idea of this generalization has been considered previously in attempts to solve some stochastic differential equations by transforming them into differential equation for linear operators on a Hilbert space.[5,6] Here we do make the generalization, applying mathematical rigor. We also describe two application.

I. THE GENERALIZED WEYL CORRESPONDENCE

Definition 1: Let J be an arbitrary index set and $(X_j)_{j\in J}$ and $(\hat{x}_j)_{j\in J}$ be, respectively, two families of real random variables and linear operators on a separable Hilbert space $\mathscr{H}$. If it is certain that there exists a nuclear operator ρ (finite trace linear operator) such that

$$E\left(e^{i(\alpha_{j_1}X_{j_1} + \cdots + \alpha_{j_n}X_{j_n})}\right) = \text{tr}\rho e^{i(\alpha_{j_1}\hat{x}_{j_1} + \cdots + \alpha_{j_n}\hat{x}_{j_n})}, \tag{1.1}$$

for any finite set of indices $\{j_1 \cdots j_n\} \subset J$ and any real numbers $\alpha_{j_1} \cdots \alpha_{j_n}$, we will say that the random variables and the linear operators are connected by the generalized Weyl correspondence.

Notice that $\text{tr}\rho e^{(\alpha_{j_1}\hat{x}_{j_1} + \cdots + \alpha_{j_n}\hat{x}_{j_n})}$ always exists because the space of nuclear operators is a two-sided ideal in $L(\mathscr{H})$ (space of linear bounded operators on $\mathscr{H}$). On the other hand, we may consider that the X_j have a nonincreasing distribution function. That means that the measure μ on the sample space $(\Omega, \mathscr{B})$ in which we are defining the X_j's is really a signed measure. But we claim that $\mu(\Omega) = 1$ as well, therefore $\text{tr}\rho = E(I) = 1$, I being the identity in the space of ran-

[a)]Present address: The University of Texas at Austin, Department of Physics, Austin, Texas 78712.

dom variables. However, we will work with the true random variables in which the distribution function increases, unless otherwise specified.

The problem of the existence of solutions of the generalizerd Weyl correspondence has three different aspects. The first is the so-called direct problem, in which we start with a set of random variables $(X_j)_{j\epsilon J}$, and our goal is to find a separable Hilbert space $\mathscr{H}$, a family $(\hat{x}_j)_{j\epsilon J}$ of linear operators on $\mathscr{H}$, and a nuclear operator ρ such that the relationship (1.1) is fulfilled. Connected with this there is the problem of representation of random variable algebras (commutative) on operator algebras (generally noncommutative). A particular instance may be found in Ref. 6.

The so-called inverse problem works reciprocally. Given the separable Hilbert space $\mathscr{H}$, the family of linear operators $(\hat{x}_j)_{j\epsilon J}$, and the positive nuclear operator ρ, we propose to find the family of random variables $(X_j)_{j\epsilon J}$, which fulfills (1.1), assuming its existence. (Here solutions may be found with a nonincreasing distribution function.) An interesting example of this will be commented on in the second part of the present paper.

Finally, we may start with $\mathscr{H}$, $(\hat{x}_j)_{j\epsilon J}$, and the family of variables $(X_j)_{j\epsilon J}$, considered only as measurable functions from Ω to R. We have then to find the set of probability measures [or signed measures with $\mu(\Omega)=1$] on (Ω,β) and the set of positive nuclear operators ρ on $\mathscr{H}$, which fulfil (1.1). A typical example of this is the representation of quantum mechanical states on phase space. (Note that the Wigner function, like the joint density function of X_j's, may not be positively defined.) The study of this example is given in Refs. 7,8.

Our next goal is to find necessary conditions for the existence of solutions of relation (1.1). We shall assert that $\mathrm{tr}\rho = 1$. However our definition is too general, so we have to introduce new assumptions in order to reach interesting conclusions. The first two results are given here.

Theorem 1: Let us asume that there is solution to (1.1) with the following properties:

(a) $\hat{x}_j$ is a self-adjoint linear operator on $\mathscr{H}$ (not necessarily bounded);

(b) ρ is a positive nuclear operator.

Then,

(a) If $G_j(x)$ is the resolution of the identity associated with $\hat{x}_j$ and $F_j(x)$ the distribution function of the random variable X_j, we shall show that

$$\mathrm{tr}\rho G_j(x) = F_j(x). \tag{1.2}$$

(b) If we call $\sigma(X_j)$ the set of strictly monotonic points of the distribution function $F_j(x)$, and $\sigma(\hat{x}_j)$ is the spectrum of $\hat{x}_j$, then

$$\sigma(X_j) \subset \sigma(\hat{x}_j). \tag{1.3}$$

(c) If $\hat{x}_j$ and X_j are bounded

$$\mathrm{tr}\rho\hat{x}_j^n = E(X_j^n) \quad \forall n \in N. \tag{1.4}$$

Proof:

(a) Consider $\alpha_i = 0$ if $i \neq j$;

$$\mathrm{tr}\rho e^{i\alpha_j\hat{x}_j} = E(e^{i\alpha_j X_j}). \tag{1.5}$$

Since $e^{i\alpha_j\hat{x}_j}$ is a unitary operator depending on the real number α_j, we can write, according to Stone's theorem.

$$e^{i\alpha_j\hat{x}_j} = \int_{-\infty}^{\infty} e^{i\alpha_j x}\, dG_j(x). \tag{1.6}$$

This integral is defined in the uniform sense, ρ is a continuous operator, and the trace is a linear and continuous functional. (The continuity holds with the usual operator norm.) Therefore, we can write

$$\begin{aligned}\mathrm{tr}\rho e^{i\alpha_j\hat{x}_j} &= \mathrm{tr}\rho \sum_{s=-\infty}^{s=\infty} \int_{s-1}^{s} e^{i\alpha_j x} dG_j(x) \\ &= \mathrm{tr}\rho \sum_{s=-\infty}^{s=\infty} \lim_{k\to\infty} \sum_{n=1}^{k} e^{i\alpha_j x'_n}[G(x_n) - G(x_{n-1})], \\ &\qquad \sup|x_n - x_{n-1}| \to 0, \end{aligned} \tag{1.7}$$

with $s-1 = x_0 \leqslant x_1 \leqslant \cdots \leqslant x_k = s;\ x'_n \epsilon (x_{n-1}, x_n)$. Then

$$\begin{aligned}\sum_{s=-\infty}^{s=\infty} \lim_{k\to\infty} \sum_{n=1}^{k} e^{i\alpha_j x'_n}[\mathrm{tr}\rho G(x_n) - \mathrm{tr}\rho G(x_{n-1})] \\ = \mathrm{tr}\rho e^{i\alpha_j\hat{x}_j}, \quad \sup|x_n - x_{n-1}| \to 0,\end{aligned} \tag{1.8}$$

and since $G(x) \leqslant G(x')$ when $x \leqslant x'$ and ρ is positive, then $\mathrm{tr}E(x)$ is a monotonic function. Hence the following Stieltjes integral makes sense.

$$\int_{-\infty}^{\infty} e^{i\alpha_j x}\, d\,[\mathrm{tr}\rho G_j(x)], \tag{1.9}$$

which is just the preceding limit. By using (1.5)

$$\int_{-\infty}^{\infty} e^{i\alpha_j x}\, dF_j(x) = \int_{-\infty}^{\infty} e^{i\alpha_j x}\, d\,[\mathrm{tr}\rho G_j(x)]. \tag{1.10}$$

Hence,

$$F_j(x) = \mathrm{tr}\rho G_j(x). \tag{1.11}$$

(b) This follows from (a). Let y be a discontinuous point in $F_j(x)$; then y will be a discontinuous point in $G_j(x)$ as well, because of the continuity of $\mathrm{tr}\rho(-)$. Hence y belongs to the point spectrum of $\hat{x}_j$. On the other hand, assume y is a continuous point in $F_j(x)$, but there is not any neighborhood of y in which $F_i(x)$ is constant. This also happens with $G_j(x)$ for the same reason as before, so y will belong to the continuous spectrum of $\hat{x}_j$. Hence $\sigma(X_j) \subset \sigma(\hat{x}_j)$.

(c) Assume X_j and $\hat{x}_j$. are bounded; then

$$E(e^{i\alpha_j X_j}) = E\left[\sum_{n=0}^{\infty} \frac{(i\alpha_j X_j)^n}{n!}\right] = \sum_{n=0}^{\infty} \frac{(i\alpha_j)^n}{n!} E(X_j^n), \tag{1.12}$$

by the continuity on the mean, and

$$\mathrm{tr}\rho e^{i\alpha_j\hat{x}_j} = \sum_{n=0}^{\infty} \frac{(i\alpha_j)^n}{n!} \mathrm{tr}\rho\hat{x}_j^n, \tag{1.13}$$

and finally by (1.5) and due to the fact that α_j is any real number, we get

$$E(X_j^n) = \mathrm{tr}\rho\hat{x}_j^n. \tag{1.14}$$

Theorem 2: Let us assume the existence of a solution of (1.1) with the following properties:

(a) the index set J is a finite one,

(b) all the operators x_j $(j \in J)$ are self-adjoint,

(c) ρ is a positive trace class operator

(d) the variables X_j as well as the operators $\hat{x}_j$ are bounded $(\forall j \in J)$,

(e) The range of X_j's is infinite; and

(f) The X_j's are linearly independent and so are their powers—we may obtain them by choosing the sample space Ω as the cartesian product of ranges of X_j's.

Let A be the complex algebra spanned by the variables X_j, their powers, and products. Let B be the algebra spanned by the $\hat{x}_j$ under the same conditions. A and B have the identites and they are not necessarily complete.

Then

(a) There is a linear mapping g between A and B considered as linear spaces, such that

$$g(X_j) = \hat{x}_j, \quad \forall j \in J. \tag{1.15}$$

(b) (1.1) is equivalent to

$$\mathrm{tr}\rho g(y) = E(y), \quad \forall y \in A. \tag{1.16}$$

(c) Let $y = X_1^{k_1} \cdots X_N^{k_N}$, then

$$g(y) = (\hat{x}_1^{k_1} \cdots \hat{x}_N^{k_N})_S, \tag{1.17}$$

where $(\ )_S$ means the symmetrized product of the operators contained in the bracket, that is to say, the sum of the products, carried out in all possible ways, of the $k = k_1 + \cdots + k_N$ operators divided by the number of such products, i.e.,

$$\frac{k!}{k_1! \cdots k_N!}. \tag{1.18}$$

(d) Let (a_{ij}) be a nonsingular real matrix, and let us assume that there is a solution to (1.1) with X_j and $\hat{x}_j$; $j = 1, \ldots, N = \text{card } J$, then there is another solution with the same ρ and with

$$Y_i = \sum_{j=1}^{N} a_{ij} X_j; \; \hat{y}_i = \sum_{j=1}^{N} a_{ij} \hat{x}_j, \tag{1.19}$$

instead of X_j and $\hat{x}_j$, $j = 1, \ldots, N$.

Proof: First of all, we give without proof the following lemma (see Appendix A).

Lemma: Any $y \in A$ may be written as a linear combination of elements on the form $\left(\sum_{k=1}^{N} \lambda_k X_k \right)^l$, where l is a natural number and λ_k are complex numbers.

Next, we prove the theorem.

(a) We shall prove that g is a linear mapping. Linearity is proved in (b). Here we show that it is really a mapping. In order to do that, we have to show that when two different representations of $y \in A$ as linear combinations of vectors, as in the lemma, are given, we obtain only one operator $\hat{y}$ such that $\hat{y} = g(y)$. g is linear in its arguments, so it will be enough to prove our assertion on vectors in A which are of the form $X_1^{k_1} \cdots X_N^{k_N}$, because any other vector must be a linear combination of those vector.

Let

$$\begin{aligned} u = X_1^{k_1} \cdots X_n^{k_N} &= \sum_{i=1}^{n} \lambda_i (a_{i1} X_1 + \cdots + a_{iN} X_N)^{k_k + \cdots + k_N = k} \\ &= \sum_{i=1}^{n} \lambda_i \Big(a_{i1}^k x_1^k + \cdots + \frac{k!}{k_1! \cdots k_N!} a_{i1}^{k_1} \cdots a_{iN}^{k_N} X_1^{k_1} \cdots X_N^{k_N} \\ &\quad + \cdots + a_{iN}^k X_N^k \Big). \end{aligned} \tag{1.20}$$

Due to statments (e) and (f) of Theorem 2, in order for this identity to be true, if $(k_1', \ldots, k_N')$ is a multi-index different from $(k_1, \ldots, k_N)$, but both with modules k, we must have

$$\sum_{i=1}^{n} \lambda_i \frac{k!}{k'_1! \cdots k'_N!} a_{i1}^{k'_1} \cdots a_{iN}^{k'_N} = 0 \Rightarrow \sum_{i=1}^{n} \lambda_i a_{i1}^{k'_1} \cdots a_{iN}^{k'_N} = 0, \tag{1.21}$$

and also

$$\sum_{i=1}^{n} \lambda_i \frac{k!}{k_1! \cdots k_N!} a_{i1}^{k_1} \cdots a_{iN}^{k_N} = 1 \Rightarrow \sum_{i=1}^{n} \lambda_i a_{i1}^{k_1} \cdots a_{iN}^{k_N} = \frac{k_1! \cdots k_N!}{k!}. \tag{1.22}$$

On the other hand,

$$\begin{aligned} g(y) = \hat{y} &= \sum_{i=1}^{n} \lambda_i (a_{i1} \hat{x}_1 + \cdots + a_{iN} \hat{x}_N)^k \\ &= \sum_{i=1}^{n} \lambda_i \big[a_{i1}^k \hat{x}_1^k + \cdots + a_{i1}^{k_1} \cdots a_{iN}^{k_N} (\hat{x}_1^{k_1} \cdots \hat{x}_N^{k_N})_p \\ &\quad + \cdots + a_{iN}^k \hat{x}_N^k \big], \end{aligned} \tag{1.23}$$

where $(-)_p$ means that inside the bracket we have to write the sum of all products obtained by permuting in all possible ways the operators in the monomial $\hat{x}_1^{k_1} \cdots \hat{x}_N^{k_N}$. Then by using the above equations on the coefficients we obtain

$$g(y) = \hat{y} = \frac{k_1! \cdots k_N!}{k!} (\hat{x}_1^{k_1} \cdots \hat{x}_N^{k_N})_p = (\hat{x}_1^{k_1} \cdots \hat{x}_N^{k_N})_S, \tag{1.24}$$

but this result is independent of the spanning of y on vectors of the form as in the lemma.

(b) Let $y \in A$ be as given in the lemma,

$$y = \sum_{\lambda=1}^{M} c_\lambda y_\lambda,$$

where

$$y_\lambda = \left(\sum_{i=1}^{N} a_{i\lambda} X_i \right)^{k_\lambda}. \tag{1.25}$$

We define g as

$$g(y) = \sum_{\lambda=1}^{M} c_\lambda g(y_\lambda), \tag{1.26}$$

$$g(y_\lambda) = \left(\sum_{i=1}^{N} a_{i\lambda} \hat{x}_i \right)^{k_\lambda}. \tag{1.27}$$

With this definition the linearity of g is guaranteed. On the other hand, it is easy to show that $O(1.1)$ involves

$$\mathrm{tr}\rho \left(\sum_{i=1}^{N} a_{i\lambda} \hat{x}_i \right)^{k_\lambda} = E \left(\sum_{i=1}^{N} a_{i\lambda} X_i \right)^{k_\lambda}, \tag{1.28}$$

or equivalently, $\mathrm{tr}\rho g(y_\lambda) = E(y_\lambda)$. By linearity $\mathrm{tr}\rho g(y) = E(y)$ follows. Writing $y_\lambda = X_j$, we have $g(X_j) = \hat{x}_j$.

The converse is straightforward.

(c) We have just shown it.

(d) We must prove that

$$E\left(e^{i(\alpha_1 Y_1 + \cdots + \alpha_N Y_N)}\right) = \mathrm{tr}\rho e^{i(\alpha_1 \hat{y}_1 + \cdots + \alpha_n \hat{y}_n)}, \tag{1.29}$$

which is straightforward.

Corollary: When, under the conditions of Theorem 2, there is a linear mapping g from A to B, considered as linear spaces, such that

$$E(y) = \mathrm{tr}\rho g(y) \quad \forall y \in A,$$

with

$$g(X_k^{k_1} \cdots X_N^{k_N}) = (\hat{x}_1^{k_1} \cdots \hat{x}_N^{k_N})_s; \; J = \{1, 2, \ldots, N\}, \tag{1.30}$$

then,

$$E\left(e^{i(\alpha_1 X_1+\cdots+\alpha_N X_N)}\right)=\operatorname{tr}\rho e^{i(\alpha_1 \hat{x}_1+\cdots+\alpha_N \hat{x}_N)}. \tag{1.31}$$

We do not include the proof, which is straightforward.

II. APPLICATIONS

Application I: Representations of quantum mechanical states as probability measures in a sample space by using countable dichotomic random variables

In the following we are going to show the possibility of representing states in nonrelativistic quantum mechanics (usually considered as nuclear positive operators with trace one) as probability measures in a sample space.

Let us consider the sample space $\Omega=\Pi_{i\in N}\Omega_i$ where $\Omega_i=\{0,1\}$. Here Ω is the set of all sequences $i\in N$ of zeros and ones. Let us consider the complex completed algebra $\tilde{A}$ generated by the identity and the projections $\chi_i:\Omega\rightarrow\Omega_i$, with a supremum norm. As a consequence of the Stone-Weisstrass theorem, $\tilde{A}$ is the set of all continuous functions over Ω. In addition, we have to construct an operator algebra on a Hilbert space, *-congruent to the function algebra $\tilde{A}$. We start by considering an infinite-dimensional separable Hilbert space $\mathscr{Y}$, along with the Grassmann algebra $\mathscr{H}=\oplus_{p=0}^{\infty}\tilde{a}(\mathscr{Y}^{\oplus})^p$, where $\tilde{a}$ denotes the antisymmetric operator in $(\mathscr{Y}^{\oplus})^p$. If $(e_i)_{i\in N}$ is an orthonormal basis in $\mathscr{Y}$, an orthonormal basis in $\mathscr{H}$ will be made up of elements of the form

$$\chi_{i_1,\ldots,i_n}=\sqrt{n!}\,(e_{i_1}\wedge\cdots\wedge e_{i_n}), \tag{2.1}$$

where $i_1<i_2<\cdots<i_n$. Since the number of $\chi_{i_1\cdots i_n}$ is countable, $\mathscr{H}$ is infinite-dimensional and separable. Now we define $b^+(e_i)$ and $b^-(e_i)$ to be the usual creation and annihilation operators. Such operators can be extended to the whole space $\mathscr{H}$ by linearity and continuity, and they are indeed a representation of canonical anticommutation relations.

Let us now consider the set of bounded operators $N_j=b^+(e_j)b^-(e_j)$ associated with each e_j in the basis chosen in $\mathscr{Y}$. Their well-known properties are $N_j^+=N_j$, $N_j^2=N_j$, and $N_jN_k=0$ $(j\neq k)$. Moreover $I, N_1, N_2,\ldots,N_p, N_1N_2, N_2N_3,\ldots,N_1\ldots N_p,\ldots$, are linearly independent. If we consider the closure (with the usual norm in $L(\mathscr{H})$ of the linear space spanned by this basis, we reach a complete algebra, which we shall call $\tilde{B}$. We shall call A and B, respectively, the algebras generated linearly by projections and operators, having identities but lacking completeness.

In the following we are going to give several lemmas. We shall omit their proofs, which are in all the cases straightforward.

Lemma I: $\tilde{A}$ and $\tilde{B}$ are *-congruent, i.e., they are *-isomorphic and this isomorphism in bicontinuous.

Lemma II: The above case admits a solution of generalized Weyl correspondence if and only if

$$E(X)=\operatorname{tr}\rho g(X),\quad \forall X\in\tilde{A}. \tag{2.2}$$

Let us consider now a state ρ on $\mathscr{H}$. Such an object is a linear positive operator with trace one (density operator). Our aim is to show that one can find a unique probability measure on Ω (endowed with the usual Borel σ-algebra) in correspondence with ρ. The more general form of a density operator on $\mathscr{H}$ is,

$$\rho=\sum_{i=1}^{\infty}\lambda_i|\psi_i\rangle\langle\psi_i|,\ \text{with}\ \sum_{i=1}^{\infty}\lambda_i=1, \tag{2.3}$$

where $|\psi_i\rangle\in\mathscr{H}$ is the eigenvector of ρ with eigenvalue $\lambda_i\geqslant 0$.

Let us consider the open set in Ω of the form $D=D_{i_1\cdots i_p}=(i_1,\ldots,i_p)\times\Pi_{i=p+1}^{\infty}\Omega_i$; $p\in N$. The collection of such sets makes up a countable basis in the topology on Ω. That D is also closed implies the topology in Ω is nonconnected. Consider the characteristic function δ_D of D. $[\delta_D(\omega)=1$ if $\omega\in D$; $\delta_D(\omega)=0$ if $\omega\notin D$.$]$ Therefore, δ_D is a continuous function. On the other hand, because D is dependent on a finite number of coordinates, one may write δ_D as a linear combination of some elements in A, i.e, $I, X_1, X_2,\ldots,X_p, X_1X_2,\ldots,X_1X_2,\ldots,X_1X_2\ldots X_p$.

Hence $\delta_D\in A$ and $\hat{\delta}_D=g(\delta_D)\in B$ exists and is the same linear combination on $I, N_1, N_2,\ldots, N_1N_2\ldots N_p$ as δ_D in the linear basis in A.

Lemma III: The eigenvectors of $\hat{\delta}_A$ are those making up the basis $\chi_{k_1\cdots k_n}=\sqrt{n!}\,e_{k_1}\wedge\cdots\wedge e_{k_n}$. $(k_1<k_2<\cdots<k_n$, so such numbers are indeed ordering numbers). If we call $\{j_1,\ldots,j_l\}$ $(l\leqslant p)$ the ordering numbers of the places occupied by the nonzero coordinates in $(i_1,\ldots,i_p)$, we have

$$\begin{aligned}&\hat{\delta}_D\chi_{k_1\cdots k_n}=\chi_{k_1\cdots k_n},\ \text{if}\ \{j_1,\ldots,j_l\}\subset\{k_1,\ldots,k_n\},\\&\hat{\delta}_D\chi_{k_1\cdots k_n}=0,\ \text{otherwise.}\end{aligned} \tag{2.4}$$

We are now ready to construct a probability measure μ on Ω, corresponding to ρ through the use of the generalized Weyl correspondence, which by Lemma II and Riesz's theorem exists and is unique. To this end we must show

$$E(\delta_D)=\operatorname{tr}\rho\hat{\delta}_D, \tag{2.5}$$

whence it follows that

$$\begin{aligned}E(\delta_D)&=\int_D d\mu=\mu(D)=\mu\left[(i_1,\ldots,i_p)\times\Pi_{i=p+1}^{\infty}\Omega_i\right]\\&=\mu\{X_1=i_1;X_2=i_2,\ldots,X_P=i_P\},\end{aligned} \tag{2.6}$$

giving us the probability of all measurable cylinders, hence the probability of all measurable sets.

Let us calculate $\operatorname{tr}\rho\delta_D$.

$$\begin{aligned}\operatorname{tr}\rho\hat{\delta}_D&=\sum_{j_1<\cdots<j_k k\in N\cup\{0\}}\langle\chi_{j_1\cdots j_k}|\rho\hat{\delta}_D|\chi_{j_1\cdots j_k}\rangle\\&=\sum_{j_1<\cdots<j_k k\in N\cup\{0\}}\left\langle\chi_{j_1\cdots j_k}\Big|\sum_{i=1}^{\infty}\lambda_i|\psi_i\rangle\right\rangle\langle\psi_i|\hat{\delta}_D|\chi_{j_1\cdots j_k}\rangle,\end{aligned} \tag{2.7}$$

with

$$|\psi_i\rangle=\sum_{i_1<\cdots<i_n}\lambda_{i,i_1,\ldots,i_n}\chi_{i_1,\ldots,i_n}. \tag{2.8}$$

Letting $D=D_{i_1\cdots i_p}$, by the foregoing lemma

$$\operatorname{tr}\rho\hat{\delta}_D=\sum_{i=1}^{\infty}\sum_{\{k_1,\ldots,k_n\}\subset J}\lambda_i|\lambda_{i,k_1,\ldots,k_n}|^2=\mu(D). \tag{2.9}$$

The second sum is extended to any set of ordered numbers containing the set $J=\{j_1,j_2,\ldots,j_l\}$ of the orders in which are placing the ones in $(i_1,\ldots,i_p)$ [for instance if $(i_1,\ldots,i_p)=(1,0,0,1,0,1)$, $J=\{1,4,6\}$.]

We write now the sequence of sets in Ω.

$$\begin{aligned}&D_0=D,\\&D_1=(i_1,\ldots,i_p,0)\times\Pi_{i=p+2}^{\infty}\Omega_i,\ldots,\\&D_n=(i_1,\ldots,i_p,0,\ldots,0)\times\Pi_{i=p+n+1}^{\infty}\Omega_i,\end{aligned} \tag{2.10}$$

$$\mu(D_n) = \sum_{i=1}^{\infty} \lambda_i |\lambda_{i,j_1,\dots,j_e}|^2 + \sum_{i=1}^{\infty} \lambda_i \sum_{\substack{j_k < \cdots < j_n \\ i_{p+n+1} < j_k}} |\lambda_{ij_1,\dots,j_e j_k,\dots,j_h}|^2. \quad (2.11)$$

We note that $D_0 \supset D_1 \supset \cdots \supset D_n \supset \cdots$. Then

$$\mu\Big[\bigcap_{n=0}^{\infty} D_n\Big] = \mu(\lim_{n\to\infty} D_n) = \lim_{n\to\infty} \mu(D_n).$$

Let us calculate such a limit. The expression which gives us $\mu\ (D_n)$ is the sum of two different series. The former is common to any $\mu\,(D_n)$. The second one is a double series which drops out as n goes to infinity. Hence

$$\mu\Big[\bigcap_{n=0}^{\infty} D_n\Big] = \sum_{i=1}^{\infty} \lambda_i |\lambda_{ij_1,\dots,j_l}|^2. \quad (2.12)$$

On the other hand, $\cap_{n=1}^{\infty} D_n$ has only one point, namely $(i_1,\dots,i_p,\ 0,0,\dots,0,\dots)$. Furthermore, by adding all contributions to the measure of all points of this kind we ge

$$\sum_{i=1}^{\infty} \sum_{j_1 < \cdots < j_l} \lambda_i |\lambda_{ij_1,\dots,j_l}|^2 = \sum_{i=1}^{\infty} \lambda_i \|\psi_i\|^2 = \sum_{i=1}^{\infty} \lambda_i = \mathrm{tr}\rho = 1. \quad (2.13)$$

This implies that the measure is concentrated on points which contain a finite number of ones and is zero on the others. We have thus obtained a necessary condition for a state on A to correspond to a state on $\mathscr{H}$. Any other state on A which gives us (according to Riesz's theorem) a measure which does not satisfy this condition, cannot correspond with a state on $\mathscr{H}$.

We now prove the converse. Let μ be a probability measure on Ω, which is of the type described above. Then we can choose a state on $\mathscr{H}$ (but not uniquely) which corresponds to μ according to Weyl generalized correspondence.

Lemma IV: Let the set of the form $D = D_{i\cdots i_p} = (i_1,\dots,i_p) \times \Pi_{i=p+1}^{\infty} \Omega_i$. The set of characteristic functions δ_D of these sets span the algebra by linear combinations and products.

A straightforward consequence of this lemma is that proving $\mathrm{tr}(\rho\hat{\delta}_D) = E\,(\delta_D)$ is equivalent to proving the existence of a solution of the generalized Weyl correspondence (by Lemma II).

Now let us write

$$\mu\,(i_1,\dots,i_n,0,\dots,0\dots) = \lambda_{j_1,\dots j_l}, \quad (2.14)$$

where $j_1,\dots,j_l$ are the orders of the i_p which are different from zero, and let

$$\rho = \sum_{j_1 < \cdots < j} \lambda_{j_1,\dots,j_l} |\chi_{j_1,\dots,j_l}\rangle\ \langle\chi_{j_1,\dots,j_l}|. \quad (2.15)$$

A straightforward caculation of $\mathrm{tr}\rho\hat{\delta}_D$ gives immediately $\mu(D)$ and hence our result, which we summarize in the next theorem.

Theorem: To a state E on A, i.e., a positive definite continuous functional, there corresponds a state ρ on $\mathscr{H}$ according to the generalized Weyl correspondence if and only if the measure μ on Ω associated with the state E (by Riesz's theorem) is concentrated in the points in of the form $(i_1,\dots,i_n,\ 0,\dots,0,\dots)$.

Remarks:

(1) This correspondence between states is not one-to-one. To a state E on A there may correspond more than one ρ on $\mathscr{H}$. Let us take $\rho = \Sigma_{i=1}^{\infty} \lambda_i |\psi_i\rangle |\psi_i|, |\psi_i$ being an orthonormal set in $\mathscr{H}$ with the condition that ρ does not commute with all elementary projections associated with the vectors $\chi_{j_1,\dots,j_n}$. In correspondence with ρ we will have a probability measure μ which will be concentrated on the points of the form $(i_1,\dots,i_p,0,\dots,0)$. By means of the above arguments we can associate with μ another density operator $\rho'(\rho \neq \rho'$ because ρ does not commute with all projections in which ρ' is spanned).

(2) The measure μ on Ω which corresponds to a density operator ρ on $\mathscr{H}$ is not independent of the choice of the basis in $\mathscr{V}$. The reason for this is simple; while $\mathrm{tr}\rho\hat{\delta}_D$ is independent of any change of basis, the correspondence $g\ (\delta_D) = \hat{\delta}_D$ is not.

(3) We can extend this correspondence in the theorem to all density matrices defined in any separable Hilbert space. Remember that two separable Hilbert spaces with the same dimensions are congruent. The case in which we have a finite-dimensional space is described in Ref. 9 and 10.

(4) If we have a pure state for $\hat{A}$, then the corresponding measure is concentrated in only one point. Then we can put it in correspondence with a pure state on $\mathscr{H}$. The converse is not always true.

(5) $\tilde{B}$ is not a von Neumann algebra, except if $\mathscr{V}$ is finite-dimensional

(6) The possibility of translating physical states into a new representation does not mean the possibility of translating physics at all. First, because we have the same "translation" for several states; and second, because we have no "translation" for most meaningful observables like position, momentum, and Hamiltonians.

Nevertheless, there are some particular cases in which we can find physical applications for this construction. One of these is described in Ref. 10.

Application II: Comparison between the states of the free radiation field given by stochastic and quantum electrodynamics with the Coulomb gauge

(a) Stochastic free radiation field

As is well-known, the free radiation field is characterized by a vector potential, which in the Coulomb gauge and under periodic boundary conditions is shown to be, at any space–time point,

$$\vec{A}(\vec{r},t) = \frac{1}{\sqrt{L^3}} \sum_j \sum_\alpha \big[c_{j\alpha}\vec{\epsilon}^{\alpha} e^{i(\vec{k}_j\vec{r} - \omega_j t)} + c_{j\alpha}^{*}\vec{\epsilon}^{\alpha} e^{-i(\vec{k}_j\vec{r} - \omega_j t)}\big], \quad (2.16)$$

where L is the length of the side of a cubic box in which we have defined the boundary condition, $\vec{\epsilon}_\alpha$ is the polarization vector, and $\vec{k}_j$ and ω_j are, respectively, the jth propagation vector and frequency of the jth plane wave, which form the wave packet. Finally $\alpha = 1,2$, and j extends to all natural numbers because

$$k_{xj}, k_{yj}, k_{zj} = 2\pi n/L. \quad (2.17)$$

The quantities $c_{j\alpha}$ are mathematical objects to be defined. In the stochastical theory $c_{j\alpha}$ are independent Gaussian random variables, so $\vec{A}\,(r,t)$ is a stochastic field. A study of such fields is given, for instance, in Ref. 11. If we define $Q_{j\alpha}$ and $P_{j\alpha}$ as real random variables such that

$$c_{j\alpha} = \frac{c}{2}\left[Q_{j\alpha} + \frac{i}{\omega_j}P_{j\alpha}\right], \tag{2.18}$$

c being the speed of light in vacuum, we have the vector potential described in terms of functions of real random variables. The states of the stochastic radiation field will be the joint distribution functions fo the variables $Q_{j\alpha}$ and $P_{j\alpha}$. If we further assume that $Q_{j\alpha}$ and $P_{j\alpha}$ are independent, we find that they are Gaussian also. Nevertheless we can assume from the beginning that the $Q_{j\alpha}$ and $P_{j\alpha}$ are jointly Gaussian.

The states of the stochastic radiation field can be also described by probability measures on a sample space in the following way. Consider $\Omega = \Pi_{i=1}^{\infty}\Omega_{i,\alpha}$ where $\Omega_{i\alpha}$ are identical copies of the real plane. Since $\omega \in \Omega_{i\alpha} = (q_{i\alpha}, p_{i\alpha})$, $Q_{i\alpha}$ and $P_{i\alpha}$ are the projections $Q_{i\alpha}(q_{i\alpha}, p_{i\alpha}) = q_{i\alpha}$ and $P_{i\alpha}(q_{i\alpha}, p_{i\alpha}) = p_{i\alpha}$. If Q_i and P_i are jointly Gaussian, we can define a probability measure $\mu_{i\alpha}$ on the Borel sets $B_{i\alpha}$ of $\Omega_{i\alpha}$ by

$$\mu_{i\alpha}[B_{i\alpha}] = \frac{1}{\gamma}\int_{B_{i\alpha}} \rho(Q_{i\alpha}, P_{i\alpha})\, dQ_{i\alpha}\, dP_{i\alpha}, \tag{2.19}$$

where

$$\rho(Q_{i\alpha}, P_{i\alpha}) = e^{-a(Q_{i\alpha}^2 + P_{i\alpha}^2 + \beta Q_{i\alpha}P_{i\alpha} + \lambda Q_{i\alpha} + \tau P_{i\alpha} + \delta)}; \quad a>0, \tag{2.20}$$

and

$$\gamma = \int_{\Omega} \rho(Q_{i\alpha}, P_{i\alpha})\, dQ_{i\alpha}\, dP_{i\alpha}. \tag{2.21}$$

By means of the hypothesis that $c_{i\alpha}$ are independent we conclude that if $B = B_{i_1\alpha_1} \times \cdots \times B_{i_n\alpha_n} \times \Pi_{n+1}^{\infty}\Omega_i$ is a measurable rectangle in Ω,

$$\mu(B) = \mu_{i_1\alpha_1}(B_{i_1\alpha_1}) \times \cdots \times \mu_{i_n\alpha_n}(B_{i_n\alpha_n}), \tag{2.22}$$

and therfore the measure μ can be extended in a unique way to all Borel sets in Ω.

Finally note that the Hamiltonian becomes

$$H = \tfrac{1}{2}\sum_j \sum_\alpha [P_{j\alpha}^2 + \omega_j Q_{j\alpha}^2], \tag{2.23}$$

so it is the sum of an infinite set of independent random variables, any of which represents a stochastic harmonic oscillator. Remember that in the classical theory the free radiation field, with the Coulomb gauge, can be regarded as a countably infinite sequence of noncoupled harmonic oscillators.

(*b*) *Formulation of the states in quantum mechanics*

We are looking for a formulation of the quantum free field which enables us to compare it with the stochastic one. The basic idea is that written above: consider the free radiation field as a sequence of noncoupled harmonic oscillators.

Let A be the subset of the product $\Pi_{i=1}^{\infty}E_i$ (we should call them $E_{i\alpha}$ but we do not do that in order to avoid unnecessary complications in the notation) with $E_i = \mathscr{L}_2(R)$, built up out of elements of the form $(v_{i_1},\dots,v_{i_n},\phi_0,\dots,\phi_0\cdots)$ where $v_{i_k} \in \mathscr{L}_2(R)$ and

$$\phi_0 = \left[\frac{\omega_i}{\pi\hbar}\right]^{1/4} e^{-\omega_i x^2/2\hbar}, \tag{2.24}$$

is the ground state of the harmonic oscillator. Then a sequence of square-summable functions belongs to A if and only if all its terms, except possibly a finite number of them are equal to ϕ_0.

We now construct the tensorial product in the usual way: build up the $\mathbb{C}$-free module T over A. T is the linear space spanned by the points in A regarded as linearly independent vectors. Let us pick in T elements of the form

$$\begin{aligned}&(v_{i_1},\dots,v_{i_p} + v_{i_p},\dots,v_{i_k},\phi_0,\dots,\phi_0,\dots)\\ &\quad + (-1)(v_i,\dots,v_{i_p},\dots,v_{i_k},\phi_0,\dots)\\ &\quad + (-1)(v_{i_1},\dots,v_{i_p},\dots,v_{i_k},\phi_0,\dots),\end{aligned} \tag{2.25}$$

and

$$(v_{i_1},\dots,v_{i_p},\dots,v_{i_k},\phi_0,\dots,\phi_0,\dots) + (v_{i_1},\dots,v_{i_p},\dots,v_{i_k},\phi_0,\dots). \tag{2.26}$$

These elements span a subspace in T which we will call T_1.

The factor space T/T_1 is the infinite tensorial product defined by E_i and ϕ_0, and we can prove easily that these elements behave like those in a usual tensorial product. The element $(v_1,\dots,v_P,\phi_0,\dots,\phi_0,\dots) + T_1 \in T/T_1$ will be called a tensorial product of $v_1,\dots,v_P,\phi_0,\dots,\phi_0,\dots$, and we will write

$$v_1 \otimes \cdots \otimes v_P \otimes \phi_0 \otimes \cdots \otimes \phi_0 \otimes \cdots. \tag{2.27}$$

The inner product is defined on T/T_1 as usual. If

$$\begin{aligned}x &= x_1 \otimes x_2 \otimes \cdots \otimes x_P \otimes \phi_0 \otimes \cdots \otimes \phi_0 \otimes \cdots, \quad \text{(for instance } p \geqslant q)\\ y &= y_1 \otimes y_2 \otimes \cdots \otimes y_P \otimes \phi_0 \otimes \cdots \otimes \phi_0 \otimes \cdots,\end{aligned} \tag{2.28}$$

then

$$\langle x|y\rangle = \langle x_1|y_1\rangle \cdots \langle x_q|x_q\rangle \cdots \langle x_p|\phi_0\rangle\langle\phi_0|\phi_0\rangle\cdots\langle\phi_0|\phi_0\rangle\cdots, \tag{2.29}$$

which is well defined and it is in fact an inner product in T/T_1. The complexation of this space will be a Hilbert space, which we will call $\mathscr{H}$. Furthermore it can be shown that $\mathscr{H}$ is separable.

Let us now define observables on $\mathscr{H}$, and in particular, the positions and momenta.

By definition

$$\hat{Q}_j \cdot v = \hat{Q}_j(v_1 \otimes \cdots \otimes v_j \otimes \cdots \otimes v_n \otimes \phi_0 \cdots)$$

$= v_1 \otimes \cdots \otimes xv_j \otimes \cdots \otimes v_n \otimes \phi_0 \cdots$ and we extend the definition to the linear combinations of vectors in this form [of course $xv_j(x)$ must be square-summable].

On the other hand, the jth momentum operator will be

$$\hat{P}_j v = v_1 \otimes \cdots \otimes (-i\hbar)\frac{d}{dx}v_p \otimes \cdots \otimes v_n \otimes \phi_0 \cdots, \tag{2.30}$$

and extensions by linearity [$v_P(x)$ must be an absolutely continuous function with square-summable first derivative].

Since $\hat{Q}_j$ and $\hat{P}_j$ are essentially self-adjoint operators, they can be extended to a bigger domain in which they are self-adjoint. Furthermore, in the common domain $\mathscr{D}Q_j \cap \mathscr{D}P_j$ (which is dense in $\mathscr{H}$) we have

$$[\hat{Q}_j, \hat{P}_j] < i\hbar I. \tag{2.31}$$

We also may define the ladder operators in the usual way. They are a representation of the canonical commutation relations over $\mathscr{H}$, with vacuum state given by $\phi_0 = \phi_0 \otimes \phi_0 \otimes \cdots \otimes \phi_0 \otimes \cdots$ as can be shown.

$\phi_0 = \phi_0 \otimes \phi_0 \otimes \cdots \otimes \phi_0 \otimes \cdots$ as can be shown.

(c) *Relationship between both formalisms*: We are going to compare both formalisms and look for the states which they have in common. The means of such a comparison will be our Weyl generalized correspondence.

Pick a state ρ on $\mathscr{H}$. A case of particular interest is that in which

$$\rho = \rho_1 \otimes \cdots \otimes \rho_n \otimes \rho_0 \otimes \cdots \otimes \rho_0 \cdots, \quad \text{with} \, \rho_0 = |\phi_0\rangle \langle\phi_0|, \tag{2.32}$$

because then

$$\begin{aligned} &\operatorname{tr}\rho e^{i(\alpha_1 \hat{Q}_1 + \beta_1 \hat{P}_1 + \cdots + \alpha_n \hat{Q}_n + \beta_n \hat{P}_n)} \\ &= \operatorname{tr}\rho_1 e^{i(\alpha_1 \hat{Q}_1 + \beta_1 \hat{P}_1)} \times \cdots \operatorname{tr}\rho_n e^{i(\alpha_n \hat{Q}_n + \beta_n \hat{P}_n)} \\ &= \left[\int_{R^2} e^{i(\alpha_1 Q_1 + \beta_1 P_1)} d\mu_1\right] \times \cdots \left[\int_{R^2} e^{i(\alpha_n Q_n + \beta_n P_n)} d\mu_n\right] \end{aligned} \tag{2.33}$$

$$= \int_\Omega e^{i(\alpha_1 Q_1 + \cdots \beta_n P_n)} d\mu, \quad \text{(by Fubini's theorem)}. \tag{2.34}$$

(The caret means we are working with operators, if it is not used we deal with random variables.)

μ is uniquely defined by $\mu_1 \cdots \mu_n \cdots$ but is no longer a probability measure, because the μ_i's are not. Therefore the quantum states described by ρ will not ordinarily be in correspondence with a "stochastic state." Furthermore, we know that only if every μ_i is jointly Gaussian in Q_i and P_i will we have a "stochastic state." Conversely, if we start with a probability measure on Ω which is the product of probability measures on Ω_i (product of ranges of Q_i and P_i), we will find by using the generlized Weyl correspondence that a state ρ on $\mathscr{H}$ of the above form is $\mu = \mu_1 \times \cdots \times \mu_n \times \Pi_{k=n}^{\infty} \mu_k$, where $\mu_k = \mu_0$, μ_0 being the measure corresponding to ρ_0, if $\Delta P_i \Delta Q_i \geqslant \hbar/2, \forall i$. (If μ_i is jointly Gaussian in P_i and Q_i, we are sure that the last inequality holds.) States different from these will be very difficult to compare.

Let us begin with the ground state of the quantum field: $\rho = |\phi\rangle \langle\phi|$ with $\phi = \phi_0 \otimes \phi_0 \otimes \cdots$ (or $\rho = \rho_0 \otimes \rho_0 \otimes \cdots$). Our questions are: What is the corresponding probability measure μ on Ω, given a state for the stochastic field, and what is the physical meaning of such measure in the context of stochastic electrodynamics? The answer is easy to reach, taking into account that our problem reduces to the one-dimensional problem, as we can infer from the above formula by writing $\rho_j = \rho_0$.

The problem in one dimension is well-known and in the present case correspnds to the classical Weyl correspondence. If we define the Wigner function (in one dimension) corresponding to the state ρ_k as

$$\int e^{i(\alpha_k Q_k + \beta_k P_k)} d_{\mu_k} = \int e^{i(\alpha_k Q_k + \beta_k P_k)} f(Q_k, P_k) dQ_k dP_k, \tag{2.35}$$

it is well known that if ρ_k is a pure state, i.e., $\rho_k = |\phi_k\rangle \langle\phi_k|$, then

$$f(Q_k, P_k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\tau P_k} \phi_k(Q_k - \tfrac{1}{2}\hbar\tau) \phi_k^*(Q_k + \tfrac{1}{2}\hbar\tau) d\tau, \tag{2.36}$$

and, dealing with pure states, $f(Q_k, P_k)$ is defined positive if and only if ϕ_k is a Gaussian function in its argument (Hudson's theorem).[4]

Coming back to the case in which every ρ_k equals ρ_0 we compute the Wigner function and find

$$f(Q_k P_k) = \frac{1}{\hbar\pi} e^{-P_k^2/\omega_k\hbar} e^{-Q_k^2\omega_k/\hbar}, \tag{2.37}$$

and then $d\mu_k = f(Q_K, P_k) dQ_k dP_k$ is jointly Gaussian in Q_k and P_k for all k. That means we have found a physical state of the stochastical radiation field. Furthermore, if we compute

$$\begin{aligned} E|c_k|^2 &= \frac{c^2}{4} E(Q_k^2) + \frac{1}{\omega_k} E(P_k^2) \\ &= \frac{c^2}{4} \frac{\hbar}{2\omega_k} + \frac{1}{\omega_k} \frac{\omega_k \hbar}{2} = \frac{1}{4} \frac{\hbar c^2}{\omega_k}, \end{aligned} \tag{2.38}$$

we see that this state behaves like background electromagnetic radiation, i.e., the ground state of the stochastic field.[11]

If we now consider this ground state plus a planewave, whose potential vector is given by

$$A(\vec{r}, t) = (\hbar c^2/2V\omega_j)^{\frac{1}{2}} 2b\epsilon^\lambda \cos(\mathbf{k}_j \cdot \mathbf{r} - \omega_j t + \delta), \tag{2.39}$$

the new variables $c_{k\lambda}$ become

$$c'_{j\lambda} = c_{k\lambda} + b\left(\frac{\hbar c}{2\omega_j}\right)^{1/2} e^{i\delta}. \tag{2.40}$$

That implies that the new Wigner function is

$$\begin{aligned} f(Q_k, P_k) = \frac{1}{\pi\hbar} \exp\Bigg\{ &- \frac{Q_k - [(2\hbar/\omega_k)^{1/2} b \cos\delta]^2 \omega_k}{\hbar} \\ &- \frac{[P_k - (2\hbar\omega_k)^{1/2} \sin\delta]^2}{\omega_k \hbar} \Bigg\}. \end{aligned} \tag{2.41}$$

It can be shown that to this Wigner function there corresponds a pure state in the quantum formalism, which is

$$\begin{aligned} \phi_k(x) = \left(\frac{\omega_k}{\pi\hbar}\right)^{1/2} \exp\Bigg\{ &- \frac{\omega_k}{2\hbar}[x - (2\hbar/\omega_k)^{1/2} b\cos\delta)]^2 \\ &+ i(2\omega_k/\hbar)^{1/2} b\sin\delta\, [x - (2\hbar/\omega_k)^{1/2} b\cos\delta]\Bigg\}. \end{aligned} \tag{2.42}$$

If now the operator

$$\begin{aligned} \hat{a}_k &= \frac{1}{\sqrt{2}}\left[\sqrt{\frac{\omega_k}{\hbar}}\hat{x} + \sqrt{\frac{\hbar}{\omega_k}}\frac{\partial}{\partial x}\right] \\ &= \frac{1}{\sqrt{2}}\sqrt{\frac{\omega_k}{\hbar}}\hat{Q}_k + i\sqrt{\frac{\hbar}{\omega_k}}\hat{P}_k, \end{aligned} \tag{2.43}$$

acts over $\phi_k(x)$ we get finally $a_k \phi_k = be^{i\delta}\phi_k$, which is a coherent state of the one-dimensional oscillator. Thus we see that the stochastic field state, made up of the background radiation plus a plane wave (well known), corresponds by the generalized Weyl correspondence to the state

$$\psi = \phi_0 \otimes \cdots \phi_0 \otimes \phi_k \otimes \phi_0 \otimes \cdots \otimes \phi_0 \otimes \cdots. \tag{2.44}$$

If instead of one we have a finite number of plane waves, we get

$$\phi = \phi_1 \otimes \cdots \otimes \phi_n \otimes \phi_0 \cdots \otimes \phi_0 \cdots, \tag{2.45}$$

with $\hat{a}_k \phi_k = b_k e^{i\delta_k} \phi_k$. Then ϕ is a coherent state of the radiation in the quantum theory.

We may try to improve this result by trying to put in correspondence more general states in both formalisms. But it does not seem possible. We do not know, for instance, how to represent the most general coherent state, and we are afraid that the quantum formalisms proposed here prevents us from further generalizations. E. Santos[11] proved in a nonrigorous way that any coherent quantum state can be put in correspondence with a state of the stochastical field, and we conjecture that the coherent states are the only ones for which this correspondence is possible.

APPENDIX A: PROOF OF THE LEMMA IN THEOREM 2 OF SECTION I

The essential points of the proof are the statements (e) and (f) of Theorem 2.

Let us first prove the lemma for A generated by $X_1 = x$ and $X_2 = y$. A is spanned by elements of the form $x^{k_1}y^{k_2}$, k_1 and k_2 being integers.

If $k = k_1 + k_2$, we have

$$x^{k_1}y^{k_2} = \lambda_1(a_1 x + b_1 y)^k + \cdots + \lambda_l(a_l x + b_l y)^k. \quad \text{(A1)}$$

Here l is the number of terms in the span of $(a+b)^k$.

$$\begin{aligned} x^{k_1}y^{k_2} &= \lambda_1\left[a_1^k x^k + \binom{k}{1} a_1^{k-1} b_1 x^{k-1} y + \cdots + b_1^k y^k\right] \\ &\quad + \lambda_l\left[a_l x^k + \binom{k}{1} a_l^{k-1} b_l x^{k-1} y + \cdots + b_l^k y^k\right] \\ &= (\lambda_1 a_1^k + \lambda_2 a_2^k + \cdots + \lambda_l a_l^k) x^k \\ &\quad + \left[\lambda_1 a_1^{k-1} b_1 \binom{k}{1} + \lambda_2 a_2^{k-1} b_2 \binom{k}{1} \right. \\ &\quad \left. + \cdots + \lambda_l a_l^{k-1}\binom{k}{1}\right] x^{k-1} y \\ &\quad + \cdots + (\lambda_1 b_1^k + \cdots + \lambda_l b_l^k) y_l^k. \end{aligned} \quad \text{(A2)}$$

x and y are linearly independent and so are their powers and products of these powers by assumptions (e) and (f) in Theorem 2. From (A2) we obtain the following system of equations in $\lambda_1,\ldots,\lambda_l$:

$$\begin{aligned} &\lambda_1 a_1^k + \lambda_2 a_2^k + \cdots + \lambda_l a_l^k = 0, \\ &\lambda_1 a_1^{k-1} b_1 \binom{k}{1} + \cdots \lambda_l a_l^{k-1} b_l \binom{k}{1} = 0, \\ &\lambda_1 a_1^{k_1} b_2^{k_2} \binom{k}{k_1} + \cdots + \lambda_l a_l^{k_1} b_l^{k_2} \binom{k}{k_1} = 1, \\ &\lambda_1 b_1^k + \cdots + \lambda_l b_l^k = 0, \end{aligned} \quad \text{(A3)}$$

which has a solution if and only if

$$\binom{k}{1}\binom{k}{2}\cdots\binom{k}{k-1} \begin{vmatrix} a_1^k & a_2^k & \cdots & a_l^k \\ a_1^{k-1} b_1 & a_2^{k-2} b_2 & \cdots & a^{k-1} b_e \\ b_1^k & b_2^k & \cdots & b^k \end{vmatrix} \neq 0. \quad \text{(A4)}$$

Now if we take $b_1 = b_2 = \cdots = b_l = 1$, the last determinant is a Vandermonde which is nonzero if and only if all the a_i are different. We can always do this. Hence our problem has at least one solution (in fact more than one), which can be found by applying Rouche's theorem and substituting the values found in to (A1).

The result is now proved by induction on N. We assume

$$X_1^{k_1}\cdots X_{N-1}^{k_N} = \sum_{l=1}^{M} \lambda_l \Big(\sum_{j=1}^{N-1} a_{jl} X_j\Big)^{k_l}. \quad \text{(A5)}$$

Now write

$$W_l = \sum_{j=1}^{N-1} a_{jl} X_j. \quad \text{(A6)}$$

Then

$$\begin{aligned} X_1^{k_1}\cdots X_{N-1}^{k_N} X_N^{k_N} &= \Big(\sum_{l=1}^{M} \lambda_l W_l^{k_l}\Big) X_N^{k_N} \quad \text{(A7)} \\ &= \sum_{l=1}^{M} \lambda_l W_l^{k_l} X_N^{k_N} \\ &= \sum_{l=1}^{M} \lambda_l \sum_{i=1} \mu_i (a_i W_l + b_i X_N)^{k_l + k_N}. \quad \text{(A8)} \end{aligned}$$

(A8) follows from the independence of W_l^j and X_N^h for all j and h. Hence (A7) and (A8) imply the lemma.

APPENDIX B: PROOF OF THE LEMMAS IN SECTION II

Lemma I: If we can show this result for A and B, we have proved the lemma, because of the uniqueness of the closure of a normal *-algebra. Let F be an element of A.

$$F \in A \text{ if } F = \lambda_0 I_A + \sum_{i_1,\ldots,i_k=1}^{n_1,\ldots,n_k} \lambda_{i_1,\ldots,i_k} X_i \cdots X_{i_k}, \quad \text{(B1)}$$

where I_A is the identity in A.

We define now the mapping from A onto B in the following way.

$$g(F) = \lambda_0 I_B + \sum_{i_1,\ldots,i_k=1}^{n_1,\ldots,n_k} \lambda_{i_1,\ldots,i_k} N_{i_1} \cdots N_{i_k}, \quad \text{(B2)}$$

where $I_B = g(I_A)$ is the identity in B.

It can be easily proved that g is an *-isomorphism between the algebras. We claim that, in fact, g carries the topology in both directions. We prove it if we show that

$$\begin{aligned} &= \|\lambda_0 I_A + \sum_{i_1,\ldots,i_k=1}^{n_1,\ldots,n_k} \lambda_{i_1,\ldots,i_k} X_i \cdots X_i\| \\ &= \|\lambda_0 I_B + \sum_{i_1,\ldots,i_k=1}^{n_1,\ldots,n_k} \lambda_{i_1} \cdots_{i_k} N_{i_1} \cdots N_{i_k}\|. \end{aligned} \quad \text{(B3)}$$

The former of these two norms is the supremum of F. The variables $X_{i_1},\ldots,X_{i_n}$ may take only the values zero and one. Such a supremum must be the maximum of the positive numbers obtained with all possible sums of the following:

$$(|\lambda_{i_1,\ldots,i_k}|)_{i_1,\ldots,i_k=1}^{n_1,\ldots,n_k}. \quad \text{(B4)}$$

It is only a matter of calculus to show that this is also equal to the right hand side of (B3). Then the lemma follows.

Lemma II: This is straightforward if we take account of the continuity of expectation and trace.

Lemma III: Let δ_D be equal to $F(I, X_1, X_2,\ldots, X_1\ldots X_P$, where F must be linear in its arguments. Then

$$\begin{aligned} &\delta_D(D_{i_1,\ldots,i_p}) = 1 = F(I, X_1, X_2,\ldots,X_1\ldots X_p), \\ &\Big[(i_1,\ldots,i_p)\times\prod_{p+1}^{\infty}\Omega_i = F(1, i_1, i_2,\ldots,i_1\ldots i_n)\Big]. \end{aligned} \quad \text{(B5)}$$

On the other hand,

$$F(1,l_1.l_2,..,l_1...l_p) = 0, \text{ if } (l_1,l_2,...l_p) \neq (i_1,i_2,...,i_p). \quad \text{(B6)}$$

Then

$$\begin{aligned} &F(I,N_1,N_2,..,N_1...N_p)\chi_{h_1,...,h_r} \\ &= F(\chi_{h_1,...,h_r}, N_1\chi_{h_1,...,h_r}, N_1...N_p\chi_{h_1,...,h_r}) \\ &= F(1,\delta_{1h_1},\delta_{2h_2},..,\delta_{1h},\delta_{2h_2}...\delta_{ph_p}), \end{aligned} \quad \text{(B7)}$$

wherever $r \geqslant p$. If $r < p$ then

$$\begin{aligned} &N_{r+1}\chi_{h_1,...,h_r} = 0, \\ &N_p\chi_{h_1,...,h_r} = 0. \end{aligned} \quad \text{(B8)}$$

Accordingly, the first row in (B7) must be

$\chi_{h_1,...,h_r}$ if $\delta_{ih_1} = i_1,...,\delta_{ph_p} = i_p$ or if $r < p$,

$i_r + 1 = \cdots = i_p = 0$,

or 0 otherwise.

Lemma IV: Let $\mathscr{C}$ be the algebra spanned by the fuctions of the form δ_D. Obviously $\mathscr{C} \subset A$.

Consider now X_k:

$$\begin{aligned} &X_k\left[(i_1,...,i_{k-1},1) \times \prod_{i=k+2}^{\infty} \Omega_i\right] = 1, \\ &X_k\left[(i_1,...,i_{k-1},0) \times \prod_{i=k+2}^{\infty} \Omega_i\right] = 0. \end{aligned} \quad \text{(B9)}$$

Then

$$X_k = \sum_{i_1,...,i_{k-1}=0}^{1} {}^{\delta}D_{i_1,...,i_{k-1},1} \quad \text{(B10)}$$

hence $X_k \in \mathscr{C}$. The random variables X_k span A. Consequently $A \subset \mathscr{C} \Rightarrow A = \mathscr{C}$.

ACKNOWLEDGMENT

We acknowledge partial financial support to Instituto de Estudios Nucleares, Madrid.

[1]G. S. Agarwal and E. Wolf, Phys. Rev. D **2**, 161 (1970).

[2]H. Weyl, *The theory of Groups and Quantum Mechanics* (Dover, New York, 1931).

[3]J. G. Krüger and A. Poffyn, Physica A **85**, 84 (1976).

[4]R. L. Hudson, Rep. Math. Phys. **6**, 249 (1974).

[5]V. Frisch and R. Bourret, J. Math. Phys. **11**, 364 (1969).

[6]E. Santos, J. Math. Phys. **15**, 1954 (1974). That paper is partially wrong because the generalized Weyl correspondence (named there "quantumlike formalism") was used for the solution of some stochastic differential equations without any correct proof of the validity of the method.

[7]S. de Groot, *La transformation de Weyl et la fonction de Wigner: une forms alternative de la mécanique quantique* (Université de Montreal, 1974).

[8]H. J. Groenewold, Physics XII, 7, 405 (1946).

[9]M. Gadella, Thesis, University of Santander, Spain (1979).

[10]M. Gadella, submitted to Nuovo Cimento.

[11]E. Santos, Nuovo Cimento B **22**, 201 (1977).

[12]J. J. Sakurai, *Advaced Quantum Mechanics* (Addison-Wesley, Reading, Mass., 1967).

Invariant inner products on spaces of solutions of the Klein–Gordon and Helmholtz equations

Stanly Steinberg
Department of Mathematics and Statistics, The University of New Mexico, Albuquerque, New Mexico 87131
Kurt Bernardo Wolf
Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, México 20, D. F., Mexico

(Received 15 January 1981; accepted for publication 10 April 1981)

We construct sesquilinear forms which are invariant under the similarity groups of the Klein–Gordon and Helmholtz equations. These give rise to positive definite inner products on subspaces of solutions of these equations. For the Klein–Gordon case, the known Poincaré-invariant inner product for positive-energy solutions is recovered. For the Helmholtz case, a new Euclidean-invariant inner product is presented which involves the function and its normal derivative to a line, integrated nonlocally over that line.

PACS numbers: 03.65. – w, 11.30. – j

1. INTRODUCTION

The Klein–Gordon and Helmholtz equations can be written in the form

$$[\partial_t^2 + \sigma\partial_q^2 + m^2] f(q,t) = 0, \tag{1}$$

where $\sigma = -1$ for the Klein–Gordon case and $\sigma = +1$ for the Helmholtz case. The use of t as a "time" variable in the Helmholtz equation is somewhat unusual, but it will be helpful to think of it as generating an evolution in one direction in the q, t plane.

Each of these equations can be rewritten as a system,

$$\mathbb{T}_2^{(\sigma,m)} \mathbb{f}(q,t) = \partial_t \mathbb{f}(q,t), \tag{2a}$$

$$\mathbb{T}_2^{(\sigma,m)} = \begin{pmatrix} 0 & 1 \\ -\sigma\partial_q^2 - m^2 & 0 \end{pmatrix}, \quad \mathbb{f}(q,t) = \begin{pmatrix} f(q,t) \\ f_t(q,t) \end{pmatrix}. \tag{2b}$$

For many applications it is important to have a positive definite inner product on the space of initial data $\mathbb{f}(q,0)$ in which the evolution described by (2) will be unitary. For the Schrödinger equation of quantum mechanics, for instance, such a time-translation invariant sesquilinear form is simply the $\mathscr{L}^2(\mathscr{R})$ inner product, which gives rise to a positive definite norm and which allows us to regard time evolution as a unitary process.

For (1) and (2) we will consider inner products defined through sesquilinear forms,

$$C^\sigma(\mathbb{f},\mathbb{g}) = \int_{-\infty}^{\infty} dq \int_{-\infty}^{\infty} dq' (f(q,t)^* f_t(q,t)^*) \times \begin{pmatrix} M_{11}^\sigma(q,q') & M_{12}^\sigma(q,q') \\ M_{21}^\sigma(q,q') & M_{22}^\sigma(q,q') \end{pmatrix} \begin{pmatrix} g(q',t) \\ g_t(q',t) \end{pmatrix} \tag{3}$$

with a metric kernel $\|M_{kk'}^\sigma(q,q')\|$.

In Sec. 2, the problems we address are:

(a) To find all metric kernels such that (3) is invariant under the similarity group of transformations—including, of course, time evolution—which leave (1) invariant.

(b) To find all subspaces of the solution space of (1) for which the invariant (3) is positive definite.

The similarity group of (1) will then be isometric in (3), meaning in particular that the double-integration line in (3) may be translated and rotated in the q, t plane (in the Euclidean or Lorentz sense, for the Helmholtz or Klein–Gordon case, respectively), without changing the value of the integral.

If

$$\hat{f}(\omega,t) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} dq\, f(q,t) e^{-i\omega q} \tag{4a}$$

is the Fourier transform of $f(q,t)$ with respect to q, then the solution of (1) can be explicitly written as

$$f(q,t) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} d\omega\, e^{i\omega q} \{\hat{f}(\omega,0)\cos([m^2 - \sigma\omega^2]^{1/2} t) + \hat{f}_t(\omega,0)[m^2 - \sigma\omega^2]^{-1/2} \sin([m^2 - \sigma\omega^2]^{1/2} t)\}. \tag{4b}$$

For the Klein-Gordon case ($\sigma = -1$) this formula shows that the initial value problem is well posed. For the Helmholtz case ($\sigma = 1$), on the other hand, the possibly complex argument of the trigonometric functions implies that the initial value problem for (1) is ill posed.[1] The evolution defined by (2), consequently, cannot be unitary in the usual sense. However, if $\hat{f}(\omega,t) = 0, \hat{f}_t(\omega,t) = 0$ for $|\omega| > m$, then the initial value problem *is* well posed. Such solutions will be called "oscillatory" (or finite-energy) solutions.

In Secs. 3 and 4 we condense the results for the Klein-Gordon and Helmholtz equations, respectively. Some further comments on the interest of conserved sesquilinear forms in the context of group theory and integral transforms are offered in Sec. 5.

2. SIMILARITY ALGEBRAS AND INVARIANT SESQUILINEAR FORMS

The differential equations (1) of second order in time, when written in a two-component first-order form (2a), display the operator $\mathbb{T}_2^{(\sigma,m)}$ in (2b) as generating translations in

 0022-2488/81/081660-04$01.00 © 1981 American Institute of Physics

the second argument of the solution two-vector $\mathbb{f}(q,t)$. Translations in the first argument of $\mathbb{f}(q,t)$ are generated by

$$\mathbb{T}_1 = \begin{pmatrix} \partial_q & 0 \\ 0 & \partial_q \end{pmatrix}. \tag{5}$$

Lorentz transformations ($\sigma = -1$) or Euclidean rotations ($\sigma = 1$) in the q, t plane are generated by

$$\mathbb{R}^{(\sigma,m)} = \begin{pmatrix} -\sigma t \partial_q & q \\ -\frac{1}{2}\{q, m^2 + \sigma \partial_q^2\}_+ & -\sigma t \partial_q \end{pmatrix}, \tag{6}$$

where $\{A,B\}_+ = AB + BA$ is the anticommutator of A and B. These three operators obey the commutation relations

$$[\mathbb{T}_1, \mathbb{T}_2^{(\sigma,m)}] = 0, \quad [\mathbb{R}^{(\sigma,m)}, \mathbb{T}_1] = -\mathbb{T}_2^{(\sigma,m)},$$
$$[\mathbb{R}^{(\sigma,m)}, \mathbb{T}_2^{(\sigma,m)}] = \sigma \mathbb{T}_1. \tag{7}$$

We thus identify the three operators as the generators of the component of the identity of the Lie group IO_2^σ [to mean $\mathrm{IO}(1,1)^\uparrow_+$ for $\sigma = -1$, and ISO(2) for $\sigma = 1$]. In addition to these we have the discrete generators of space inversion and, for $\sigma = -1$, time inversion as well. These complete the Poincaré and Euclidean invariance groups of Eqs. (1) and (2). The Lie algebra (7) is the similarity algebra of these equations. Its elements commute with $\mathbb{T}_2^{(\sigma,m)} - \partial_t$, and hence map solutions into solutions.[2] Initial conditions, of course, need not be invariant nor covariant under these transformations.

In order that an IO_2^σ transformation, $\exp(\alpha\,\mathbb{A})$, be isometric under a nondegenerate sesquilinear form (3), the generator $\mathbb{A}$ must be skew-Hermitian under it, i.e., $C^\sigma(\mathbb{A}\mathbb{f},\mathbb{g}) = C^\sigma(\mathbb{f},\mathbb{A}\mathbb{g})$. Through integration by parts, sufficient differential and boundary conditions can be found to guarantee this. The metric kernel matrix elements $M^\sigma_{kk'}(q,q')$, $k,k' = 1,2$ in (3) are then determined by differential equations from the matrix elements $A_{kk'}(q,\partial_q)$ of $\mathbb{A}$, as

$$\sum_{k''}\{A_{k''k}(q,\partial_q)^{\mathrm{T}*} M^\sigma_{k''k'}(q,q') + A_{k''k'}(q',\partial_{q'})^{\mathrm{T}} M^\sigma_{kk''}(q,q')\} = 0, \tag{8}$$

where $^{\mathrm{T}}$ is the operator transpose: $q^{\mathrm{T}} = q$, $\partial_q^{\mathrm{T}} = -\partial_q$ and $(AB)^{\mathrm{T}} = B^{\mathrm{T}}A^{\mathrm{T}}$.

Sufficient boundary conditions—which will determine suitable classes of function pairs $\mathbb{f}(q,t)$—are (deleting the argument t)

$$\int_{-\infty}^{\infty} dq \sum_k f_k(q)^* \sum_{k'k''} \mathscr{B}\{M^\sigma_{kk''}(q,\cdot), A_{k''k'}, g_{k'}\} = 0, \tag{9a}$$

$$\int_{-\infty}^{\infty} dq' \sum_{k'}\sum_{kk''} \mathscr{B}\{M^\sigma_{k''k'}(\cdot,q'), A^*_{k''k}, f^*_k\} g_{k'}(q') = 0, \tag{9b}$$

where $f_1(q)$ and $f_2(q)$ are $f(q)$ and $f_t(q)$, respectively, and

$$\mathscr{B}\{u,A,v\} = \int_{-\infty}^{\infty} dx\, [u(x)A(x,\partial_x)v(x) - v(x)A(x,\partial_x)^{\mathrm{T}}u(x)]. \tag{9c}$$

Equations (9) will be analyzed once the consequences of the formal equation (8) are drawn.

We shall examine now the implication of (8) for the IO_2^σ algebra generators. Substitution of $\mathbb{T}_1$ into (8) yields

$$(\partial_q + \partial_{q'})M^\sigma_{kk'}(q,q') = 0, \quad k,k' = 1,2 \tag{10a}$$

and hence

$$M^\sigma_{kk'}(q,q') = M^\sigma_{kk'}(q-q'), \quad k,k' = 1,2, \tag{10b}$$

i.e., the in general nonlocal metric kernel must be diagonal in the integration variables q and q'. Invariance under space inversions may be seen separately to lead to the requirement that $M^\sigma_{kk'}(|q-q'|)$ be *even* functions of their argument.

Invariance under t-translations results from substituting $\mathbb{T}_2^{(\sigma,m)}$ from (2) into (8), leading to

$$M^\sigma_{12}(q,q') = -M^\sigma_{21}(q,q'), \tag{11a}$$

$$[\sigma\partial_q^2 + m^2]M^\sigma_{22}(q,q') = M^\sigma_{11}(q,q') = [\sigma\partial_{q'}^2 + m^2]M^\sigma_{22}(q,q'), \tag{11b}$$

$$(\partial_q^2 - \partial_{q'}^2)M^\sigma_{kk'}(q,q') = 0, \quad k,k' = 1,2. \tag{11c}$$

The last equations mean that all $M^\sigma_{kk'}(q,q')$ can be written as a function of $q - q'$ plus a function of $q + q'$. The latter is zero due to (10).

Finally, invariance under boosts/rotations (6) leads, together with (10) and (11), to

$$\left[\frac{d^2}{dz^2} + \frac{1}{z}\frac{d}{dz} + \sigma m^2\right]M^\sigma_{22}(z) = 0, \tag{12a}$$

$$M^\sigma_{11}(z) = -\frac{\sigma}{z}\frac{d}{dz}M^\sigma_{22}(z), \tag{12b}$$

$$zM^\sigma_{12}(z) = 0, \tag{12c}$$

for $z = q - q'$. Equation (12a) is the Bessel ($\sigma = 1$) or the modified Bessel ($\sigma = -1$) differential equation of order zero, while (12b) displays the raising operator for these functions. The solution of (12c) is a Dirac δ.

Invariance under the Poincaré or Euclidean group thus determines the metric kernel, for arbitrary constants $a^\pm$, $b^\pm$, $c^\pm$ as follows. For the Klein–Gordon equation case

$$M^{-1}_{11}(z) = m^2[a^- I_1(mz)/mz - b^- K_1(mz)/mz], \tag{13a}$$

$$M^{-1}_{22}(z) = a^- I_0(mz) + b^- K_0(mz), \tag{13b}$$

$$M^{-1}_{12}(z) = c^-\delta(z) = -M^{-1}_{21}(z), \tag{13c}$$

while for the Helmholtz equation case, it is

$$M^{+1}_{11}(z) = m^2[a^+ J_1(mz)/mz + b^+ Y_1(mz)/mz], \tag{14a}$$

$$M^{+1}_{22}(z) = a^+ J_0(mz) + b^+ Y_0(mz), \tag{14b}$$

$$M^{+1}_{12}(z) = c^+\delta(z) = -M^{+1}_{21}(z). \tag{14c}$$

Now, if the solutions $\mathbb{f}(q,t)$ and $\mathbb{g}(q,t)$ are only assumed to be differentiable and $\mathscr{L}^2(\mathscr{R})$ functions of q, then the boundary conditions (9) for (13) and (14) disqualify the modified Bessel function term, on the basis of its exponentially growing asymptotic behavior. This forces us to set $a^- = 0$. Under the same assumptions of the solutions, the requirement that (3) be finite implies that $M^\sigma_{kk'}(z)$ be in $\mathscr{L}^1(\mathscr{R})$. This disqualifies the Macdonald and Neumann function terms $K_1(z)/z$ and $Y_1(z)/z$, which have a z^{-2} singularity at the origin, thus forcing $b^- = 0$ and $b^+ = 0$.

3. THE KLEIN–GORDON CASE

For the Klein–Gordon case we are left with the antidiagonal term (13c). Choosing $c^- = -i$ to insure that $C^{-1}(\mathrm{f},\mathrm{f})$ be real, we obtain the unique Poincaré-invariant sesquilinear form

$$C^{K-G}(\mathrm{f},\mathrm{g}) = -i\int_{-\infty}^{\infty} dq\,(f(q,t)^* f_t(q,t)^*) \times\begin{pmatrix} 0 & 1\\ -1 & 0\end{pmatrix}\begin{pmatrix} g(q,t)\\ g_t(q,t)\end{pmatrix}. \tag{15}$$

leading to an inner product where the three generators of $\mathrm{IO}(1,1)^{\prime}_{+}$ are skew-Hermitian in the space of $\mathscr{L}^2$ solutions. This invariant inner product for the Klein–Gordon equation is known.[3] It appears to have been first recognized by Klein and Gordon,[4] who considered conserved current densities derived from the continuity condition.

As to the positivity of (15), the Fourier transform with respect to t of the solution space $\mathscr{S}^{\mathrm{K-G}}$ of the Klein–Gordon equation, is the space of $\mathscr{L}^2(\mathscr{R})$ functions of a variable E, whose physical interpretation is that of energy, with support on $(-\infty, -m]\cup[m,\infty)$. Finite $\mathrm{IO}(1,1)^{\prime}_{+}$ transformations map this space onto itself. Further, under space inversion, $\mathscr{S}^{\mathrm{K-G}}$ maps onto itself and (15) is also invariant under this transformation, and hence under $\mathrm{IO}(1,1)^{\prime}$. Time inversion, on the other hand, is *not* an invariance transformation of the sesquilinear form (15). For $t=0$, time derivatives will change sign and (15) does so likewise. Time inversion exchanges the two parts of $\mathscr{S}^{\mathrm{K-G}}$ having Fourier transforms with support on $[m,\infty)$ and $(-\infty,-m]$. We call them $\mathscr{S}^{\mathrm{K-G}}_{\pm}$, respectively: Positive- and negative-energy solution spaces. The sesquilinear form (15) on $\mathscr{S}^{\mathrm{K-G}}_{+}$ is positive definite, i.e., $C^{K-G}(\mathrm{f},\mathrm{f})>0$ for $0\neq \mathrm{f}\in\mathscr{S}^{\mathrm{K-G}}_{+}$. The usual proof of this statement[3] makes use of the Fourier decomposition of $\mathrm{f}(q,t)$ with respect to t. A positive definite sesquilinear form on the function space $\mathscr{S}^{\mathrm{K-G}}_{+}$ allows for the introduction of a nondegenerate inner product defining a complex Hilbert space and an associated norm. The known inner product (15) for the Klein–Gordon solution space has been thus rederived and validated as the only such $\mathrm{IO}(1,1)^{\prime}$ invariant form.

4. THE HELMHOLTZ CASE

Regarding the Helmholtz equation case, the requirement of skew-Hermiticity in $\mathscr{L}^2(\mathscr{R})$ of the Euclidean algebra generators $\mathbb{T}_1$, $\mathbb{T}_2^{(1,m)}$, and $\mathbb{R}^{(1,m)}$, and invariance under space inversions led to the forms (3)–(14) with $b^+=0$. The exponentiation of these generators to finite Euclidean transformations, however, differs in two respects from the Klein–Gordon case. First, time inversion, multiplied by space inversion, is an element of the finite group: A rotation by π. The antidiagonal part (13c) of the metric kernel changes sign under this operation and must therefore be absent from a Euclidean-invariant form. This leaves us with the only possible sesquilinear form

$$C^{H}(\mathrm{f},\mathrm{g}) = c\int_{-\infty}^{\infty} dq\int_{-\infty}^{\infty} dq'\,(f(q,t)^*\ f_t(q,t)^*) \times\begin{pmatrix} m^2\dfrac{J_1(m[q-q'])}{m[q-q']} & 0\\ 0 & J_0(m[q-q'])\end{pmatrix}\begin{pmatrix} g(q',t)\\ g_t(q',t)\end{pmatrix}. \tag{16}$$

For real c, $C^H(\mathrm{f},\mathrm{f})$ is real.

Second, the solution space $\mathscr{S}^{H}$ of the Helmholtz equation contains both oscillatory and "exponential" solutions. The former have a Fourier transform, with respect to any line in the q,t plane, which is in $\mathscr{L}^2(\mathscr{R})$ with support on the interval $[-m,m]$. They are elements of a subspace which we denote by $\mathscr{S}^{H}_{o}$. The latter, whose behavior is exponentially growing in some direction on the q,t plane, are elements of a subspace we denote by $\mathscr{S}^{H}_{e}$. The intersection of $\mathscr{S}^{H}_{o}$ and $\mathscr{S}^{H}_{e}$ is empty, and each is invariant under Euclidean transformations. The Euclidean algebra generators are thus skew-Hermitian under (16) for $\mathscr{S}^{H}_{o}$ and this sesquilinear form is invariant under all IO(2) transformations.

Finally, the form (16) is positive definite on $\mathscr{S}^{H}_{o}$, as can be verified noting that the Fourier transforms of $\mathscr{L}^1(\mathscr{R})$ functions $m^2J_1(mz)/mz$ and $J_0(mz)$ have support on $[-m,m]$ and are, respectively $(2/\pi)^{1/2}(m^2-p^2)^{\pm 1/2}$ which are positive definite functions. The positive definite sesquilinear form (16) on $\mathscr{S}^{H}_{o}$ (with $c>0$) thus allows for the introduction of a nondegenerate Euclidean-invariant inner product defining a Hilbert space and an associated norm. In contradistinction to (15), (16) appears to be new.

5. DEFORMATION OF $\mathrm{IO_2}$ AND NEW REALIZATIONS OF SO(2,1)

As all separable Hilbert spaces are unitarily equivalent, the IO_2 generators given in (2b), (4), and (6), in the Hilbert spaces $\mathscr{S}^{\mathrm{K-G}}$ and $\mathscr{S}^{H}_{o}$, may be mapped unitarily on the usual realization of the IO_2 generators on the hyperbola and circle and $\mathscr{L}^2$ spaces thereupon. In the latter spaces $\mathbb{R}^{(\sigma,m)}$ is realized as $d/d\phi$, where for the Klein–Gordon case $\phi\in\mathscr{R}\dotplus\mathscr{R}$ (the two branches of a hyperbola), while for the Helmholtz case $\phi\in S_1$ (the circle). The two other generators $\mathbb{T}_1$ and $\mathbb{T}_2^{(\sigma,m)}$ become, respectively, operators with hyperbolic and trigonometric functions in ϕ. The corresponding Hilbert spaces are $\mathscr{L}^2(\mathscr{R})\dotplus\mathscr{L}^2(\mathscr{R})$ and $\mathscr{L}^2(S_1)$, respectively. The intertwining operators are easy to find as generating functions built out of the generalized plane-wave eigenbases of $\mathbb{T}_1$ and $\mathbb{T}_2^{(\sigma,m)}$, and Dirac δ's for $\mathscr{L}^2(\mathscr{R})\dotplus\mathscr{L}^2(\mathscr{R})$ and $\mathscr{L}^2(S_1)$. The Helmholtz case was analyzed in Ref. 5, while the Klein–Gordon case is just as simple. The IO_2-invariant inner products (15) and (16) are then obtained from the corresponding $\mathscr{L}^2$ inner products with measure $d\phi$. {Note that for the Helmholtz case, the disqualified weight function in (14c) corresponds to the Euclidean noninvariant measure $\mathrm{sign}\,(\phi)d\phi$, $\phi\in(-\pi,\pi]$}.

In Ref. 5 we set out to deform IO_2 through

$$\mathbb{M}_1 = -im^{-1}\mathbb{R}\mathbb{T}_2+\tau\mathbb{T}_1,\qquad \mathbb{M}_2=\sigma im^{-1}\mathbb{R}\mathbb{T}_1+\tau\mathbb{T}_2,\quad \mathbb{M}_3=\mathbb{R}, \tag{17}$$

obtaining (for $\sigma = 1$) 2×2 matrices of up to third order operators, generators of a group SO(2,1) with the well-known commutation relations

$$[\mathbf{M}_1,\mathbf{M}_2] = \sigma \mathbf{M}_3$$
$$[\mathbf{M}_2,\mathbf{M}_3] = -\sigma \mathbf{M}_1, \quad [\mathbf{M}_3,\mathbf{M}_1] = -\mathbf{M}_2 \qquad (18)$$

We also found the exponentiated action of (17) on $\mathscr{S}_o^H$, as a group of integral transforms on the two-component space function. The operators (17) for $\tau = \rho + i\sigma/2m, \rho \in \mathscr{R}$ are now skew-Hermitian under the inner product (15) or (16) for $\sigma = -1$ or 1, respectively. The corresponding group of integral transforms [5] will be isometric under the same inner product. The group representation obtained in this way belongs to Bargmann's continuous principal series[6] C_p^0 with $p = \frac{1}{4} + \rho^2 m^2, \rho \in \mathscr{R}$. The striking feature of this realization is that the kernel corresponding to the group unit element reduces, not to a Dirac δ as is usual for group realizations on $\mathscr{L}^2$ spaces, but to the reproducing kernel under (2). For the Helmholtz inner product (16), it is a diagonal matrix with elements $\sin(m[q-q'])/\pi(q-q')$.

Under the action of the group of integral transforms generated by (17), we unitarily map solutions of (1) into solutions of the same equation. This is another example of transformations more general than the Lie transformations.[7] In the classical theory, we recall, groups of transformations may be generated only by first-order differential operators.

It should be stressed that the realization of the SO(2,1) covering group given in this section remains to be studied in detail, on a par with the Bargmann realization[6] of this group on S_1 (local measure for the continuous principal series, non local measure for the continuous exceptional[6] and discrete[8] series) and the integral-transform realizations on $\mathscr{R}^+ : \mathscr{L}^2(\mathscr{R}^+)$ for the discrete series[9] and 2×2 matrices on the same space for the continuous series.[10]

[1] L. E. Payne, *Improperly Posed Problems in Partial Differential Equations* (SIAM, Philadelphia, 1975).
[2] G. W. Bluman and J. D. Cole, *Similarity Methods for Differential Equations*, Applied Mathematics Series, Vol. 13 (Springer, Berlin, 1974).
[3] J. D. Bjorken and S. D. Drell, *Relativisic Quantum Mechanics* (McGraw-Hill, New York, 1964), pp. 6 and 186; P. Roman, *Advanced Quantum Theory* (Addison-Wesley, Reading, Mass., 1964), pp. 118–119.
[4] W. Gordon, Z. Phys. **40**, 117 (1926); O. Klein, *ibid.* **41**, 407 (1927).
[5] S. Steinberg and K. B. Wolf, Nuovo Cimento A **53** 149 (1979), Sec. 5.
[6] V. Bargmann, Ann. Math. **48**, 568 (1974).
[7] V. I. Fushchich and A. G. Nikitin, J. Phys. A **12**, 747 (1979).
[8] C. P. Boyer and K. B. Wolf, J. Math. Phys. **16**, 1493 (1975).
[9] K. B. Wolf, J. Math. Phys. **15**, 2102 (1974).
[10] K. B. Wolf, J. Math. Phys. **21**, 680 (1980).

Representations of a local current algebra in nonsimply connected space and the Aharonov–Bohm effect [a]

G. A. Goldin
Department of Mathematical Sciences, Northern Illinois University, DeKalb, Illinois, 60115 and Theoretical Division, Los Alamos Scientific Laboratory, University of California, Los Alamos, New Mexico 87545

R. Menikoff and D. H. Sharp
Theoretical Division, Los Alamos Scientific Laboratory, University of California, Los Alamos, New Mexico 87545

(Received 9 December 1980; accepted for publication 20 February 1981)

A recent paper established technical conditions for the construction of a class of induced representations of the nonrelativistic current group $\mathscr{S} \wedge \mathscr{K}$, where $\mathscr{S}$ is Schwartz's space of rapidly decreasing C^∞ functions, and $\mathscr{K}$ is a group of C^∞ diffeomorphisms of $\mathbb{R}^s$. Bose and Fermi N-particle systems were recovered as unitarily inequivalent induced representations of the group by lifting the action of $\mathscr{K}$ on an orbit $\Delta \subseteq \mathscr{S}'$ to its universal covering space $\tilde{\Delta}$. For $s \geqslant 3$, $\tilde{\Delta}$ is the coordinate space for N particles, which is simply connected. In two-dimensional space, however, the coordinate space is multiply connected, implying induced representations other than those describing the usual Bose or Fermi statistics; these are explored in the present paper. Likewise the Aharonov–Bohm effect is described by means of induced representations of the local observables, defined in a nonsimply connected region of $\mathbb{R}^s$. The vector potential plays no role in this description of the Aharonov–Bohm effect.

PACS numbers: 03.65.Bz

I. INTRODUCTION

Nonrelativistic quantum mechanics can be described by means of the local operators $\rho(\mathbf{x})$, the number density of particles, and $\mathbf{J}(\mathbf{x})$, the particle flux. When integrated with test functions having components in Schwartz' space $\mathscr{S}$ (C^∞ functions of rapid decrease), these operators form a Lie algebra. We define $\rho(f) = \int \rho(\mathbf{x}) f(\mathbf{x})\, d\mathbf{x}$ and $J(\mathbf{g}) = \int \mathbf{J}(\mathbf{x}) \cdot \mathbf{g}(\mathbf{x})\, d\mathbf{x}$; then the commutation relations (at fixed time) become

$$[\rho(f_1), \rho(f_2)] = 0, \tag{1.1}$$

$$[\rho(f), J(\mathbf{g})] = i\rho(\mathbf{g}\cdot\nabla f), \tag{1.2}$$

$$[J(\mathbf{g}_1), J(\mathbf{g}_2)] = iJ([\mathbf{g}_1, \mathbf{g}_2]), \tag{1.3}$$

where $[\mathbf{g}_1, \mathbf{g}_2] = \mathbf{g}_2\cdot\nabla\mathbf{g}_1 - \mathbf{g}_1\cdot\nabla\mathbf{g}_2$ is the Lie bracket of the vector fields $\mathbf{g}_1$ and $\mathbf{g}_2$. Exponentiation of the current commutators leads to the consideration of continuous unitary representations of the semidirect product group $\mathscr{S} \wedge \mathscr{K}$; where $\mathscr{S}$ is Schwartz's space under addition, $\mathscr{K}$ is a group of diffeomorphisms of Euclidean space under composition, and the group law is given by $(f_1, \psi_1)\cdot(f_2, \psi_2) = (f_1 + f_2\circ\psi_1, \psi_2\circ\psi_1)$ for $f_1, f_2 \in \mathscr{S}$ and $\psi_1 \psi_2 \in \mathscr{K}$.[1–3]

The formalism of Gel'fand and Vilenkin[4] describes a representation of $\mathscr{S} \wedge \mathscr{K}$ by means of a measure μ in $\mathscr{S}'$ (the space of tempered distributions), quasi-invariant under the action of $\mathscr{K}$. For representations describing finitely many identical particles, μ is concentrated on a single orbit Δ in $\mathscr{S}'$. In a recent paper the authors established technical conditions permitting the construction of a class of induced representations for the case of a nonlocally compact group such as the diffeomorphism group.[5] Induced representations of $\mathscr{S} \wedge \mathscr{K}$ are obtained by lifting the action of $\mathscr{K}$ on an orbit Δ to its universal covering space $\tilde{\Delta}$. In this way Bose and Fermi N-particle representations are recovered as induced representations on the same orbit, and it appears that representations describing parastatistics are similarly obtained.[6] Thus the representations of $\mathscr{S} \wedge \mathscr{K}$ depend importantly on the connectedness (more specifically, the homotopy) of the orbit on which the measure is concentrated.

In three or more dimensions, the coordinate space for N particles is simply connected, even after removal of the set in which two particles have the same coordinates. In two-dimensional space, however, the coordinate space is multiply connected, leading to induced representations other than the usual Bose or Fermi representations. These are described in Sec. II of the present paper.

In Sec. III we apply our results to describe the Aharonov–Bohm effect[7] for a single boson or fermion exclusively in terms of observables. The reader who wishes to bypass the mathematical description of induced representations can proceed directly to this section. Excluding the particle from access to the region of nonvanishing magnetic field results in a nonsimply connected orbit, and consequently a one-parameter family of inequivalent induced representations of the current group. Our prescription leads to the choice of an irreducible representation which is equivalent to that obtained by requiring a phase shift λ in the wave function when the excluded region is circled once, where λ is proportional to the magnetic flux inside the excluded region *at the instant that the particle is excluded*. The vector potential for the magnetic field plays no role in this description of the Aharonov–Bohm effect.

Our conclusions are discussed in Sec. IV.

II. INDUCED REPRESENTATIONS OF THE CURRENT GROUP DESCRIBING PARTICLES IN TWO-DIMENSIONAL SPACE

For a representation of $\mathscr{S} \wedge \mathscr{K}$ describing N particles in s-dimensional, space, we have the $\mathscr{K}$-orbit

[a] Work supported by the U.S. Department of Energy.

 © 1981 American Institute of Physics

$$\Delta_N^{(s)} = \left\{ F \in \mathscr{S}' \middle| F = \sum_{j=1}^{N} F_{\mathbf{x}_j}, \quad \mathbf{x}_i \neq \mathbf{x}_j \ (\forall i \neq j) \right\},$$

where $\mathbf{x}_i \in \mathbb{R}^s$, $\mathscr{S}'$ is the space of continuous linear functionals on Schwartz's space $\mathscr{S}(\mathbb{R}^s)$, $\mathscr{K}$ is the diffeomorphism group of $\mathbb{R}^s$ obtained by exponentiation of the current commutators, and $F_{\mathbf{x}}$ denotes the evaluation functional $(F_{\mathbf{x}}, f) = f(\mathbf{x})$ for $f \in \mathscr{S}$. For $\psi \in \mathscr{K}$ we have ψ: $\mathscr{S}' \to \mathscr{S}'$ given by $(\psi^* F, f) = (F, f \circ \psi)$ in the notation of earlier papers.[2,3,5] Then

$$\psi^* \sum_{j=1}^{N} F_{\mathbf{x}_j} = \sum_{j=1}^{N} \psi^* F_{\mathbf{x}_j} = \sum_{j=1}^{N} F_{\psi(\mathbf{x}_j)}$$

establishes $\Delta_N^{(s)}$ as a $\mathscr{K}$-orbit. $\Delta_N^{(s)}$ may be identified with the *configuration space* $\Gamma_N^{(s)}$ consisting of all (unordered) sets of N distinct points in $\mathbb{R}^s$.

In Ref. 5 we showed how induced representations of $\mathscr{S} \wedge \mathscr{K}$ could be obtained by lifting the action of $\mathscr{K}$ on an orbit Δ to its universal covering space $\tilde{\Delta}$. For $s \geq 3$, the universal covering space of $\Delta_N^{(s)}$ is the *coordinate space* $\mathbb{R}^{sN} \setminus D$ consisting of ordered N-tuples of distinct coodinates in $\mathbb{R}^s$. Here D is the set $\{(\mathbf{x}_1, \cdots, \mathbf{x}_N) \in \mathbb{R}^{sN} | \mathbf{x}_i = \mathbf{x}_j \text{ for some } i \neq j\}$. The projection p: $\mathbb{R}^{sN} \setminus D \to \Delta_N^{(s)}$ is given by $p(\mathbf{x}_1, \ldots, \mathbf{x}_N) = F_{\mathbf{x}_1} + \cdots + F_{\mathbf{x}_N}$. The fundamental group $\pi(\Delta_N^{(s)}, F)$ for $F \in \Delta_N^{(s)}$ is just the symmetric group for N objects, S_N. Let $\mathscr{K}_F$ denote the stability group of F; i.e., $\mathscr{K}_F = \{\psi \in \mathscr{K} | \psi^* F = F\}$. Then there is a natural homomorphism from $\mathscr{K}_F$ to S_N. The two distinct one-dimensional representations of S_N induce representations of $\mathscr{S} \wedge \mathscr{K}$ corresponding to bosons (representation of S_N by $+1$) or fermions (representation of S_N by ± 1).

For $s = 2$, however, the covering space $\mathbb{R}^{2N} \setminus D$ is not the universal covering space of $\Delta_N^{(2)}$. This is easily seen by considering the case of two particles with coordinates $\mathbf{x}_1$, $\mathbf{x}_2 \in \mathbb{R}^2$. With $\mathbf{y}_1 = \frac{1}{\sqrt{2}}(\mathbf{x}_1 + \mathbf{x}_2)$ and $\mathbf{y}_2 = \frac{1}{\sqrt{2}}(\mathbf{x}_1 - \mathbf{x}_2)$, the condition $\mathbf{x}_1 \neq \mathbf{x}_2$ is the condition $\mathbf{y}_2 \neq (0,0)$, while $\mathbf{y}_1$ is unrestricted. Thus $\mathbb{R}^{2N} \setminus D$ is the space $\mathbb{R}^4$ without the two-dimensional subspace $\mathbf{y}_2 = (0,0)$. This space has the connectedness of the Euclidean plane without the origin—a closed loop shrinks to a point if and only if the loop in $\mathbb{R}^2$ defined by its $\mathbf{y}_2$-coordinates does not enclose the origin. Consequently S_N for $N = 2$ is not the fundamental group for $\Delta_2^{(2)}$, and there will be induced representations of $\mathscr{S} \wedge \mathscr{K}$ other than the Bose and Fermi representations.

For the case $N = 2$, denote the universal covering space of $\mathbb{R}^{2N} \setminus D$ by C. Then C is the product of $\mathbb{R}^2$ (the $\mathbf{y}_1$-coordinates) with a helical covering space H of $\mathbb{R}^2 \setminus \{(0,0)\}$. Writing $\mathbf{y}_2$ in polar coordinates (r, θ), with $0 \leq \theta < 2\pi$, an element of H may be written $\tilde{\mathbf{y}}_2 = (r, \tilde{\theta})$, with $-\infty < \tilde{\theta} < \infty$. The projection $\tilde{p}$ is given by $\tilde{p}(\mathbf{y}_1, r, \tilde{\theta}) = (\mathbf{y}_1, r, \theta)$ with $\theta \cong \tilde{\theta} \pmod{2\pi}$. Thus we have $C \xrightarrow{\tilde{p}} \mathbb{R}^4 \setminus D \xrightarrow{p} \Delta_2^{(2)}$. The fundamental group $\pi(\Delta_2^{(2)}, F_{\mathbf{x}_1} + F_{\mathbf{x}_2})$ is isomorphic to the additive group of integers $\mathbb{Z}$; the closed path γ based at the (unordered) pair $\{\mathbf{x}_1, \mathbf{x}_2\}$ in $\Gamma_2^{(2)}$ is associated with $n \in \mathbb{Z}$, where n is the (signed) number of half-turns of the vector $\mathbf{x}_1 - \mathbf{x}_2$ as the path is traversed. The fundamental group $\pi(\mathbb{R}^4/D, (\mathbf{x}_1, \mathbf{x}_2))$ is isomorphic to the additive group $\mathscr{E}$ of even integers. A one-dimensional representation of $\mathbb{Z}$ induces a representation of $\mathscr{S} \wedge \mathscr{K}$ corresponding to bosons or fermions only when the subgroup $\mathscr{E}$ is represented by unity. In that case we have a representation of $\mathbb{Z}/\mathscr{E} = \mathbb{Z}_2$ (the integers mod 2); i.e., the permutation group S_2.

Let us look at the representation of $\mathscr{S} \wedge \mathscr{K}$ induced by a general one-dimensional representation of $\mathbb{Z}$. Let $T_\lambda(n) = e^{i\lambda n}$ for $0 \leq \lambda < 2\pi$; thus $\lambda = 0$ corresponds to our Bose representation and $\lambda = \pi$ to our Fermi representation. The Hilbert space $\mathscr{H}$ for the induced representation is the space of measurable and square-integrable functions $\tilde{\Psi}(\mathbf{y}_1, r, \tilde{\theta})$ which transform in accordance with the representation T_λ of $\mathbb{Z}$—that is, $\tilde{\Psi}(\mathbf{y}_1, r, \tilde{\theta} + n\pi) = e^{i\lambda n}\tilde{\Psi}(\mathbf{y}_1, r, \tilde{\theta})$; while the product $\bar{\tilde{\Psi}} \cdot \tilde{\Psi}$ (independent of n and thus defining a function on the orbit $\Delta_2^{(2)}$) must be square integrable with respect to the normalized quasi-invariant measure μ on $\Delta_2^{(2)}$ (locally equivalent to Lebesgue measure).

In writing down the representation $U(f)V(\psi)$ of $\mathscr{S} \wedge \mathscr{K}$ in $\mathscr{H}$, we recall that if ψ_t is a continuous path connecting the identity $\mathbf{e}$ to ψ in $\mathscr{K}$, for $0 \leq t \leq 1$, then the path $\psi^*_t(F_{\mathbf{x}_1} + F_{\mathbf{x}_2})$ in $\Delta_2^{(2)}$ lifts to a unique path in C commencing at $(\mathbf{y}_1, r, \tilde{\theta})$, where $(\mathbf{y}_1, r, \tilde{\theta})$ is any point in C such that $(p \circ \tilde{p})(\mathbf{y}_1, r, \tilde{\theta}) = F_{\mathbf{x}_1} + F_{\mathbf{x}_2}$. We denote by $\psi^*(\mathbf{y}_1, r, \tilde{\theta})$ the terminal point of such a path in C. The fact that this terminal point is independent of the choice of ψ_t follows from Ref. 5 (Lemma 1, p. 657). Then the representation is given by

$$U(f)\tilde{\Psi}(\mathbf{y}_1, r, \tilde{\theta}) = e^{i(F_{\mathbf{x}_1} + F_{\mathbf{x}_2}, f)}\tilde{\Psi}(\mathbf{y}_1, r, \tilde{\theta}), \tag{2.1}$$

and

$$V(\psi)\tilde{\Psi}(\mathbf{y}_1, r, \tilde{\theta}) = \tilde{\Psi}(\psi^*(\mathbf{y}_1, r, \tilde{\theta})) \left[\frac{d\mu_\psi}{d\mu}(F_{\mathbf{x}_1} + F_{\mathbf{x}_2}) \right]^{1/2}, \tag{2.2}$$

where $F_{\mathbf{x}_1} + F_{\mathbf{x}_2} = (p \circ \tilde{p})(\mathbf{y}_1, r, \tilde{\theta})$, μ_ψ is the transformed measure on $\Delta_2^{(2)}$ given by $\mu_\psi(X) = \mu(\psi^* X)$ for a measurable set X, and $d\mu_\psi / d\mu$ is the Radon–Nikodym derivative.

Let us examine what the representation of Eqs. (2.1) and (2.2) implies about symmetry under the exchange of particle coordinates. Suppose that ψ either leaves fixed or exchanges the points $\mathbf{x}_1$ and $\mathbf{x}_2$ in $\mathbb{R}^2$. The path ψ_t from $\mathbf{e}$ to ψ allows us to keep track of the number of times $\mathbf{x}_1$ and $\mathbf{x}_2$ are "wound around" each other by ψ. This is not a well-defined quantity for $\mathbb{R}^s$ with $s > 2$, because then $\mathbb{R}^{sN} \setminus D$ is simply connected, and for $N = 2$, any path describing an exchange of $\mathbf{x}_1$ and $\mathbf{x}_2$ can be continuously deformed into any other. Now if ψ exchanges $\mathbf{x}_1$ and $\mathbf{x}_2$ in $\mathbb{R}^2$ by means of (let us say) a counterclockwise rotation by π, then $\psi^*(\mathbf{y}_1, r, \tilde{\theta}) = (\mathbf{y}_1, r, \tilde{\theta} + \pi)$, and $\tilde{\Psi}^*(\psi(\mathbf{y}_1, r, \tilde{\theta})) = e^{i\lambda}\tilde{\Psi}(\mathbf{y}_1, r, \tilde{\theta})$ where λ is not necessarily 0 or π. Thus in two-dimensional space, quantum mechanics permits species of particles which are neither bosons nor fermions.

In polar cordinates (r, θ), the angular momentum operator about the origin is $(1/i)(\partial/\partial\theta)$ (in appropriate physical units). Since θ is the angle describing $\frac{1}{\sqrt{2}}(\mathbf{x}_1 - \mathbf{x}_2)$ in polar cordinates, $(1/i)(\partial/\partial\theta)$ describes the center of mass angular momentum of our two-particle, two-dimensional system. Now in $L^2([0,\pi]), (1/i)(\partial/\partial\theta)$ may be defined on the domain of absolutely continuous functions Φ whose derivatives belong to L^2 and which are zero at 0 and π. This operator has deficiency indices (1,1), and hence has a one-parameter family of self-adjoint extensions, corresponding to domains of functions satisfying the boundary conditions $\Phi(\pi) = e^{i\lambda}\Phi(0)$, $0 \leqslant \lambda < 2\pi$ (Ref. 8). This establishes the condition required above on $\tilde{\Psi}(\tilde{\theta})$. {Because in our case we are considering indistinguishable particles $(1/i)(\partial/\partial\theta)$ is defined in $L^2([0,\pi])$ rather than in $L^2([0,2\pi])$, as it would be if the particles were distinguishable.} For each distinct choice of self-adjoint extension, a different and inequivalent representation of $\mathscr{S} \wedge \mathscr{K}$ is obtained. The infinitesimal generators of one-parameter unitary subgroups in each representation describe local particle densities and currents, and satisfy the same equal-time current algebra. In each representation the local operators for angular momentum and energy density may be written as functions of the local particle densities and currents.

For $0 \leqslant \lambda < 2\pi$, the eigenfunctions of $(1/i)(\partial/\partial\theta)$ are $\Psi_n(\theta) = e^{i(2n+\lambda/\pi)\theta}$ and $(1/i)(\partial/\partial\theta)\Psi_n = (2n + \lambda/\pi)\Psi_n$. So for $\lambda = 0$ (spinless bosons) we obtain an angular momentum spectrum of even integers, for $\lambda = 1$ (spinless fermions) a spectrum of odd integers, and for other values of λ a spectrum shifted from the even integers by λ/π. Similarly $-(\partial^2/\partial\theta^2)\Psi_n = (2n + \lambda/\pi)^2\Psi_n$ and the energy spectrum is also changed. So in two-dimensional space the quantum mechanics of local currents, as interpreted by means of induced representations of the local current group, permits species of particles with various local angular momentum and energy spectra which would not be permitted in higher-dimensional space. We shall observe that the mathematical framework used to describe the situation in two-dimensional space permits the interpretation of the Aharonov–Bohm effect in terms of the choice of representation of the local current algebra.

If the group $\mathscr{K}$ were enlarged to include global rotations of $\mathbb{R}^2$, then the condition needed for Lemma 1, Ref. 5, would not hold. That is, if ψ_t described a rotation by $2\pi t$ for $0 \leqslant t \leqslant 1$, we would have a closed loop in $\mathscr{K}$ for which $\psi_t(F_{x_1} + F_{x_2})$ could not be shrunk to a point in $\Delta_2^{(2)}$. Then $\psi^*(y_1, r, \tilde{\theta})$ would not be uniquely defined in C. In accordance with Ref.5 we must then identify the terminal points of all such paths, obtaining the covering space $\mathbb{R}^4 \setminus D$ instead of C. So one physical condition eliminating representations with $\lambda \neq 0$ or π (in the two- particle case) is the existence of a unitary representation of global rotations for which the total angular momentum operator is the infinitesimal generator. With three or more particles, however, even including global rotations does not eliminate the additional (non-Fermi and non-Bose) representations. For example, with three particles in two-dimensional space, one can still keep track of the number of times $\mathbf{x}_1$ "passes between" $\mathbf{x}_2$ and $\mathbf{x}_3$ given that $\{\psi(\mathbf{x}_1), \psi(\mathbf{x}_2), \psi(\mathbf{x}_3)\} = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$. The coordinate space is then $\mathbb{R}^6 \setminus D$, where

$D = \{(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) | \mathbf{x}_1 = \mathbf{x}_2\} \cup \{\mathbf{x}_2 = \mathbf{x}_3\} \cup \{\mathbf{x}_1 = \mathbf{x}_3\}$ is the union of three four-dimensional subspaces. The universal covering space is considerably more complicated than in the two-particle case, and there exist additional induced representations of $\mathscr{S} \wedge \mathscr{K}$.

In the next section we examine the Aharonov–Bohm effect from the viewpoint of induced representations of the local current group.

III. THE AHARONOV-BOHM EFFECT IN TERMS OF LOCAL CURRENTS

In this section we consider an infinite closed cylindrical region $\mathscr{Z}$ in $\mathbb{R}^3$, in which there is a magnetic field directed along the axis of the cylinder. Outside $\mathscr{Z}$ it is assumed that there is no magnetic field. The important feature of $\mathscr{Z}$ of course is that the region $\mathbb{R}^3 \setminus \mathscr{Z}$ is multiply connected. Thus, in the usual statement of the Aharonov-Bohm "paradox," there exist distinct choices for the magnetic vector potential $\mathbf{A}(\mathbf{x})$ which describes different magnetic fields $\mathbf{B}(\mathbf{x}) = \nabla \times \mathbf{A}(\mathbf{x})$ inside $\mathscr{Z}$, but which have vanishing curl outside $\mathscr{Z}$. If a charged particle is excluded from $\mathscr{Z}$ by means of a potential barrier, the usual quantum-mechanical equations of motion still have physically different solutions for different values of the magnetic flux inside $\mathscr{Z}$. This effect can be interpreted locally as caused by the supposedly "unobservable" field $\mathbf{A}$, or alternatively as a consequence of an inevitable overlap between the region of nonvanishing wave functions and the region of nonvanishing $\mathbf{B}$ field.[7,9] Here we shall apply the results of the previous section to determine the representation of the algebra of local currents describing this situation. Since the local currents form a complete set of observables, we obtain a resolution of the apparent paradox.

The equal-time current algebra describing charged particles in the presence of an external (c-number) magnetic field $\mathbf{B}(\mathbf{x})$ is, in appropriate units,

$$[\rho(f_1), \rho(f_2)] = 0. \tag{3.1}$$

$$[\rho(f), J(\mathbf{g})] = i\rho(\mathbf{g}\cdot\nabla f), \tag{3.2}$$

$$[J(\mathbf{g}_1), J(\mathbf{g}_2)] = iJ([\mathbf{g}_1, \mathbf{g}_2]) + i\rho(\mathbf{B}\cdot(\mathbf{g}_1 \times \mathbf{g}_2)), \tag{3.3}$$

as described in Ref. 10.

In taking account of the fact that the particle is excluded from $\mathscr{Z}$, we restrict attention to test functions with support in $\mathbb{R}^3 \setminus \mathscr{Z}$, so that $\mathbf{B}\cdot(\mathbf{g}_1 \times \mathbf{g}_2) = 0$ in Eq. (3.3). Thus it will be appropriate to consider representations of $\mathscr{S} \wedge \mathscr{K}$, where $\mathscr{S}$ is the additive group of Schwartz space functions which vanish in $\mathscr{Z}$, and $\mathscr{K}$ is the group of diffeomorphisms of $\mathbb{R}^3$ leaving points in $\mathscr{Z}$ fixed. For a single particle, such a representation is described by a measure concentrated on the orbit in $\mathscr{S}'$ given by $\Delta = \{F_\mathbf{x} | \mathbf{x} \in \mathbb{R}^3 \setminus \mathscr{Z}\}$. Since Δ is not simply connected, there exists a family of inequivalent induced representations of $\mathscr{S} \wedge \mathscr{K}$ analogous to that found in the preceding section for the two-particle orbit of the diffeomorphism group of $\mathbb{R}^2$. As before, the fundamental group $\pi(\Delta, F_\mathbf{x})$ is isomorphic to the additive group $\mathbb{Z}$ of integers, and representations of $\mathbb{Z}$ determine induced representations of $\mathscr{S} \wedge \mathscr{K}$. The inducing construction gives a Hilbert space of wave functions on the universal covering space $\tilde{\Delta}$, which is now parameterized by cylindrical coordinates (r, θ, z) with $-\infty < \theta < \infty$. The wave functions transform in accordance with a representation of $\mathbb{Z}$:

$$\Psi(r, \theta + 2\pi n, z) = e^{i\lambda n}\Psi(r, \theta, z) \tag{3.4}$$

To make the correct physical choice of λ, we consider the situation before the potential barrier excluding the particle from $\mathscr{T}$ has been erected. Then we have expressions for $\rho(f)$ and $J(\mathbf{g})$ as differential operators in $L^2(\mathbb{R}^3)$, without excluding any regions, satisfying Eqs. (3.1)–(3.3). These are

$$\rho(f)\Psi(\mathbf{x}) = f(\mathbf{x})\Psi(\mathbf{x}), \tag{3.5}$$

$$J(\mathbf{g})\Psi(\mathbf{x}) = \left\{\frac{1}{2i}[\mathbf{g}(\mathbf{x})\cdot\nabla + \nabla\cdot\mathbf{g}(\mathbf{x})] - \mathbf{g}(\mathbf{x})\cdot\int d\mathbf{y}\frac{(\operatorname{curl}\mathbf{B})(\mathbf{y})}{4\pi|\mathbf{x}-\mathbf{y}|}\right\}\Psi(\mathbf{x}). \tag{3.6}$$

Writing $\Psi = \Psi(r,\theta,z)$ and imposing the usual boundary condition $\Psi(r,0,z) = \Psi(r,2\pi,z)$ on functions in the domain of $J(\mathbf{g})$, we have a representation of Eqs. (3.1)–(3.3) by self-adjoint operators. The last term in Eq. (3.6) may be thought of as arising from the choice of the Coulomb gauge,

$$\mathbf{A}(\mathbf{x}) = \int d\mathbf{y}\frac{(\operatorname{curl}\mathbf{B})(\mathbf{y})}{4\pi|\mathbf{x}-\mathbf{y}|}.$$

However, any other choice of gauge leads to a representation satisfying Eqs. (3.1)–(3.3) that is unitarily equivalent to the representation given by Eqs. (3.5) and (3.6).

Next we ask which induced representation of $\mathscr{S}\wedge\mathscr{K}$ is determined by Eqs. (3.5) and (3.6) when the test functions are restricted to have support in $\mathbb{R}^3\setminus\mathscr{T}$. Define the unitary multiplication operator Q in $L^2(\mathbb{R}^3)$ by

$$(Q\Psi)(\mathbf{x})\begin{cases} = \Psi(\mathbf{x}), \quad \text{for } \mathbf{x}\in\mathscr{T} \\ = \exp\left[-i\int_\Gamma^{\mathbf{x}} d\mathbf{l}\cdot\int d\mathbf{y}\,\frac{(\operatorname{curl}\mathbf{B})(\mathbf{y})}{4\pi|\mathbf{y}'-\mathbf{y}|}\right]\Psi(\mathbf{x}), \\ \qquad\qquad \text{for } \mathbf{x}\in\mathbb{R}^3\setminus\mathscr{T}, \end{cases} \tag{3.7}$$

where $\mathbf{y}'$ moves along a path Γ from infinity to $\mathbf{x}$. In order to make $Q\Psi$ a well-defined function of $\mathbf{x}$, we specify a particular path for one value of $\mathbf{x}$, and let the path vary continuously with $\mathbf{x}$ without passing through the region $\mathscr{T}$. Now if $\mathbf{g}$ has support in $\mathbb{R}^3\setminus\mathscr{T}$ it is easy to see, using the fact that

$$\mathbf{B}(\mathbf{y}') = \nabla\times\int d\mathbf{y}\,\frac{(\operatorname{curl}\mathbf{B})(\mathbf{y})}{4\pi|\mathbf{y}'-\mathbf{y}|},$$

that $J'(\mathbf{g}) = QJ(\mathbf{g})Q^{-1}$ is a self-adjoint operator in $L^2(\mathbb{R}^3)$ represented by

$$J'(\mathbf{g})\Psi(\mathbf{x}) = \frac{1}{2i}\{\mathbf{g}(\mathbf{x})\cdot\nabla + \nabla\cdot\mathbf{g}(\mathbf{x})\}\Psi(\mathbf{x}). \tag{3.8}$$

The domain of $J'(\mathbf{g})$ consists of functions satisfying the boundary condition

$$\Psi(r,2\pi,z) = e^{-i\int\int\mathbf{B}(\mathbf{x})\cdot d\mathbf{S}}\Psi(r,0,z),$$

in $\mathbb{R}^3\setminus\mathscr{T}$.

The representation $J'(\mathbf{g})$ is easily seen to be a one-particle induced representation with $\lambda = -\int\int\mathbf{B}(\mathbf{x})\cdot d\mathbf{S}$, if we extend Ψ from $0\leqslant\theta\leqslant 2\pi$ to $-\infty<\theta<\infty$ by $\Psi(\theta+2\pi) = e^{i\lambda}\Psi(\theta)$.

Thus we see that in the situation where the magnetic field is confined to $\mathscr{T}$, but the particle is *not* excluded from $\mathscr{T}$, we have a unitary representation of a subgroup $\mathscr{S}\wedge\mathscr{K}$ of the full set of observables, obtained by restricting the support of the test functions to $\mathbb{R}^3\setminus\mathscr{T}$. Since this discussion has taken place under the assumption that the particle can penetrate $\mathscr{T}$, there is not yet any Aharonov–Bohm paradox (nor does $\mathscr{S}\wedge\mathscr{K}$ generate a complete set of observables). The existence of induced representations of the current group for which λ is not an integral multiple of 2π is a direct consequence of the nonsimple connectedness of $\mathbb{R}^3\setminus\mathscr{T}$. However the particular choice of λ determining the choice of representation can be regarded as a boundary condition imposed by the penetration of the particle into the region where $\mathbf{B}\neq 0$, in accordance with Ref. 9.

Thus a physical system in which $\mathbf{B}$ is confined to a cylindrical or toroidal region establishes an induced representation of $\mathscr{S}\wedge\mathscr{K}$ in the remaining region characterized by λ. If we then introduce a potential barrier, the value of λ remains constant for all values of the potential V and we have the statement that however large the barrier becomes, physical measurements outside it indicate the presence of the $\mathbf{B}$ field.

If we permit V to become *actually infinite*, the representation of $\mathscr{S}\wedge\mathscr{K}$ becomes a representation of the full set of local observables, and no longer merely a representation of a subgroup. Such a representation having been established, changes in $\mathbf{B}$ behind the infinite barrier can *no longer* effect a change to a unitarily inequivalent representation. There is no paradox; the outcomes of physical measurements simply depend on the representation of the current algebra, which in turn depends on the *history* of the system. Likewise there is no contradiction between the viewpoint that the effect should be described in terms of residual penetration of the barrier by the particle, and the viewpoint that the effect should be described in terms of the topology of the space outside the barrier.

Remarks:

1. Casati and Guarneri[11] have indicated that an examination of hydrodynamic variables (c-number currents) can explain the Aharonov-Bohm effect without ascribing extra physical significance to $\mathbf{A}$. In this section we have shown how a similar conclusion emerges naturally from the rigorous representation theory of local current operators.

2. The induced representation theory described here can be applied directly to the case of more complicated nonsimply connected spaces, with more than one cylinder or torus (possibly intertwined), or with knotted regions, within which there is a magnetic field and outside of which the field vanishes. The one-dimensional irreducible representations of the fundamental group will be described by means of an ordered set of parameters (λ_j) associated with windings about the respective excluded regions. These representations will induce representations of $\mathscr{S}\wedge\mathscr{K}$ describing the physical situation outside the region of nonvanishing $\mathbf{B}$.

In the case of intertwined or knotted regions, higher-dimensional irreducible representations of the fundamental group may induce representations of $\mathscr{S}\wedge\mathscr{K}$ in which there is no vector cyclic for the representation of $\mathscr{S}$ alone (i.e., for the particle density operators $\rho(f)$ alone). It appears that such higher-dimensional induced representations could not be prepared physically by confining a magnetic field to the knotted region and erecting a knotted potential barrier. Consequently we do not offer a physical interpretation of these representations at the present time.

IV. CONCLUSION

We have examined induced representations of the group obtained by exponentiating the infinite-dimensional

Lie algebra of local nonrelativistic currents. In particular these representations arise when the particles are restricted to move in two-dimensional space, or when the test functions are restricted to have support in a multiply connected region. In the latter case, the representation theory of local currents gives us a useful perspective on the Aharonov-Bohm effect.

ACKNOWLEDGMENT

It is a pleasure for the authors to thank Professor A. S. Wightman for his valuable comments on this work.

[1] R. F. Dashen and D. H. Sharp, Phys. Rev. **165**, 1857 (1968).
[2] G. A. Goldin, J. Math. Phys. **12**, 462 (1971).
[3] G. A. Goldin, J. Grodnik, R. T. Powers, and D. H. Sharp, J. Math Phys. **15**, 88 (1974).
[4] I. Gel'fand and N. Vilenkin, *Generalized Functions* (Academic, New York, 1964), Vol. 4.
[5] G. A. Goldin, R. Menikoff, and D. H. Sharp, J. Math. Phys. **21**, 650 (1980).
[6] A. A. Dicke and G. A. Goldin (in preparation).
[7] W. Ehrenburg and R. E. Siday, Proc. Phys. Soc. London, Sect. B **62**, 8 (1949); Y. Aharonov and D. Bohm, Phys. Rev. **115**, 485 (1959).
[8] F. Riesz and B. Sz. -Nagy, *Functional Analysis* (Ungar, New York, 1955), pp. 308–13.
[9] F. Strocchi and A. S. Wightman, J. Math. Phys. **15**, 2198 (1974).
[10] R. Menikoff and D. H. Sharp, J. Math. Phys. **18**, 471 (1977).
[11] G. Casati and I. Guarneri, Phys. Rev. Lett. **42**, 1579 (1979).

Hypervirial calculation of energy eigenvalues of a bounded centrally located harmonic oscillator

Francisco M. Fernández and Eduardo A. Castro
INIFTA, Sección Química Teórica, Suc. 4-C.C.16, La Plata 1900, Argentina

(Received 15 July 1980; accepted for publication 27 March 1981)

The diagonal hypervirial equations for enclosed quantum systems which obey boundary conditions $\phi(a)=\phi(b)=0$ are applied to calculate energy eigenvalues of a bounded centrally located harmonic oscillator. Hypervirial equations were previously derived by us [F. M. Fernández and E. A. Castro, Int. J. Quantum Chem. (in press)], and recurrence rules are easier to deal with than previous formulas based on the roots of the hypergeometric series. The comparison of numerical results with those given by Vawter [R. Vawter, J. Math. Phys. **14**, 1864 (1973)] shows the greater accuracy of the hypervirial method.

PACS numbers: 03.65.Db

I. INTRODUCTION

We have deduced the diagonal and off-diagonal hypervirial equations for enclosed quantum systems,

$$-\tfrac{1}{2}\phi''(x)+V(x)\phi(x)=E\phi(x),\quad \phi(a)=\phi(b)=0, \tag{1}$$

in a previous work.[1]

For the diagonal hypervirial theorem we have deduced the equations

$$\langle[H,f]\rangle=0, \tag{2}$$

$$\langle[H,fD]\rangle=-f(b)\frac{\partial E}{\partial b}-f(a)\frac{\partial E}{\partial a}, \tag{3}$$

where $f\equiv f(x)$ is a differentiable function and $D\equiv\partial/\partial x$. When the open interval (a,b) is symmetric with respect to the coordinates origin (i.e., $a=-b<0$) and $V(x)$ is an even function, Eq. (3) is simplified to

$$\langle[H,fD]\rangle=-f(b)\frac{\partial E}{\partial b}. \tag{4}$$

Replacing $f(x)$ by x^N in Eqs. (2) and (4), we can eliminate the terms D and D^2 according to the Swenson and Danforth method.[2] In particular, when the potential function have the general form

$$V(x)=cx^{2m},\quad m=1,2,\cdots, c\text{ constant}, \tag{5}$$

we obtain a recurrence relationship which relates average values of the coordinate powers with eigenvalues E and their derivatives with respect to b:

$$+\tfrac{1}{4}N(N-1)(N-2)A^{N-3}+2NEA^{N-1}-2(N+m)cA^{N+2m-1}=-b^N\frac{\partial E}{\partial b}, \tag{6}$$

where

$$A^N\equiv\langle x^N\rangle. \tag{7}$$

As a particular case, Eq. (6) yields the virial theorem[3,4] when $N=1$

$$2E-2(m+1)cA^{2m}=-b\frac{\partial E}{\partial b}. \tag{8}$$

From the expansion of E and A^N in power series of c,

$$E=\sum_{s=0}^{\infty}E^{(s)}c^s,\quad A^N=\sum_{s=0}^{\infty}A_s^Nc^s, \tag{9}$$

and from the Hellmann–Feynman theorem[5,6]

$$\frac{\partial E}{\partial c}=A^{2m}, \tag{10}$$

we can obtain a relation which determines $E^{(M)}$ as a function of b:

$$2(1-(m+1)M)E^{(M)}=-b\frac{\partial E^{(M)}}{\partial b}. \tag{11}$$

Eq. (11) assures us that a perturbational series for the energy can be written

$$E=\sum_{s=0}^{\infty}c^sk_sb^{2((m+1)s-1)}, \tag{12}$$

where

$$k_0(n)=E_n^{(0)}(b=1)=(n+1)^2\pi^2/8\equiv e_n^0, \tag{13}$$

$$k_M(n)=E_n^{(M)}(b=1)=A_{M-1}^{2m}(b=1)/M,\quad M>0. \tag{14}$$

The elimination of $\partial E/\partial b$ from (6) and (8), the expansion of E and A^N in power series of c, and the application of Eq. (10), allows us to deduce a set of equations which let the calculation of the whole set of matrix elements $(b=1)$ [1]:

$$A_0^N=\frac{1}{N+1}-\frac{N(N-1)}{8E^{(0)}}A_0^{N-2},\quad A_0^0=1, \tag{15}$$

$$A_M^N=\frac{(1-(m+1)M)}{M(N+1)E^{(0)}}A_{M-1}^{2m}-\frac{N(N-1)}{8E^{(0)}}A_M^{N-2}+\frac{(M+m+1)}{(N+1)E^{(0)}}A_{M-1}^{N+2m}-\frac{1}{E^{(0)}}\sum_{s=1}^{M}\frac{A_{s-1}^{2m}A_{M-s}^N}{s},$$
$$A_M^0=0,\quad M>0. \tag{16}$$

The equations corresponding to the harmonic oscillator model are gotten at once by setting $m=1$. The polynomial (12) adopts the followiong form $(c=\tfrac{1}{2})$:

$$E=\sum_{s=0}^{\infty}\frac{k_s}{2^s}b^{4s-2}. \tag{17}$$

Taking into account the asymptotic behavior of the eigenvalues $E_n(b)$ associated with the harmonic oscillator, and considering the limiting properties of coth$z$, Vawter[7] proposed the approximation of such eigenvalues by way of the formula

$$E_n(b)=(n+\tfrac{1}{2})\coth F(b^2), \tag{18}$$

 0022-2488/81/081669-03$01.00 © 1981 American Institute of Physics

TABLE I. Comparison of the exact energy eigenvalues of a bounded harmonic oscillator with other two approximate methods.

L	R_0^{exact}	$R_0(0,L)$	$R_0(1,L)$	$R_0(2,L)$	R_0(Vawter, Ref. 7)
1.0	9.90225	9.90336	9.90226	9.90226	9.90335
1.5	4.45979	4.46222	4.45979	4.45979	4.46221
2.0	2.59691	2.60104	2.59688	2.59692	2.60103
2.5	1.77893	1.78479	1.77879	1.77894	1.78478
3.0	1.37786	1.38497	1.37749	1.37788	1.38496
3.5	1.17497	1.18232	1.17425	1.175025	1.18232
4.0	1.07492	1.08133	1.07387	1.07506	1.08132
4.5	1.028293	1.03358	1.02776	1.02920	1.03358
5.0	1.00990	1.01269	1.00890	1.01028	1.01269
5.5	1.00297	1.00436	1.00233	1.00339	1.00436
6.0	1.00076	1.00136	1.000473	1.00115	1.00138

where

$$F(x) = \sum_{k=0}^{\infty} C_k x^{2k+1}. \tag{19}$$

The coefficients C_k are calculated immediately by expanding $\coth F(b^2)$ in the relation (18) in power series of b and replacing $E_n(b)$ by the polynomial (17). The formula for the first members of the set is

$$C_0 = \frac{n+\frac{1}{2}}{k_0},$$
$$C_1 = \frac{C_0^3}{3} - \frac{k_1 C_0^2}{n+\frac{1}{2}}, \tag{20}$$
$$C_2 = \frac{C_0^5}{5} - \frac{k_1 C_0^4}{n+\frac{1}{2}} - \frac{k_2 C_0^4}{n+\frac{1}{2}} + \frac{k_1^2 C_0^3}{n+\frac{1}{2}}.$$

Vawter[7] applied the roots of the confluent hypergeometric function to calculate the coefficients k_s, which constitutes a very long and tedious procedure, Instead of that we propose in this work the employment of Eqs. (15) and (16) (with $m = 1$). Our procedure has several advantages with respect to Vawter's method: (a) Eqs. (15) and (16) are easier to manage than the roots of the hypergeometric function, from an analytical as well as from a computational standpoint, and (b) the confluent hypergeometric function can be used only for the harmonic oscillator model, while Eqs. (15) and (16) are appropriate to calculate coefficients k_s of any potential function like (5).

II. CALCULATION AND RESULTS

Owing to the possibility of calculating just a finite number (s) of coefficients C_k,

$$F_s(x) = \sum_{k=0}^{s} C_k x^{2k+1}, \tag{21}$$

we can get the eigenvalues up to a certain degree of approximation

$$E_n(s,b) = (n+\tfrac{1}{2})\coth F_s(b^{-2}). \tag{22}$$

The utilization of Eqs. (15) and (16) (with $m = 1$) allows us to determine at once as many coefficients k_t as we need. The first three members of the set are

$$k_0(n) = e_n^0,$$
$$k_1(n) = \frac{1}{6} - \frac{1}{8e_n^0}, \tag{23}$$
$$k_2(n) = \frac{1}{180e_n^0} - \frac{5}{96(e_n^0)^2} + \frac{7}{128(e_n^0)^3}.$$

TABLE II. First eigenvalue for the enclosed harmonic oscillator R_0^* calculated from Eq. (24).

$s\backslash b$	0.5	$2^{-1/2}$	0.75	1.0	1.25	1.50
1	9.902277144	5.000147687	4.460004516	2.598092070	1.783341345	1.390677394
2	9.902258636	4.999999622	4.459793698	2.596907559	1.778822791	1.377185065
3	9.902258647	5.000000003	4.459794385	2.596919752	1.778936356	1.377888226
4	9.902258647	5.000000002	4.459794383	2.596919668	1.778934440	1.377863625
5	9.902258647	5.000000002	4.459794383	2.596919668	1.778934432	1.377863413
6	9.902258647	5.000000002	4.459794383	2.596919668	1.778934434	1.377863525
7	9.902258647	5.000000002	4.459794383	2.596919668	1.778934433	1.377863516
E_0^{exact}	9.90225	5.00000	4.45979	2.59691	1.77893	1.37786

$s\backslash b$	1.75	2.00	2.25	2.5
1	1.2059231	1.13961	1.1490	1.21
2	1.1719005	1.06381	0.9953	0.92
3	1.1751854	1.07629	1.0358	1.04
4	1.1749725	1.07491	1.0287	1.01
5	1.1749691	1.07487	1.0283	1.01
6	1.1749724	1.07494	1.0292	1.013
7	1.1749719	1.07492	1.0289	1.008
R_0^{exact}	1.17497	1.07492	1.20893	1.0099

Replacing (23) in formulae (20) we get the coefficients C_k $(k = 0,1,2)$, and with them the eigenvalues with three degrees of accuracy: $E_n(s,b)$, $s = 0,1,2$. In Table I we present the numerical values $R_0(s,L) = 2E_0(s,b)$ as a function of $L = 2b$, together with exact values (R_0^{exact}) reported by Vawter.[7] The data in the third column $[R_0(0,L)]$ correspond to that given by Vawter, and they are obtained by the sole inclusion of C_0 in the expansion (22). The two remaining columns show clearly the increase of accuracy when more coefficients are used.

Finally we wish to show in numerical way, the convergence of the series (17) by calculating the partial sums

$$R_n^s = \sum_{t=0}^{s} \frac{k_t}{2^t} b^{4t-2}. \tag{24}$$

For this purpose we have computed the coefficients k_t from Eqs. (13)–(16) up to $s = 7$. In Table II we display the comparison of eigenvalues R_0^s calculated from Eq. (24) with respect to those given by Vawter for different choices of $b = L/2$. For $b < 2$ our eigenvalues are the most accurate ones which have been calculated up to now. Especially, when $b = 2^{-1/2}$ the exact value of R_0^s is 5, which means that in this neighborhood, the hypervirial-perturbational scheme of calculation is extremely good (exact up to the eighth decimal place). Higher eigenvalues are calculated with a similar ease. According to the previous discussion, Eqs. (12)–(16) permit the determination of any potential with a general form (5). In particular, we have recently presented the results for a quartic oscillator.[1]

[1]F. M. Fernández and E. A. Castro, Int. J. Quantum Chem. (in press).
[2]R. Swenson and S. Danforth, J. Chem. Phys. **57**, 1734 (1972).
[3]T. L. Cottrell and S. Paterson, Philos. Mag. **42**, 391 (1951).
[4]W. Byers-Brown, J. Chem. Phys. **28**, 522 (1958).
[5]H. Hellmann, *Einführung in die Quantechemie* (Deuticke, Leipzig, 1937).
[6]R. P. Feynman, Phys. Rev. **56**, 340 (1939).
[7]R. Vawter, J. Math. Phys. **14**, 1864 (1973).

The mathematical structure of arrangement channel quantum mechanics

J. W. Evans
Ames Laboratory and Department of Chemistry, Iowa State University, Ames, Iowa 50011

(Received 5 February 1980; accepted for publication 5 January 1981)

A non-Hermitian matrix Hamiltonian **H** appears in the wavefunction form of a variety of many-body scattering theories. This operator acts on an arrangement channel Banach or Hilbert space $\mathscr{C} = \oplus_\alpha \mathscr{H}$ where $\mathscr{H}$ is the N-particle Hilbert space and α are certain arrangement channels. Various aspects of the spectral and semigroup theory for **H** are considered. The normalizable and weak (wavelike) eigenvectors of **H** are naturally characterized as either physical or spurious. Typically **H** is scalar spectral and "equivalent" to H on an **H**-invariant subspace of physical solutions. If the eigenvectors form a basis, by constructing a suitable biorthogonal system, we show that **H** is scalar spectral on $\mathscr{C}$. Other concepts including the channel space observables, trace class and trace, density matrix and Möller operators are developed. The sense in which the theory provides a "representation" of N-particle quantum mechanics and its equivalence to the usual Hilbert space theory is clarified.

PACS numbers: 03.65.Nk,03.65.Db

I. INTRODUCTION

In the theory of many-particle scattering, formal difficulties arise because the kernels of the standard Lippman–Schwinger scattering equations are badly behaved (Weinberg[1]) for $N > 2$ particles. This leads to problems with the solvability of both the T-matrix and wavefunction forms of these equations and with the uniqueness of scattering solutions of the latter (Redish[2]). The approaches adopted to combat this problem implement various forms of channel decomposition for the potentials, T-matrices, etc. In much of this work a channel is defined as partition of the particle labels into clusters corresponding to physically stable states together with a specification of the bound-state quantum numbers for these. However, for the arrangement channel theory discussed in detail here, the channels are defined as any partition of the particle labels (independent of stability of the clusters) are are denoted $\alpha, \beta, \cdots$.

Most of these scattering theories result in T-matrix scattering equations which have the generic form

$$T_{b,a} = B_{b,a} + \sum_c K_{b,c} T_{c,a}, \qquad (1.1)$$

where the $T_{b,a}$ are channel T-matrices, the $K_{b,c}$ are the components of the kernel, and the $B_{b,a}$ are inhomogeneous terms. The indices are either partition (i.e., channel) labels or chains of partitions (see Vanzani *et al.*[3]). Examples of the former are the Bencze–Redish–Sloan (BRS)[4] equations and the Baer–Kouri–Levin–Tobocman (BKLT)[5] equations. The latter include the Yakubovskii,[6] Alt–Grassberger–Sandhas (AGS),[7] and Vanzani[8] equations. The wave function form of the scattering equations for suitably defined components has a similar Fredholm structure.

The object of these methods is to obtain exact sets of coupled scattering equations having the property that some finite iterate of the kernel is "connected". Polyzou and Redish[9] have given the following definition of α connectivity for an operator B on the N-particle Hilbert space where α is a partition of the N-particle labels with n_α clusters. B has connectivity α if B commutes with the $3n_\alpha$ parameter unitary group of translations that describe the motion of the clusters of α; and these are the only translations with which B commutes. B is "connected" if it has connectivity $(12\cdots N)$, i.e., it only commutes with N-particle center of mass translations. One also may think of an α connected operator as one that vanishes (in the coordinate representation) as any particles in a single cluster of α are asymptotically separated. A technical description of this condition is given by Polyzou[9] and uses the strong operator topology. There is also a diagrammatic description of this property in terms of graph connectivity (Redish[2]). Connectivity of an iterate of the kernel of some coupled scattering equations then means that each component of that iterate is a connected operator in the sense described above.

Clearly, the homogeneous form of a connected kernel equation for the wavefunction components can not possess any wavelike solutions (as demonstrated by iteration of this equation), so uniqueness of scattering solutions is assured. The standard operating philosophy in this work is that connectivity for some iterate of the kernel leads to compactness for this or a higher iterate. This is termed the Fiber Compactness Assumption (FCA) by Polyzou and Redish[9] and from the extended form of Fredholm or Riesz–Schauder theory (Yosida[10]), we know that it guarantees solvability of the corresponding scattering equations (rigorously, at least for the T-matrix equations). The discussion here is somewhat oversimplified since in general it is only reasonable to assume the FCA for points of analyticity of the kernel.[9] On the scattering cut, the operators may no longer be compact but unbounded solutions of the scattering equations still exist in general. T-matrices only make sense when applied to a dense set of initial states. So an analytic Fredholm theory is constructed by choosing a Banach space norm to exclude other states. A proof of the modified form of the FCA has been given for a general class of "relatively bounded" potentials for the three-particle, two-cluster Faddeev equations.[11]

[a]Operated for the U. S. Department of Energy by Iowa State University under contract No. W-7405-Eng-82. This research was supported by the Director for Energy Research, Office of Basic Energy Science, WPAS-KC-01-03-01-2.

 © 1981 American Institute of Physics

However, no proof exists in general except for more restricted Rollnik classes of potentials (Reed and Simon[12]).

Vanzani[13] has considered in detail the relationships between some of the different types of scattering equations. This analysis leads to an understanding of the origin of spurious solutions of the homogeneous form of (1.1) or the corresponding wavefunction equations. These are solutions which do not correspond to any physical solution of the N-particle Schrödinger equation. If no such solutions exist, then the equations are said to satisfy a constrained Fredholm alternative.

A somewhat different approach has been adopted by Chandler and Gibson[14] using two-Hilbert space theory. Unlike other approaches, their more recent work[15] does not rely on the FCA but demonstrates rigorously that the kernel of their scattering equations is A-solvable (Petryshyn[16]). This constraint is weaker than the FCA, but sufficient to guarantee solvability of the scattering equations. For a more detailed review of all these aspects of multiparticle scattering, we refer the reader to Kowalski's[17] article.

Some of the above approaches lead to an unusual "representation" of N-particle quantum mechanics characterized by a non-Hermitian channel space Hamiltonian **H**.[18,19] This operator acts on a channel space which is the direct sum (over certain channels α) of copies of the N-particle Hilbert space. Aspects of these time dependent and time independent wavefunction theories termed "arrangement channel quantum mechanics" were analyzed by Kouri, Küger, and Levin[20] for the BKLT choices of **H**. The Faddeev[11] equations, a transposed form of the BRS[21] equations and certain hybrid schemes[22] also fall into this category. One way to view these theories is that they exploit the extra degree of flexibility in the non-Hermitian choice of the matrix **H** to achieve connected kernel scattering equations. The theory has not only been applied to the determination of multichannel S and T matrices, but also to bound-state calculations.[23] Hoffman, Kouri, and Top[24] have shown how to define a channel space density matrix and have derived the corresponding von Neumann equation and BBGKY hierarchy.

These theories are not in general expected to satisfy the constrained Fredholm alternative (except for the Faddeev case[11]). Chandler[21] has demonstrated the possible existence of normalizable spurious solutions of the three-particle, two-cluster BKLT equations by first considering a limiting singular choice of potentials where the equations become algebraic and then using Rouche's theorem to extend the result to a neighboring class of potentials. Glöckle and Adhikari[25] have further elucidated the origin of these solutions. A discussion for the BRS equations is given by Vanzani.[13] These normalizable spurious solutions will manifest themselves as eigenvectors of **H** (not corresponding to physical Schrödinger equation solutions) with possibly complex eigenvalues.

In Sec. II, some observations are made immediately on the relationship of the arrangement channel to the normal quantum theory, particularly, regarding the imbedding of physical solutions and the nature of spurious solutions. The latter are defined somewhat more generally here to include wavelike solutions and these play a useful role in analyzing the structure of **H**. For a rigorous treatment, we choose a topology on the arrangement channel space (given by a Banach space norm) which is naturally induced by that of the original N-particle Hilbert space. Certain results for the non-Hermitian **H** follow directly from spectral and semigroup theory. General channel space observables are discussed.

Unlike other treatments, we consider first the arrangement channel theory for the spatially confined system in Sec. III. This is useful for applications to the statistical mechanics of reactive systems[24,26] as well as for the introduction of various mathematical concepts. Typically **H** is real eigenvalue scalar spectral (*-self-adjoint) and "equivalent" to H on an **H**-invariant subspace of physical solutions (with a mild technical assumption). If the eigenvectors form a basis, by constructing a biorthogonal system from these and their duals, we show that **H** is scalar spectral on $\mathscr{C}$ (with a simple functional calculus). Other concepts such as channel space projection operators, trace class, trace, and density matrices are developed in a Banach space framework. The sense in which the theory provides a "representation" of N-particle quantum mechanics and its equivalance to the usual Hilbert space theory is clarified.

These ideas are extended in Sec. IV to spatially infinite systems where the wavelike scattering solutions must be considered. They are placed on a firmer mathematical basis by the introduction of the appropriate generalization of the Gel'fand triplet. In Sec. V some remarks are made on time-dependent scattering theory and statistical mechanics. In particular, we discuss the existence of channel space Möller operators and certain trace operator topologies.

II. THE MATHEMATICAL FORMULATION

We begin with the Schrödinger equation for a system of N distinguishable particles

$$(\lambda - H)|\Psi\rangle = 0 \text{ (time-independent)}, \tag{2.1}$$

$$\left(i\hbar\frac{\partial}{\partial t} - H\right)|\Psi\rangle = 0 \text{ (time dependent)}, \tag{2.2}$$

where the center of mass motion is removed from the Hamiltonian H for spatially infinite systems. As previously, the channels (partitions of the particle labels) are denoted by α, β, $\cdots$. The most cruicial step in the development is the decomposition of the Hamiltonian H into channel components $H_{\alpha\beta}$ such that[24]

$$\sum_{\alpha} H_{\alpha\beta} = H, \quad \forall\beta \tag{2.3}$$

(where α,β belong to the subset of channels of interest). The corresponding matrix structured operator denoted by **H** is called the channel space Hamiltonian (*so* $[\mathbf{H}]_{\alpha\beta} = H_{\alpha\beta}$). A vector in the channel space upon which this operator acts is denoted by $\underline{\Psi}$ with components $|\Psi_\alpha\rangle$ (*so* $[\underline{\Psi}]_\alpha = |\Psi_\alpha\rangle$).

The standard structure of **H** for a spatially infinite system with no external potential is described below. First decompose $H = K + V$, where K is the kinetic energy and V the interparticle potentials. Let V_α denote the sum of the interparticle potentials between particles in the same α clusters so $H_\alpha = K + V_\alpha$ is the corresponding channel Hamiltonian.[17] The sum of residual interactions between particles in different clusters is denoted by $V^\alpha = V - V_\alpha$, so $H = H_\alpha + V^\alpha$ for all α. Typically $\mathbf{H}$ is chosen so that $H_{\alpha\alpha} = H_\alpha$ for all α. For this class of decompositions, we write

$$\mathbf{H} = \mathbf{H}_0 + \mathbf{V}, \tag{2.4}$$

where $[\mathbf{H}_0]_{\alpha\beta} = \delta_{\alpha\beta} H_\beta$ and $\mathbf{V}$ is off-diagonal. For the cases mentioned in the introduction, the $\mathbf{V}$ is chosen with the potentials in V^α distributed between $[\mathbf{V}]_{\beta\alpha}$ for different $\beta \neq \alpha$, so that each component of $\mathbf{V}^n$ [or $(\mathbf{G}_0(z)\mathbf{V})^n$ where $\mathbf{G}_0(z) = (z - \mathbf{H}_0)^{-1}$] is connected[9] for some n. Such choices of $\mathbf{V}$ (or $\mathbf{H}$) are usually referred to as "connected".

For the Faddeev choice $\alpha, \beta, \cdots$ are chosen from (12)(3),(13)(2), and (1)(23), and $\mathbf{V}$, for pairwise interactions is given by

$$[\mathbf{V}]_{\alpha\beta} = (1 - \delta_{\alpha,\beta})V_\alpha. \tag{2.5}$$

The potentials of the three-particle problems have the property that $\Sigma_\alpha V_\alpha = V$; hence, (2.3) is satisfied by the decomposition (2.5). The BKLT choice is given by

$$[\mathbf{V}]_{\alpha\beta} = W_{\alpha\beta} V^\beta, \tag{2.6}$$

where $W_{\alpha\beta}$ is any of $(N_{\text{ch}} - 1)!$ channel permuting arrays[5] corresponding to single cycle permutations of the N_{ch} channel indices Thus (2-3) is satisfied. The transposed BRS[21] choice

$$[\mathbf{V}]_{\alpha\beta} = (-1)^{n_\alpha}(n_\alpha - 1)! V^\beta_\alpha \tag{2.7}$$

is more complicated, though it still satisfies (2.3). Here n_α is the number of clusters in channel α, V^β_α consists of those potentials in H_α and not in H_β, and α,β range over all channels. The $N = 3$ case reduces to the Faddeev equations after setting the (1)(2)(3) channel wavefunction component identically zero.

The channel space Schrödinger equation may now be written down as[20]

$$(\lambda - \mathbf{H})\underline{\Psi} = \underline{0} \text{ (time-independent)} \tag{2.8}$$

$$\left(i\hbar \frac{\partial}{\partial t} - \mathbf{H}\right)\underline{\Psi} = \underline{0} \text{ (time-dependent).} \tag{2.9}$$

If $\underline{\Psi}$ satisfies (2.8) [resp. (2.9)], then, summing over components and using (2.3), it follows that[24]

$$\sum_\alpha |\Psi_\alpha\rangle = |\Psi\rangle \tag{2.10}$$

either satisfies (2.1) [resp. (2.2)] or $|\Psi\rangle = 0$. We return to this point later.

For the formulation in terms of channel space to be useful, it is necessary for the channel space Schrödinger equation to exhibit at least some properties characteristic of the associated Hilbert space equation. The possibility of an imbedding of certain solutions of (2.1) into those of (2.8) is certainly suggested by (2.10), and in previous work has been assumed at least for scattering solutions with two cluster asymptotic conditions (and sometimes more generally). An imbedding of any scattering solution $|\Psi^\pm_\beta\rangle$ of energy E with channel β asymptotic clustering $|\phi_\beta\rangle$ [where β consists of stable clusters and is contained in the decomposition (2-3)] is obtained from the solution to the equation [cf., (1.1)]

$$\underline{\Psi}^\pm_\beta = \underline{\phi}_\beta + \mathbf{G}_0^\pm(E)\mathbf{V}\underline{\Psi}^\pm_\beta \tag{2.11}$$

(should it exist). The asymptotic form $\underline{\phi}_\beta$ of this solution satisfies

$$[\underline{\phi}_\beta]_\alpha = \delta_{\alpha,\beta}|\phi_\beta\rangle, \quad (E - \mathbf{H}_0)\underline{\phi}_\beta = \underline{0}. \tag{2.12}$$

Here $\mathbf{G}_0(\lambda) = (\lambda\mathbf{I} - \mathbf{H}_0)^{-1}$, so $[\mathbf{G}_0(\lambda)]_{\alpha\beta} = \delta_{\alpha,\beta}G_\beta(\lambda)$, where $G_\beta(\lambda) = (\lambda - H_\beta)^{-1}$. Also $\mathbf{G}_0^\pm(E) = \text{"lim"}_{\epsilon\to 0}\mathbf{G}_0(E \pm i\epsilon)$ and $+(-)$ denotes a choice of pre- (post-) collisional asymptotic condition. The association of channel component with particle clustering is also clear.

Let us now develop the mathematical concepts appropriate to the theory of arrangement channel quantum mechanics. We impose a rigorous mathematical structure in the following way. Suppose that the wavefunctions Ψ may be regarded as elements of a separable Hilbert space $\mathscr{H}$. In the usual Dirac notation, the elements of $\mathscr{H}$ are denoted by the kets $|\Psi\rangle$. The Hamiltonians H and H_α are taken as unbounded self-adjoint operators (Kato[27]) which are assumed asymptotically complete with appropriate eigenfunction expansions. As a linear vector space, the channel space $\mathscr{C}$ has a decomposition of the form

$$\mathscr{C} = \bigoplus_\alpha \mathscr{H} \tag{2.13}$$

where the direct sum is over channels of interest, e.g., $\mathscr{H} \oplus \mathscr{H} \oplus \mathscr{H}$ for Faddeev. An element of this space is denoted by $\underline{\Psi}$ with components

$$[\underline{\Psi}]_\alpha = |\Psi_\alpha\rangle \in \mathscr{H}. \tag{2.14}$$

The ℓ^p-type norms may be defined in the following way. Let $\langle\ |\ \rangle$ denote the inner product on $\mathscr{H}$ and let $\|\text{—}\|_{\mathscr{H}}$ denote the corresponding norm, so

$$\|\,|\Psi\rangle\|^2_{\mathscr{H}} = \langle\Psi|\Psi\rangle. \tag{2.15}$$

Then for $1 \leqslant p < \infty$, define

$$\|\underline{\Psi}\|_p = \Big(\sum_\alpha \|\,|\Psi_\alpha\rangle\|^p_{\mathscr{H}}\Big)^{1/p}. \tag{2.16}$$

Using the completeness of $\|\text{—}\|_{\mathscr{H}}$ and the finiteness of the number of channels, it is immediate that each of the norms (2.16) is complete on $\mathscr{C}$. So the space $\mathscr{C}_p = (\mathscr{C}, \|\text{—}\|_p)$ of channel vectors with finite $\|\text{—}\|_p$ norm is a Banach space under the $\|\text{—}\|_p$ norm. For the case $p = 2$, $\mathscr{C}_2$ is also a Hilbert space with inner product

$$\langle\underline{\Psi}, \underline{\phi}\rangle = \sum_\alpha \langle\Psi_\alpha|\phi_\alpha\rangle. \tag{2.17}$$

The finiteness of the number of channels is not essential to these results. The proof of uniform equivalence of norms on a finite dimensional space may be adapted here to show that the norms $\|\text{—}\|_p$ are uniformly equivalent. So they induce the same topology $\mathscr{C}$ and we may write

$$\mathscr{C}_r = \mathscr{C}_s \text{ for } 1 \leqslant r,s < \infty \tag{2.18}$$

in the sense that they contain the same elements. Finally, we note that the separability of $\mathscr{H}$ implies that $\mathscr{C}_p$ is separable.

The channel space Hamiltonian $\mathbf{H}$ is now regarded as an unbounded operator with domain dense in $\mathscr{C}_p$. However, for the cases described $\mathbf{H}$ is not self-adjoint or even normal on $\mathscr{C}_2$. Let us write $\mathbf{H} = \tilde{\mathbf{H}}_0 + \tilde{\mathbf{V}}$, where $(\tilde{\mathbf{H}}_0)_{\alpha\beta} = \delta_{\alpha\beta} K$. Then, if the components of $\tilde{\mathbf{V}}$ are bounded operators on $\mathscr{H}$, it follows that since $\mathbf{H}_0$ is closed, so is $\mathbf{H}$. For various applications this condition on $\tilde{\mathbf{V}}$ is too severe; however, we may usually assume that $\tilde{\mathbf{V}}$ is $\tilde{\mathbf{H}}_0$-bounded with relative bound <1:

$$\|\tilde{\mathbf{V}}\underline{\Psi}\|_p \leqslant a\|\tilde{\mathbf{H}}_0\underline{\Psi}\|_p + b\|\underline{\Psi}\|_p \quad (2.19)$$

and least upper bound $(a) < 1$,

where a and b may be chosen independent of the value of p. It is easily verified from Kato's[27] work that (2.19) is valid for a large class of unbounded pairwise and external potentials, e.g., L^2_{loc} and bounded at infinity in which case the $\tilde{\mathbf{H}}_0$ bound is zero (but b becomes larger as a $\rightarrow 0$). That $\mathbf{H}$ is closed now follows from a standard stability theorem for closed operators under relatively bounded perturbations.[27]

Since the solutions to the dual of the channel space Schrödinger equation are important for the functional calculus of $\mathbf{H}$ and for the statistical mechanics, the appropriate mathemetical concepts are developed here. The dual space $\mathscr{C}'_p$ of $\mathscr{C}_p$ is the space of bounded linear functionals acting on $\mathscr{C}_p$. We adopt the convenient representation of elements of $\mathscr{C}'_p$ by vectors $\underline{\zeta}'$ with components $\langle\zeta_\alpha| \in \mathscr{H}'$. The action of $\underline{\zeta}' \in \mathscr{C}'_p$ on $\underline{\Psi} \in \mathscr{C}_p$ is given by

$$(\underline{\zeta}', \underline{\Psi}) = \sum_\alpha \langle\zeta_\alpha|\Psi_\alpha\rangle \in \mathbb{C}. \quad (2.20)$$

The norm (and thus topology) of the dual $\mathscr{C}'_p$ is induced by that on $\mathscr{C}_p$. For $1<p<+\infty$, the Banach space $\mathscr{C}'_p$ is associated with the ℓ^q-type norm

$$\|\underline{\zeta}'\|_q = \sup_{\|\underline{\Psi}\|_p=1} |(\underline{\zeta}', \underline{\Psi})| = \Big(\sum_\alpha \|\langle\zeta_\alpha|\,\|_{\mathscr{H}'}{}^q\Big)^{1/q}, \quad (2.21)$$

where $1/p + 1/q = 1$, and for $p = 1$, the Banach space $\mathscr{C}'_1$ is associated with the ℓ^∞-type norm

$$\|\underline{\zeta}'\|_\infty = \sup_{\|\underline{\Psi}\|_p=1} |(\underline{\zeta}', \underline{\Psi})| = \sup_\alpha(\|\langle\zeta_\alpha|\,\|_{\mathscr{H}'}). \quad (2.22)$$

The spaces $\mathscr{C}'_p$ are all equal in the sense that they contain the same elements.

For an operator $\mathbf{A}$ acting on $\mathscr{C}_p$ with domain dom($\mathbf{A}$) dense in $\mathscr{C}_p$, we define the Banach space dual $\mathbf{A}'$ as follows (Yosida[10]). If

$$(\underline{\zeta}', \mathbf{A}\underline{\Psi}) = (\underline{\eta}', \underline{\Psi}), \quad \underline{\Psi} \in \mathrm{dom}(\mathbf{A}), \quad (2.23)$$

then $\underline{\eta}'$ is determined uniquely by $\underline{\zeta}'$, and we define

$$\mathbf{A}' \quad \text{on} \quad \mathscr{C}'_p \quad \text{by} \quad \underline{\eta}' = \mathbf{A}'\underline{\zeta}'. \quad (2.24)$$

$\mathbf{A}'$ is defined on dom($\mathbf{A}'$), the totality of $\underline{\zeta}' \in \mathscr{C}'_p$ such that there exists $\underline{\eta}'$ satisfying (2.23). For the case $p = 2$, $\mathscr{C}_2$ is a Hilbert space and $\mathbf{A}'$ defined as above is simply related to the usual Hilbert space adjoint of $\mathbf{A}$.

We previously elucidated the connection between solutions $\underline{\Psi}$ of (2.8) and those of the Hilbert space equation (2.1). We call those solutions for which $|\Psi\rangle = \Sigma_\alpha|\Psi_\alpha\rangle \neq 0$ "physical" solutions since $|\Psi\rangle$ is a solution of (2.1), also with eigenvalue λ ($= E$)$\in R$. Those for which $\Sigma_\alpha|\Psi_\alpha\rangle = 0$ are called "spurious". For these λ need not be real[21] as was first discovered by Federbusch and others[28] for a different class of scattering equations. This classification applies to all solutions normalizable and unnormalizable. If we denote the class of channel vectors with components summing to zero in $\mathscr{H}$ by $\mathscr{S}$, then clearly $\mathscr{S}$ is a subspace of $\mathscr{C}_p$ containing the spurious solutions. Any two physical solutions, the difference of which lies in $\mathscr{S}$, are necessarily degenerate and may be regarded physically as the same. Thus they may be replaced by linear combinations, one of which is physical and one spurious. In this way the physical eigenvectors may be chosen as "distinct." Note that the quotient space $\mathscr{C}_p/\mathscr{S} \cong \mathscr{H}$.

It is possible to state some very general results about imbedding of Hilbert space into channel space solutions and completeness of physical and spurious solutions. By completeness of a set of vectors which may be normalizable (in $\mathscr{C}_p$) and/or unnormalizable, we mean that any vector in $\mathscr{C}_p$ can be approximated in norm as a possibly partly continuous linear combination of these. Note that for nonorthogonal sets, there exists a distinction between completeness and the basis property. The latter is stronger requiring that any vector can be represented as a unique, possibly partly continuous linear combination of basis vectors (convergence in norm implied).

Theorem 1: If the physical and spurious solutions are complete in $\mathscr{C}_p$, then it is necessary for all (strictly almost all) solutions of the Hilbert space equation (2.1) to be imbedded into physical solutions of (2.8).

Proof: The spurious solutions can at most span $\mathscr{S}$ and the physical solutions lie outside $\mathscr{S}$. Now suppose that we have a complete set of physical $\underline{\Psi}^{\mathrm{p}}_k$ and spurious $\underline{\Psi}^{\mathrm{s}}_{k'}$ eigenvectors of $\mathbf{H}$ labeled by k,k' (possibly partly continuous). Then these may be chosen to be distinct, e.g., $\Sigma_\alpha[\underline{\Psi}^{\mathrm{p}}_k]_\alpha \neq \Sigma_\alpha[\underline{\Psi}^{\mathrm{p}}_{k'}]_\alpha$ for $k \neq k'$. Set $\mathscr{P} = \mathrm{span}\{\underline{\Psi}^{\mathrm{p}}_k\}$ so $\mathscr{C}_p = \overline{\mathscr{P} \oplus \mathscr{S}}$ by hypothesis. Suppose that some H eigenvector with discrete eigenvalue and/or H eigenvectors with continuous eigenvalue for a range of k of nonzero measure are not imbedded into the set of eigenvectors of H. Let $|\Psi_0\rangle \neq 0$ be in the subspace spanned by these vectors and let $|\Psi^{\mathrm{p}}\rangle$ be a linear combination of the other (imbedded) eigenvectors of H, then from orthogonality

$$0 < \|\,|\Psi_0\rangle\|_{\mathscr{H}} \leqslant \|\,|\Psi_0\rangle - |\Psi^{\mathrm{p}}\rangle\|_{\mathscr{H}}. \quad (2.25)$$

Consequently, if $\underline{\Psi}_0$ is any channel vector such that $\Sigma_\alpha[\underline{\Psi}_0]_\alpha = |\Psi_0\rangle$, $\underline{\Psi}^{\mathrm{p}}$ is a linear combination of the $\underline{\Psi}^{\mathrm{p}}_k$ (so $\Sigma_\alpha[\underline{\Psi}^{\mathrm{p}}]_\alpha = |\Psi^{\mathrm{p}}\rangle$, a vector described above) and $\underline{\Psi}^{\mathrm{s}} \in \mathscr{S}$, then

$$0 < \|\,|\Psi_0\rangle\|_{\mathscr{H}} \leqslant \|\,|\Psi_0\rangle - |\Psi^{\mathrm{p}}\rangle\|_{\mathscr{H}} \leqslant \|\underline{\Psi}_0 - (\underline{\Psi}^{\mathrm{p}} + \underline{\Psi}^{\mathrm{s}})\|_1. \quad (2.26)$$

This contradicts the hypothesis that $\mathscr{C}_p = \overline{\mathscr{P} \oplus \mathscr{S}}$.

There is also a partial converse to this result.

Theorem 2: Suppose that all the eigenvectors of H are imbedded into physical eigenvectors $\underline{\Psi}^{\mathrm{p}}_k$ of $\mathbf{H}$ necessarily in a one to one fashion. Since the eigenvectors of H are complete, any vector in $\mathscr{H}$ can be approximated in norm by a certain class of linear combinations of these. Suppose firstly that the corresponding linear combinations of the $\underline{\Psi}^{\mathrm{p}}_k$ are convergent in $\mathscr{C}_p$ (trivial where all eigenvalues are discrete), then we show that $\overline{\mathscr{P} \oplus \mathscr{S}} = \mathscr{C}_p$.

Proof: Take any $\underline{\Psi} \in \mathscr{C}_p$, $\Sigma_\alpha[\underline{\Psi}]_\alpha$ may be approximated in $\mathscr{H}$ by a linear combination of eigenvectors of H. Let $\underline{\Psi}^{\mathrm{p}}$

be the corresponding linear combination of eigenvectors of $\mathbf{H}$ (convergent by assumption), so

$$\sum_\alpha [\underline{\Psi}]_\alpha - \sum_\alpha [\underline{\Psi}^{\rm p}]_\alpha = |\Psi^\epsilon\rangle, \quad \text{where } \| \, |\Psi^\epsilon\rangle \|_{\mathscr{H}} < \epsilon. \tag{2.27}$$

We may therefore write

$$\underline{\Psi} - \underline{\Psi}^{\rm p} = \begin{pmatrix} |\Psi^\epsilon\rangle \\ 0 \\ 0 \\ \vdots \end{pmatrix} + \underline{\Psi}^{\rm s} \quad \text{for some } \underline{\Psi}^{\rm s} \in \mathscr{S},$$

so

$$\|\underline{\Psi} - (\underline{\Psi}^{\rm p} + \underline{\Psi}^{\rm s})\|_p = \left\| \begin{pmatrix} |\Psi^\epsilon\rangle \\ 0 \\ \vdots \end{pmatrix} \right\|_p < \epsilon \tag{2.28}$$

as required. From (2.28) it is clear that $\underline{\Psi}_k^{\rm p}$ together with any complete set in $\mathscr{S}$ is complete in $\mathscr{C}_p$.

Theorem 3: Suppose that for any convergent linear combination of eigenvectors of H, the corresponding linear combination of $\underline{\Psi}_k^{\rm p}$ is convergent. Then, under the assumption of Theorem 2, $\underline{\Psi}_k^{\rm p}$ together with any basis for $\mathscr{S}$ form a basis for $\mathscr{C}_p = \overline{\mathscr{P}} \oplus \mathscr{S}$.

Proof: Uses arguments similar to Theorem 2.

Such linear combinations are, of course, exactly the ℓ^2 for discrete eigenvalues and L^2 for continuous ones.

We now develop some spectral theoretic results for $\mathbf{H}$. It is easily verified that for real potentials,

$$P\sigma(\mathbf{H}) = P\sigma(\mathbf{H})^*, \quad C\sigma(\mathbf{H}) = C\sigma(\mathbf{H})^*,$$
$$R\sigma(\mathbf{H}) = R\sigma(\mathbf{H})^*. \tag{2.29}$$

Next, observe that for any choice of potentials such that $\widetilde{\mathbf{V}}$ has $\widetilde{\mathbf{H}}_0$ bound < 1, it is possible to show that

$$\sigma(\mathbf{H}) \subseteq \left[\lambda \in \mathbb{C}: \operatorname{Re}\lambda > 0, |\operatorname{Im}\lambda| \leqslant \left\{\frac{(b + a \operatorname{Re}\lambda)}{1-a}\right\} \cup \left\{\lambda \in \mathbb{C}: |\lambda| \leqslant \frac{b}{1-a}\right\}\right\}, \tag{2.30}$$

where a and b appear in (2.19) (see Appendix I). If $\widetilde{\mathbf{V}}$ is strictly bounded (by b) then (2.30) holds with $a = 0$. It is also useful to examine the solutions of the dual channel space Schrödinger equation

$$(\mathbf{H}' - \lambda)\underline{\xi}' = \underline{0}'. \tag{2.31}$$

There are a class of solutions of (2.31) which are of particular interest to us. These are constructed as follows. Let $|\Psi_E\rangle$ be a solution of (2.1) with $\lambda = E$, and choose $\underline{\xi}'_E$ so that

$$[\underline{\xi}'_E]_\alpha = \langle \Psi_E|, \quad \forall \alpha. \tag{2.32}$$

Then it is easy to show that

$$(\mathbf{H}' - E)\underline{\xi}'_E = \underline{0}'. \tag{2.33}$$

This was first established by Kouri and Levin.[29] This construction applies to both the normalizable (bound state) and scattering solutions.

This construction guarantees the spectral inclusions

$$P\sigma(H) \subseteq P\sigma(\mathbf{H}'), \quad C\sigma(H) \subseteq \pi(\mathbf{H}')$$

so

$$\sigma(H) = P\sigma(H) \dot{\cup} C\sigma(H) \subseteq \pi(\mathbf{H}') \subseteq \sigma(\mathbf{H}'). \tag{2.34}$$

Here $\pi(\ )$ is the approximate point spectrum (Appendix A) and $\dot{\cup}$ means "disjoint union." Approximate eigenvectors for $\mathbf{H}'$ are constructed just as for H by modulation of the weak scattering eigenfunctions. However, if $\mathbf{H}$ is a closed linear operator, we have that (Yosida[10])

$$\sigma(\mathbf{H}') = \sigma(\mathbf{H}), \tag{2.35}$$

so

$$\sigma(H) \subseteq \sigma(\mathbf{H}) \text{ and } P\sigma(H) \subseteq P\sigma(\mathbf{H}') = \Gamma(\mathbf{H}) \subseteq P\sigma(\mathbf{H}) \cup R\sigma(\mathbf{H}),$$

where $\Gamma(\ )$ is the compression spectrum (see Appendix A). The possible appearance of a residual spectrum $R\sigma(\mathbf{H})$ here results from the fact that $\mathbf{H}$ is chosen to be nonnormal. In certain circumstances, however, we can guarantee that $R\sigma(\mathbf{H}) = \varnothing$.

Theorem 4: If $\mathbf{G}_0(\lambda)\mathbf{V}$ can be extended to a bounded operator, and $[\mathbf{G}_0(\lambda)\mathbf{V}]^n$ can be extended to a compact operator for some n and for all λ, then $R\sigma(\mathbf{H}) = \varnothing$ follows from the Fredholm alternative.

Proof: See Appendix B.

Note the appearance of a fiber compactness type assumption here. In those cases where $R\sigma(\mathbf{H}) = \varnothing$, from (2.34) and (2.35), we have

$$P\sigma(H) \subseteq P\sigma(\mathbf{H}') \subseteq P\sigma(\mathbf{H}), \quad C\sigma(H) \subseteq C\sigma(\mathbf{H}) \cup P\sigma(\mathbf{H}|\mathscr{S}). \tag{2.36}$$

If also $\mathbf{V}\mathbf{G}_0(\lambda)$ can be extended to a bounded operator and $[\mathbf{V}\mathbf{G}_0(\lambda)]^n$ can be extended to a compact operator for some n and for all λ, a similar argument shows that $R\sigma(\mathbf{H}') = \varnothing$ (see Appendix B). Then, for a finite number of channels, since $\mathscr{C}_p$ is reflective and $\mathbf{H}''$ is a closed extension of (and thus equal to) $\mathbf{H}$, we have in addition to (2.36) that

$$P\sigma(\mathbf{H}'') = P\sigma(\mathbf{H}) = \Gamma(\mathbf{H}') \subseteq P\sigma(\mathbf{H}')$$

so

$$P\sigma(\mathbf{H}') = P\sigma(\mathbf{H}) \quad \text{and} \quad C\sigma(\mathbf{H}') = C\sigma(\mathbf{H}). \tag{2.37}$$

Even if $P\sigma(H) \subseteq P\sigma(\mathbf{H})$, $C\sigma(H) \subseteq C\sigma(\mathbf{H}) \cup P\sigma(\mathbf{H}|\mathscr{S})$, we have not proved that all solutions of (2.1) are imbedded into the physical solutions of (2.8). However, clearly in this case, if a nondegenerate normalizable physical solution is missing, it must be replaced by a normalizable spurious solution.

Some further properties of the operator $\mathbf{H}$ can be obtained from the general semigroup theory for closed operators on Banach space. One question answered partially here is the nature of time evolution in this formalism, e.g., the existence and behavior of solutions to (2.9). We thus naturally ask if $i\mathbf{H}$ generates a group $\{e^{i/\hbar \mathbf{H}t}: t \in R\}$. Since $\mathbf{H}$ is not self-adjoint, Stone's theorem can not be implemented. However, in the case where the components of $\mathbf{V}$ are bounded on $\mathscr{H}$ we may utilize the following result (Peters, Pazy[30]).

Lemma 5: If $\mathbf{A}$ is the infinitesimal generator of a C_0 semigroup $\mathbf{T}(t)$ satisfying $\|\mathbf{T}(t)\| \leqslant me^{wt}$ and if $\mathbf{B}$ is a bounded operator, then $\mathbf{A}+\mathbf{B}$ is the infinitesimal generator of a C_0 semigroup $\mathbf{S}(t)$ satisfying $\|\mathbf{S}(t)\| \leqslant me^{(w+m\|\mathbf{B}\|)t}$.

Theorem 5: For V bounded, $i\mathbf{H}$ generates a C_0 group of bounded operators $\mathbf{U}(t)=e^{i/\hbar \mathbf{H}t}$, and

$$\|\mathbf{U}(t)\|_{(2)} \leqslant e^{\|\mathbf{V}\|_{(2)}|t|/\hbar}, \quad t \in R. \tag{2.38}$$

Proof: In Lemma 5 choose $\mathbf{A}= \pm i/\hbar\, \mathbf{H}_0$, $\mathbf{B}= \pm i/\hbar\, \mathbf{V}$, so in the $\|-\|_{(2)}$ norm $m=1$ and $w=0$ (where $\|-\|_{(2)}$ is the uniform operator norm for $p=2$).

Such exponential growth may be associated with spurious eigenvectors with nonreal eigenvalue. The following result is available for much weaker conditions on the potentials:

Theorem 6: Suppose that $\tilde{\mathbf{V}}$ is $\tilde{\mathbf{H}}_0$-bounded with relative bound "0", then $\mathbf{H}$ generates a holomorphic (C_0) semigroup $e^{-\lambda \mathbf{H}}$ in the sector $\{\lambda : |\arg\lambda| < \pi/2\}$.

Proof: For $\mathbf{V}$ bounded, the result follows from a standard perturbation theorem.[30] The more general result is proved in Appendix C.

This result will be useful in the discussion of channel space equilibrium density matrices. The existence of a time evolution operator $\mathbf{U}(t)$ may alternatively be considered in terms of the existence of suitable boundary values for $e^{-\lambda \mathbf{H}}$ defined on the open right half-plane. A different approach to this existence question is taken in the following sections.

Next we comment briefly on the general structure of the operators that will naturally appear in this formulation. It is assumed that for each (Hilbert space) quantum mechanical operator A of interest, there is a natural decomposition into a channel space operator $\mathbf{A}$ which satisfies the summation property

$$\sum_{\alpha} A_{\alpha\beta} = A \quad \text{for all } \beta \tag{2.39}$$

(Hoffman *et al.*[24]). If we consider those bounded operators satisfying (2.39) whose domain is the whole channel space, then they also form a Banach subalgebra of the bounded operators on $\mathscr{C}_p$ as may be seen from the identity

$$\sum_{\alpha} (\mathbf{AB})_{\alpha\beta} = AB. \tag{2.40}$$

Denote the algebra of such bounded operators by $\mathscr{A}$. This property is expected to be significant in an algebraic formulation of the theory. The operators on $\mathscr{C}_p$ corresponding to the bounded observables must be contained in the Banach "field" algebra $\mathscr{A}$. They will be associated with a von Neumann or C^*-algebra in $\mathscr{A}$ but not with the usual $* = +$ involution (i.e., adjoint) for operators on $\mathscr{C}_2$ as may be anticipated from (2.39).[31,32]

III. SPATIALLY CONFINED SYSTEMS: DISCRETE SPECTRA AND NORMALIZABLE EIGENVECTORS

Let us consider a system of N distinguishable particles confined to a region $\mathscr{R}$ by an external potential going to infinity at the boundary $\partial\mathscr{R}$ of $\mathscr{R}$ and described by a Hamiltonian H (the full kinetic energy is included here). Then this external potential may be chosen to appear along the diagonal of $\mathbf{H}$. Alternatively, we could assume that the wavefunctions are defined on the interior of a box satisfying the appropriate equations there and impose periodic boundary conditions at the walls. If the potentials have a range greater than the box size, then they are regarded as suitably treated. These problems are important for applications to the statistical mechanics of reactive systems where a convenient representation is first required for a confined system of a finite number of particles.[24,26] It is anticipated that some aspects of the interpretation of the channel space wavefunction components will carry over from the spatially infinite case.[3,20,26] Many of the mathematical concepts used to describe the structure of $\mathbf{H}$ are most conveniently introduced first here because of the following result.

Theorem 7: For spatially confined systems and for relatively bounded potentials of the form (2.19), $\mathbf{H}$ has only discrete eigenvalues with normalizable eigenvectors. Consequently, $\sigma(H) = P\sigma(H) \subsetneq P\sigma(\mathbf{H}) = \sigma(\mathbf{H})$ since $\mathbf{H}$ is closed.

Proof: This follows essentially as a consequence of the compactness of $(\lambda - \mathbf{H})^{-1}$ for certain $\lambda \in \rho(\mathbf{H})$ (see Appendix D).

To facilitate the spectral analysis of $\mathbf{H}$, it is convenient to introduce the concept of a biorthogonal system in the Banach space $\mathscr{C}_p$ (Singer,[33] Dunford and Schwartz[34]). Such a system is defined by a pair of sequences $\{\underline{\eta}'_n, \underline{\phi}_n\}$, where $\underline{\eta}'_n \in \mathscr{C}'_p$, $\underline{\phi}_n \in \mathscr{C}_p$, and

$$(\underline{\eta}'_m, \underline{\phi}_n) = \delta_{m,n}, \quad \forall m,n. \tag{3.1}$$

We suppose that some of the $\underline{\phi}_n$ are chosen as linearly independent eigenvectors of $\mathbf{H}$.[24] These are not orthogonal in $\mathscr{C}_2$ since $\mathbf{H}$ is not normal. We shall denote the physical eigenvectors by $\underline{\Psi}^{\mathrm{p}}_n$ where $\Sigma_\alpha [\underline{\Psi}^{\mathrm{p}}_n]_\alpha = |\Psi^{\mathrm{p}}_n\rangle$ are distinct orthonormal H eigenvectors (with eigenvalues λ^{p}_n) and the spurious eigenvectors by $\underline{\Psi}^{\mathrm{s}}_n$ (with eigenvalues λ^{s}_n). For the physical eigenvectors, we define the corresponding dual vectors $\underline{\xi}'^{\mathrm{p}}_n$ using the prescription outlined in the previous section, i.e.,

$$[\underline{\xi}'^{\mathrm{p}}_n]_\alpha = \langle \Psi^{\mathrm{p}}_n|, \ \forall\alpha. \tag{3.2}$$

As mentioned previously, $\underline{\xi}'^{\mathrm{p}}_n$ so defined satisfies

$$\mathbf{H}'\underline{\xi}'^{\mathrm{p}}_n = \lambda^{\mathrm{p}}_n \underline{\xi}'^{\mathrm{p}}_n. \tag{3.3}$$

With such a choice (3.1) is satisfied for $\underline{\phi}_n \in \{\underline{\Psi}^{\mathrm{p}}_n, \underline{\Psi}^{\mathrm{s}}_n\}$ and $\underline{\eta}'_m \in \{\underline{\xi}'^{\mathrm{p}}_m\}$. The object here is to extend this set to obtain a "complete biorthogonal system," if possible, in a way providing a useful functional calculus for $\mathbf{H}$.

Suppose first that $\{\underline{\Psi}^{\mathrm{p}}_n, \underline{\Psi}^{\mathrm{s}}_n\}$ form a basis for the channel space in which case it is immediate from (2.10) that exactly all ℓ^2 linear combinations of $\underline{\Psi}^{\mathrm{p}}_n$ converge in norm. It is only possible for the normalizable eigenvectors to form a basis for spatially confined systems since we know that the scattering solutions must be included in any completeness discussion for the spatially infinite case. Completeness of $\{\underline{\Psi}^{\mathrm{p}}_n, \underline{\Psi}^{\mathrm{s}}_n\}$ has been proved for the Faddeev equations in a spatially confined region (see Evans and Hoffman[35]). Since $\{\underline{\Psi}^{\mathrm{p}}_n, \underline{\Psi}^{\mathrm{s}}_n\}$ is assumed to be a basis, there exists a unique associated sequence of coefficient functionals[33] $\{\underline{\eta}'^{\mathrm{p}}_n, \underline{\eta}'^{\mathrm{s}}_n\}$ satisfying (3.1). A biorthogonal system formed from a basis is called regular. It necessarily follows that $\eta_n'^{\mathrm{P}} = \xi_n'^{\mathrm{P}}$. Next, we show that $\eta_n'^{\mathrm{S}} = \xi_n'^{\mathrm{S}}$, so defined, satisfies the dual Schrödinger equation. From the relations

$$
\begin{aligned}
(\xi_n'^{\mathrm{S}},\mathbf{H}\underline{\Psi}_m^{\mathrm{S}}) &= \lambda_m^{\mathrm{S}}(\xi_n'^{\mathrm{S}},\underline{\Psi}_m^{\mathrm{S}}) = \lambda_m^{\mathrm{S}}\delta_{n,m} \\
&= \lambda_n^{\mathrm{S}}(\xi_n'^{\mathrm{S}},\underline{\Psi}_m^{\mathrm{S}}), \qquad (3.4)\\
(\xi_n'^{\mathrm{S}},\mathbf{H}\underline{\Psi}_m^{\mathrm{P}}) &= \lambda_m^{\mathrm{P}}(\xi_n'^{\mathrm{S}},\underline{\Psi}_m^{\mathrm{P}}) = 0 \\
&= \lambda_n^{\mathrm{S}}(\xi_n'^{\mathrm{S}},\underline{\Psi}_m^{\mathrm{P}}),
\end{aligned}
$$

and since $\{\underline{\Psi}_n^{\mathrm{P}},\underline{\Psi}_n^{\mathrm{S}}\}$ is a basis, we conclude that

$$\mathbf{H}'\xi_n'^{\mathrm{S}} = \lambda_n^{\mathrm{S}}\xi_n'^{\mathrm{S}}. \qquad (3.5)$$

The existence of a regular biorthogonal system of eigenvectors for $\mathbf{H}/\mathbf{H}'$ is significant in developing a functional calculus for $\mathbf{H}$ analogous to that for normal operators.

Theorem 8: If $\{\underline{\Psi}_n^{\mathrm{P}},\underline{\Psi}_n^{\mathrm{S}}; n = 1,2,\cdots\}$ is a basis for $\mathscr{C}_p$ and $\{\xi_n'^{\mathrm{P}},\xi_n'^{\mathrm{S}}; n = 1,2,\cdots\}$ The associated sequence of coefficient functionals, then

$$\mathbf{E}(\delta) = \sum_{\lambda_n^{\mathrm{P}}\in\delta} \underline{\Psi}_n^{\mathrm{P}}\xi_n'^{\mathrm{P}} + \sum_{\lambda_n^{\mathrm{S}}\in\delta} \underline{\Psi}_n^{\mathrm{S}}\xi_n'^{\mathrm{S}} \qquad (3.6)$$

(where δ are Borel sets in the complex plane) is a countably additive resolution of the identity[36] for $\mathbf{H}$. Furthermore, $\mathbf{H}$ is scalar spectral[36] and given by

$$\mathbf{H}\underline{\Psi} = \lim_{N\to\infty}\left(\sum_{n=1}^{N}\lambda_n^{\mathrm{P}}(\xi_n'^{\mathrm{P}},\underline{\Psi})\underline{\Psi}_n^{\mathrm{P}} + \sum_{n=1}^{N}\lambda_n^{\mathrm{S}}(\xi_n'^{\mathrm{S}},\underline{\Psi})\underline{\Psi}_n^{\mathrm{S}}\right) \qquad (3.7)$$

for $\underline{\Psi}\in\mathrm{dom}(\mathbf{H})$.

Proof: Consider first (3.7). Define

$$\widetilde{\mathbf{H}}_N\underline{\Psi} = \sum_{n=1}^{N}\lambda_n^{\mathrm{P}}(\xi_n'^{\mathrm{P}},\underline{\Psi})\underline{\Psi}_n^{\mathrm{P}} + \sum_{n=1}^{N}\lambda_n^{\mathrm{S}}(\xi_n'^{\mathrm{S}},\underline{\Psi})\underline{\Psi}_n^{\mathrm{S}}, \quad \forall\underline{\Psi}\in\mathscr{C}_p, \qquad (3.8)$$

and let $\widetilde{\mathbf{H}}\underline{\Psi} = \lim_{N\to\infty}\widetilde{\mathbf{H}}_N\underline{\Psi}$ for those $\underline{\Psi}$ where the limit exists. From the basis property, for any $\underline{\Psi}\in\mathscr{C}_p$,

$$\underline{\Psi} = \sum_{n=1}^{\infty}c_n^{\mathrm{P}}\underline{\Psi}_n^{\mathrm{P}} + \sum_{n=1}^{\infty}c_n^{\mathrm{S}}\underline{\Psi}_n^{\mathrm{S}} \qquad (3.9)$$

for some $c_n^{\mathrm{P}},c_n^{\mathrm{S}}$. So, define

$$\underline{\Psi}_N = \sum_{n=1}^{N}c_n^{\mathrm{P}}\underline{\Psi}_n^{\mathrm{P}} + \sum_{n=1}^{N}c_n^{\mathrm{S}}\underline{\Psi}_n^{\mathrm{S}}. \qquad (3.10)$$

Then, if $\underline{\Psi}\in\mathrm{dom}(\mathbf{H})$,

$$\widetilde{\mathbf{H}}_N\underline{\Psi} = \widetilde{\mathbf{H}}\underline{\Psi}_N = \mathbf{H}\underline{\Psi}_N \to \mathbf{H}\underline{\Psi} \quad \text{as } N\to\infty \qquad (3.11)$$

since $\mathbf{H}$ is closed. Thus, $\widetilde{\mathbf{H}}\underline{\Psi}$ exists and

$$\widetilde{\mathbf{H}}\underline{\Psi} = \mathbf{H}\underline{\Psi}, \qquad (3.12)$$

which proves (3.7). Clearly the $\mathbf{E}(\delta)$ commute with $\mathbf{H}$ and define a countably additive projection valued measure. Also $\sigma(\mathbf{H}|\mathbf{E}(\delta))\subseteq\bar{\delta}$, the closure of δ. Boundedness of these possibly infinite sums follows from the basis property and the Banach–Steinhaus theorem. Thus the first part of the theorem is proved.

A functional calculus may be defined for $\mathbf{H}$ (under these conditions) by

$$
\begin{aligned}
f(\mathbf{H})\underline{\Psi} &= \lim_{n\to\infty}\int_{\Delta(\mathbf{H})} f_n(\lambda)\mathbf{E}(d\lambda)\underline{\Psi} \\
&= \lim_{N\to\infty}\left(\sum_{n=1}^{N} f(\lambda_n^{\mathrm{P}})(\xi_n'^{\mathrm{P}},\underline{\Psi})\underline{\Psi}_n^{\mathrm{P}}\right. \\
&\quad \left. + \sum_{n=1}^{N} f(\lambda_n^{\mathrm{S}})(\xi_n'^{\mathrm{S}},\underline{\Psi})\underline{\Psi}_n^{\mathrm{S}}\right), \qquad (3.13)
\end{aligned}
$$

for those $\underline{\Psi}$ where the right-hand side is convergent. Here f is measurable, $\Delta(\mathbf{H})$ is an open set containing $\sigma(\mathbf{H})$, and

$$f_n(\lambda) = \begin{cases} f(\lambda), & |f(\lambda)|\leqslant n \\ 0, & |f(\lambda)| > n \end{cases} \qquad (3.14)$$

(see Dunford and Schwartz[36]). Suppose that $f(\ )$ is also bounded. Then since $(\xi_n'^{\mathrm{P}},\underline{\Psi}) = \langle\xi_n^{\mathrm{P}}|\Sigma_\alpha\Psi_\alpha\rangle$, clearly $\{f(\lambda_n^{\mathrm{P}})(\xi_n'^{\mathrm{P}},\underline{\Psi})\}_{n=1}^{\infty}\in\ell^2$, so the first term in (3.13) is convergent for all $\underline{\Psi}$. Convergence of the second term for all $\underline{\Psi}$ follows for example if the basis is Besselian and Hilbertian[33], i.e., exactly the ℓ^2 linear combinations of eigenvectors, and thus, in particular, of spurious eigenvectors, converge in norm (e.g., the Faddeev case[35]). The choice $f(\lambda) = e^{i/\hbar\lambda t}$ provides an alternative approach to the analysis of the time evolution operator $\mathbf{U}(t)$.

Secondly, we consider the case where $\{\underline{\Psi}_n^{\mathrm{P}},\underline{\Psi}_n^{\mathrm{S}}\}$ do not form a basis. For example, there could be a normalized nondegenerate eigenvector $|\Psi_{n^*}^{\mathrm{P}}\rangle$ of H which is not imbedded into the set of physical eigenvectors of $\mathbf{H}$, but we still have $P\sigma(H)\subseteq P\sigma(\mathbf{H})$ (Theorem 7). Degeneracies for a spatially confined system are, in general, not expected if the boundary conditions are chosen to break all geometric symmetries. Then for the dual eigenvector $\xi_{n^*}'^{\mathrm{P}}:(\xi_{n^*}'^{\mathrm{P}})_\alpha = \langle\underline{\Psi}_{n^*}^{\mathrm{P}}|$, for all α with eigenvalue $\lambda_{n^*}^{\mathrm{P}}$, there corresponds at least one spurious eigenvector $\underline{\Psi}_{n^*}^{\mathrm{S}}$ with the same eigenvalue. However, this correspondence is not in the biorthogonal sense since

$$(\xi_{n^*}'^{\mathrm{P}},\underline{\Psi}_{n^*}^{\mathrm{S}}) = 0. \qquad (3.15)$$

Furthermore, without the basis property, there is no need for the physical duals to be uniquely specified by the equal component form although this is still a valid (and the most natural) choice. If $\{\underline{\Psi}_n^{\mathrm{P}},\underline{\Psi}_n^{\mathrm{S}}\}$ can be extended by $\{\phi_n\}$ in $\mathscr{N}' = \{\underline{\Psi}\in\mathscr{C}_p:(\xi_m'^{\mathrm{P}},\underline{\Psi}) = 0, \forall m\}$ to form a basis, then a regular biorthogonal system may be constructed and the coefficient functionals associated with $\underline{\Psi}_n^{\mathrm{P}}$ are given by $\xi_n'^{\mathrm{P}}$. However, those associated with the $\underline{\Psi}_n^{\mathrm{S}}$ are not expected to all satisfy the dual channel space Schrödinger equation. This problem will be discussed in later work. Again the appropriate linear combinations of the $\underline{\Psi}_n^{\mathrm{P}}$ are exactly the ℓ^2 ones. Further results incorporating this case are described below.

To show the "equivalence" of the channel and Hilbert space formulations, it is necessary to demonstrate an agreement in corresponding expectation values calculated from different theories. Since in the channel space theory only the physical eigensolutions contribute, it is appropriate to construct of projection operator $\mathbf{P}$ corresponding to the subspace spanned by these, preferably, so that $[\mathbf{P},\mathbf{H}] = 0$. If the physical eigenvectors of $\mathbf{H}$ can be extended in $\mathscr{N}'$ to form a basis $\mathscr{C}_p$ then we define $\mathbf{P}$ to be the projection operator associated with the decomposition $\mathscr{C}_p = \overline{\mathrm{span}(\underline{\Psi}_n^{\mathrm{P}})}\oplus\mathscr{N}'$. $\mathbf{P}$ is then not self-adjoint but has the simple representation

$$\underline{\mathbf{P}\Psi} = \sum_n \underline{\Psi}_n^{\mathrm{P}}(\underline{\xi}_n'^{\mathrm{P}},\underline{\Psi}), \quad \forall \underline{\Psi} \in \mathscr{C}_p. \tag{3.16}$$

$\Sigma_\alpha P_{\alpha\beta} = P$ is the projection operator onto the corresponding physical states in $\mathscr{H}$. Clearly, $\mathbf{H}$ restricted to the $\mathbf{H}$-invariant subspace Range($\mathbf{P}$) is real eigenvalue scalar spectral and is *-self-adjoint with a suitable choice of involution.[31] This amounts to defining self-adjointness with respect to a duality mapping $D:\mathscr{C}_p \to \mathscr{C}'_p$ naturally induced by any basis that includes the $\underline{\Psi}_n^{\mathrm{P}}$ and the $\underline{\xi}_n'^{\mathrm{P}}$ as their associated coefficient functionals, i.e., D is conjugate linear and maps all basis vectors to their corresponding coefficient functionals. Then

$$(D\underline{\Psi},(\mathbf{H}\ \mathbf{P})\underline{\phi}) = (D(\mathbf{H}\ \mathbf{P})\underline{\Psi},\underline{\phi}), \quad \forall \underline{\Psi},\underline{\phi} \in \mathscr{C}_p \tag{3.17}$$

[cf., the theory of accretive operators on Banach space (Pazy[30])]. Further, a simple construction shows that $\mathbf{H}$ on Range ($\mathbf{P}$) is equivalent (via a similarity transformation) to H on Range (P).[35] If $\mathbf{H}$ is also scalar spectral on $\mathscr{C}_p$, then $\mathbf{P}$ is naturally obtained from the resolution of the identity and $P = I$.

The injection operator imbedding H eigenvectors into physical $\mathbf{H}$ eigenvectors may be defined in terms of $\mathbf{P}$ as

$$J|\Psi\rangle = \mathbf{P}\underline{\theta}|\Psi\rangle, \tag{3.18}$$

where $\underline{\theta}$ is a numerical vector and $\Sigma_\alpha \theta_\alpha = 1$. The boundedness of J follows from that of $\mathbf{P}$. Consequently, the $\underline{\Psi}_n^{\mathrm{P}}$ are uniformly bounded with respect to n. This last property follows directly from the assumption that $\underline{\Psi}_n^{\mathrm{P}}$ form part of a basis, since then there exists a constant M (independent of n) such that[33]

$$1 \leqslant \|\underline{\Psi}_n^{\mathrm{P}}\|_p \|\underline{\xi}_n'^{\mathrm{P}}\|_q \leqslant M, \tag{3.19}$$

and we have chosen $\|\underline{\xi}_n'^{\mathrm{P}}\|_q = N_{\mathrm{ch}}^{1/q}$ for all n.

To calculate expectation values for general (mixed) states, an appropriate concept of "trace class" and "trace" is needed. Since $\mathscr{C}_p = \mathscr{C}_2$, we could try using the standard definition for operators on Hilbert space. However, a more natural definition for our purposes has been developed by Ruston[37] in the Banach space framework. Here we consider first finite rank operators of the form $\Sigma_{i=1}^n \underline{\phi}_i \underline{\eta}_i'$ with $\underline{\phi}_i \in \mathscr{C}_p, \underline{\eta}_i' \in \mathscr{C}'_p$. Then $\Sigma_{i=1}^m \hat{\underline{\phi}}_i \hat{\underline{\eta}}_i'$ is said to be equivalent to $\Sigma_{i=1}^n \underline{\phi}_i \underline{\eta}_i'$ if it represents the same operator. We define the cross norm

$$\gamma_p\Big(\sum_{i=1}^n \underline{\phi}_i \underline{\eta}_i'\Big) = \inf\Big(\sum_{i=1}^m \|\hat{\underline{\phi}}_i\|_p \|\hat{\underline{\eta}}_i'\|_q\Big), \tag{3.20}$$

with $1/p + 1/q = 1$ where the infimum is taken over $\Sigma_{i=1}^m \hat{\underline{\phi}}_i \hat{\underline{\eta}}_i'$ in the above class. The closure of the finite rank operators in the $\gamma_p(\ )$ norm is called the trace class of the Banach space $\mathscr{C}_p$ (actually independent of p here). Clearly,

$$\gamma_p\Big(\sum_{i=1}^n \underline{\phi}_i \underline{\eta}_i'\Big) \geqslant \Big\|\sum_{i=1}^n \underline{\phi}_i \underline{\eta}_i'\Big\|_{(p)}, \tag{3.21}$$

where $\|\text{—}\|_{(p)}$ is the uniform operator norm. If $\mathbf{A}$ is a bounded operator on $\mathscr{C}_p$ and $\mathbf{T}$ is in the trace class, then $\mathbf{AT}$ and $\mathbf{TA}$ are in the trace class and

$$\gamma_p(\mathbf{AT}) \leqslant \gamma_p(\mathbf{T})\|\mathbf{A}\|_{(p)}, \quad \gamma_p(\mathbf{TA}) \leqslant \gamma_p(\mathbf{T})\|\mathbf{A}\|_{(p)}. \tag{3.22}$$

For an operator of finite rank, we define the trace "tr" by

$$\mathrm{tr}\Big(\sum_{i=1}^n \underline{\phi}_i \underline{\eta}_i'\Big) = \sum_{i=1}^m (\hat{\underline{\eta}}_i',\hat{\underline{\phi}}_i). \tag{3.23}$$

Equation (3.23) is independent of the choice $\Sigma_{i=1}^m \hat{\underline{\phi}}_i \hat{\underline{\eta}}_i'$ of representation. Clearly,

$$\Big|\mathrm{tr}\Big(\sum_{i=1}^n \underline{\phi}_i \underline{\eta}_i'\Big)\Big| \leqslant \gamma_p\Big(\sum_{i=1}^n \underline{\phi}_i \underline{\eta}_i'\Big), \tag{3.24}$$

so by a standard limiting procedure we can uniquely define the trace of any operator in the trace class. It is easily proved that for $\mathbf{A},\mathbf{T}$ as above,

$$\mathrm{tr}(\mathbf{AT}) = \mathrm{tr}(\mathbf{TA}). \tag{3.25}$$

For any regular biorthogonal system $(\underline{\xi}_i', \underline{\Psi}_i)$ we may alternatively express the trace of an operator $\mathbf{T}$ (in the trace class) as

$$\mathrm{tr}(\mathbf{T}) = \sum_i (\underline{\xi}_i', \mathbf{T}\underline{\Psi}_i). \tag{3.26}$$

The proof of the invariance of the rhs of (3.26) for different choices of a biorthogonal system follows easily from their regularity. This definition of course agrees with the usual one where an orthonormal basis is used.

We are now in a position to define a class of operators corresponding to the density matrices for mixtures of discrete states. As observed by Hoffman *et al.*[24] (assuming $\mathbf{P}$ may be defined as previously), if $\mathbf{A} \in \mathscr{A}$ and

$$\mathbf{PAP}(= \mathbf{P}A\,\mathbf{P}) = \mathbf{A}, \tag{3.27}$$

then $(\mathbf{A})_{\alpha\beta}$ is independent of β. In fact $\mathscr{A}_{\mathbf{P}} = \{\mathbf{A} \in \mathscr{A} : \mathbf{PAP} = \mathbf{A}\}$ is a subalgebra of $\mathscr{A}$. The physical channel space density matrices are elements $\boldsymbol{\rho}$ of this subalgebra of the form

$$\boldsymbol{\rho} = \mathbf{P}\rho\mathbf{P}, \tag{3.28}$$

where ρ is a positive self-adjoint trace class Hilbert space density matrix for a mixture of the "imbedded" physical states, so $P\rho P = \rho$. Normalization is chosen so that $\mathrm{Tr}\,\rho = 1$. Using an explicit representation for $\rho = \Sigma_m \rho_m |\Psi_m\rangle\langle\Psi_m|$, where $\rho_m \geqslant 0$ and $\Sigma_m \rho_m = 1$, one may show that $\boldsymbol{\rho}$ is trace class using the boundedness of J, and that

$$\mathrm{tr}\,\boldsymbol{\rho} = 1. \tag{3.29}$$

Of course $\boldsymbol{\rho}$ so defined is not "self-adjoint" or "positive" in the usual sense but is *-self-adjoint and *-positive with respect to a suitably defined involution*.[31] In terms of the duality mapping D described previously,

$$(D\underline{\Psi}, \boldsymbol{\rho}\underline{\phi}) = (D\,\boldsymbol{\rho}\underline{\Psi},\underline{\phi}), \quad \forall \underline{\Psi},\ \underline{\phi} \in \mathscr{C}_p, \tag{3.30}$$

and

$$(D\underline{\Psi}, \boldsymbol{\rho}\underline{\Psi}) \geqslant 0, \quad \forall \underline{\Psi} \in \mathscr{C}_p.$$

The equilibrium form of the channel density matrix is given by

$$\begin{aligned}\boldsymbol{\rho}^{\mathrm{eq}} &= \mathbf{P}e^{-\beta H}\,\mathbf{P}/Z_c\\ &= \mathbf{P}e^{-\beta \mathbf{H}}/Z_c,\end{aligned} \tag{3.31}$$

where $Z_c = \mathrm{Tr}(\,Pe^{-\beta H})$ and

$$\mathbf{P}e^{-\beta\mathbf{H}} = \sum_n e^{-\beta\lambda_n^{\mathrm{P}}}\underline{\Psi}_n^{\mathrm{P}}\underline{\xi}_n'^{\mathrm{P}}. \tag{3.32}$$

Note that the existence of $e^{-\beta \mathbf{H}}$ follows without any basis property assumption and for quite general potentials (Theorem 6).

Finally, we show how the channel theory makes contact with the Hilbert space quantum theory in the calculation of expectation values. In any representation, the observables correspond to the *-self-adjoint elements of a Von-Neumann or C^* algebra where * is the appropriate involution. Real expectation values are then calculated by acting on them with *-positive linear functionals, e.g., $\mathrm{Tr}(\rho\cdot)$,$\mathrm{tr}(\boldsymbol{\rho}\cdot)$. The agreement of expectation values of the Hilbert and channel space theories is guaranteed by the special summation structure (2.39) of the Banach algebra containing the observables together with the extra structure of $\boldsymbol{\rho}$ (independence of $\rho_{\alpha\beta}$ on β). This correspondence may be written as[24]

$$\langle \mathbf{A} \rangle = \mathrm{tr}(\mathbf{A}\,\boldsymbol{\rho}) = \mathrm{Tr}(A\,\rho) = \langle A \rangle. \tag{3.33}$$

Note the restriction $P\rho P = \rho$ (we have not proved that $P = I$ in general).

IV. SPATIALLY INFINITE SYSTEMS: WEAK (SCATTERING) EIGENVECTORS

For a spatially infinite system, we expect that $\mathbf{H}$ will have a variety of unnormalizable (weak) scattering eigenfunctions. Here we shall assume the existence of these in a suitably large auxiliary space (Amrein, Jauch, and Sinha[38]) rather than attempting to prove their existence, e.g., from the integral equations and appropriate technical assumptions. These are associated with $C\sigma(\mathbf{H})$. Here the center of mass kinetic energy is removed from the Hamiltonians.

It is first useful to characterize the (weak) solutions of the equation

$$(H_\alpha - \lambda)|\Psi\rangle = 0. \tag{4.1}$$

This is done as follows. The channel indices α were defined to be certain partitions of the labels $1,2,...,N$ for the N particles of the system into appropriate clusters. We say $\alpha \subset \beta$ or $\beta \supset \alpha$ if α can be obtained from β by breaking up some (possibly 0) clusters of α. This relation is a partial order. We can also define the meet $\alpha \cap \beta$ as the coarest partition satisfying $\alpha \cap \beta \subset \alpha$ and $\alpha \cap \beta \subset \beta$. Similarly the join $\alpha \cup \beta$ is the finest partition satisfying $\alpha \cap \beta \supset \alpha$ and $\alpha \cup \beta \supset \beta$. $\cap$ and $\cup$ endow the set of partitions with a lattice structure (see Polyzou and Redish[9]). Now the solutions of (4.1) are labelled by

$$|\phi^{\pm}_{\alpha,\alpha';m_j,k_i}\rangle, \tag{4.2}$$

where $+$ $(-)$ refers to a precollisional (postcollisional) choice of asymptotic condition, and $\alpha' \subset \alpha$ represents an asymptotic clustering of particles into stable (bound-state) clusters α'. The $\underline{k}_i$ are a suitably chosen set of relative momentum labels for the asymptotic motion of the clusters in this scattering state. If μ_i are the corresponding reduced masses for these clusters and if m_j label the bound-states of energy E_{m_j} of those clusters of more than one particle, then

$$\lambda = \lambda_{\alpha';m_j,k_i} = \sum_j E_{m_j} + \sum_i \underline{k}_i^2/2\mu_i. \tag{4.3}$$

Consider now the equation $(\mathbf{H} - \lambda)\underline{\Psi} = \underline{0}$. One may look for scattering solutions $\underline{\Psi} = \underline{\Psi}^{\pm}_{\alpha,\alpha';m_j,k_i}$, $\alpha' \subset \alpha$, which have the asymptotic structure $[\underline{\Psi}^{\pm}]_\beta \sim \delta_{\alpha\beta}|\phi^{\pm}_{\alpha,\alpha';m_j,k_i}\rangle$ in the appropriate regions of coordinate space (i.e., the α' and also the α tubes). For fixed $\alpha',m_j,\underline{k}_i$, these correspond to the same H-eigensolution $|\Psi^{\pm}_{\alpha';m_j,k_i}\rangle = \Sigma_\beta[\underline{\Psi}^{\pm}_{\alpha,\alpha';m_j,k_i}]_\beta$ and satisfy the integral equations

$$\underline{\Psi}^{\pm}_{\alpha,\alpha';m_j,k_i} = \underline{\phi}^{\pm}_{\alpha,\alpha';m_j,k_i} + \mathbf{G}_0^{\pm}(\lambda_{\alpha';m_j,k_i})\mathbf{V}\underline{\Psi}^{\pm}_{\alpha,\alpha';m_j,k_i}, \tag{4.4}$$

where $[\underline{\phi}^{\pm}_{\alpha,\alpha';m_j,k_i}]_\beta = \delta_{\alpha\beta}|\phi^{\pm}_{\alpha,\alpha';m_j,k_i}\rangle$.

These equations generalize (2.10). There may also be solutions of the differential equations corresponding to a totally bound cluster of the N particles with quantum number m. In the above notation these would be represented as $\underline{\Psi}_{(1,2\cdots N),m}$. Clearly the $\underline{\Psi}^{\pm}_{\alpha,\alpha';m_j,k_i}$ do not lie in $\mathscr{C}_p$ but the components are expected to be bounded wavelike functions. A rigorous proof could involve an analysis of (4.4) with suitable conditions on the potentials. A set of eigenvectors for $\mathbf{H}'$ with eigenvalues $\lambda_{\alpha';m_j,k_i}$, denoted $\underline{\zeta}^{\pm\prime}_{\alpha',m_j,k_i}$, are given by

$$[\underline{\zeta}^{\pm\prime}_{\alpha';m_j,k_i}]_\beta = \langle\Psi^{\pm}_{\alpha';m_j,k_i}|, \quad \forall\beta, \tag{4.5}$$

where $|\Psi^{\pm}_{\alpha';m_j,k_j}\rangle$ are also assumed δ-function orthonormal. Clearly $\underline{\zeta}^{\pm\prime}_{\alpha';m_j,k_i}$ are not in $\mathscr{C}'_p$.

In a discussion of completeness and spectral theory, it is convenient to partition the eigenvectors into distinct physical ones together with the remaining set of spurious eigenvectors. So from the $\underline{\Psi}^{\pm}_{\alpha,\alpha';m_j,k_i}$ for different $\alpha \supset \alpha'$, we pick one, say $\alpha = \alpha^*$, to be the physical eigenvector and from the rest construct weak (unnormalizable) spurious eigenvectors as follows. For any partition α' consisting of stable (bound-state) clusters, assuming $\underline{\Psi}^{\pm}_{\alpha,\alpha';m_j,k_i}$ exist, we set

$$\underline{\Psi}^{\pm s}_{\alpha';m_j,k_i} = \sum_{\alpha\supset\alpha'} \Theta_\alpha \underline{\Psi}^{\pm}_{\alpha,\alpha';m_j,k_i}, \tag{4.6}$$

with $\Sigma_{\alpha\supset\alpha'}\Theta_\alpha = 0$. Then $\underline{\Psi}^{\pm s}$ is a spurious weak eigensolution of $\mathbf{H}$ with eigenvalue $\lambda_{\alpha',m_j,k_i}$. These are obtainable directly in the Faddeev case.[35]

These ideas are presented most naturally in a mathematical framework which generalizes the Gel'fand triplet (see Amrein, Jauch, and Sinha[38]). We introduce an auxiliary Hilbert space E dense in $\mathscr{H}$ so that its Banach space dual E^{d} contains the bounded measurable functions. Imbedding $\mathscr{H}$ into E^{d} (using the duality of $\mathscr{H}$) we write

$$E \subset \mathscr{H} \subset E^{\mathrm{d}}. \tag{4.7}$$

Channel spaces E_p and $(E^{\mathrm{d}})_p$ may be constructed as in (2.16), but using the appropriate Hilbert space norms. Their elements are also denoted by $\underline{\Psi}$ with components $|\Psi_\alpha\rangle$. E'_p [resp.$(E^{\mathrm{d}})'_p$] is the dual of E_p [resp.$(E^{\mathrm{d}})_p$] where the action of the dual is given by

$$(\underline{\zeta}',\underline{\Psi}) = \sum_\alpha \langle\zeta_\alpha|\Psi_\alpha\rangle_E \quad (\text{resp.} = \sum_\alpha \langle\zeta_\alpha|\Psi_\alpha\rangle_{E^{\mathrm{d}}}). \tag{4.8}$$

Note that $(E^{\mathrm{d}})'_p = (E_p)'$ where the action of an element of $(E_p)'$ on E_p is given by (2.20). We have the inclusion relations (analogous to the usual Gel'fand triplet):

$$
\begin{array}{ccc}
(E^{\rm d})_p = & (E^{\rm d})'_p & \\
\cup & \cup & \\
\mathscr{C}_p = & \mathscr{C}'_p & \quad (4.9) \\
\cup & \cup & \\
E_p = & E'_p &
\end{array}
$$

where for the equalities we have made the identification of $|\Psi\rangle$ with $\langle\Psi|$. The countably normed space $\mathscr{S}$, dense in $\mathscr{H}$ of C^∞ fast decreasing functions (Gel'fand and Shilov[39]), could be used rather than E. The elements of its dual $\mathscr{S}^{\rm d}$ (continuous in the countably normed topology) are the generalized functions, cf., the rigged Hilbert space treatment of quantum mechanics (Gel'fand and Vilenkin[40], Böhm[41]).

The operator $\mathbf{H}$ may be extended uniquely to a dense subspace of $(E^{\rm d})_p$ or $(\mathscr{S}^{\rm d})_p$ which includes the wavelike solutions discussed previously. So these are solutions of the channel equation in the larger space, termed as weak solutions of the original equations. Similarly $\mathbf{H}'$ may be extended to a dense subspace of $(E^{\rm d})'_p$ or $(\mathscr{S}^{\rm d})'_p$ including such vectors as $\underline{\xi}^{\pm'}_{\alpha';m_j,k_i}$.

It is useful to introduce the concept of a generalized biorthogonal system $\{\underline{\eta}'_{\underline{k}^\nu},\underline{\phi}_{\underline{k}^\nu}\}$ where normalizable $\underline{\phi}_{\underline{k}^\nu}\in\mathscr{C}_p$, unnormalizable $\underline{\phi}_{\underline{k}^\nu}\in(E^{\rm d})_p$, normalizable $\underline{\eta}'_{\underline{k}^\nu}\in\mathscr{C}'_p$, unnormalizable $\underline{\eta}'_{\underline{k}^\nu}\in(E^{\rm d})'_p$. $\{\underline{k}^\nu\}$ is a discrete and/or continuous label for the vectors, e.g., $\underline{k}^\nu=\{\alpha';m_j,k_i\}$ and $\Sigma_\nu\int d\underline{k}^\nu$ represents the corresponding sum and/or integral. The biorthogonality condition may be written (loosely) as

$$(\underline{\eta}'_{\underline{k}^\nu},\underline{\phi}_{\underline{k}^{\nu'}}) = \delta_{\nu,\nu'}\delta(\underline{k}^\nu-\underline{k}'^\nu). \tag{4.10}$$

Again we choose some of the $\underline{\phi}_{\underline{k}^\nu}$ to be distinct physical eigenvectors of $\mathbf{H}$ denoted $\underline{\Psi}^{\rm p}_{\underline{k}}$ with eigenvalues $\lambda^{\rm p}_{\underline{k}^\nu}$. The corresponding $\underline{\eta}'_{\underline{k}^\nu}$, denoted $\underline{\xi}'^{\rm p}_{\underline{k}^\nu}$, are constructed as in (4.5). Some of the remaining $\underline{\phi}_{\underline{k}^\nu}$ are chosen as spurious eigenvectors of $\mathbf{H}$ denoted $\underline{\Psi}^{\rm s}_{\underline{k}^\nu}$, with eigenvalues $\lambda^{\rm s}_{\underline{k}^\nu}$, and the object as previously is to suitably complete this set.

Suppose first that these form a basis in $\mathscr{C}_p$. Specifically, we mean that any vector in $\mathscr{C}_p$ can be represented as a suitable linear combination of basis vectors. Convergence in norm of the integrals in the limit-in-mean sense is implied. Further we suppose that the subspaces of such linear combinations with eigenvalues in specified Borel subsets of the complex plane are closed. For the physical scattering eigenvectors, these linear combinations are precisely the L^2 ones. If the spurious wavelike eigenvectors are constructed as in (4.6), then any convergent linear combination must also be L^2 (i.e., the basis is "generalized" Hilbertian) and we expect that all L^2 linear combinations are required ("generalized" Besselian). This is certainly true where α' is the complete breakup channel.

We may then automatically define (strictly almost everywhere) a generalized associated sequence of coefficient functionals $\underline{\eta}'_{\underline{k}^\nu}$ denoted $\{\underline{\xi}'^{\rm p}_{\underline{k}^\nu},\underline{\xi}'^{\rm s}_{\underline{k}^\nu}\}$ (using obvious notation). The $\underline{\xi}'^{\rm p}_{\underline{k}^\nu}$ are defined as in (4.5) and thus lie in $(E^{\rm d})'_p$ under suitable conditions on the potentials.[38] Suppose that all the wavelike $\underline{\Psi}^{\rm s}_{\underline{k}^\nu}$ are generated by the procedure described in (4.6). Then certain wavelike features of the asymptotic structure of the $\underline{\xi}'^{\rm s}_{\underline{k}^\nu}$ are easily determined from those of the corresponding $\underline{\Psi}^{\rm s}_{\underline{k}^\nu}$ and $\underline{\Psi}^{\rm p}_{\underline{k}^\nu}$ using biorthogonality (Appendix E). This suggests that the $\underline{\xi}'^{\rm s}_{\underline{k}^\nu}$ also lie in $(E^{\rm d})'_p$. Further, knowing these features of the asymptotic structure of the dual vectors, we can easily write down the appropriate integral equations for them and analyze the solutions of these equations directly. The eigenvectors and their duals provide in this case a generalized regular biorthogonal system which leads to the following results:

Theorem 9: If $\{\underline{\Psi}^{\rm p}_{\underline{k}^\nu},\underline{\Psi}^{\rm s}_{\underline{k}^\nu}\}$ is a basis for $\mathscr{C}_p$ and $\{\underline{\xi}'^{\rm p}_{\underline{k}^\nu},\underline{\xi}'^{\rm s}_{\underline{k}^\nu}\}$ the generalized associated sequence of coefficient functionals, then

$$\mathbf{E}(\delta) = \sum_\nu \int_{\lambda^{\rm p}_{\underline{k}^\nu}\in\delta} d\underline{k}^\nu \underline{\Psi}^{\rm p}_{\underline{k}^\nu}\underline{\xi}'^{\rm p}_{\underline{k}^\nu} + \sum_\nu \int_{\lambda^{\rm s}_{\underline{k}^\nu}\in\delta} d\underline{k}^\nu \underline{\Psi}^{\rm s}_{\underline{k}^\nu}\underline{\xi}'^{\rm s}_{\underline{k}^\nu} \tag{4.11}$$

(where δ are Borel sets in the complex plane) is a countably additive resolution of the identity for $\mathbf{H}$. Furthermore, $\mathbf{H}$ is scalar spectral and given by

$$\mathbf{H}\underline{\Psi} = \lim_{N\to\infty}\Big(\sum_\nu \int_{|\underline{k}^\nu|\le N} d\underline{k}^\nu \lambda^{\rm p}_{\underline{k}^\nu}(\underline{\xi}'^{\rm p}_{\underline{k}^\nu},\underline{\Psi})\underline{\Psi}^{\rm p}_{\underline{k}^\nu} + \sum_\nu \int_{|\underline{k}^\nu|\le N} d\underline{k}^\nu \lambda^{\rm s}_{\underline{k}^\nu}(\underline{\xi}'^{\rm s}_{\underline{k}^\nu},\underline{\Psi})\underline{\Psi}^{\rm s}_{\underline{k}^\nu}\Big) \tag{4.12}$$

for $\underline{\Psi}\in\mathrm{dom}(\mathbf{H})$. Here $|\underline{k}^\nu|$ is defined in some natural way.

Proof: Formally the proof is the same as for Theorem 8. Boundedness of the $\mathbf{E}(\delta)$ follows from closure of the corresponding subspace. Uniform boundedness follows from their countable additivity.[36]

A functional calculus may be defined for $\mathbf{H}$ (under these conditions) by

$$\begin{aligned} f(\mathbf{H})\underline{\Psi} &= \lim_{n\to\infty}\int_{\Delta(\mathbf{H})} f_n(\lambda)\mathbf{E}(d\lambda)\underline{\Psi} \\ &= \lim_{N\to\infty}\Big(\sum_\nu \int_{|\underline{k}^\nu|\le N} d\underline{k}^\nu f(\lambda^{\rm p}_{\underline{k}^\nu})(\underline{\xi}'^{\rm p}_{\underline{k}^\nu},\underline{\Psi})\underline{\Psi}^{\rm p}_{\underline{k}^\nu} \\ &\quad + \sum_\nu \int_{|\underline{k}^\nu|\le N} d\underline{k}^\nu f(\lambda^{\rm s}_{\underline{k}^\nu})(\underline{\xi}'^{\rm s}_{\underline{k}^\nu},\underline{\Psi})\underline{\Psi}^{\rm s}_{\underline{k}^\nu}\Big) \end{aligned} \tag{4.13}$$

for all $\underline{\Psi}$ where the right-hand side is convergent. Here $f_n(\ )$ and $\Delta(\mathbf{H})$ are defined as previously. If $\underline{\Psi}\in E_p$, then $(\underline{\xi}'^{\rm p(s)}_{\underline{k}^\nu},\underline{\Psi})$ is well defined. For a rigorous theory this definition must be extended to $\underline{\Psi}\in\mathscr{C}_p$ adopting a suitable limit-in-mean interpretation of the integral.[38] The integral in (4.13) is similarly treated. In later work[35] we give the explicit form for the Faddeev case of the expressions presented here.

Even if the eigenvectors of $\mathbf{H}$ are not complete on $\mathscr{C}_p$, it is of interest to construct an $\mathbf{H}$-invariant projection operator $\mathbf{P}$ onto the imbedded physical solutions (where $P=\Sigma_\alpha P_{\alpha\beta}$ is the corresponding Hilbert space operator). Suppose the physical eigenvectors can be extended in $\mathscr{N}=\{\underline{\Psi}\in\mathscr{C}_p:(\underline{\xi}'^{\rm p}_{\underline{k}^\nu},\underline{\Psi})=0,\forall\underline{k}^\nu\}$ to form a basis for $\mathscr{C}_p$. Then $\bar{\mathbf{P}}$ is associated with the decomposition $\mathscr{C}_p=\overline{\mathrm{span}(\underline{\Psi}^{\rm p}_{\underline{k}^\nu})}\oplus\mathscr{N}$ and has a simple representation analogous to (3.16). Again $\mathbf{H}$ restricted to Range($\mathbf{P}$) is real eigenvalue scalar spectral, *-self-adjoint (with a suitable choice of involution), and equivalent (via a similarity transformation) to H on Range(P). Physical channel space density matrices may be defined in a fashion analogous to the previous section and the corresponding results on agreement of expectation values established.

V. TIME DEPENDENT SCATTERING THEORY AND NONEQUILIBRIUM STATISTICAL MECHANICS

The Möller operators play a central role in scattering theory as they relate the asymptotic form of the wavefunction to the full scattered state wavefunction. The channel space Möller operators may be expressed formally as[20]

$$\mathbf{\Omega}^{\pm} = \text{“}\lim_{t\to\mp\infty}\text{”}\, e^{+i\mathbf{H}t}e^{-i\mathbf{H}_0 t}, \tag{5.1}$$

assuming that $i\mathbf{H}$ generates a C_0 group and that the limits exist in some sense. Note that $\mathscr{C}_p$ may be decomposed into a direct sum of orthogonal subspaces associated with $\mathbf{H}_0$ and with projection operators given by $\mathbf{E}^{\pm}_{\alpha,\alpha'}$, where

$$(\mathbf{E}^{\pm}_{\alpha,\alpha'})_{\beta\gamma} = \delta_{\beta\gamma}\delta_{\alpha\beta}E^{\pm}_{\alpha,\alpha'}. \tag{5.2}$$

Here $\alpha' \subset \alpha$ corresponds to stable clusters and

$$E^{\pm}_{\alpha,\alpha'} = \sum_{m_j}\int d\underline{k}_i\, |\phi^{\pm}_{\alpha,\alpha';m_j,\underline{k}_i}\rangle\langle\phi^{\pm}_{\alpha,\alpha';m_j,\underline{k}_i}|. \tag{5.3}$$

If $\mathbf{\Omega}^{\pm}_{\alpha,\alpha'} = \mathbf{\Omega}^{\pm}\mathbf{E}^{\pm}_{\alpha,\alpha'}$ is the corresponding restriction of $\mathbf{\Omega}^{\pm}$, then formally,

$$\underline{\Psi}^{\pm}_{\alpha,\alpha';m_j,\underline{k}_i} = \mathbf{\Omega}^{\pm}\underline{\phi}^{\pm}_{\alpha,\alpha';m_j,\underline{k}_i} = \mathbf{\Omega}^{\pm}_{\alpha,\alpha'}\underline{\phi}^{\pm}_{\alpha,\alpha';m_j,\underline{k}_i}. \tag{5.4}$$

If there exist nonreal eigenvalue spurious solutions, then (5.1) can not be used to represent the operators in (5.4), as a divergence results in the $t \to \pm\infty$ limit. The inclusion of the $\mathbf{E}^{\pm}_{\alpha,\alpha'}$ is not expected to circumvent this problem. In the Hilbert space theory existence of the Möller operators in the sense of the strong limit may be demonstrated in some cases by a combination of Cook's method and a stationary phase analysis (see Reed and Simon[42]). With an extra assumption, this approach may be adopted here.

Theorem 10: (Modified Cook's method): Suppose $i\mathbf{H}$ generates a uniformly bounded C_0 group, and suppose there is a set $\mathscr{D} \subset \mathrm{dom}(\mathbf{H}_0) \cap \mathbf{E}^{\pm}_{\alpha,\alpha'}\mathscr{C}_p$ which is dense in $\mathbf{E}^{\pm}_{\alpha,\alpha'}\mathscr{C}_p$, so that for any $\underline{\phi}\in\mathscr{D}$, there is T_0 satisfying:

(a) for $|t| > T_0$, $e^{-i\mathbf{H}_0 t}\underline{\phi}\in\mathrm{dom}(\mathbf{H}_0)$,

(b) $\displaystyle\int_{T_0}^{\infty} dt\, \|\mathbf{V}e^{\mp i\mathbf{H}_0 t}\underline{\phi}\|_p < +\infty.$

Then $\mathbf{\Omega}^{\pm}_{\alpha,\alpha'}$ exist in the sense of the "strong limit" on $\mathscr{C}_p$.

Proof: Analogous to the Hilbert space version[42], except the uniformly bounded C_0 group assumption rather than a unitary group property is used.

This result may also be extended in the direction of the Kupsch–Sandhas theorem.[42] If there exist any spurious solutions with nonreal eigenvalues, then the conditions of the theorem are not satisfied since $e^{i\mathbf{H}t}$, if it exists, will not be uniformly bounded. It remains to control $\|\mathbf{V}e^{\pm i\mathbf{H}_0 t}\underline{\phi}\|_p$. In the Hilbert space theory, this is where a stationary phase analysis is used together with some extra "rate of decay" assumptions on the potential. This approach should extend to the channel space problem at least for the case where we consider only two cluster channels. This problem will be dealt with in more detail in later work.

Let us now present the fundamental equation of nonequilibrium statistical mechanics. Define a physical channel space density matrix $\boldsymbol{\rho} = \mathbf{P}\rho\mathbf{P}$ as previously, so $\mathbf{P}\boldsymbol{\rho}\mathbf{P} = \boldsymbol{\rho}$ [where $\mathbf{P}$ is the projection operator onto the physical solutions defined in Secs. III and IV and $\rho = P\rho P$ is a Hilbert space density matrix with $P = \Sigma_\alpha(P_{\alpha\beta})$]. The channel space von Neumann equation[24]

$$i\hbar\partial/\partial t\, \boldsymbol{\rho} = [\mathbf{H}, \boldsymbol{\rho}] \tag{5.5}$$

may be easily derived from the Hilbert space version[43] and the commutation relation $[\mathbf{P},\mathbf{H}] = 0$. The algebra $\mathscr{A}_{\mathbf{P}} = \{\mathbf{A}\in\mathscr{A}: \mathbf{PAP} = \mathbf{A}\}$ is invariant under (5.5), however, there will be a more general class of solutions.

The calculation of expectation values is of central importance in the theory. As noted by Jauch, Misra, and Gibson,[44] the distance between two states manifests itself in the calculation of expectation values of observables: the states are "close" if the expectation values are close. This dictates the physical choice of topology for the density matrix states to be the weakest in which we have continuity of expectation values. It is convenient to define, for some algebra $\mathscr{B}$, the corresponding trace operator topology

$$\zeta^{p}_{\mathscr{B}}(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2) = \sup_{\substack{A\in\mathscr{B}\\ \|\mathbf{A}\|_{(p)}=1}} |\mathrm{tr}(\mathbf{A}(\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2))|, \tag{5.6}$$

with $\boldsymbol{\rho}_1, \boldsymbol{\rho}_2$ in the trace class of $\mathscr{C}_p$. Then the physical topology is given by $\zeta^{p}_{\mathscr{A}_{\mathbf{P}}}(\ )$. Clearly, using (3.25), $\zeta^{p}_{\mathscr{A}_{\mathbf{P}}}(\ )$ only involves $\mathbf{P}\,\boldsymbol{\rho}_i\,\mathbf{P}$ as expected.

In the formulation of the asymptotic condition of scattering theory [prescribing the "physical" sense of the limit in (5.1)],[44] it is necessary to consider approximations to $\boldsymbol{\rho}$ of the form $e^{+i\mathbf{H}t}e^{-i\mathbf{H}_0 t}\boldsymbol{\rho}^{\mathrm{asym}}e^{+i\mathbf{H}_0 t}e^{-i\mathbf{H}t}$ not in $\mathscr{A}_{\mathbf{P}}$. The sense in which this is close to $\boldsymbol{\rho}$ seems more appropriately described by a stronger topology, say $\zeta^{p}_{\mathscr{A}}(\ )$. It is easily verified that

$$\zeta^{p}_{\mathscr{A}_{\mathbf{P}}}(\boldsymbol{\rho},\mathbf{0}) \leqslant \zeta^{p}_{\mathscr{A}}(\boldsymbol{\rho},\mathbf{0}) \leqslant \gamma_p(\boldsymbol{\rho}). \tag{5.7}$$

In a Hilbert space formulation, the distinction between the appropriately defined metric and trace norm does not arise. For in that case, there is no restriction corresponding to $\mathbf{A}\in\mathscr{B}$ in (5.6). Also the density matrices are positive self-adjoint (in the usual sense) so their trace and the (appropriately defined) trace norm are equal.

We have noted previously that, essentially since $\mathbf{H}$ is not normal, there is no significant advantage in the Hilbert space choice $p = 2$ of channel space norm. In fact the choice $p = 1$ is in a sense more natural. This is suggested by the summation property of operators in $\mathscr{A}$ [since $\|\mathbf{A}\|_{(1)} = \sup_\beta(\Sigma_\alpha\|(\mathbf{A})_{\alpha\beta}\|_{(\mathscr{H})})$, where $\|-\|_{(\mathscr{H})}$ is the operator norm on $\mathscr{H}$] and by the equal component nature of the physical dual eigenvectors ($p = 1$ gives an ℓ^∞ structure for the dual norm and $\|\underline{\zeta}^{\prime p}_n\|_\infty = \|\,|\underline{\Psi}^{p}_n\rangle\|_{\mathscr{H}} = 1$). Although there is little significance in the choice of p for a small number of channels, as the number increases, only $\|\underline{\zeta}^{\prime p}_n\|_q$ for $q = \infty$ (corresponding to a choice $p = 1$) remains bounded. This is presumably also true only for $\gamma_p(\ )$ restricted to $\mathscr{A}_{\mathbf{P}}$ for $p = 1$. This regime is important in the consideration of the thermodynamic limit for spatially confined systems.[26]

ACKNOWLEDGMENTS

The author would like to express his appreciation to Professor D. K. Hoffman for several useful discussions on this work. Ames Laboratory is operated for the U. S. Department of Energy by Iowa State University under Contract

No. W-7405-Eng-82. This research was supported by the Director for Energy Research, Office of Basic Energy Sciences, WPAS-KC-01-03-01-2.

APPENDIX A

Some of the operators appearing in this theory are unbounded on the channel space $\mathscr{C}_p$. Futhermore, these operators are characteristically not self-adjoint or even normal when regarded as acting on $\mathscr{C}_2$. A definition of the resolvent and spectrum suitable for handling such operators is given below (cf. Yosida[10]).

We say that λ is in the resolvent $\rho(\mathbf{A})$ of $\mathbf{A}$ if Range $(\lambda\,\mathbf{I}-\mathbf{A})$ is dense in $\mathscr{C}_p$ and a continuous inverse, denoted $(\lambda\,\mathbf{I}-\mathbf{A})^{-1}$, exists. All complex numbers λ not in $\rho(\mathbf{A})$ form the spectrum $\sigma(\mathbf{A})$ of $\mathbf{A}$. The spectrum is decomposed into disjoint sets, $P\sigma(\mathbf{A})$, $C\sigma(\mathbf{A})$, and $R\sigma(\mathbf{A})$ with the following properties:

$P\sigma(\mathbf{A})$ is the totality of complex numbers for which $\lambda\,\mathbf{I}-\mathbf{A}$ does not have an inverse. For such λ we can find $\underline{\Psi}\neq 0$ in $\mathscr{C}_p$ such that $(\lambda\,\mathbf{I}-\mathbf{A})\underline{\Psi}=\underline{0}\cdots$ the point spectrum.

$C\sigma(\mathbf{A})$ is the totality of complex numbers λ for which $\lambda\,\mathbf{I}-\mathbf{A}$ has a discontinuous inverse with domain dense in $\mathscr{C}_p$ ($\lambda\,\mathbf{I}-\mathbf{A}$ must have a range dense in $\mathscr{C}_p$) $\cdots$ the continuous spectrum.

$R\sigma(\mathbf{A})$ is the totality of complex numbers λ for which $(\lambda\,\mathbf{I}-\mathbf{A})$ has an inverse whose domain is not dense on $\mathscr{C}_p$ (the range of $\lambda\,\mathbf{I}-\mathbf{A}$ is a proper subset of $\mathscr{C}_p$) $\cdots$ the residual spectrum.

Define the compression spectrum

$$\Gamma(\mathbf{A})=\{\lambda\in\mathbb{C}:\ \overline{\mathrm{Range}(\lambda\,\mathbf{I}-\mathbf{A})}\neq\mathscr{C}_p\}.$$

It is clear that $R\sigma(\mathbf{A})\subseteq\Gamma(\mathbf{A})\subseteq P\sigma(\mathbf{A})\cup R\sigma(\mathbf{A})$. Now from the Hahn–Banach theorem $\Gamma(\mathbf{A})\subseteq P\sigma(\mathbf{A}')$. A more usual statement of this result is given for operators on Hilbert space by Weinberg.[1] It is easy to show that also $P\sigma(\mathbf{A}')\subseteq\Gamma(\mathbf{A})$, so these sets are equal. Thus $R\sigma(\mathbf{A})$ is the set of λ which are eigenvalues of $\mathbf{A}'$ [$\lambda\in P\sigma(\mathbf{A}')$] but not of $\mathbf{A}$ [$\lambda\notin P\sigma(\mathbf{A})$]. It is also useful to define the approximate point spectrum

$$\pi(\mathbf{A})=\{\lambda\in\mathbb{C}:\exists\underline{\Psi}_n\ \text{ with }\ \|\underline{\Psi}_n\|_p=1\ \text{ and }\ \|(\lambda-\mathbf{A})\underline{\Psi}_n\|_p\to 0\},$$

which in general overlaps $P\sigma(\mathbf{A})$, $C\sigma(\mathbf{A})$, and $R\sigma(\mathbf{A})$.

Let us now try to locate $\sigma(\mathbf{H})$ for $\widetilde{\mathbf{V}}$ as in (2.19), i.e., $\widetilde{\mathbf{H}}_0$-bounded with relative bound <1. First consider the operator $\widetilde{\mathbf{V}}(\widetilde{\mathbf{H}}_0-\lambda)^{-1}$. Note that $\widetilde{\mathbf{H}}_0$ is a positive self-adjoint operator on $\mathscr{C}_2$.

For $\mathrm{Re}\lambda\geqslant 0, \mathrm{Im}\lambda\neq 0$, we have for any $\underline{\Psi}\in\mathscr{C}_p$,

$$\begin{aligned}&\|\widetilde{\mathbf{V}}(\widetilde{\mathbf{H}}_0-\lambda)^{-1}\underline{\Psi}\|_2\\ &\quad\leqslant a\|\widetilde{\mathbf{H}}_0(\widetilde{\mathbf{H}}_0-\lambda)^{-1}\underline{\Psi}\|_2+b\,\|(\widetilde{\mathbf{H}}_0-\lambda)^{-1}\underline{\Psi}\|_2\\ &\quad\leqslant[a(1+\mathrm{Re}\lambda/|\mathrm{Im}\lambda|)+b/|\mathrm{Im}\lambda|]\|\underline{\Psi}\|_2,\end{aligned}$$

so $\|\widetilde{\mathbf{V}}(\widetilde{\mathbf{H}}_0-\lambda)^{-1}\|_{(2)}<1$ for $|\mathrm{Im}\lambda|>(b+a\,\mathrm{Re}\lambda)/(1-a)$.

For $\mathrm{Re}\lambda\leqslant 0$,

$$\begin{aligned}&\|\widetilde{\mathbf{V}}(\widetilde{\mathbf{H}}_0-\lambda)^{-1}\underline{\Psi}\|_2\\ &\quad\leqslant a\|(\widetilde{\mathbf{H}}_0-\ \mathrm{Re}\lambda)(\widetilde{\mathbf{H}}_0-\lambda)^{-1}\underline{\Psi}\|_2+b\,\|(\widetilde{\mathbf{H}}_0-\lambda)^{-1}\underline{\Psi}\|_2\\ &\quad\leqslant(a+b/|\lambda|)\|\underline{\Psi}\|_2;\end{aligned}$$

using that since $\widetilde{\mathbf{H}}_0$ is positive self-adjoint, $\|\widetilde{\mathbf{H}}_0\underline{\Phi}\|_2\leqslant\|(\widetilde{\mathbf{H}}_0-\ \mathrm{Re}\lambda)\underline{\Phi}\|_2$, so $\|\widetilde{\mathbf{V}}(\widetilde{\mathbf{H}}_0-\lambda)^{-1}\|_{(2)}<1$ for $|\lambda|>b/(1-a)$.

Consequently for $\lambda\in\mathbb{C}$ subject to the above constraints, $\mathbf{1}+\widetilde{\mathbf{V}}(\widetilde{\mathbf{H}}_0-\lambda)^{-1}$ has a bounded inverse defined on all of $\mathscr{C}_p$, so for such λ,

$$(\mathbf{H}-\lambda)^{-1}=(\widetilde{\mathbf{H}}_0-\lambda)^{-1}[1+\widetilde{\mathbf{V}}(\widetilde{\mathbf{H}}_0-\lambda)^{-1}]^{-1}$$

exists as a bounded operator defined on all of $\mathscr{C}_p$, i.e., such $\lambda\in\rho(\mathbf{H})$.

APPENDIX B

Let us consider the equation

$$(\lambda-\mathbf{H})\underline{\Psi}=\underline{\phi},$$

for $\underline{\Psi},\underline{\phi}$ in $\mathscr{C}_p$. This equation may be rewritten in integral form as

$$\underline{\Psi}=\mathbf{G}_0(\lambda)\mathbf{V}\underline{\Psi}+\mathbf{G}_0(\lambda)\underline{\phi},$$

where $\mathbf{G}_0(\lambda)=(\lambda-\mathbf{H}_0)^{-1}$ is the inverse or the semi-inverse of $\lambda-\mathbf{H}_0$ on $\mathscr{C}_p$. If λ is in the resolvent of $\mathbf{H}_0$ then the above equation is valid for all $\underline{\phi}$ in $\mathscr{C}_p$, and if λ is in the continuous spectrum, then it is only valid for $\underline{\phi}\in\mathrm{Range}(\lambda-\mathbf{H}_0)$, a dense subset of $\mathscr{C}_p$.

We shall assume that $(\mathbf{G}_0(\lambda)\mathbf{V})^n$ can be extended to a compact operator for some $n=n^*$, for all λ. Consequently, $\sigma((\mathbf{G}_0(\lambda)\mathbf{V})^{n^*})$ is either a finite set or a countably infinite set accumulating at 0. Since it is easy to check that $(P\sigma(\mathbf{G}_0(\lambda)\mathbf{V}))^{n^*}\subseteq P\sigma((\mathbf{G}_0(\lambda)\mathbf{V})^{n^*})\subseteq\sigma((\mathbf{G}_0(\lambda)\mathbf{V})^{n^*})$, the same is true of $P\sigma(\mathbf{G}_0(\lambda)\mathbf{V})$.

On iterating the integral equation for $\underline{\Psi}$ "n^*-1" times, we obtain

$$(\mathbf{1}-(\mathbf{G}_0(\lambda)\mathbf{V})^{n^*})\underline{\Psi}=\sum_{m=0}^{n^*-1}(\mathbf{G}_0(\lambda)\mathbf{V})^m\mathbf{G}_0(\lambda)\underline{\phi},$$

where $\underline{\phi}$ is in the domain $\mathscr{D}$ of $\Sigma_{m=0}^{n^*-1}(\mathbf{G}_0(\lambda)\mathbf{V})^m\mathbf{G}_0(\lambda)$ which is either all or a dense subset of $\mathscr{C}_p$. Since $(\mathbf{G}_0(\lambda)\mathbf{V})^{n^*}$ can be extended to a compact operator, it follows from the Fredholm alternative that either the above equation has a unique solution for all $\underline{\phi}$ in $\mathscr{D}$, or the corresponding homogeneous equation

$$(\mathbf{1}-(\mathbf{G}_0(\lambda)\mathbf{V})^{n^*})\underline{\Psi}=\underline{0}$$

has a solution.

Let us suppose first that the former is the case. This solution is written as

$$\underline{\Psi}=(\mathbf{1}-(\mathbf{G}_0(\lambda)\mathbf{V})^{n^*})^{-1}\sum_{m=0}^{n^*-1}(\mathbf{G}_0(\lambda)\mathbf{V})^m\mathbf{G}_0(\lambda)\underline{\phi},$$

where $(\mathbf{1}-(\mathbf{G}_0(\lambda)\mathbf{V})^{n^*})^{-1}$ denotes the inverse of the operator $(\mathbf{1}-(\mathbf{G}_0(\lambda)\mathbf{V})^{n^*})$. We must determine whether this solution satisfies the noniterated equation. Therefore we calculate

$$\begin{aligned}
&(\mathbf{1}-(\mathbf{G}_0(\lambda)\mathbf{V})^{n^*})[\mathbf{G}_0(\lambda)\mathbf{V}\underline{\Psi}+\mathbf{G}_0(\lambda)\underline{\phi}-\underline{\Psi}]\\
&=[\mathbf{1}-(\mathbf{G}_0(\lambda)\mathbf{V})^{n^*}][\mathbf{G}_0(\lambda)\mathbf{V}-\mathbf{1}][\mathbf{1}-(\mathbf{G}_0(\lambda)\mathbf{V})^{n^*}]^{-1}\sum_{m=0}^{n^*-1}(\mathbf{G}_0(\lambda)\mathbf{V})^m\mathbf{G}_0(\lambda)\underline{\phi}+[\mathbf{1}-(\mathbf{G}_0(\lambda)\mathbf{V})^{n^*}]\mathbf{G}_0(\lambda)\underline{\phi}\\
&=\left[(\mathbf{G}_0(\lambda)\mathbf{V}-\mathbf{1})\sum_{m=0}^{n^*-1}(\mathbf{G}_0(\lambda)\mathbf{V})^m+(\mathbf{1}-(\mathbf{G}_0(\lambda)\mathbf{V})^{n^*})\right]\mathbf{G}_0(\lambda)\underline{\phi}\\
&=\underline{0}.
\end{aligned}$$

Since we have assumed here that the homogeneous equation has no solutions, we conclude that $\underline{\Psi}$ satisfies the noniterated equation and $\underline{\phi}$ is an arbitrary channel vector in $\mathscr{D}$. Consequently $\lambda\notin R\sigma(\mathbf{H})$.

Secondly, we consider the possibility that the homogeneous equation has a solution. By a standard extension of the usual Riesz–Schauder theory,[10] it is possible to show that $e^{2\pi i m/n^*}$, for some $m=0,1,\ldots,n^*-1$, must be in the point spectrum of $\mathbf{G}_0(\lambda)\mathbf{V}$ [where we have now used the assumption that $\mathbf{G}_0(\lambda)\mathbf{V}$ can be extended to a bounded operator, for all λ, together with the identity $(\lambda^n I-A^n)^{-1}\times(\lambda^{n-1}I+\cdots+A^{n-1})=(\lambda I-A)^{-1}$ for $n=n^*$.] If this is the case for $m=0$, then we have shown that $\lambda\in P\sigma(\mathbf{H})$. So in these circumstances $\lambda\notin R\sigma(\mathbf{H})$. If the above only holds for $m\neq 0$, then we must iterate the integral equation one further time. Using compactness of $(\mathbf{G}_0(\lambda)\mathbf{V})^{n^*+1}$, we show that either the corresponding inhomogeneous equation has a solution for all $\underline{\phi}$ [so $\lambda\notin R\sigma(\mathbf{H})$], or $\mathbf{G}_0(\lambda)\mathbf{V}$ has eigenvalues $e^{2\pi i m/n^*+1}$ for some $m=0,1,\ldots,n^*$. If this is true for $m=0$, then $\lambda\in P\sigma(\mathbf{H})$, so the result $\lambda\notin R\sigma(\mathbf{H})$ is proved. If it is only true for $m\neq 0$, we must iterate again. Repeating this procedure ad infinitum, we show either that $\lambda\notin R\sigma(\mathbf{H})$ or that $P\sigma(\mathbf{G}_0(\lambda)\mathbf{V})$ contains an infinite set of points on the unit circle. If the latter is the case, there must exist an accumulation point for these eigenvalues on the unit circle. This contradicts a previous assertion that the only accumulation point is zero. Thus we have shown $R\sigma(\mathbf{H})=\varnothing$.

A similar analysis applies for the dual of the channel Hamiltonian $\mathbf{H}'$. We can show that $R\sigma(\mathbf{H}')=0$ provided that $\mathbf{G}_0(\lambda)'\mathbf{V}'$ can be extended to a bounded operator, and $(\mathbf{G}_0(\lambda)'\mathbf{V}')^n$ can be extended to a compact operator for some $n=n^{**}$, for all λ. Since $(\mathbf{V}\mathbf{G}_0(\lambda))^{m'}=(\mathbf{G}_0(\lambda)'\mathbf{V}')^m$, it suffices to assume boundedness of $\mathbf{V}\mathbf{G}_0(\lambda)$ and compactness of $(\mathbf{V}\mathbf{G}_0(\lambda))^{n^{**}}$, since this guarantees boundedness (compactness) of the dual.[10]

APPENDIX C

We show here that if $\tilde{\mathbf{V}}$ is $\tilde{\mathbf{H}}_0$-bounded with relative bound '0', then $-\mathbf{H}$ generates a holomorphic (C_0) semigroup in the sector $\{\lambda:|\arg\lambda|<\pi/2\}$. We observe that $-\tilde{\mathbf{H}}_0$ generates such a semigroup, so if $0<w<\pi/2$ in $S=\{\lambda:|\arg\lambda|\leqslant\pi/2+w\}$, we have

$$\|(\tilde{\mathbf{H}}_0+\lambda)^{-1}\|_{(p)}\leqslant m/|\lambda|,$$

where $m>0$ depends on w. An analysis similar to that of Appendix B shows that for $\lambda\in S$,

$$\|\tilde{\mathbf{V}}(\tilde{\mathbf{H}}_0+\lambda)^{-1}\underline{\Psi}\|_p\leqslant[a(1+m)+bm/|\lambda|]\|\underline{\Psi}\|_p.$$

So choosing $0<a<1/(1+m)$, it follows that $\mathbf{1}+\tilde{\mathbf{V}}(\tilde{\mathbf{H}}_0+\lambda)^{-1}$ has a uniformly bounded inverse if $|\lambda|>bm/[1-a(1+m)-\epsilon]$, where $\epsilon>0$ and $\lambda\in S$. So in this region

$$\begin{aligned}
&\|(\mathbf{H}+\lambda)^{-1}\|_{(p)}\\
&\quad\leqslant\|(\tilde{\mathbf{H}}_0+\lambda)^{-1}\|_{(p)}\|[\mathbf{1}+\tilde{\mathbf{V}}(\tilde{\mathbf{H}}_0+\lambda)^{-1}]^{-1}\|_{(p)}\\
&\quad\leqslant m'/|\lambda|,
\end{aligned}$$

where m' depends on m,a,b. Thus, clearly, we can choose $c>bm/[1-a(1+m)-\epsilon]$, $m''>0$ such that if $\mathbf{H}_c=\mathbf{H}+c\mathbf{I}$, then $(\mathbf{H}_c+\lambda)^{-1}$ exists in $\{\lambda:|\arg\lambda|\leqslant\pi/2+w\}$ and

$$\|(\mathbf{H}_c+\lambda)^{-1}\|_{(p)}\leqslant m''/|\lambda|$$

for λ in this sector. Following Kato,[27] we define

$$\mathbf{W}(t)=\frac{1}{2\pi i}\int_\Gamma e^{\lambda t}(\mathbf{H}_c+\lambda)^{-1}\,d\lambda,$$

where $\Gamma=\{\lambda:|\arg\lambda|=\pi/2+w-\epsilon\}$ with $0<\epsilon<w$ and integrated for increasing imaginary part. $\mathbf{W}(t)$ exists for $|\arg t|<w$ (since we can deform the contour so that $|\arg(t\lambda)|>\pi/2$). Kato has shown that $\mathbf{W}(t)$ is a uniformly bounded holomorphic semigroup in the above sector with generator $-\mathbf{H}_c$. The proof uses the fact that $\mathbf{H}$ is closed. So for any w such that $0<w<\pi/2$,

$$e^{-t\mathbf{H}}=e^{tc}\mathbf{W}(t)\quad\text{for}\quad|\arg t|<w,$$

and c depends on w.

APPENDIX D

Let us consider the channel Hamiltonian $\mathbf{H}=\tilde{\mathbf{H}}_0+\tilde{\mathbf{V}}$, where $\tilde{\mathbf{V}}$ includes an external potential spatially confining the N particles to a bounded region $\mathscr{R}$ along the diagonal. Also suppose $\tilde{\mathbf{V}}$ is $\tilde{\mathbf{H}}_0$ bounded with relative bound <1.

If we choose $\lambda\in\rho(\mathbf{H})$ outside the region specified in (2.30), then we have that

$$\|\tilde{\mathbf{V}}(\lambda-\tilde{\mathbf{H}}_0)^{-1}\|_{(2)}<1,$$

so $\mathbf{1}-\tilde{\mathbf{V}}(\lambda-\tilde{\mathbf{H}}_0)^{-1}$ has a bounded inverse. So we write

$$\mathbf{G}(\lambda)=(\lambda-\mathbf{H})^{-1}=(\lambda-\tilde{\mathbf{H}}_0)^{-1}[\mathbf{1}-\tilde{\mathbf{V}}(\lambda-\tilde{\mathbf{H}}_0)^{-1}]^{-1}.$$

But $\tilde{\mathbf{H}}_0$ may be regarded as $N_{\rm ch}$ copies of the Dirichlet–Laplacian on the bounded region $\mathscr{R}$. Since $\lambda\in\rho(\tilde{\mathbf{H}}_0)$ it follows that $(\lambda-\tilde{\mathbf{H}}_0)^{-1}$ is compact (Reed and Simon[12]). Therefore $\mathbf{G}(\lambda)$ is compact as the product of a compact and a bounded operator. So $\sigma(\mathbf{G}(\lambda))=\{X_n(\lambda)\}$, where $X_n(\lambda)\to 0$ as $n\to\infty$ [set $X_\infty(\lambda)=0$]. Each X_n except $X_\infty=0$ is an isolated eigenvalue of finite multiplicity. So there exists $\underline{\Psi}_n\in\mathscr{C}_p$ such that

$$\mathbf{G}(\lambda)\underline{\Psi}_n=X_n\underline{\Psi}_n,\quad n=1,2,\cdots.$$

We now determine $\sigma(\mathbf{H})$ using this information about $\mathbf{G}(\lambda)$. Since $\underline{\Psi}_n \in \mathrm{dom}(\mathbf{H})$ $[\mathbf{G}(\lambda):\mathscr{C}_p \to \mathrm{dom}(\mathbf{H})]$,

$$\mathbf{H}\underline{\Psi}_n = (\lambda - X_n^{-1})\underline{\Psi}_n, n = 1,2,3,\cdots.$$

So $\{\lambda - X_n^{-1}(\lambda): n = 1,2,\cdots\} \subseteq P\sigma(\mathbf{H})$. Now choose $X \notin \{0, X_n : n = 1,2,\cdots\}$. Then $(X - \mathbf{G}(\lambda))$ is bounded invertible and has a bounded inverse. So it is one to one and onto. Suppose $\lambda - 1/X$ is an eigenvalue of $\mathbf{H}$ with eigenvector $\underline{\Psi}$. Then $\mathbf{G}(\lambda)\underline{\Psi} = X\underline{\Psi}$ (a contradiction). So $\lambda - 1/X \notin \mathbf{P}\sigma(\mathbf{H})$. Consider next Range$(\lambda - 1/X - \mathbf{H})$. The equation

$$(\lambda - 1/X - \mathbf{H})\underline{\Psi} = \underline{\phi}$$

may be rewritten [since $\mathbf{G}(\lambda)$ has an inverse] as

$$[X - \mathbf{G}(\lambda)]\underline{\Psi} = XG(\lambda)\underline{\phi},$$

and since $X - \mathbf{G}(\lambda)$ is onto, this equation has a solution $\underline{\Psi}$ $[\in \mathrm{dom}(\mathbf{H})]$ for all $\underline{\phi} \in \mathscr{C}_p$. So Range $(\lambda - 1/X - \mathbf{H}) = \mathscr{C}_p$. Finally, we note that $\mathbf{G}(\lambda - 1/X) = (\lambda - 1/X - \mathbf{H})^{-1}$ is bounded since

$$\mathbf{G}(\lambda - 1/X) = X\,\mathbf{G}(\lambda)[X - \mathbf{G}(\lambda)]^{-1}$$

We conclude that

$$\sigma(\mathbf{H}) = P\sigma(\mathbf{H}) = \{\lambda - X_n(\lambda)^{-1} : n = 1,2,\cdots\}.$$

For the case where the confinement is achieved by imposing a periodic boundary condition the analysis is similar. The operator $\mathbf{H}$ with the property that the resolvent is compact for some λ is called discrete.[36]

APPENDIX E

Consider the $\alpha' =$ complete breakup scattering solutions where the Hilbert space eigenvector is characterized by the planewave $|\phi_0\rangle$ in the asymptotic complete breakup region. The asymptotic part of the corresponding channel space solution can be chosen in any of the $N_{\rm ch}$ positions α since $\alpha' \subset \alpha$ for all α. Choose one of these to be the physical eigenvector, say $\alpha = \alpha^*$. This solution is naturally associated with the numerical vector $\underline{\theta}^1$ where $\theta^1_\alpha = \delta_{\alpha,\alpha^*}$. Choose $\underline{\theta}^2, \underline{\theta}^3, \ldots, \underline{\theta}^{N_{\rm ch}}$ to be linearly independent numerical spurious vectors. Spurious wavelike solutions may then be constructed by taking corresponding linear combinations of the above $N_{\rm ch}$ physical scattering solutions. Thus in the asymptotic complete breakup region, the planewave part of these solutions is

$$\underline{\theta}^i|\phi_0\rangle, \quad i = 1,2,\ldots,N_{\rm ch}.$$

Let $\underline{\Phi}^{1\prime} = (1,1,\ldots), \underline{\Phi}^{2\prime}, \ldots \underline{\Phi}^{N_{\rm ch}\prime}$ be the set of numerical vectors biorthogonal to the $\underline{\theta}^i$. Clearly, the only way to achieve δ-function biorthogonality of wavefunctions is if the plane wavelike part of the dual of the channel vector associated with $\underline{\theta}^i$ in the asymptotic complete breakup region is

$$\underline{\Phi}^{i\prime} \langle\phi_0|, \quad i = 1,2,\ldots,N_{\rm ch}.$$

Of course, for the physical dual,

$$\underline{\Phi}^{1\prime}\langle\phi_0| = (\langle\phi_0|, \langle\phi_0|, \cdots).$$

If we consider scattering solutions for general α', corresponding to stable clusters characterized asymptotically by the planewave $|\phi_{\alpha'}\rangle$, then the analysis is similar. This time the asymptotic part of the physical channel space scattering solutions can be chosen in any of $M \leqslant N_{\rm ch}$ (say) positions α for which $\alpha' \subset \alpha$. Consequently, here suitable M component vectors $\underline{\theta}^i$ and their biorthogonal duals $\underline{\Phi}^{i\prime}$ $i = 1,2,\ldots,M$ are constructed. These determine the asymptotic planewave part of the physical and spurious solutions, and their duals which are confined to the M channels described. The dual vectors may also have this planewave asymptotic structure in other channels (the physical dual has equal components in all channels). This structure for the spurious duals is not determined from biorthogonality.

The inhomogeneous term $\underline{\xi}'$ in the integral equations satisfied by the dual vectors is, however, completely determined by the $\underline{\Phi}^{i\prime}$. There can clearly not be any nonzero components in $\underline{\xi}'$ outside the M channels described, or the integral equations

$$\underline{\xi}^{\pm\prime}_{k^\nu} = \underline{\xi}' + \underline{\xi}^{\pm\prime}_{k^\nu}\mathbf{V}\mathbf{G}_0^{\mp}(E)$$

would not reduce back to the differential form on applying the operator $E - \mathbf{H}_0$. Specifically, $\underline{\xi}' \sim \underline{\Phi}^{i\prime}\langle\phi_{\alpha'}|$ in the asymptotic α' tube and $\underline{\xi}'\cdot(E - \mathbf{H}_0) = \underline{0}'$.

The equations for the physical duals reduce to LS-GT equations after explicitly setting their components equal.

[1]S. Weinberg, Phys. Rev. **133**, B232 (1964).

[2]E. F. Redish, *Lectures on the Quantum Three Body Problem* (University of Maryland, College Park, Md., 1978).

[3]V. Vanzani, in *Few Body Nuclear Physics* (International Atomic Energy Commission, Vienna, 1978); G. Cattapan and V. Vanzani, Lett. Nuovo Cimento **24**, 391 (1979); L. Lovitch and V. Vanzani, *ibid.* **26**, 65 (1979).

[4]G. Bencze, Nucl. Phys. A **210**, 568 (1973); E. F. Redish, Nucl. Phys. A 235, **16** (1974); I. H. Sloan, Phys. Rev. C **6**, 1945 (1972).

[5]D. J. Kouri and F. S. Levin, Bull. Am. Phys. Soc. **19**, 489 (1974); Phys. Rev. A **10**, 1616 (1974); Phys. Lett. B **50**, 421 (1974); Nucl. Phys. A **250**, 127 (1975); **253**, 395 (1975); M. Baer and D. J. Kouri, J. Math. Phys. **14**, 1637 (1973); W. Tobocman, Phys. Rev. C **9**, 2406 (1974).

[6]O. A. Yakubovskii, Sov. J. Nucl. Phys. **5**, 937 (1967).

[7]W. Sandhas, Acta Phys. Austr. Suppl. **13**, 679 (1974); Czech. J. Phys. B **25**, 251 (1975); P. Grassberger and W. Sandhas, Nucl. Phys. B **2**, 181 (1967); E. O. Alt, P. Grassberger, and W. Sandhas, JINR Report E4-6688 (1972).

[8]V. Vanzani, Lett. Nuovo Cimento **16**, 1 (1976).

[9]W. N. Polyzou and E. Redish, Ann. Phys. **119**, 1 (1979); W. N. Polyzou, Ph.D. thesis, University of Maryland (1979).

[10]K. Yosida, *Functional Analysis* 4th ed. (Springer, New York, 1974).

[11]L. D. Faddeev, *Mathematical Aspects of the Three Body Problems in Quantum Scattering Theory* (Davey, New York, 1965).

[12]M. Reed and B. Simon, *Methods of Modern Mathematical Physics IV: Analysis of Operators* (Academic, New York, 1978).

[13]V. Vanzani, in *Few Body Systems and Nuclear Forces* I, Lecture Notes in Physics, Vol. 82 (Springer, New York, 1978); Lett. Nuovo Cimento **23**, 586 (1978).

[14]C. Chandler and A. G. Gibson, J. Math. Phys. **14**, 1328 (1973); **18**, 2336 (1977); **19**, 1610 (1978).

[15]C. Chandler and A. G. Gibson, "A Two-Hilbert-Space Formulation of Multichannel Scattering Theory" (Conference on Mathematical Methods and Applications to Scatting Theory, Washington, D. C., 1979, and University of New Mexico preprint).

[16]W. V. Petryshyn, Bull. Amer. Math. Soc. **81**, 223 (1975); Proc. Symp. Pure Math. **18**, Part 1, 206 (1970).

[17]K. L. Kowalski, in *Few Body Systems and Nuclear Forces*, Lecture Notes in Physics, Vol. **87** (Springer, New York, 1978).

[18]Y. Hahn, D. J. Kouri, and F. S. Levin, Phys. Rev. C **10**, 1615 (1974).

[19]R. Goldflam and D. J. Kouri, Chem. Phys. Lett. **34**, 594 (1975).

[20]D. J. Kouri, H. Krüger, and F. S. Levin, Phys. Rev. D **15**, 1156 (1977).

[21]C. Chandler, J. Nucl. Phys. A **301**, 1 (1978).

[22]W. Sandhas, in *Few Body Problems* (North-Holland, Delhi, 1976).

[23]F. S. Levin and H. Krüger, Phys. Rev. A **15**, 2147 (1977).

[24]D. K. Hoffman, D. J. Kouri, and Z. H. Top, J. Chem. Phys. **70**, 4640

(1979).
[25]S. K. Adhikari and W. Glöckle, Phys. Rev. C **19**, 616 (1979); **19**, 2121 (1979).
[26]J. W. Evans and D. K. Hoffman, "The Reactive Quantum Boltzmann Equation from the Arrangement Channel BBGKY Hierarchy," 14th IUPAP International Conference on Thermodynamics and Statistical Mechanics, Edmonton, 1980.
[27]T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1966); Trans. Amer. Math. Soc. **70**, 195 (1951).
[28]P. G. Federbusch, Phys. Rev. **148**, 1551 (1966); R. G. Newton, Phys. Rev. **153**, 1502 (1967).
[29]F. S. Levin, in *Few Body Problems and Nuclear Forces*, Lecture Notes in Physics, Vol. 82 (Springer, New York, 1978).
[30]J. R. Peters, private communication and Ames Laboratory (Iowa State University) Research Report; A. Pazy, *Semi-groups of Linear Operators and Applications to Partial Differential Equations*, Lecture Notes #10 (University of Maryland, College Park, Md., 1974).
[31]Let $\mathbf{P}$, P be defined as in Sec. III. For any bounded Hilbert space observable $A = PA\,P$, we may define the corresponding channel space observable, restricted to $\mathbf{P}\mathscr{C}_p$, by the real eigenvalue scalar spectral operator $\mathbf{A} = \mathbf{P}\mathrm{A}\,\mathbf{P}$. For the subalgebra $\{\mathbf{B} = \mathbf{P}B\,\mathbf{P}, B\in B(\mathscr{H})\}$, we may define an involution * by $\mathbf{B}^* = \mathbf{P}B^+\,\mathbf{P}$. Then the observables on $\mathbf{P}\mathscr{C}_p$ are *-self-adjoint. If $P = I$, we have constructed a representation of the C^*-algebra of N-particle quantum mechanics.
[32]G. G. Emch, *Algebraic Methods in Statistical Mechanics and Quantum Field Theory* (Wiley-Interscience, New York, 1972).
[33]I. Singer, *Bases in Banach Spaces I* (Springer, New York, 1970).
[34]N. Dunford and J. T. Schwartz, *Linear Operators, Part I, General Theory*, (Interscience, New York, 1967).
[35]J. W. Evans and D. K. Hoffman, "Faddeev's Equations in Differential Form: Completeness of Physical and Spurious Solutions and Spectral Properties," Ames Laboratory, USDOE, and Department of Chemistry preprint, Iowa State University.
[36]N. Dunford and J. T. Schwartz, *Linear Operators, Part III, Spectral Operators* (Interscience, New York, 1971).
[37]A. F. Ruston, Proc. Lond. Math. Soc., Ser. 2 **53**, 109 (1951).
[38]W. O. Amrein, J. M. Jauch, and K. B. Sinha, *Scattering Theory in Quantum Mechanics* (Benjamin, Reading, 1977).
[39]I. M. Gel'fand and G. E. Shilov, *Generalized Functions*, Vol. 2 (Academic, New York, 1968).
[40]I. M. Gel'fand and N. Ya. Vilenkin, *Generalized Functions*, Vol. 4 (Academic, New York, 1968).
[41]A. Böhm in *Lectures in Theoretical Physics*, Vol. 9A (Gordon and Breach, New York, 1967); *The Rigged Hilbert Space and Quantum Mechanics*, Lecture Notes in Physics, Vol. 78 (Springer, New York, 1978).
[42]M. Reed and B. Simon, *Methods of Modern Mathematical Physics III: Scattering Theory* (Academic, New York, 1979).
[43]R. P. Feynman, *Statistical Mechanics* (Benjamin, Reading, 1972).
[44]J. M. Jauch, B. Misra, and A. G. Gibson, Helv. Phys. Acta **41**, 513 (1968).

Fluctuations of $P(\phi)$ random fields

Z. Haba
Institute of Theoretical Physics, Wroclaw University, Cybulskiego 36, 50–205 Wroclaw, Poland

(Received 7 January 1980; accepted for publication 18 July 1980)

We discuss measures of certain sets in $\mathscr{S}'(R^d)$ related to the behavior of $\sup_x |\phi(g(\cdot - x))|$ (fluctuations of ϕ) for large $|x|$ in an ultraviolet regularized infinite volume $P(\phi)_d$. We investigate also whether the measure of a set is preserved, when the ultraviolet cut-off is removed.

PACS numbers: 03.70. + k, 02.50.Ey

I. INTRODUCTION

The $P(\phi)$ models of Euclidean field theory (see Refs. 1–3) correspond to a measure on $\mathscr{S}'(R^d)$ ($d<4$ if there is no ultraviolet regularization). This measure can be defined as a limit, in the sense of the convergence of the characteristic function or moments, of a finite volume Λ ultraviolet regulrized μ_κ^Λ measure defined as

$$d\mu_\kappa^\Lambda(\phi) = \left(\int \exp\left[-\int_\Lambda P(\phi_\kappa(x))\right] d\mu_0(\phi)\right)^{-1} \times \exp\left[-\int_\Lambda P(\phi_\kappa(x))\right] d\mu_0(\phi), \tag{I.1}$$

where μ_0 is the Gaussian measure with the covariance $(-\Delta + m_0^2)^{-1}(x,y)$, P is a polynomial of order $2n$ bounded from below, and $\phi_\kappa(x) = \tilde{\phi}(\omega_\kappa(p)e^{ipx})$ with a certain function ω_κ such that $\langle \phi_\kappa^2 \rangle = \int d\mu_0\, \phi_\kappa^2(x) < \infty$.

For the small coupling a convergent expansion is known for the moments of the limit measure μ ($\kappa = \infty, \Lambda = R^d$; in $P(\phi)_2$ and $(\phi^4)_3$). Such an expansion determines the measure μ uniquely. However, it is of little use when asking the question, what is the value $\mu(C)$, which the measure μ takes on a set C. The problem is of some (at least mathematical) interest. In particular, one would like to know the smallest set of measure 1, i.e., the support of μ. The support properties are known to be a useful tool when solving various (physical) problems in the theory of stochastic processes (see, e.g., Ref. 4). As the measure $\mu_\kappa(\Lambda = R^d)$ is the limit of a local perturbation of the Gaussian measure μ_0, one expects that the limit measure will have similar support properties as μ_0, when restricted to the σ–algebra of sets generated by $\phi(f)$ with $\operatorname{supp} f \subset \Lambda$. So, only "at infinity" we can get a non–Gaussian behavior of the random field $\phi(x)$ (this is known at least for $d \leqslant 2$).[2,5,6] In one dimension the behavior of $q(t)$ for $|t| \to \infty$ in $P(\phi)_1$ has been investigated in terms of lim sup by Rosen and Simon,[6] who computed this limit exactly. The corresponding results for Gaussian random fields were known before.[7,8] In this paper we get only a bound from above on $\lim_{|x|\to\infty} \sup \phi(x)(\ln|x|)^{-\gamma}$ in the infinite volume ultraviolet regularized $P(\phi)_d$. This extends the results of Rosen and Simon [who discussed briefly also $P(\phi)_2$]. The bound from above is sufficient to distinguish the interacting random field from the free one. So we can conclude that the infinite volume limit (if it exists) gives a measure which has its support disjoint with the support of the Gaussian measure.

In Euclidean field theory one is rather interested in support properties of the measure μ and its relation to the cut-off measures μ_κ. In Ref. 9 we have got some results in this direction, which are summarized below and will be applied at the end of this paper. Consider the σ–algebra Σ_g generated by the cylinder sets in $\mathscr{S}'(R^d)$ of the form

$Z_g^B(x_1,\ldots,x_n) = \{\phi \in \mathscr{S}' : (\phi_g(x_1),\ldots,\phi_g(x_n)) \subset B \subset R^n\}$,

where $\phi_g(x) = (g'\phi)(x) = \phi(g(\cdot - x))$ is a mapping $\mathscr{S}' \to C^\infty$. We introduce a $\mathscr{L}$–topology in C^∞ (which can be also considered as a topology in $\mathscr{S}'$, which which is weaker then the usual weak topology) saying that $\phi_n \to \phi$ if $\phi_g^n(x) \to \phi_g(x)$ uniformly on every compact sets $K \subset R^d$. We have proved in Ref. 9 that the set of measures $\{\nu_\gamma = g'^{-1}\mu_\gamma\}$, being the restriction of μ_γ to the σ–algebra Σ_g, is under some assumptions weakly conditionally compact.[10] We formulate this result as:

Theorem I.1: Assume (i) μ_γ are invariant under translations in R^d; (ii) $\int \exp\phi(g)\, d\mu_\gamma(\phi)$ is bounded in γ; (iii) $\lim\sup_{|u|\to\infty} |u|^{-\rho} \ln(\int d\mu_\gamma(\phi) \exp u\phi(g))$ is bounded in γ for certain ρ. Then any sequence $\{\gamma\}$ contains a subsequence $\{\gamma_n\}$ such that $\int F(\phi_g)\, d\mu_{\gamma_n}(\phi)$ converges for all bounded continuous functions F. The limit determines a measure μ on Σ_g and

(a) $\lim\sup \mu_{\gamma_n}(C) \leqslant \mu(C)$ for every set $C \in \Sigma_g$ closed in the $\mathscr{L}$–topology;

(b) $\lim\inf \mu_{\gamma_n}(U) \geqslant \mu(U)$ for every open set $U \in \Sigma_g$;

(c) $\lim \mu_{\gamma_n}(A) = \mu(A)$ for a set $A \in \Sigma_g$ such that there exist an opens set U and a closed set C, $U \subset A \subset C$, with $\mu(C - U) = 0$.

In application to the μ_κ measures (I.1), the assumptions (i) and (iii) are fulfilled if we use regularizations preserving Osterwalder–Schrader (O.S.) positivity[11] [$\omega_\kappa(p)$ independent of p_1 or lattice approximation] and if the translation invariant infinite volume limit of μ_κ^Λ exists. The infinite volume limit of μ_κ^Λ is expected to exist on the basis of the uniform bound resulting from the chessboard estimates.[5,12] It seems that the unique (and translation invariant) infinite volume limit of μ_κ^Λ could be obtained for the weak coupling using the cluster expansion.[13] Finally, if the pressure α_∞^κ is bounded in κ then (ii) is fufillled.

II. FLUCTUATIONS OF RANDOM FIELDS

The L^2 properties[14,15] of the random field $\phi_g(x) = \phi(g(\cdot - x))$, $g \in \mathscr{S}(R^d)$ show that when $|x| \to \infty$ it behaves in the L^2 sense like a bounded function. However, as is known for the Gaussian[7,8] as well as for the $P(\phi)$ fields[6], the sample functions $\phi_g(x)$ fluctuate and the range of fluctuations behaves asymptotically like $(\ln|x|)^\gamma$, when $|x| \to \infty$.

 0022-2488/81/081687-05$01.00 © 1981 American Institute of Physics

For Gaussian case and for $P(\phi)_1$[6] the limit as $|x|\to\infty$ of these fluctuations has been computed exactly. We will be able to give here only a bound from above under an assumption, which will be checked for ultraviolet regularized $P(\phi)_d$ fields in the next section. Define first an average over $\phi_g(x)$, which is an analog of $\int_t^{t+1} q(s)\,ds$ from Ref. 6

$$\phi_g(Q_R) = \left(\frac{1}{|Q_R|}\right)\int_{Q_R}\phi_g(x)\,dx, \tag{II.1}$$

where Q_R is a ring in R^d of radius R and volume $|Q_R|$. We denote $E[\;] = \int d\mu\,[\;]$ and assume that the measure μ is *translation invariant*. We have then

Theorem II.1: Assume there exists $\rho > 1$ such that $\lim_{|u|\to\infty}\sup|u|^{-\rho}\ln|E[\exp u\phi(g)]| < \infty$; then (with $g\in C_0^\infty(R^d)$) the following sets have measure 1:

(a) $\{\phi: \sup_R |\phi_g(Q_R)|[\ln(2+R)]^{(1-\rho)/\rho} < \infty\}$ if $|Q_R|$ is fixed;

(b) $\{\phi: \sup_{x_i}|\phi_g(x)|[\ln(2+|x|)]^{(1-\rho)/\rho} < \infty\}$

if the remaining $d-1$ variables are fixed.

We need first the following:

Lemma II.2: If

$$\lim_{|u|\to\infty}\sup|u|^{-\rho}\ln|E[\exp u\phi(g)]| < M(g), \tag{II.2}$$

then $\exp(r\phi(g))^\beta$ is integrable for all r if $\beta < \rho/(\rho-1)$ and for $r\leqslant r_0 = (a(\rho))^{-1}(M(g))^{-1/\rho}$ with $a(\rho) = \rho(\rho-1)^{-1+1/\rho}$ if $\beta = \rho/(\rho-1)$.

This lemma was proved in Refs. 16 and 9 using the inequality $\exp|x|^{\rho/(\rho-1)}\leqslant b\int dy\,\exp(-|y|^\rho)\exp(ay|x|)$, true for $a\leqslant a(\rho)$.

Proof of Theorem II.1: (a) From the Jensen inequality

$$\begin{aligned}\exp\left|\frac{r}{|Q_R|}\int_{Q_R}dx\,\phi_g(x)\right|^{\rho/(\rho-1)} &\leqslant \frac{1}{|Q_R|}\int_{Q_R}dx\,\exp|r\phi_g(x)|^{\rho/(\rho-1)}\\ &\leqslant \frac{1}{|Q_R|}\int_{K_R}dx\,\exp|r\phi_g(x)|^{\rho/(\rho-1)}\\ &\equiv \frac{|K_R|}{|Q_R|}C_R,\end{aligned} \tag{II.3}$$

where K_R is a ball containing Q_R. C_R is integrable by virtue of Lemma II.2 for $r\leqslant r_0$. Let $|Q_R|$ be fixed. If $n-1\leqslant R\leqslant n$ then

$$\frac{|K_{n-1}|}{|K_R|}C_{n-1}\leqslant C_R\leqslant\frac{|K_n|}{|K_R|}C_n.$$

By the ergodic theorem $\lim_{n\to\infty} C_n$ exists, hence C_R is bounded by an almost surely (a.s.) finite random variable C. We can now rewrite Eq. (II.3) in the form, which implies (a),

$$|\phi_g(Q_R)|\leqslant r_0^{-1}\left(\ln\frac{|K_R|}{|Q_R|}C\right)^{\rho-1/\rho}; \tag{II.4}$$

(b) Consider first a sequence of points $x_n = (n,\mathbf{x}_{d-1})$ $\mathbf{x}_{d-1} = (x_2,\ldots,x_d)$. From Lemma II.2 and the ergodic theorem

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\exp|r\phi_g(x_n)|^{\rho/(\rho-1)} = \mathscr{D}(g,\mathbf{x}_{d-1}) \tag{II.5}$$

exists with probability 1 for $r\leqslant r_0$. Hence for every n

$$(1/n)\exp|r\phi_g(x_n)|^{\rho/(\rho-1)}\leqslant\mathscr{D}(g,\mathbf{x}_{d-1}). \tag{II.6}$$

Then (b) follows from the inequality

$$\begin{aligned}|\phi_g(x)| &\leqslant |\phi_g(x_n)| + |\phi_g(x_n)-\phi_g(x)|\\ &\leqslant r_0^{-1}(\ln\mathscr{D}(g,\mathbf{x}_{d-1})n)^{\rho-1/\rho}\\ &\quad + |\phi||g(\cdot-x)-g(\cdot-x_n)|_{\mathscr{S}},\end{aligned} \tag{II.7}$$

where the $\mathscr{S}$ norm is less than $K|x-x_n|$.

Remark: We could prove the theorem without the assumption that $g\in C_0^\infty(R^d)$ using the results of Ref. 16 on Hölder continuity of sample paths. We can get in a similar way,

Theorem II.3: Let T_i be a unit translation in the ith direction. Assume (II.2) and that the measure μ is ergodic under translations $T = \Pi_{i=1}^d T_i^{n_i}$; then with probability 1

$$\begin{aligned}&\sup_{|n_i|>2}|\phi_g(n)|\left(\sum_{i=1}^d\ln|n_i|\right)^{(1-\rho)/\rho}\\ &\leqslant r^{-1}(1+\ln\mathscr{D}_r(g))^{(\rho-1)/\rho}\end{aligned} \tag{II.8}$$

for any $0<r<r_0$ with $\mathscr{D}_r(g) = E[\exp|r\phi(g)|^{\rho/(\rho-1)}]$ and $n = (n_1,\ldots,n_d)$.

Proof: From the ergodic theorem

$$\lim_{N_i\to\infty}\left(\prod_{i=1}^d\frac{1}{N_i}\right)\sum_{k_1=1}^{N_1}\cdots\sum_{k_d=1}^{N_d}\exp|r\phi_g(k)|^{\rho/(\rho-1)}$$

exists and is a constant equal to $E[\exp|r\phi(g)|^{\rho/(\rho-1)}]$. Then Eq. II.8 follows by the same argument as in the proof of Theorem II.2.

Let us still rewrite Eqs. (II.4) and (II.6) in the following form:

Theorem II.4: Under the assumptions of Lemma II.2 with $|Q_R|$ and $\mathbf{x}_{d-1}$ fixed the following sets have measure 1:

(a) $\{\phi: \lim_{R_n\to\infty}\sup|\phi_g(Q_{R_n})|(\ln R_n)^{(1-\rho)/\rho} \leqslant a(\rho)(M(g))^{1/\rho}\}$;

(b) $\{\phi: \lim_{|x_n|\to\infty}\sup|\phi_g(x_n)|(\ln|x_n|)^{(1-\rho)/\rho} \leqslant a(\rho)(M(g))^{1/\rho}\}$

for any increasing sequence of radii and

$$x_n = (n,x_2,\ldots,x_d).$$

Remark: With $|Q_R|$ being fixed $\phi_g(Q_R)$ tends to the mean value of $\phi_g(x)$ on a sphere of radius R. That is why $\phi_g(x_n)$ and $\phi_g(Q_R)$ have the same asymptotic behavior.

Let us note finally that the theorems proved above are true also *on a lattice* (see Ref. 2) i.e., if the measure μ is a measure on sequences $\phi_\delta(n)$, $n\in Z^d$. Denote

$$\phi_\delta(g) = \Sigma_{n\in Z^d}\delta^d g(n\delta)\phi_\delta(n)$$

$$\phi_\delta(g;n) = \sum_r\delta^d g(r\delta-n\delta)\phi_\delta(r)$$

and assume (II.2) for $\phi_\delta(g)$. Then Theorems II.1–II.4 (b) are true for $\phi_\delta(g;x)$ with $x\in Z^d$. If $\phi_\delta(n)=\phi_\kappa(n\delta)$ with regularization κ such that $\phi_\kappa(x)$ is a continuous function, then $\phi_\delta(g) \xrightarrow{\delta\to 0} \phi_\kappa(g) \xrightarrow{\kappa\to\infty} \phi(g)$.

III. BOUNDS ON FLUCTUATIONS OF $P(\phi)$ FIELDS

We shall show first that Eq. (II.2) is true for regularized $P(\phi)$ fields and their powers with $M(g)$ independent of the cutoff. Then the bounds of Theorems II.1 and II.4 hold true with probability 1 with respect to every μ_κ. This does not mean that they are true in the lim $\kappa\to\infty$, as the sets of measure 1 of Theorems II.1 and II.4 are not closed in the $\mathscr{L}$-topology (see Sec. I). We will discuss the lim $\kappa\to\infty$ at the end of this section.

As we have mentioned in Sec. I, we consider regularizations $\omega_\kappa(p)$ independent of p_1 in order to preserve the O.S. positivity.[11] In such a case we have the following bound in $P(\phi)$ theory (I.1) coming from the chessboard estimates[5,12]

$$E_\kappa[\exp:Q_\kappa:(\mathbf{g})] \leqslant \exp\int dx\,[\alpha_\infty^\kappa(P-Q(\mathbf{g}(x))-\alpha_\infty^\kappa(P)] \quad \text{(III.1)}$$

for any polynomial $Q(\mathbf{g},y)=\Sigma_{j=1}^{2n-1}g_jy^j$ if $g_j\in C_0^\infty(R^d)$. Here

$$\alpha_\infty^\kappa(P)=\lim_{\Lambda\to R^d}\frac{1}{|\Lambda|}\ln\int\exp\left(-\int:P(\phi_\kappa(x)):dx\right)\times d\mu_0(\phi) \quad \text{(III.2)}$$

and $P—Q(\mathbf{g})$ means that the interaction $P(y)=\Sigma_{j=1}^{2n-1}a_jy^j$ has been replaced by $P(y)-\Sigma_{j=1}^{2n-1}g_jy^j$. The existence of the limit (III. 2) is again the result of chessboard estimates[5,17] i.e. a consequence of O.S. positivity.

If we introduce the lattice approximation[2] i.e., replace $:P(\phi):$ by $\Sigma_{r\in\Lambda}\delta^d:P(\phi_\delta(r)):$ (with $\phi_\delta(r)=\phi(h_{r,\delta})$, $h_{r,\delta}$ being a certain test function) and impose periodic boundary conditions, then assuming that the infinite volume limit of the lattice theory with periodic boundary conditions exists[18] we can get an analog of Eq. (III.1),[17,19,20]

$$E_\delta[\exp:Q_\delta(\mathbf{g}):] \leqslant \exp\sum_r\delta^d\left[\alpha_\infty^\delta\left(P-Q(\mathbf{g}(r\delta))\right)-\alpha_\infty^\delta(P)\right], \quad \text{(III.3)}$$

where

$$:Q(\mathbf{g};\phi_\delta(r)):=\sum_r\delta^d:\phi_\delta^j:(r)g_j(r\delta)$$

and

$$\alpha_\infty^\delta(P)=\lim_{\Lambda\to R^d}\frac{1}{|\Lambda|}\times\ln\int\exp\left[-\sum_{r\in\Lambda}\delta^d:P(\phi_\delta(r)):\right]d\mu_0^p.$$

μ_0^p is the Gaussian measure with the covariance $(-\Delta+m_0^2)_p^{-1}$ with periodic boundary conditions on Λ.

Then we have the following "Wick lower bound" Ref. 21, (Lemma VII.11):

Lemma. III.1: (Guerra, Rosen, and Simon): Let $P(y)=\Sigma_{j=1}^{2n}a_jy^j$ $(a_{2n}>0)$ then there exists a constant A (depending only on n) such that:

$$:P(\phi_\kappa(x)):-:Q(\mathbf{g}(x);\phi_\kappa(x)):\geqslant -a_{2n}A\,[\langle\phi_\kappa^2\rangle^n+\sigma(a,\mathbf{g}(x))], \quad \text{(III.4)}$$

where

$$\sigma(a,\mathbf{g})=\sum_{j=1}^{2n-1}(|a_{2n-j}|/|a_{2n}|)^{2n/j}+(|g_{2n-j}|/|a_{2n}|)^{2n/j}, \quad \text{(III.5)}$$

and an analogous bound for the lattice theory.

Lemma III.2: There exists a function N of κ and a_j such that

$$\alpha_\infty^\kappa(P-Q(\mathbf{g}))\leqslant N_\kappa(a)+Aa_{2n}\sum_{j=1}^{2n-1}\left(\frac{|g_{2n-j}|}{|a_{2n}|}\right)^{2n/j}. \quad \text{(III.6)}$$

Proof: This follows immediately from the definition (III.2) and the inequality (III.4).

Asume now that $g_{2n-j}=hb_{2n-j}$ $j=1,\ldots,2n-2$ and $g_1=hb_1+g$. From the arithematic-geometric mean inequality we get

$$(|hb_{2n-j}|)^{2n/j}\leqslant\frac{r-1}{r}(|h|^{2n/j})^{r/(r-1)}+\frac{1}{r}(|b_{2n-j}|)^{2nr/j},$$

and from the Hölder inequality

$$|g_1|^{2n/(2n-1)}\leqslant n^{1/(2n-1)}(|hb_1|^{2n/(2n-1)}+|g|^{2n/(2n-1)}).$$

From these two inequalities we obtain the following bound:

$$\alpha_\infty^\kappa(P-Q(\mathbf{g}))\leqslant\tilde N_\kappa(a,b,\epsilon)+A_1(a_{2n})^{-1/(2n-1)}|g|^{2n/(2n-1)}+A_2a_{2n}\sum_{j=1}^{2n-1}\delta_j\left(\frac{|h|}{|a_{2n}|}\right)^{(1+\epsilon)2n/(2n-j)} \quad \text{(III.7)}$$

for any $\epsilon>0$ and

$$\delta_j=\begin{cases}1 & \text{if } b_j\neq 0\\ 0 & \text{for } b_j=0\end{cases}.$$

We now allow b_j and a_j to depend on the cut-off κ assuming only that $a_{2n}(\kappa)\geqslant\tilde a_{2n}>0$. Denote

$$N(\phi_\kappa^r(x))=\sum_{j=1}^r b_j(\kappa):\phi_\kappa^j:(x) \quad \text{(III.8)}$$

and

$$N(\phi_\kappa^r)(h)=\int_N(\phi_\kappa^r(x))h(x)\,dx. \quad \text{(III.9)}$$

From Eq. (III.1) we have $(r<2n)$

$$E_\kappa[\exp\{u\phi_\kappa(g)+vN(\phi_\kappa^r)(h)\}] \leqslant\exp\int dx\,[\alpha_\infty^\kappa(P-ug(x)\phi-vN(\phi^r(x))h(x))-\alpha_\infty^\kappa(P)]. \quad \text{(III.10)}$$

Applying the estimate (III.6) to Eq. (III.10) we get:

Theorem III.3: There exist $A_1A_2,K_\kappa(|\Omega|,a,b,\epsilon)$ such that if h and g have their support in Ω (and $|\Omega|<\infty$), then for any $\epsilon>0$, $|v|>1$, and $r<2n$

$$E_\kappa[\exp\{u\phi_\kappa(g)+vN(\phi_\kappa^r)(h)\}]$$
$$\leqslant K_\kappa(|\Omega|,a,b,\epsilon)\exp\left[A_1(\tilde a_{2n})^{-1/(2n-1)}|u|^{2n/(2n-1)}\int dx\,|g(x)|^{2n/(2n-1)}\right]$$
$$\times\exp\left[A_2\tilde a_{2n}|v|^{(1+\epsilon)2n/(2n-r)}\sum_{j=1}^{r}\delta_j\int dx\left(\frac{|h(x)|}{|\tilde a_{2n}|}\right)^{(1+\epsilon)2n/(2n-j)}\right]. \tag{III.11}$$

The same estimate is true for the lattice theory with

$$\int dx\,F(x)\to\sum_n\delta^dF(n\delta).$$

Proof: The estimate (III.11) follows from Eqs. (III.7) and (III.10) if we note that integration over x in Eq. (III.10) is only over the support of g and h. We have also applied the assumption that $a_{2n}(\kappa)\geqslant\tilde a_{2n}>0$ for all κ.

For $d<3$, K_κ can be matched with $\phi_\kappa-\phi^{21}$ by means of the Duhamel expansion in such a way that (III.11) remains true for $\kappa=\infty$ with a certain constant K. Comparing Eq. (III.11) with Eq. (II.2) we get:

Corollary III.4: The sets of Theorem II.1 and II.4 with $(\rho-1)/\rho=1/2n$ and

$$(\rho-1)/\rho=\frac{r}{2n}+\left(1-\frac{r}{2n}\right)\frac{\epsilon}{1+\epsilon},$$

$$M(g)=A_1(\tilde a_{2n})-1/(2n-1)\int dx|g(x)|^{2n/(2n-1)}$$

and

$$M(g)=A_2\tilde a_{2n}\sum_{j=1}^{r}\delta_j\int dx\left(\frac{|h(x)|}{|\tilde a_{2n}|}\right)^{(1+\epsilon)2n/(2n-j)}$$

(M being independent of κ) for ϕ and $N(\phi^r)$ correspondingly have for each κ, μ_κ measure 1. If $d<3$, the μ measure ($\kappa=\infty$) of these sets is also 1.

The sets in question are not closed in the $\mathscr{L}$-topology, therefore we cannot apply Theorem I.1 to show that the μ measure of these sets is 1 if $d>2$. The bounds on fluctuations depend on the asymptototic behavior of Eq. (II.2), which implies integrability of $\exp|r\phi(g)|^{\rho/(\rho-1)}$. We could show the *inverse statement* that integrability implies the bound (II.2). Then one expects integrability of $\exp(v\phi(g))^{2n}$ with $g\in C_0^\infty(R^d)$, when P is of order $2n$, if $\lim\kappa\to\infty$ is nontrivial, because of the damping factor $\exp(-a_{2n}\phi^{2n})$.

Let us consider now Theorem II.3. Applying Theorem I.1 we get:

Theorem III.5: The set

$$C_R=\left\{\phi:\sup_{|n_i|>2}|\phi_g(n)|\left(\sum_{i=1}^{d}\ln|n_i|\right)^{(1-\rho)/\rho}\leqslant R\right\}$$

if closed in the $\mathscr{L}$-topology. Hence, if $E_\kappa[\exp|v\phi_\kappa(g)|^{\rho/(\rho-1)}]$ is bounded in κ, then under the assumptions of Theorem I.1 there exists an R such that $\mu(C_R)=1$.

Proof: That C_R is closed is a consequence of the inequality

$$|\phi_g(n)|\leqslant|\phi_g^N(n)-\phi_g(n)|+|\phi_g^N(n)|$$

for $\phi^N\to\phi$. Then, $\mu(C_R)=1$ follows from Theorems I.1 and II.3.

Let us consider now $N(\phi_\kappa^r)(x)$ as defined in Eq. (III.8). By smearing out we can get a map $h':N(\phi_\kappa^r)(x)$, $N(\phi_\kappa^r)(h(\cdot-x))=\chi(x)$ of $\mathscr{S}'(R^d)$ into $C^\infty(R^d)$ equipped with the $\mathscr{L}$-topology (cf. Sec. I). We define a measure $\tilde\nu_\kappa=h'\mu_\kappa$ on $C^\infty(R^d)$ by means of its finite dimensional distributions

$$\int d\tilde\nu_\kappa(\chi)\prod_{j=1}^{l}\exp iu_j\chi(x_j)$$
$$=\int d\mu_\kappa(\phi)\prod_{j=1}^{l}\exp[iu_jN(\phi_\kappa^r)(h(\cdot-x_j))]$$
$$=\mathscr{F}_\kappa^h(u_1;x_1,\dots;u_l,x_l). \tag{III.12}$$

Then assume:

(i). There exists a choice of functions $a_j(\kappa)$ ($0\leqslant j\leqslant 2n$, $a_{2n}(\kappa)\geqslant\tilde a_{2n}>0$) and $b_j(\kappa)$ ($j\leqslant r<2n$) such that $\alpha_\infty^\kappa(P)$ and $\alpha_\infty^\kappa(P')$ [where $P'(\phi)=P(\phi)-hN(\phi^r)$] are bounded in κ. Using (i) and Eq. (III.10) we can prove an analog of Theorem I.1: There exists a sequence of cut-offs κ_n and a measure $\tilde\nu$ on $C^\infty(R^d)$ such that $\tilde\nu_{\kappa_n}\to\tilde\nu$ weakly and

$$\tilde\nu(C)\geqslant\lim_{\kappa_n\to\infty}\sup\tilde\nu_{\kappa_n}(C) \tag{III.13}$$

for any set C closed in the $\mathscr{L}$-topology. The measure $\tilde\nu$ can be defined by its finite dimensional distributions $\lim_{\kappa_n\to\infty}\mathscr{F}_{\kappa_n}(u_1x_1;\dots;u_nx_n)$. This defines a random field, which we shall denote $N(\phi^r)_h(x)$ ($N(\phi^r)_h$ if $x=0$) having the characteristic function

$$\mathscr{F}^h(u)=\int\exp iu\chi(0)d\tilde\nu(\chi)$$
$$\equiv E[\exp iuN(\phi^r)(h)]. \tag{III.14}$$

We know nothing about the continuity of $\mathscr{F}^h$ in h, therefore we cannot apply the Minlos theorem in order to determine from Eq. (III.14) a generalized random field $N(\phi^r)(h)$ over $\mathscr{S}'(R^d)$. [If $d=2$ the continuity of $\mathscr{F}^h$ in the $\mathscr{S}$ norm follows from Ref. 22. Then Eq. (III.14) defines a generalized random field: $\phi^r{:}(h)$. However, if $d>2$, $:\phi_\kappa^r{:}(h)$ is not convergent to $:\phi^r{:}(h)$ in the L_p sense even in the free case. For some results on $N(\phi^r)$ in $d=3$, see Refs. 23 and 24.]

Consider now the set $C_L=\{|\chi(0)|\leqslant L\}$. We have

$$\int e^{N(\phi_\kappa^r)(h)}d\mu_\kappa(\phi)=\int e^{\chi(0)}d\tilde\nu_\kappa(\chi)\geqslant e^L\tilde\nu_\kappa\{\chi(0)>L\}$$

and

$$\int e^{-N(\phi_\kappa^r)(h)}d\mu_\kappa(\phi)$$
$$=\int e^{-\chi(0)}d\tilde\nu_\kappa(\chi)\geqslant e^L\tilde\nu_\kappa\{-\chi(0)>L\}.$$

The left-hand side of these inequalities is bounded in κ by Eq. (III.10) and the assumption (i); hence

$\tilde{\nu}_\kappa\{C_L\} \geqslant 1 - Ke^{-L}$.

The set C_L is closed in the $\mathscr{S}'$-topology. So we get from Eq. (III.13) that

$$\tilde{\nu}(C_L) = \nu\{\,|\chi(0)| \leqslant L\,\} \geqslant 1 - Ke^{-L}. \tag{III.15}$$

From Eq. (III.15) we can get:

Lemma III. 6: $E\,[\exp|\lambda N(\phi^r)_h|\,] < \infty$ if $|\lambda| < 1$, and for any $R > 1$ there exist C_n and M such that if $p > M$

$E\,[(N(\phi^r)_h)^{2p}] \leqslant C_n(2p)!R^{2p}$.

Proof:

$$E\,[\exp|\lambda N(\phi^r)_h|\,] = \int \exp|\lambda\chi(0)|\,d\tilde{\nu}(\chi)$$
$$= -\int \exp|\lambda|x\,d\tilde{\nu}\{\,|\chi(0)| > x\}$$
$$= 1 + |\lambda|\int_0^\infty \exp|\lambda|x$$
$$\times \tilde{\nu}\{\,|\chi(0)| > x\}\,dx < \infty$$

by Eq. (III.15). Next, due to Lebesgue monotonic convergence,

$$E\,[\mathrm{ch}(|\lambda|N(\phi^r)_h)]$$
$$= \sum_{p=0}^{\infty}\frac{|\lambda|^{2p}}{(2p)!}E\,[(N(\phi^r)_h)^{2p}] < \infty.$$

This implies the bound on correlations in the theorem.

Next, as $\int d\mu_0(\phi)\,(:\phi^r:(h))^{2p}((2p)!)^{r/2}$ we have:

Corollary III. 7: $N(\phi^r)_h$ for $r > 2$ is not a Wick power of a free field of order higher than two.

Now, making use of Lemma III.6 and repeating the proofs of Sec. II. We can get a weaker form of the statements of Corollary III.4.

Theorem III.8: The sets

(a) $\{\,\sup_R |N(\phi^r)_h(Q_R)|\,[\ln(2+R)]^{(1-p)/p} < \infty\,\}$;

(b) $\{\,\sup_{x_i} |N(\phi^r)_h(x)|\,[\ln(2+|x_i|)]^{(1-p)/p} < \infty, \mathbf{x}_{d-1}$ fixed$\,\}$;

(c) $\left\{\,\sup_{|n_i|>2} |N(\phi^r)_h(n)|\left(\sum_1^d \ln|n_i|\right)^{(1-p)/p} \leqslant A\,\right\}$,

have $\tilde{\nu}$ measure 1 (for some A large enough in c).

Summarizing our results on the behavior of $\phi(x)$ for large x we can see that this behavior depends on the damping factor $\exp(-a_{2n}\phi^{2n})$ of the interacting measure (I.1). It would be useful if we could say something about the dependence of the measure μ on a_{2n} after the removal of the cutoffs Λ and κ. From Corollary III.4 we can see that $a_{2n}^{-1/2n}$ bounds the fluctuations of ϕ_κ. Further statements of this section concerning the $\kappa\to\infty$ limit are a weak form of these bounds on fluctuations and were proven under rather strong assumptions. It seems that it should be possible to characterize the support of μ extracting properly only the fact that $a_{2n}(\kappa) \geqslant \tilde{a}_{2n} > 0$.

ACKNOWLEDGMENT

The author wishes to thank Professor T. Balaban for a discussion.

[1] *Constructive Quantum Field Theory*, Lecture Notes in Physics, Vol. 25, edited by G. Velo and A. Wightman (Springer, Berlin, 1973).
[2] F. Guerra, L. Rosen and B. Simon, Ann. Math. **101**, 111 (1975).
[3] J. Feldman and K. Osterwalder, Ann. Phys. **97**, 80 (1976); J. Magnen and R. Sénéor, Ann. Inst. Henre Poincaré **24**, 95 (1976).
[4] H. Cramer and M. R. Leadbetter, *Stationary and Related Stochastic Processes* (Wiley, New York, 1967).
[5] J. Fröhlich and B. Simon, Ann. Math. **105**, 493 (1977).
[6] J. Rosen and B. Simon, Ann. Prob. **4**, 155 (1976).
[7] P. Colella and O. E. Lanford III, in Ref. 1.
[8] G. A. Hunt, Trans. Amer. Math. Soc. **71**, 38 (1951).
[9] Z. Haba, "Continuity of sample paths and weak compactness of measures" in Euclidean field theory, preprint, 1979.
[10] Yu. V. Prokhorov, Theor. Prob. Appl. **1**, 157 (1956).
[11] K. Osterwalder and R. Schrader, Commun. Math. Phys. **33**, 83 (1973); **42**, 281 (1975).
[12] E. Seiler and B. Simon, Ann. Phys. **97**, 470 (1976).
[13] J. Glimm and A. Jaffe, in Ref. 1.
[14] M. Reed and L. Rosen, Commun. Math. Phys. **36**, 123 (1974).
[15] J. Rosen and B. Simon, Duke Math. J. **42**, 51 (1975).
[16] Z. Haba, Commun. Math. Phys. **69**, 247 (1979).
[17] J. Fröhlich, R. Israel, E. H. Lieb and B. Simon, Commun. Math. Phys. **62**, 1 (1978).
[18] G. A. Baker, Jr., Math. Phys. **16**, 1324 (1975).
[19] F. Guerra, Meth. Math. Th. du Champ Quan., Proceed, Marseille Conf. 1975.
[20] Y. M. Park, J. Math. Phys. **17**, 1143 (1976).
[21] F. Guerra, L. Rosen and B. Simon, Ann. Inst. Henri Poincaré, **A25**, 232 (1976).
[22] J. Glimm and A. Jaffe, Commun. Math. Phys. **44**, 293 (1975).
[23] J. S. Feldman and R. Rączka, Ann. Phys. **108**, 212 (1977).
[24] J. P. Eckmann and H. Epstein, Commun. Math. Phys. **64**, 95 (1979).

Strange solutions to field theories in one spatial dimension

Daniel C. Mattis and Bill Sutherland
Physics Department, University of Utah, Salt Lake City, Utah 84112

(Received 11 December 1980; accepted for publication 13 February 1981)

Many models of interacting particles rely heavily for their solution on restriction to one-dimensional motion and a linearized kinetic energy. We examine this in detail, and find that the linearization can lead to patently strange and possible spurious solutions in first quantization. The usual, correct solutions are obtained only in second quantization. The strange solutions do not reduce to the usual plane wave determinantal solutions, even when the interactions are extinguished, and have the character of a condensed phase — a sort of Wigner lattice—for arbitrary interactions.

PACS numbers: 03.70. + k, 11.10.Ef

We examine a class of rather simple fermion Hamiltonians in one spatial dimension. These are exactly soluble, in the sense that one can exhibit explicitly their "correct," physically acceptable solutions. We also discover a totally new class of solutions for which no simple physical interpretation exists, which we label "strange." It has been very popular recently to linearize the kinetic energy, so as to obtain easy and convenient solutions of difficult problems. We shall examine this premise. Let us start with the simplest case, a one-component theory for N fermions characterized by the following Hamiltonian:

$$H = -i\int_0^L dx\, \psi^\dagger(x)\partial_x\psi(x) + \lambda_1 \int dx \int dx' \psi^\dagger(x)\,\psi(x)\, V(x-x')\psi^\dagger(x')\psi(x') = \mathrm{KE} + \mathrm{PE}. \tag{1}$$

Here $\psi(x)$ is a fermion field operator:

$$\{\psi(x), \psi^\dagger(x')\} = \delta(x-x'), \quad \{\psi(x), \psi(x')\} = 0. \tag{2}$$

In terms of momentum operators c_k, ψ may be written

$$\psi(x) = L^{-1/2}\sum_k e^{ikx}c_k. \tag{3}$$

We also define the particle-density (or "current") operators:

$$\rho_q = \sum_k c^\dagger_{k+q}c_k, \quad \rho_{-q} = \rho_q^+, \quad \text{and } N = \rho_0. \tag{4}$$

While the c_k's anticommute ($\{c_k, c^\dagger_{k'}\} = \delta_{k,k'}$), the ρ_q's commute:

$$[\rho_q, \rho_{q'}] = 0, \quad \text{all } q,q', \tag{5a}$$

as long as the number N is finite. If however, the Fermi–Dirac sea is filled, thus $N=\infty$, the ρ_q's satisfy a different set of commutation relations[1]:

$$[\rho_{-q}, \rho_{q'}] = (qL/2\pi)\,\delta_{q,q'}. \tag{5b}$$

[It is very easy to verify (5b) when both sides of the equation operate on the Fermi–Dirac sea, the state in which all $k<k_F$ are occupied and all $k>k_F$ are unoccupied. The proof of (5b) as an operator identity is given elsewhere.[1]] In connection with (5b) we define a set of Bose operators a_q:

$$\rho_q = (|q|L/2\pi)^{1/2}\times\begin{cases} a_{|q|} & \text{for } q<0 \\ a_q^+ & \text{for } q>0 \end{cases} \tag{6}$$

the a_q's— now defined only for $q>0$— satisfy the standard Bose–Einstein commutation relations:

$$[a_q, a_{q'}] = 0, \quad [a_q, a^+_{q'}] = \delta_{q,q'}. \tag{7}$$

Before proceeding to the exact solution of (1), we require one more algebraic identity, concerning the KE, which takes on the following form:

$$-i\int dx\ \psi^\dagger(x)\,\partial_x\psi(x) = \sum k c_k^+ c_k. \tag{8}$$

After subtraction of the energy of the Fermi–Dirac sea, it can be written as[2]

$$\sum k c_k^+ c_k - \sum_{k<k_F} k = \frac{2\pi}{L}\sum_{q>0}\rho_q^+\rho_q = \sum_{q>0} q a_q^+ a_q. \tag{9}$$

The normal-mode operators a_q were, in fact, first introduced by Tomonaga.[3] It is seen that the free-fermion kinetic energy can be written as the energy of the decoupled normal modes. The solution of (1) comes from the realization that the same situation holds for the PE, which is written as follows:

$$\lambda_1\int dx\int dy\ \psi^+(x)\,\psi(x)\,V(x-y)\,\psi^+(y)\psi(y) = (\lambda_1/L)\sum_{q>0}V_q(\rho_q^+\rho_q + \rho_q\rho_q^+) = (\lambda_1/\pi)\sum_{q>0}(qV_q)(a_q^+a_q + \tfrac{1}{2}), \tag{10}$$

in which V_q is the Fourier transform of $V(x)$, presumed real. Combining the above, we obtain H in the form:

$$H = E_0 + \sum_{q>0}\omega_q a_q^+ a_q, \tag{11}$$

where E_0 is the sum of the ground state and renormalization energies,

$$E_0 = \sum_{k<k_F} k + (\lambda_1/2\pi)\sum_{q>0} qV_q, \tag{12}$$

and the renormalized normal-mode energies are

$$\omega_q = q(1+\lambda_1 V_q/\pi). \tag{13}$$

We now turn to the strange solutions of (1). These are evidently confined to the case $N=$ finite, in which the Dirac–Fermi sea is unfilled. Nevertheless, such a case is frequently considered in conjunction with a cutoff k_c below which no fermions are to be permitted.[4] The limit $k_c \to -\infty$

 © 1981 American Institute of Physics

is taken after the solution is obtained, under the assumption that many of the properties are cutoff independent.

The energy eigenstates of (1) are written as

$$\Psi = F(x_1,x_2,\cdots,x_N)\prod_{n=1}^{n}\psi^+(x_n)|0), \tag{14}$$

where $F(x_1,\cdots)$ satisfies the partial differential equation

$$-i\sum_{n=1}^{N}\frac{\partial F}{\partial x_n}+\lambda_1\sum_{n,m}V(x_n-x_m)F=EF(x_1,\ldots) \tag{15}$$

and E is the energy eigenvalue. F is recognized as the ordinary wavefunction in first quantization. The anticommutation relations applied to (14) limit us to solutions of (15) that are totally antisymmetric, i.e., that change sign under the interchange of any pair of coordinates x_n, x_m. It is these solutions that we now determine. The key observation here is that the KE operator, identical to the total momentum operator,

$$-i\sum_n\frac{\partial}{\partial x_n} \tag{16}$$

commutes with the PE operator

$$\lambda_1\sum_{n,m}V(x_n-x_m). \tag{17}$$

An additional simplification comes from the fact that the KE involves only first derivatives. If $G(x_1,\cdots)$ is defined as the symmetric free field ($\lambda_1=0$) solution of (15) belonging to energy E_1 and $F(x_1,\cdots)$ is the solution of (15) with energy E_2, then it is easily seen that $G(x_1,\cdots)F(x_1,\cdots)$ is also a solution of (15) belonging to energy eigenvalue $E=E_1+E_2$.

Taken together, these remarks determine what can be shown to be, in fact, the most general form of the solution; the argument goes as follows: the kinetic energy depends upon a single variable—the center of mass coordinate; the potential energy does not depend upon it at all. Thus, we may write the eigenfunction as

$$F=\exp\left\{i\sum x_j\left[E-\lambda_1\sum V(x_n-x_m)\right]\Big/N\right\}$$
$$\cdot f(x_N-x_{N-1},\ldots,x_2-x_1). \tag{18}$$

Since the exponential phase is symmetric under permutation of particles, the function f of the relative coordinates will be antisymmetric. As it stands, this is an eigenstate for any value of energy; these states are the scattering states of our system. We now impose periodic boundary conditions on the system; this has the effect of fixing the phase so that $[E-\lambda_1\Sigma V(x_n-x_m)]L\equiv KL=\pi x$ (even integers) for N odd and πx (odd integers) for N even. This, however, can only hold if $\Sigma\, V(x_n-x_m)$ is constant, and thus the function f of the relative coordinates is required, by periodic boundary conditions, to be a product of delta functions fixing the relative separations $x_n-x_{n-1}=r_n$. We then arrive at our final form:

$$F=e^{iKx_1}\prod_{n=2}^{N}\delta(x_n-x_{n-1}-r_n) \tag{19}$$

for $x_1<x_2\cdots<x_N$. For any permutation on this ordering one introduces a factor $(-1)^P$ to satisfy the Pauli principle. Thus, as previously derived, the periodic boundary condition requires $KL=\pi\times$(even integer) for $N=$ odd, and $\pi\times$(odd integer) for $N=$ even. The r_n's are a set of arbitrary nearest neighbor separations, subject only to $d\equiv\Sigma r_n<L$, and can be used to compute x_n-x_m, which we may define as r_{nm}, constants of the motion. (Note: $r_{N1}=L-d$, by periodic boundary conditions.) The energy eigenvalue corresponding to the solution (18) is

$$E=K+\lambda_1\sum_{n,m}V(r_{nm}). \tag{20}$$

In a calculation of the partition function, the kinetic and potential energies contribute separately just as in classical physics; quantum mechanics seems to play a negligible role in discretizing K. The spectrum of energies (20) bears no discernible relation to (11)–(13).

These "strange" solutions may be viewed as the condensation of fermions into a "Wigner lattice" [the value of the r_{nm} which minimizes the total energy (20)].

The "strangeness" is compounded when it is seen that these solutions do *not* reduce to the expected determinantal wavefunctions,

$$F_0=(N!)^{-1/2}\det\|e^{ik_jx_n}\| \tag{21}$$

when $\lambda_1\to 0$. Such determinantal functions are only valid at precisely $\lambda_1=0$, where they can be constructed by taking a linear combination of F's belonging to different sets of $\{r_n\}$, which are degenerate only when the PE is extinguished. So the "intuitively obvious" eigenstates (21) of the noninteracting system are not even the natural limiting functions! This is a particularly clear example of the "tracks of the vanished dinosaurs" that Klauder[5] has remarked in various examples of field theory: The influence of interactions persists in the strange form of the solution even after the coupling constant vanishes.

Our second example concerns Luttinger's model[6] of a two-component field theory. This example is nontrivial in the sense that KE and PE do not now commute. Nevertheless, a set of "strange solutions" persists, differing qualitatively from Luttinger's own solutions[6] in the same way that the strange solutions found above differed from the determinantal functions. But this discrepancy may be academic, for *none* of these are physically acceptable; the physically acceptable solutions are obtained quite differently, by first filling the Fermi–Dirac sea and then examining the operators in the manner first prescribed by Mattis and Lieb.[1] The model Hamiltonian is now

$$\begin{aligned}H=&-i\int dx\,[\psi^+(x)\partial_x\psi(x)-\phi^+(x)\partial_x\phi(x)]\\&+\lambda_1\int dx\,[dx'\,\psi^+(x)\,\psi(x)\,V(x-x')\,\psi^+(x')\,\psi(x')]\\&+\lambda_1\int dy\int dy'\,\phi^+(y)\,\phi(y)\,V(y-y')\,\phi^+(y')\,\phi(y')\\&+2\lambda_2\int dx\int dy\,\psi^+(x)\,\psi(x)\,U(x-y)\,\phi^+(y)\,\phi(y).\end{aligned} \tag{22}$$

The exact eigenstates—found by first filling the Fermi–Dirac seas ($k<k_{F1}$ for the ψ particles, $k>k_{F2}$ for the ϕ particles) and then following the prescriptions of Eqs. (5b)–(13)—

now involve a complete set of q's (positive for the density fluctuations in the ψ particles, negative for ϕ's). The relevant Hamiltonian is

$$H = E_0' = \sum_{q>0} q(1 + \lambda_1 V_q/\pi)(a_q^+ a_q + a_{-q}^+ a_{-q})$$
$$+ (\lambda_2/\pi) \sum_{q>0} U_q q(a_q^+ a_{-q}^+ + \text{H.c.}), \tag{23}$$

where

$$E_0' = \sum_{k<k_{F1}} k - \sum_{k>k_{F2}} k + (\lambda_1/\pi) \sum_{q>0} V_q q. \tag{24}$$

The final step is a Bogoliubov transformation to a new set of normal modes diagonalizing (23):

$$a_q = \cosh u_q b_q + \sinh u_q b_{-q}^+. \tag{25}$$

With

$$\tanh 2u_q = \frac{-\lambda_2 U_q/\pi}{1 + \lambda_1 V_q/\pi} \tag{26}$$

we obtain

$$H = E_0'' + \sum_{\text{all}} \omega_q' b_q^+ b_q, \tag{27}$$

where

$$\omega_q' = |q| \, [(1 + \lambda_1 V_q/\pi)^2 - (\lambda_2 U_q/\pi)^2]^{1/2} \tag{28}$$

and

$$E_0'' = E_0' + \sum_{q>0} (\omega_q' - \omega_q). \tag{29}$$

Turning next to finite N_1 and N_2, we find only solutions of the strange variety. First, write the eigenstates in the form

$$\Psi = F(x_1,\ldots,x_{N1}; y_1,\ldots,y_{N2}) \prod \psi^+(x_1) \cdots \phi^+(y_1) \cdots |0\rangle, \tag{30}$$

and study the eigenvalue equations for F

$$-i \sum \frac{\partial F}{\partial x_n} + i \sum \frac{\partial F}{\partial y_m} + \lambda_1 VF + 2\lambda_2 UF = EF, \tag{31}$$

where

$$V \equiv \sum_{n,m} [V(x_n - x_m) + V(y_n - y_m)],$$
$$U \equiv \sum_{n,m} U(x_n - y_m).$$

Borrowing from the previous procedure, we fix $x_2,\ldots,x_{N1}$ relative to x_1 and similarly for the y's:

$$x_2 = x_1 + r_2, \; x_3 = x_1 + r_2 + r_3, \cdots,$$
$$y_2 = y_1 + r_2', \; y_3 = y_1 + r_2' + r_3', \cdots,$$

where the r's are constants of the motion. Thus, V, which depends only on the r's, is itself a constant of the motion while U depends on the coordinates only through $x_1 - y_1$. We write this dependence explicitly as $U(x_1 - y_1)$.

The partial differential equation (31) reduces then to the simpler problem:

$$-i\left(\frac{\partial}{\partial x_1} - \frac{\partial}{\partial y_1}\right)F + 2\lambda_2 U(x_1 - y_1)F = (E - \lambda_1 V)F. \tag{32}$$

With

$$W(z) \equiv \int^z dx \, U(x) \tag{33}$$

we obtain the most general solution:

$$F = e^{i(k_1 x_1 + k_2 y_1)} e^{-i\lambda_2 W(x_1 - y_1)}$$
$$\times \Pi \delta(x_n - x_{n-1} - r_n) \, \delta(y_m - y_{m-1} - r_m'). \tag{34}$$

Periodic boundary conditions on x yield the magnitude of k_1:

$$k_1 L - \lambda_2 W(L) = p_1 \pi, \tag{35}$$

where p_1 is an even/odd integer depending on whether N_1 is odd/even. A similar equation is constructed for k_2, and the energies are found to be

$$E = k_1 - k_2 + \lambda_1 V \tag{36}$$

independent of λ_2 except through the quantization condition (35). Note the interesting consequence that whenever λ_2 is increased such that $\lambda_2 W(L)$ increases by a multiple of 2π, there is no effect on the eigenstates save a relabeling. Again, the strange solutions appear to have no discernible physical interpretation. The more plausible states guessed by Luttinger[6] can be obtained *only if* $\lambda_1 \equiv 0$. Luttinger was, of course, careful to make this restriction in his original paper[6] as well as the simplifying assumption $W(L) = 0$. By lifting these restrictions, we have pointed out some of the difficulties that might not otherwise have been apparent.

The analysis can be generalized further, in two significant ways, which preserve the dichotomy between acceptable and strange solutions.

First, we can relax the requirements on the potential energy that it depend on differences $x_n - x_m$ or $x_n - y_m$ of the coordinates, and consider two-body forces that depend in an arbitrary way on the coordinates x_n and y_m. On the one hand, the Hamiltonian remains a quadratic form in Tomonaga operators regardless of the nature of the two-body forces, and can always be diagonalized by standard methods. On the other, the strange solutions can always be found because in the $(N_1 + N_2)$-dimensional coordinate space, the kinetic energy is effective only along one axis, viz., $\zeta_1 = (N_1 + N_2)^{-1/2}(x_1 + x_2 + \cdots - y_1 - y_2 \cdots)$, whereas it commutes with all coordinates $\zeta_2, \zeta_3, \ldots, \zeta_{N1+N2}$ along the orthogonal axes. Therefore, the eigenvalue equation for F can always be transformed into an ordinary first-order, linear differential equation in the variable ζ_1 and solved exactly (albeit, with a physically unacceptable result!)

Second, we can introduce spin as a variable. The particles are then labeled according to their spin component (up or down) as well as their motion (right- or left-going). As long as the numbers of particles in each component are conserved, the models remain soluble in both second and first quantization, with the strange solutions found in the latter.

We have found no simple prescription that "heals" the strange solutions. The introduction of a cutoff does not restore a proper form to the eigenstates nor a proper set of dispersion relations to the excitations. Proceeding to the limit $N \to \infty$ does not help the situation.

In a future paper we plan to explore other facets of this interesting area in mathematical physics, and analyze models which are somewhat more complex and interesting than the above, having applications in field theory as well as in condensed matter theory.

ACKNOWLEDGMENTS

We thank David Campbell for stimulating discussions, and AFOSR for partial support of this work.

[1]D. Mattis and E. Lieb, J. Math. Phys. **6**, 304 (1965); E. Lieb and D. Mattis, *Mathematical Physics in One Dimension: Exactly Soluble Models of Interacting Particles* (Academic, New York, 1966); D. Mattis, Phys. Rev. Lett. **36**, 483 (1976).

[2]Interestingly, no physical parameters depend on the magnitude of k_F. The rhs of (9) is proved in Refs. 1.

[3]S. Tomonaga, Progr. Theoret. Phys. (Jpn.) **5**, 349 (1950).

[4]H. Sharatchandra and P. Mitra, Ann. Physics (N.Y.) **123**, 24 (1979).

[5]J. R. Klauder, Ann. Phys. (N.Y.) **117**, 19 (1979).

[6]J. M. Luttinger, J. Math. Phys. **4**, 1154 (1963).

Restrictions on relativistically rotating fluids [a)]

J. Schendel [b)] and J. Winicour
Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

(Received 25 November 1980; accepted for publication 6 February 1981)

We develop an extensive set of inequalities which apply to the surface of a relativistically rotating fluid with asymptotically flat exterior. We explore the physical content of these inequalities by examining the restrictions they impose on the existence of rotating fluid models with Kerr exteriors. In that case, the dominant set of inequalities can be expressed in a simple analytic form. We find for all models with Kerr parameter $a > m$ that there is a finite maximum redshift between observers at the fluid surface and at infinity. However, for all models with $0 < a/m \leq 1$ there is no upper bound to the redshift. In the static limit, as $a/m \rightarrow 0$, a finite redshift maximum emerges in a discontinuous manner. The value of this maximum depends upon the moment of inertia of the static background fluid. We discuss the implications toward the possibility of a high redshift quasar model.

PACS numbers: 04.20.Cv

I. INTRODUCTION

According to general relativity, astrophysical systems with rotational velocities close to the velocity of light should exhibit distinctly non-Newtonian behavior in the analogous way that magnetic systems differ from electrostatic. Of course, the nonexistence of negative mass precludes the drastic case of a purely "gravimagnetic" system whose only sources are pure matter currents having no net mass. Yet, for extremely relativistic, rotating systems, one might still expect such matter currents to play a role of equal importance as mass. Various theoretical investigations have borne this out, although observational evidence is lacking. In the relativistic regime, an increase in the angular velocity of a rotating fluid leads to a decrease in its eccentricity, just opposite to the Newtonian effect. New phenomena, such as ergoregions, arise. For reviews, see Refs. 1–4.

One way or another, it would appear probable that rotation leads to distinctly relativistic effects for systems with sufficient mass to necessitate collapse. If angular momentum were conserved, then rotational velocities would eventually become relativistic with the ultimate collapse to a black hole, at which stage dragging-of-inertial-frame effects become overwhelming. If angular momentum were not conserved, as might be expected in view of the instabilities present in all rotating fluid systems,[5] then general relativistic effects would appear in the form of gravitational radiation which carries off angular momentum. In reality some mixture of these cases is likely to occur.

Stationary, axisymmetric, rotating fluid models have provided one means of investigating such systems, the underlying idea being that sequences of such models should approximate quasistationary evolutionary stages. Many such models, incorporating various equations of state, have been constructed, some by slow motion perturbation theory[6–11] and some by fully relativistic numerical solutions.[12–20] In this paper, rather than constructing detailed models, we investigate some general restrictions which limit the range of such models. These restrictions arise from Einstein's equations under certain global assumptions, namely asymptotic flatness and Euclidean spatial topology. Without global considerations, the construction of a local fluid model in some neighborhood of any pressure-free fluid boundary is always possible.[21] However, such models, unless they could be extended globally, would not be of physical interest. We further simplify the problem by considering only the case of a rigidly rotating fluid whose boundary has spherical topology.

In Sec. II we discuss the inequalities which restrict the range of these models. In addition to previously known results,[22–25] we formulate a new class of integral inequalities. All these inequalities apply directly to the exterior geometry at the fluid boundary. They assume no details of the equation of state of the fluid interior other than positive energy density and pressure.

In Sec. III we examine the physical content of these inequalities, determining which are strongest, in a specialization to fluid models with Kerr exteriors. It would be interesting to known how the ensuing results would be modified by other choices of exterior geometry. Choice of a Kerr exterior leads to tremendous mathematical simplification compared, say, to the more general Tomimatsu–Sato exterior. A Kerr exterior might seem unphysical for describing astrophysical objects because of the special relationship[26] $Q = J^2/M$ it implies between the quadrupole moment Q, the mass M, and the angular momentum J. However, several investigations indicate that a Kerr exterior, or at least the above Kerr relationship, is attained by astrophysical systems in the extreme relativistic limit.[8–13,20] Even in the slowly rotating case, there are no known arguments, including the restrictions of the present paper, which prohibit a Kerr exterior.

In Sec. IV we present graphs which display the allowed parameter values for these models and we discuss the physical significance of the restrictions. Given some quasistatic evolutionary process, the motion of a systems position on these graphs determines whether the quasistatic evolution

[a)]Supported in part by the NSF Contract PHY-800823.
[b)]Part-time graduate research assistant at the Theoretical Applications Division, Los Alamos Scientific Laboratories.

 © 1981 American Institute of Physics

must end before reaching an extreme relativistic stage. Our most interesting results concern the maximum surface redshift allowed by the inequalities. For rotating fluids with Kerr parameter $a \leqslant m$, we find, in the extreme relativistic limit, that the angular velocity must approach the angular velocity of the Kerr black hole which would result were the fluid absent. As a consequence, the fluid boundary must approach the Kerr horizon and the surface redshift is unbounded. However, in the nonrotating limit $a \to 0$, a finite redshift maximum emerges in a discontinuous manner. The value of this maximum depends upon the moment-of-inertia of the nonrotating fluid. All fluid models with Kerr parameter a $> m$ also have a finite redshift maximum.

II. INEQUALITIES

We consider stationary, axisymmetric, asymptotically flat space–times with Euclidean topology, whose matter source is a rigidly rotating fluid having positive energy density μ, positive pressure p, and constant (positive) angular velocity Ω. Einstein's equation may be written in terms of the scalars constructed from the Killing vectors $T^a + \Phi^a$:

$$\lambda_\rho = (\lambda_{00}, \lambda_{01}, \lambda_{11}) = (T^a T_a, T^a \Phi_a, \Phi^a \Phi_a), \tag{1}$$

where $T^a \nabla_a = \partial/\partial t$ and $\Phi^a \nabla_a = \partial/\partial\phi$ define the time coordinate t and azimuthal coordinate ϕ. Einstein's equation then reduces to[24,25,27]

$$D_m(\tau^{-1}\lambda_{[\rho} D^m \lambda_{\sigma]}) = 8\pi(\mu + p)\tau\psi^{-1}\lambda_{[\rho}S_{\sigma]} \tag{2}$$

and

$$(\mu + p)\psi^{-1} D_m \psi = -2D_m p, \tag{3}$$

where D_m denotes covariant differentiation in the 2-space orthogonal to T^a and Φ^a, the bracket denotes antisymmetrization, $\psi^\rho = S^\rho \lambda_\rho$ with $S^\rho = (1, 2\Omega, \Omega^2)$, $\tau^2 = -\lambda^\rho \lambda_\rho = 2(\lambda_{01}{}^2 - \lambda_{00}\lambda_{11})$, and $S_\rho = (\Omega^2, -2\Omega, 1)$. (There is a natural metric $g_{\alpha\beta}$ in the space of Killing scalars[27] which serves to raise and lower Greek indices, e.g., $S_\rho = g_{\rho\sigma} S^\sigma$.) In the Newtonian limit, with gravitational potential V, we have $\lambda_{00} \to -1 - 2V/c^2$, $\lambda_{01} \to 0$, $\lambda_{11} \to r^2 \sin^2\theta$, $\tau \to (2)^{1/2} r \sin\theta$, and $\psi \to -1 - 2V/c^2 + \Omega^2 r^2 \sin^2\theta/c^2$, with respect to speherical coordinates. (In the remainder of this paper we set $G = c = 1$.)

Certain inequalities follow directly from the requirements on the spacelike or timelike character of the Killing vectors:

$$\lambda_{11} \geqslant 0 \text{ and } \tau^2 \geqslant 0, \tag{4}$$

with equality holding, in both cases, on and only on the symmetry axis.[25] The former of these inequalities expresses the demand that the closed circular orbits of Φ^a be spacelike and the latter, that the 2-space spanned by T^a and Φ^a be timelike. Note that λ_{00} may be positive since we do not exclude the possiblity of an ergoregion. Our conditions do exclude the existence of a black hole. (Otherwise τ^2 would vanish on the horizon as well as on the axis.)

As recognized by Boyer,[22] (3) implies that ψ is constant on the pressure-free boundary of the fluid. In the Newtonian limit this reduces to the well-known result that, on isobaric surfaces, the fluid particles have a constant difference between their kinetic and potential energies. Boyer established an inequality on ψ by noting that $\psi = (T^a + \Omega\phi^a)(T_a + \Omega\Phi_a)$, where $T^a + \Omega\Phi^a$ is tangent to the fluid world lines. The timelike character of the worldlines then implies

$$\psi < 0 \tag{5}$$

throughout the fluid. The Boyer inequality actually assumes a stronger form for an exterior Kerr metric, as discussed in the next section. From (3), (5), and the assumption of positive pressure, we also have on the fluid boundary

$$n^a D_a \psi \leqslant 0, \tag{6}$$

where n^a is the outward normal.

An additional set of inequalities follow from the elliptic nature of (2). Contracting (2) with an arbitrary constant "bivector" $A^{[\rho}B^{\sigma]}$, we obtain

$$D_m[\tau^{-1}\alpha^2 D^m(\beta/\alpha)] = 8\pi(\mu + p)\tau\psi^{-1}(\alpha B^\rho + \beta A^\rho)S_\rho, \tag{7}$$

where $\alpha = A^\rho \lambda_\rho$ and $\beta = B^\rho \lambda_\rho$. A choice of A^ρ and B^ρ for which the right side of (7) is positive (negative) rules out the possibility of a local maximum (minimum) for the corresponding function β/α. The inequalities follow from combining this result with the boundary conditions on β/α. The major results are[24,25] (for $\Omega > 0$):

$$\lambda_{01} \leqslant 0, \quad \eta = \lambda_{01} + \Omega\lambda_{11} \geqslant 0, \quad \text{and}$$
$$v = \lambda_{00} + \Omega\lambda_{01} < 0, \tag{8}$$

where, for λ_{01} and η, equality holds on and only on the axis. These three inequalities are trivially satisfied in the Newtonian limit. The first inequality states that the direction of inertial frame dragging agrees uniformly with the direction of Ω. The second makes the same statement about the sense of the fluid's angular momentum density. The third states that the fluid elements must have positive energy, even in the presence of ergoregions. All three inequalities apply throughout the space–time as well as in the fluid interior.

A new set of integral inequalities also results from (7), in the following way. Multiply (7) by $F(\beta/\alpha)$ where F is, for the moment, an arbitrary function of β/α. Next, integrate this product over the volume of the fluid minus a small tube containing the axis, where the integrand might be singular. After taking advantage of the axial symmetry to carry out the ϕ-integration, this gives

$$\int_A F D^m[\tau^{-1}\alpha^2 D_m(\beta/\alpha)]\, dA = 8\pi\int_A F(\mu + p)\tau\psi^{-1} (\alpha B^\rho - \beta A^\rho)S_\rho\, dA,$$

where A is the cross-sectional area of the fluid in the 2-space orthogonal to T^a and Φ^a, between the fluid's boundary and the curve $\tau = \epsilon$, which approaches the axis as $\epsilon \to 0$. Gauss' theorem then gives

$$\int_C F\tau^{-1}\alpha^2 D^m\left(\frac{\beta}{\alpha}\right) dl_m = \int_A F'\tau^{-1}\alpha^2 D^m\left(\frac{\beta}{\alpha}\right) D^m\left(\frac{\beta}{\alpha}\right) dA + 8\pi\int_A F(\mu + p)\tau\psi^{-1}(\alpha B^\rho - \beta A^\rho)S_\rho\, dA, \tag{9}$$

where F' is the derivative of F with respect to its argument β/α, C is the one dimensional boundary of A, and dl_m

$= \epsilon_{mn}\, dx^n$ in terms of the displacement dx^n along C and the alternating tensor ϵ_{mn} of the 2-space orthogonal to T_a and Φ^a. We now obtain an inequality for the line integral over C by choosing A^ρ and B^ρ such that the second integral on the right side of (9) is positive (negative), while choosing F to be an increasing (decreasing) function of β/α so that the first integral on the right is also positive (negative). For this inequality to be useful it must be possible to eliminate the contribution from the curve $\tau = \epsilon$ by passing to the limit $\epsilon \to 0$. Then the left side of (9) reduces to a line integral over the profile of the boundary of the fluid, which necessitates no knowledge of the fluid interior. At issue here is whether the factors of τ^{-1} in (9) permit this limit without a contribution from the axis where $\tau = 0$.

As an example, consider $A_\rho = S^\rho$ and $B^\rho = N^\rho = (0,1,\Omega)$. Then (9) reduces to

$$\int_C F\tau^{-1}\psi^2 D_m\left(\frac{\eta}{\psi}\right) dl_m = \int_A F'\tau^{-1}\psi^2 D^m\left(\frac{\eta}{\psi}\right) D_m\left(\frac{\eta}{\psi}\right) dA. \tag{10}$$

In the case, there is no matter contribution so that any increasing or decreasing function F leads to an inequality for the integral over C. The conditions necessary to eliminate the axis contribution from C follow from considering the behavior of η and ψ near the axis: $\eta = O(\tau^2)$ and $\psi = O(1)$. Then, since dl_a is parallel to $D_a\tau$ along the $\tau = \epsilon$ curve, the axis contribution vanishes provided F vanishes on the axis. Thus, for example, we may choose $F(\eta/\psi) = (-\eta/\psi)^q$, for $q > 0$. Thereupon (10) reduces to

$$\int_B (-\eta/\psi)^q \tau^{-1}\psi^2 D^m(\eta/\psi)\, dl_m \leqslant 0, \tag{11}$$

where B is the profile of the fluid's boundary. The choice $q = 1$ is equivalent to the inequality in Eq. (3.3) of the work of Abramowicz, Lasota, and Muchotrzeb.[23] The inequality (11) also holds in the limit $q \to 0$, as may be checked either by considering the sign of the axis contribution or, alternatively, choosing in (10), F to be a unit step function which vanishes for $(-\eta/\psi) < \epsilon$. This gives

$$\int_B \tau^{-1}\psi^2 D^m(\eta/\psi)\, dl_m \leqslant 0. \tag{12}$$

In the application to Kerr interiors, treated in the next section, we find that (12) is the strongest inequality which results from (10).

While a host of integral inequalities may be extracted from (9), by the analogous process that led from (9) to (10), there appears to be no simple algorithm to decide whether they might yield addition restrictions on fluid models. However, we have not found any integral inequalities, in addition to (12), which give further restrictions on models with Kerr exteriors. This seems partly due to the special nature of the Kerr metric which automatically incoporates many potential inequalities.

Some integral inequalities of special interest result from the choices of A^ρ and B^ρ, with $F = 1$, for which the C integral in (9) reduces to the Komar integral for the fluid's total mass M or total angular momentum J.

For instance, by taking a combination of these integrals, we find

$$M - 2(\Omega + \omega)J \geqslant 0, \tag{13}$$

where $1/\omega$ is given by the maximum value of $(-2\eta/\psi)$ on the fluid boundary B. (This is the optimum choice of ω for which the matter contribution, from the right side of (9), remains positive.) Abramowicz, Lasota, and Muchotrzeb[23] give a discussion of some related integral inequalities, showing how they reduce, in the Newtonian limit, to the Poincare condition and the Newtonian virial theorem.

III. APPLICATION TO KERR INTERIORS

To examine the physical content and relative dominance of the inequalities introduced in Sec. II, we now consider fluid models with Kerr exteriors. We assume, for simplicity, that the boundary of the fluid has spherical topology, corresponding to an object such as a star or galactic core. We apply the inequalities on the fluid boundary B, which is the innermost surface for which they may be determined entirely by the exterior Kerr geometry. (The scalars introduced in Sec. II, as well as their first derivatives, must be continous in a neighborhood of B.)

The Kerr metric, in Boyer–Lindquist coordinates, is[28]

$$ds^2 = \rho^2(dr^2/\Delta + d\theta^2) + (r^2 + a^2)\sin^2\theta\, d\phi^2 - dt^2 + (2mr/\rho^2)(a\sin^2\theta\, d\phi - dt)^2,$$

where $\rho^2 = r^2 + a^2\cos^2\theta$ and $\Delta = r^2 - 2mr + a^2$. The scalars, introduced in Sec. II, take the explicit form

$$\lambda_{00} = -1 + 2mr/(r^2 + a^2\cos^2\theta), \tag{14}$$

$$\lambda_{01} = -2mra\sin^2\theta/(r^2 + a^2\cos^2\theta), \tag{15}$$

$$\lambda_{11} = (r^2 + a^2)\sin^2\theta + 2mra^2\sin^4\theta/(r^2 + a^2\cos^2\theta), \tag{16}$$

$$\tau^2 = 2(r^2 - 2mr + a^2)\sin^2\theta, \tag{17}$$

$$\psi = -1 + \Omega^2(r^2 + a^2)\sin^2\theta + 2mr(1 - a\Omega\sin^2\theta)^2/(r^2 + a^2\cos^2\theta), \tag{18}$$

$$\eta = [\Omega(r^2 + a^2) + 2mar(a\Omega\sin^2\theta - 1)/(r^2 + a^2\cos^2\theta)]\sin^2\theta, \tag{19}$$

with

$$\nu = \psi - \Omega\eta. \tag{20}$$

Note that, since we exclude the existence of a black hole, the inequalities (4) are automatically satisfied.

From (18), we see that condition (5) may be enlarged. Writing K for the value of $\psi + 1$ on B, we must have

$$0 < K < 1. \tag{21}$$

We may also use (18) to write Boyer's equation for the boundary as a quartic in r,

$$P(r, \theta; m, a, \Omega, K) = 0, \tag{22}$$

where

$$P = r^4\Omega^2\sin^2\theta + r^2[a^2\Omega^2\sin^2\theta(1 + \cos^2\theta) - K] + 2mr(1 - a\Omega\sin^2\theta)^2 + a^2\cos^2\theta(a^2\Omega^2\sin^2\theta - K). \tag{23}$$

On the axis, the quartic reduces to the quadratic equation

$$r^2 - 2mr/K + a^2 = 0,$$

with the only possible positive roots

$$r_{\pm} = m/K \pm (m^2/K^2 - a^2)^{1/2}. \tag{24}$$

The quartic also has a positive root $r \to \infty$ associated with its degeneracy as $\sin\theta \to 0$. The inequality (6) is satisfied only for the root r_+. Boyer's equation, applied at the pole, thus requires $r = r_+$ with

$$m^2 \geqslant a^2K^2 \tag{25}$$

for the existence of a spherical boundary.

Similarly, on the equator, the quartic reduces to the cubic equation

$$r^3\Omega^2 + r(a^2\Omega^2 - K) + 2m(1 - a\Omega)^2 = 0,$$

having possible positive roots

$$r_k = 2a[\tfrac{1}{3}(K/a^2\Omega^2 - 1)]^{1/2} \cos\tfrac{1}{3}(\alpha + 2k\pi),$$

where $k = 0$ or 2 and

$$\cos\alpha = -m\Omega(1 - a\Omega)^2[3/(K - a^2\Omega^2)]^{3/2}, \tag{26}$$

with $\pi \leqslant \alpha \leqslant 3\pi/2$. The quartic has an additional root, $r = 0$, resulting from its degeneracy when $\cos\theta = 0$. From (25), the roots r_k are positive if and only if

$$27m^2\Omega^2(1 - a\Omega)^4 \leqslant (K - a^2\Omega^2)^3. \tag{27}$$

It is useful to note that (25) and (27) combine to give the rough inequality

$$4a^2\Omega^2 \leqslant K. \tag{28}$$

The only positive root which satisfies (6) corresponds to r_0 for the choice $k = 0$.

Thus Boyer's condition implies that a spherical boundary must pass through $r = r_+$ at the pole and $r = r_0$ at the equator with inequalities (25) and (27) giving the necessary restrictions on the fluid parameters. The existence of a continuous spherical boundary at intermediate angles would seem to imply additional inequalities, analogous to (25) and (27). However, our numerical calculations indicate that such inequalities are already containd in (24) and (26). We have attempted to use Sturm's algorithm[29] to show that (22) must have three nondegenerate, positive roots for r at all intermediate θ. This would establish the existence of three nonintersecting surfaces connecting $r = r_-$ at the pole with $r = 0$ at the equator, $r = r_+$ at the pole with $r = r_0$ at the equator, and $r = \infty$ at the pole with $r = r_2$ at the equator. This analysis became too algebraically complicated to complete, although we were able to establish the result for certain ranges of the parameters in (22). However, we have numerically checked that Boyer's condition leads to no further inequalities in the fairly exhausitive set of cases described by Figs. 2 and 3.

We now apply the inequalities (8) to the Kerr scalars λ_{01}, η, and ν. From (15), we obtain $a > 0$, which guarantees that the directions of angular velocity and angular momentum agree. From (5) and (20), it follows that the ν inequality holds provided the η inequality is satisfied. There remains the analysis of the η inequality. One might expect, on the basis of the Newtonian limit in which $\eta \to \Omega r^2 \sin^2\theta$, that η monotonically increases along the boundary as θ increases form 0 to $\pi/2$. If such a result held on the boundary of a Kerr interior it would reduce the analysis of the η inequality to a simple investigation of a neighborhood of the axis. We now establish that this is indeed the case. By differentiating (19) with respect to $\sin^2\theta$ and reexpressing the result, using (18), (19), (22), and (23), we obtain

$$\begin{aligned} &\frac{\partial}{\partial \sin^2\theta}\left[\frac{\eta(1 - a\Omega\sin^2\theta)}{\sin^2\theta}\right] \\ &= -\frac{2a\Omega r}{P'(1 - a^2\Omega^2\sin^4\theta)} \\ &\times\{(1 - K)(K - a^2\Omega^2\sin^4\theta) + [(K - 1)a\Omega\sin^2\theta \\ &+ \eta\sin^{-2}\theta(1 - a^2\Omega^2\sin^4\theta)]^2\}, \end{aligned} \tag{29}$$

where $P' = \partial P/\partial r$ is determined by (23). The right side of (29) is positive, since (27) implies $(K - a^2\Omega^2\sin^4\theta) > 0$ and (6) requires $P' \leqslant 0$. Thus we can conclude from (29) that η inequality is satisfied on the entire boundary provided

$$\partial\eta/\partial\sin^2\theta \geqslant 0 \tag{30}$$

at the pole. The η inequality then takes the simple form

$$(2m\Omega/aK^2)[m/K + (m^2/K^2 - a^2)^{1/2}] \geqslant 1. \tag{31}$$

We can also reexpress the integral inequality (12) algebraically in terms of the fluid parameters m, a, Ω, and K. At first sight, this might seem a difficult task since the equation for the fluid boundary B is a quartic. However, taking $S^{[\rho}N^{\sigma]}$ for the bivector $A^{[\rho}B^{\sigma]}$ in (7), we obtain

$$D_m[\tau^{-1}\psi^2 d_m(\eta/\psi)] = 0.$$

This equation, according to Gauss' theorem, allows us to freely deform the integration curve B in the integral inequality (12), into any curve beginning at $(r + r_+, \theta = 0)$ and ending at $(r = r_+, \theta = \pi)$. In particular, the choice $r = r_+$, for this curve, recasts the inequality (12) into a standard form for analytic integration. The resulting algebraic form of the inequality is

$$2ma\Omega^2 + (r_+ - 3m)\Omega + ma/(a^2 + r_+^2) > 0. \tag{32}$$

We may also reexpress analytically some of the other integral inequalities associated with (9). For instance, (13) takes the form

$$a(\Omega + \omega) \leqslant \tfrac{1}{2},$$

where an algebraic expression expression for ω results from recalling that, on the boundary, $\psi = -1 + K$ and that η takes its maximum at the equator, where

$$\eta_{\max} = \Omega(r_0^2 + a^2) + 2mar_0^{-1}(a\Omega - 1).$$

As already remarked, our investigations indicate that such additional integral inequalities are redundant for exterior Kerr models.

IV. DISCUSSION OF RESULTS

The graphs in Figs. 2 and 3 describe the limitations on possible Kerr interiors which arise from the Boyer conditions (25) and (27), the η inequality (31), and the integral inequality (32). As discussed in Sec. IV these inequalities appear to form a maximal set of restrictions for the existence of fluid models within the present framework of inequalities. They permit an extensive parameter range. Figure 1 gives an example of the weaker restrictions resulting from other inequalities, in this case the integral inequality (11) with various values of q, compared to the choice $q = 0$ which gives

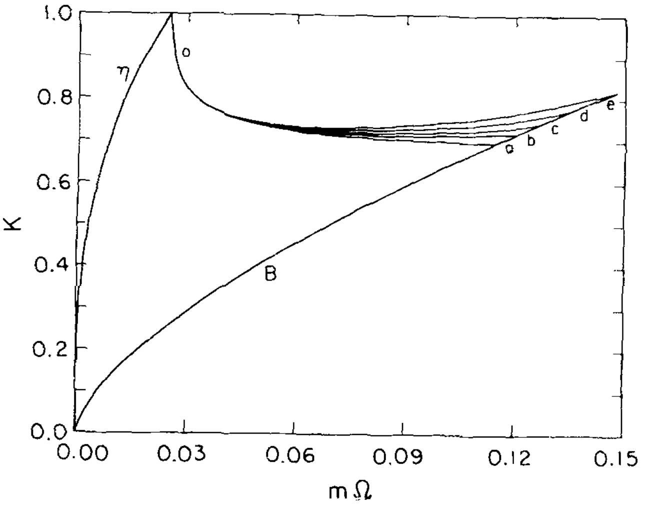


FIG. 1. The curves labelled η and B represent the η inequality and Boyer conditions, respectively. The allowed parameters region is bounded at the top by the curves *oa–oe* which, in order, represent the integral inequality (11) with choices of q equal to 0, $\frac{1}{4}$, $\frac{1}{2}$, 1, and 2. The choice $q = 0$, corresponding to (32), gives the dominant integral inequality.

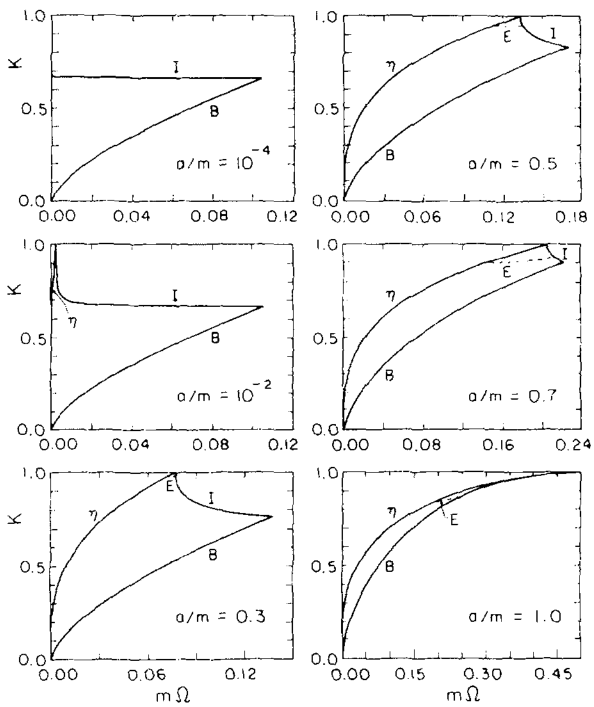


FIG. 3. For $a/m < 1$, the allowed region is three-sided, bounded by the η inequality, the Boyer condition (27), and the curve I representing the integral inequality (32). For $a/m = 1$, the curve I passes through the intersection of the curves η and B and the allowed region becomes two-sided. As $a/m \to 0$, the curve η approaches the K-axis and the curve I becomes step-shaped. Above the broken curve labelled E, the allowed models contain an ergotoroid which extends beyond the equator of the fluid boundary.

rise to (32).

Figure 2 describes the results for values $a/m > 1$. In these cases, the integral inequality (31) is extraneous and the allowed region in the $(\Omega m, K)$ plane is bounded between the η inequality and the Boyer conditions. For a/m less than approximately 5.23, the first Boyer condition (25) is also extraneous and the allowed region is two-sided. For $a/m \geqslant 5.23$, condition (25) becomes operative, introducing a third side $K \leqslant m/a$ to the allowed region. In all cases $a/m > 1$, K is bounded away from its extreme relativistic value $K = 1$. This corresponds to a finite upper bound to the redshift factor between two stationary observers, one on the fluid boundary and the other at infinity.

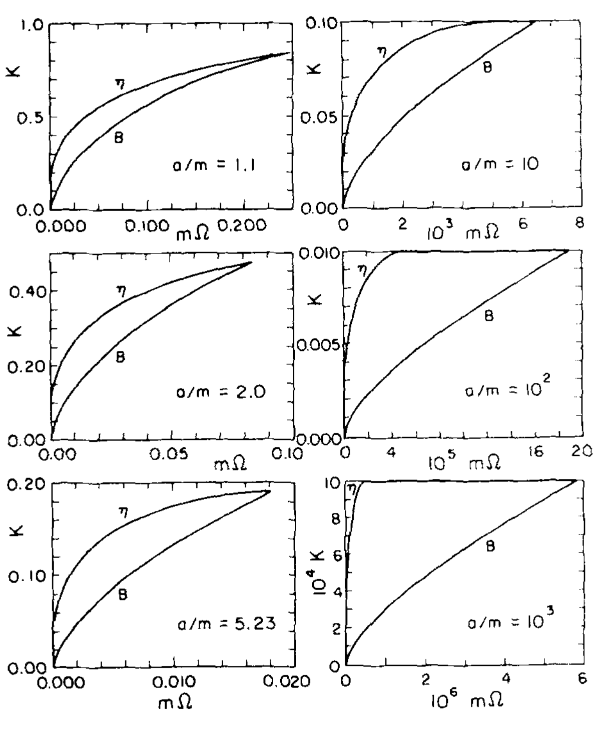


FIG. 2. For $1 < a/m \lesssim 5.23$, the allowed region is two-sided and bounded by the η inequality and the Boyer condition (27). For $a/m \gtrsim 5.23$, the Boyer condition (25) imposes an additional boundary at the top of the allowed region.

Figure 3 describes the results for values $a/m \leqslant 1$. For $a/m < 1$, the allowed parameter range forms a three-sided region in the $(\Omega m, K)$ plane, bounded by the η inequality, the integral inequality (31), and the Boyer condition (27). In the case $a = m$, the side corresponding to the integral inequality shrinks to zero. In each case, the upper vertex is located at the value $K = 1$. This implies a unique relativistic limit Ω_R for the angular velocity of the fluid. We can determine Ω_R from either (31) or (32) by setting $K = 1$. This gives

$$\Omega_{\rm R} = \Omega_{\rm H} = a/2m^2[1 + (1 - a^2/m^2)^{1/2}]^{-1},$$

where $\Omega_{\rm H}$ is the angular velocity of the corresponding Kerr black hole.[28] Consequently, the fluid boundary must approach the Kerr horizon in the extreme relativistic limit. Note that if values $\Omega_{\rm R} > \Omega_{\rm H}$ were allowed in this limit, then the fluid boundary would be exterior to the horizon except at the axis.

It is of interest to inquire whether the inequalities allow the existence of ergoregions inside which $\lambda_{(0)} > 0$. For our models, which contain no black holes, general considerations imply that an ergoregion cannot contain points on the rotation axis, so that they must have toroidal shape.[18] Also, they must intersect the fluid so that the emergence of an ergotoroid along some quasistatic sequency of models must occur inside the fluid.[25] The models of Bardeen and Wagoner,[12] of Wilson,[13] and of Buttorworth and Ipser[18] contain ergotoroids and confirm these properties. Our present investigations, which are confined to the fluid boundary, can only supply rough criteria for the presence of ergotoroids. The condition that an ergotoroid exists and extends to the equator of the fluid boundary is that $r_0 \geqslant 2m$, which corresponds to the allowed regions above the broken curves labelled E, in the graphs of Fig. 3. Our inequalities are compatible with the existence of models with ergotoroids for all $a/m < 1$, although for $a/m \ll 1$ the corresponding parameter range is too small to indicate in Fig. 3. The most interesting feature of the graph is that, for each value of a/m in the range $0 < a/m \leqslant 1$, there is an allowed model with arbitrarily high redshift between stationary observers, one at the fluid boundary and the other at infinity. We now examine what happens to this feature in the static limit of a spherically symmetric fluid with Schwarzschild exterior. In this limit, to first order in a and Ω, the η inequality (31) reduces to

$$R^3\Omega \geqslant 2ma \tag{33}$$

and the integral inequality reduces to

$$R^2(R-3m)\Omega + ma \geqslant 0, \tag{34}$$

where R represents the value, at the fluid boundary, of the Schwarzschild coordinate r. (In terms of this coordinate, $K = 2m/R$ for the background.) To analyze the content of (33) and (34), we utilize the moment of inertia of the background,

$$I = \frac{\partial J}{\partial \Omega}\Big|_{\Omega=0},$$

as formulated by Hartle.[6] We may set $I = \alpha m R^2$, where α ranges from 0 to 1, corresponding to the two extreme density distributions for the fluid ball: in one case ($\alpha = 0$) the mass concentrated at the center; and in the other ($\alpha = 1$) concentrated at the fluid boundary. The Kerr relationship $J = ma$ allows us to set $a = \alpha R^2\Omega$ and thereby eliminate the non-background quantities from (33) and (34). In this way, we find that (33) is automatically satisfied and that (34) reduces to

$$K \leqslant 2/(3-\alpha) \tag{35}$$

or equivalently, $R \geqslant (3-\alpha)m$. Thus there is a *finite* upper limit to the background redshift (except in the thin shell limit $\alpha \to 1$). This result emerges because of a step function behavior of the integral inequality in the static limit. The development of such a step function is apparent from the sequence of graphs in Fig. 3 with decreasing a/m values. As $a/m \to 0$ in this sequence, the limiting step function hits the K axis at $K = \frac{2}{3}$, corresponding to the value given by (35) for the case $\alpha = 0$. On approach to the static limit through intermediate values of α, the integral inequality manifests the same step function behavior, with intercept given by (35). Thus, for a given α, even though there is an allowed rotating model with arbitrarily high redshift, a finite upper bound to the redshifts arises discontinuously in the static limit.

Neither extreme case, $\alpha = 0$ or $\alpha = 1$, represents a physically reasonable mass distribution for an astrophysical system. The $\alpha = 0$ case with vanishing moment of inertia, would not appear possible in the static limit without negative density or pressure regions to prevent the formation of a horizon. Positive density and pressure are sufficient but not necessary conditions for the inequalities, so that the inequalities do not exclude such possiblities as long as the total mass is positive. The $\alpha = 1$ case, in the static limit, is attained with positive density and surface stress by the thin shell models of Brill and Cohn[7,8] and of Israel.[9] They find a range of slowly rotating models with unbounded redshift as the shell approaches the horizon, in full accord with the limits allowed by the inequalities.

For the more realistic case of a slowly rotating, homogeneous, spherical fluid, Chandrasekhar and Miller[10] have found that α ranges between the Newtonian value $\frac{2}{5}$ and an upper limit approximately equal to $\frac{4}{5}$. At this upper limit the (unperturbed) homogeneous sphere has radius $R = 9m/4$, which corresponds to the maximum possible redshift $z = 2$ for a fluid with inwardly increasing density in accord with stability criteria.[30,31] For this same value $\alpha = \frac{4}{5}$, K equals $\frac{8}{9}$ for the homogeneous sphere whereas $\frac{10}{11}$ is the maximum value of K allowed by the inequality (35). Thus, as might be expected, in the static limit, our restrictions based upon existence are slightly weaker than the restrictions based upon stablity. For $\alpha = \frac{4}{5}$ the density of a static model with $K = \frac{10}{11}$ must increase outward in some region.

The above considerations have some bearing, although indirect, on the possiblity of a high redshift, rotating fluid model for a quasar. The stability limit $z = 2$ for a static model, has discourged such attempts subsequent to the observations of $z > 2$ quasars. On the other hand, our results show that finite redshift limits based upon the *existence* of a static model are completely misleading in the rotating case for which arbitrarily high redshifts are allowed. Is it possible that the $z = 2$ *stability* limit is also misleading? Unfortunately, in the rapidly rotating case, there is no known generalization of the $z = 2$ limit for stability to axially symmetric perturbations. The issue is further clouded by the instability of all rotating systems to nonaxisymmetric perturbations.[5] However, our results do provide fresh motivation to investigate whether high redshift models might have appreciable lifetimes due to the relativistic effects of rotation. Is there a class of astrophysical objects whose support against gravitational collapse depends significantly on the repulsive gravimagnetic forces between corotating matter loops?

ACKNOWLEDGMENT

We gratefully acknowledge several helpful discussions with J. Friedman.

[1]K. S. Thorne, in *Gravitation and Cosmology, Proceedings of the International School of Physics Enrico Fermi*, Course 47, edited by R. K. Sachs (Academic, New York, 1971).
[2]J. M. Bardeen, in *Black Holes*, edited by C. DeWitt and B. S. deWitt (Gordon and Breach, New York, 1973).
[3]J. C. Miller and D. W. Sciama, in *General Relativity and Gravitation*, Vol. 2, edited by A. Held (Plenum, New York, 1980).
[4]J. Winicour, in *General Relativity and Gravitation*, Vol. 2, edited by A. Held (Plenum, New York, 1980).

[5]J. L. Friedman, Commun. Math. Phys. **62**, 247 (1978).
[6]J. B. Hartle, Astrophys. J. **167**, 359 (1967).
[7]D. R. Brill and J. M. Cohen, Phys. Rev. **143**, 1011 (1966).
[8]J. M. Cohen, J. Math. Phys. **8**, 1477 (1967).
[9]V. de la Cruz and W. Israel, Phys. Rev. **170**, 1187 (1968).
[10]S. Chandrasekhar and J. C. Miller, Mon. Not. R. Astron. Soc. **167**, 63 (1974).
[11]J. C. Miller, Mon. Not. R. Astron. Soc. **179**, 483 (1977).
[12]J. M. Bardeen and R. V. Wagoner, Astrophys. J. **167**, 359 (1971).
[13]J. R. Wilson, Astrophys. J. **176**, 195 (1972).
[14]J. R. Wilson, Phys. Rev. Lett. **30**, 1082 (1973).
[15]J. Pachner, Can. J. Phys. **51**, 447 (1973).
[16]S. Bonazzola and J. Schneider, Astrophys. J. **191**, 273 (1974).
[17]E. M. Buttorworth and J. R Ipser, Astrophys. J. **200**,L103 (1975).
[18]E. M. Buttorworth and J. R. Ipser, Astrophys. J. **204**, 200 (1976).
[19]E. M. Buttorworth, Astrophys. J. **204**, 561 (1976).
[20]J. R. Wilson, in *Physics and Astrophysics of Neutron Stars and Black Holes*, edited by R. Giaconni and R. Ruffini (North-Holland, Amsterdam, 1978).
[21]W. Roos, Gen. Relativ. Gravit. 7, 431 (1976).
[22]R. H. Boyer, Proc. Cambridge Philos. Soc. **68**, 199 (1965).
[23]M. A Abramowicz, J. P. Lasota, and B. Muchotrzeb, Commun. Math. Phys. **47**, 109 (1976).
[24]R. O. Hansen and J. Winicour, J. Math. Phys. **16**, 804 (1975).
[25]R. O. Hansen and J. Winicour, J. Math. Phys. **18**, 1206 (1977).
[26]W. C. Hernandez, Phys. Rev. **159**, 1070 (1967).
[27]R. Geroch, J. Math. Phys. **13**, 394 (1972).
[28]S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space–Time* (Cambridge U. P., Cambridge, 1973).
[29]I. N. Bronshtein and K. A. Semendyayev, *A Guide Book to Mathematics* (Springer, New York, 1973), p.166.
[30]H. A. Buchdahl, Phys. Rev. **116**, 1027 (1959).
[31]H. Bondi, Proc. R. Soc. London, Ser. A **282A**, 303 (1964).

Hamiltonian formulation for the gauge theory of the gravitational coupling

Riccardo Giachetti
Osservatorio Astrofisico di Arcetri, I.N.F.N.-Sezione di Firenze, Firenze, Italy
Riccardo Ricci
Istituto Matematico "U. Dini", Università di Firenze, Firenze, Italy
Emanuele Sorace
I.N.F.N.-Sezione di Firenze, Firenze, Italy

(Received 11 November 1980; accepted for publication 6 March 1981)

This paper studies the Hamiltonian mechanics of a relativistic particle interacting with a gravitational field considered as a gauge field of the Poincaré group. We follow a general method developed by Sternberg for the case of internal symmetries, that describes the interaction by a suitable modification of the symplectic form. This approach is reviewed and the explicit examples of the electromagnetic and Yang–Mills gauge interactions are widely explained in local coordinates. The peculiar features of a gauge theory of the Poincaré group are then discussed and the geometrical picture that emerges suggests the way of modifying the symplectic form for a correct description of the gravitational coupling.

PACS numbers: 04.20.Fy, 04.20.Me, 12.25. + e, 11.10.Np

I. INTRODUCTION

The procedure that starts from a Lagrangian quasi-invariant under the global action of a Lie group and, by the localization of the symmetry, leads to the introduction of the gauge fields, has become a standard tool in current theoretical physics.[1] In this framework, the essential contribution given by differential geometry for a deeper understanding and further extensions of gauge theories, which are raised to paradigm for the description of elementary processes, has been analyzed, fixed in its patterns, and completely exploited. It could therefore seem that no room has been left for the development of different and original points of view.

Nevertheless, still adopting the geometrical language, Sternberg has recently proposed a new approach that introduces the interaction with the gauge fields directly in the realm of the single particle dynamics.[2] This allows one, moreover, to connect physical and geometrical structures to date unrelated: Hamiltonian mechanics in its general form as the geometry of symplectic manifolds,[3] on one hand, and the geometry of gauge fields—i.e., the geometry of principal bundles—on the other.[4]

The fundamental result of Sternberg is that of furnishing a prescription to deform,[5] in a gauge invariant way, the symplectic structure of the original mechanical system; in this process a major role is played by a connection form on an appropriate principal bundle. It is then shown that this procedure—globally meaningful—is locally equivalent to introducing the gauge field–particle coupling by keeping unchanged the symplectic form and modifying, on the contrary, the Hamiltonian function in a gauge invariant way. A generalized version of the electromagnetic minimal coupling is thus obtained.

In the sketched approach to gauge interactions, it is evident that the whole material concerning the geometry of symplectic manifolds is of the same importance as the theory of principal bundles, although the former still appears to be *terra incognita* for many active physicists. In particular, the use of concepts like orbits of the coadjoint representation and Hamiltonian G-spaces[6,7] allows one to reach, within classical mechanics, interesting results otherwise deduced either by taking the appropriate limit on a quantum system,[8] or by means of complex or anticommuting variables[9]; we think, for instance, of the equations of motion for a particle with isotopic spin degrees of freedom in the presence of a Yang–Mills field. We also think of the equations proposed by Souriau[10] as an improvement of the Bargmann–Michel–Telegdi equation for the motion of a classical spinning particle in an electromagnetic field.

In the present paper we propose a theory of the gravitational coupling of a particle as a gauge theory of the Poincaré group. The technique consists in adapting Sternberg's approach to a somewhat different situation. The mechanical system under consideration is that of a free relativistic particle moving in Minkowski space. The peculiarity of this case is due to the fact that the gauge group is just the group of affine transformations of the configuration space and the "internal" variables are coincident with the phase-space variables. The number of degrees of freedom of the system is thus not increased by the effect of the interaction, contrary to what happens in the usual case; this is a possible source of the difficulties frequently encountered in the attempts to obtain gravitational interactions out of a gauge principle. We also observe that a gauge theory of gravitation must be able to reproduce the distinguishing features of an Einstein–Cartan theory from a connection on a principal bundle. As a matter of fact, this is impossible if a proper subgroup of the Poincaré group is used to build up the gauge theory.[11] Indeed, both the pioneering attempt of Utiyama,[12] based on the Lorentz group, and the works of Hayashi and Nakano[13] and Cho,[14] that use the translations only, must be considered unsatisfactory.

The necessary and sufficient condition for the gauge theory of the Poincaré group to be equivalent to an Einstein–Cartan theory can be expressed by requiring the maximality of the rank of a certain one-form.[15] As shown in previous

 © 1981 American Institute of Physics

papers,[16] this condition is equivalent to the nondegeneracy of the tetrad field. Using this one-form and a connection form on the bundle of the orthonormal Lorentz frames, we directly reconstruct the gravitational gauge potential, namely the connection on the bundle of the affine frames, which is a principal bundle with the Poincaré group, as structure group. The result that we then derive demonstrates the correctness of the adopted procedure, that produces the equations of motion directly in a canonical form. Moreover, the intrinsic formulation in geometrical language clarifies some conceptual problems, still left open by the fundamental paper of Kibble,[17,18] without having to resort to the artifice of interpreting the Poincaré group as a group of linear transformation of a five-dimensional space.[18]

In order to give a self-contained treatment and since the original paper[2] may appear somewhat hard to read, we present in Sec. 2 a resumé of Sternberg's approach to gauge interactions. Besides the almost trivial example of the electromagnetic interactions, we have completely worked out in local coordinates the SU (2)-gauge interaction of a particle with isotopic spin; this concrete example will certainly help to clarify the basic ideas of the method.

In Sec. 3 we set up the mechanism that produces gravitation as a gauge interaction of the Poincaré group. In particular, starting with a relativistic free particle, we give an explicit construction of the deformed symplectic structure. We proceed to prove that the result is gauge invariant and we show that this fact implies the invariance under general coordinate transformations. All the important calculations—as the evaluation of the momentum mapping—are described in detail. For the sake of clarity we also give in the Appendix the construction of the affine connection starting from a linear connection and an appropriate one-form. This is done because the same type of calculations may in general be applied to gauge theories on reductive homogeneous spaces.[19a,19b] We notice that only the proof of the existence and uniqueness of the affine connection is found in the most diffuse books on differential geometry.

We finally give some brief conclusions and suggest possible developments of the theory.

To conclude this Introduction, we make a short list of some notations used in the text, possibly indicating the references where the same symbols are used:

If $f: M \to N$ is a map of differentiable manifolds (resp.: a function if $N = \mathbb{R}$), then df denotes the tangent map (resp.: the differential) and f^* is the pullback from the exterior algebra of forms on N to the exterior algebra of forms on M.

$\lrcorner$ is the interior product of a vector field with an exterior form.[20]

Lie$G$ is the Lie algebra of the Lie group G and (Lie$G$)* is the dual vector space of Lie$G$.

exp denotes the exponential map of Lie groups, so that $\exp(t\xi)$ is the one-parameter subgroup of G generated by $\xi \in \mathrm{Lie}G$.[19]

$\mathbb{1}$ indicates the identity matrix of the Lorentz group and $\mathbb{R}^{1,3}$ is the pseudo-Euclidean arithmetic space with signature $(+, -, -, -)$.

II. GENERAL GEOMETRICAL FRAMEWORK FOR GAUGE FIELD–PARTICLE INTERACTIONS

Due to the increasingly more comprehensive formulations given to mechanics, the concept of mechanical explanation of a physical phenomenon has undergone successive abstractions up to present time, where it is widely accepted that "mechanical systems should consist of symplectic manifolds which do not necessarily admit any global interpretation as the phase space of some configuration space".[21]

As a simple example, for a free pointlike particle admitting the manifold M as configuration space, the corresponding mechanical system is the couple (T^*M, ω) given by the cotangent bundle $q: T^*M \to M$ and the canonical two-form $\omega = dp_i \wedge dx^i$. The motion of the particle is described by the equations

$$X_H \lrcorner \omega = -dH, \tag{2.1}$$

where H is the Hamiltonian function and X_H the corresponding Hamiltonian vector field.[22]

In principle, possible external fields acting on the particle should affect the Hamiltonian function by means of appropriate potential terms. Indeed this is the case of a test particle in the presence of an electromagnetic field; the field–particle coupling is introduced through the "minimal substitution," $p_i \to p_i - eA_i$, in the Hamiltonian, where the interaction strength is measured by a suitable parameter, namely the electric charge e. If the initial Hamiltonian is a free Hamiltonian $H = H(\eta^{ij} p_i p_j)$, where $\eta^{ij} = \mathrm{diag}(+1, -1, -1, -1)$ is the usual flat metric tensor, then Eq. (2.1) expressed in local coordinates gives the well-known equations

$$\begin{aligned} \dot{x}^i &= (1/m)\eta^{ij}(p_j - eA_j), \\ \dot{p}_i &= (1/m)\eta^{kj}(p_k - eA_k)\partial A_j/\partial x^i, \end{aligned} \tag{2.2}$$

with $1/m = 2H'[(p_i - eA_i)^2] = \mathrm{const}$,[23] and the prime denotes differentiation with respect to the argument.

In this case, however, it is possible to give a completely equivalent description of the interaction by keeping the Hamiltonian unchanged and adding, instead, an interaction term $e\,dA = e\,F_{ij}\,dx^i \wedge dx^j$ to the symplectic form,[24] so as to obtain as equations of the motion,

$$X_H \lrcorner (\omega + e\,dA) = -dH, \tag{2.3}$$

or in local coordinates,

$$\begin{aligned} \dot{x}^i &= (1/m)\eta^{ij} p_j, \\ \dot{p}_i &= e\,F_{ij}\dot{x}^j, \end{aligned} \tag{2.4}$$

where $1/m = 2H'(p^2) = \mathrm{const}$ and F_{ij} is the usual e.m. field tensor.

Observing that electromagnetism is a gauge theory with U(1) as gauge group, Eq. (2.3) admits a geometrical interpretation that can be generalized to any gauge theory with arbitrary gauge group.[2] In fact it has long been realized that a most natural space arising in a gauge theory is a principal bundle E whose base space is the configuration space M and whose structural group is the gauge group G. This bundle indicates how the geometries of the configuration space and of the gauge group are blended, while the gauge potentials are simply described by the components of a connection form on E.[1] The electromagnetic potential A, therefore, is a Lie U(1)-valued one-form, and the role of the electric charge e is to provide a measure for the interaction strength. This means that we may think of e as to an element of (Lie U(1))* acting on the potential by the duality pairing.

The electric charge is conserved by the electromagnetic interactions. For a general gauge theory, however, the dynamics may be dependent upon several parameters that are subject to evolve. Their domain constitutes the space F of the "internal variables" and of course an action of G on F must be defined, i.e., G has to be a symmetry group for the space F. Moreover, if any mechanical theory has to be established, we must assume that F is a Hamiltonian G-space,[6,7] namely: (i) F has a symplectic structure determined by a two-form Ω and G acts by canonical transformations, so that for any $\xi \in \mathrm{Lie}G$ the vector field Y_ξ on F generated by $\exp(t\xi)$ is globally Hamiltonian; (ii) we are given a lifting of the homomorphism $\xi \to Y_\xi$ to a homorphism λ: $\mathrm{Lie}G \to \mathscr{F}(F)$, $\lambda(\xi) = f_\xi$, where $\mathscr{F}(F)$ is the Lie algebra of the functions on F whose product is the Poisson bracket $\{\ ,\ \}$. In other words, the meaning of the above assumptions is that any local one-parameter group of canonical transformations arising from the action of G is actually global, and moreover, the generating functions of such one-parameter groups form a subalgebra of $\mathscr{F}(F)$ under Poisson bracket.

The following question now appears natural: what takes the place of the charge in a general gauge theory? From the electromagnetic discussion it emerges that the new charge must be a $(\mathrm{Lie}G)^*$-valued object. In fact, since for any Hamiltonian G-space the correspondence $\lambda: \mathrm{Lie}G \to \mathscr{F}(F): \xi \to f_\xi$ is linear in ξ at any point $y \in F$, there is an element $\mu(y) \in (\mathrm{Lie}G)^*$ such that

$$\langle \mu(y), \xi \rangle = f_\xi(y);\ y \in F,\ \xi \in \mathrm{Lie}G. \tag{2.5}$$

The mapping $\mu: F \to (\mathrm{Lie}G)^*$, called "momentum mapping,"[24] is thus naturally defined.

It is easy to see that, if H is a Hamiltonian on F, invariant under the action of G, then μ is constant along the flow of the Hamiltonian vector field Y_H. Indeed, let Y_ξ be the Hamiltonian vector field generated by ξ as in item (i) of the definition of Hamiltonian G-space, so that Y_ξ is equal to the Hamiltonian vector field Y_{f_ξ}. Then, by assumption, $Y_{f_\xi} H = 0$ for any $\xi \in \mathrm{Lie}G$. Therefore,

$$\begin{aligned} 0 = Y_{f_\xi} H &= \{ f_\xi, H \} = -\{H, f_\xi\} \\ &= -Y_H f_\xi = -Y_H \langle \mu, \xi \rangle = -\langle Y_H \mu, \xi \rangle, \end{aligned}$$

which clearly entails

$$Y_H \mu = 0. \tag{2.6}$$

From the proof we see that the condition (2.6) is equivalent to the vanishing of the Poisson bracket of H with f_ξ. In this way we recover the well-known relationship between symmetries and conserved quantities, i.e., a generalized Hamiltonian version of the classical Noether theorem.[25]

There is a rigorous argument that allows us to interpret μ as a good generalization of the charge. Indeed, the natural space for a mechanical description of the gauge field–particle coupled system is obviously the domain of both the external and internal variables, that is the associated bundle $Q = (q^*E \times F)/G$, where q^*E is the pullback of E on T^*M, and the equivalence is made with respect to the action of G, as usual.[19] A mechanical system (Q, Ω_Θ) can be defined for any gauge field, i.e., for any connection form Θ on E, since[2]: (a) there exists a closed two-form σ on Q such that

$$\pi^*\sigma = d\langle \mu, \Theta \rangle + \Omega, \tag{2.7}$$

where π: $q^*E \times F \to Q$ is the quotient projection and $\langle\ ,\ \rangle$ denotes the duality pairing in $\mathrm{Lie}G$; (b) the two-form

$$\Omega_\Theta = \omega + \sigma \tag{2.8}$$

is a symplectic form on Q. The equations of motion

$$X_H \lrcorner \Omega_\Theta = -dH, \tag{2.9}$$

where H is a free Hamiltonian as in (2.1), represent the evolution of a particle interacting with the gauge field whose potential is given by Θ. Therefore, the interaction is contained in σ and its expression makes it reasonable to look at μ as the new "charge".

Typical examples of Hamiltonian G spaces are the orbits of G in $(\mathrm{Lie}G)^*$ under the coadjoint representation, ad*, defined by

$$\langle \mathrm{ad}_g^* f, \xi \rangle = \langle f, \mathrm{ad}_{g^{-1}} \xi \rangle, \tag{2.10}$$

where $g \in G$, $f \in (\mathrm{Lie}G)^*$, $\xi \in \mathrm{Lie}G$. The natural G-invariant symplectic structure on an orbit[6] is given by the two-form Ω such that

$$\Omega(f)(Y_\xi(f), Y_\eta(f)) = \langle f, [\xi, \eta] \rangle, \tag{2.11}$$

with $f \in (\mathrm{Lie}G)^*$ and ξ, $\eta \in \mathrm{Lie}G$.

A well-known result[7] states that any homogeneous Hamiltonian G-space is a covering of an orbit of G in $(\mathrm{Lie}G)^*$. Moreover, general arguments based on cohomological theories of Lie algebras[7] demonstrate that for a semisimple Lie group G any symplectic space on which G acts by canonical transformations is actually a Hamiltonian G-space. This gives the most general choice for the space of the internal variables in any gauge theory with a semisimple Lie group.

We notice that if F is an orbit of G in $(\mathrm{Lie}G)^*$, the momentum mapping is just the imbedding of F into $(\mathrm{Lie}G)^*$. In particular, for electromagnetism, since U(1) is commutative, the orbit reduces to a single point $e \in (\mathrm{LieU}(1))^*$ and $\mu(e) = e$.

An illuminating example is provided by the classical Yang–Mills gauge theory, where the gauge group is SU(2). Since SU(2) is semisimple, its Hamiltonian spaces are entirely classified in terms of the orbits in (LieSU(2))*. To find such orbits we observe that the nondegenerate Killing form of SU(2) gives an equivalence of the coadjoint with the adjoint representation. Orbits are therefore level surfaces of the invariants under similarity transformations of any matrix of the type

$$a = b_1\tau_1 + b_2\tau_2 + b_3\tau_3, \tag{2.12}$$

where b_A, $A = 1,2,3$, are real numbers, $\tau_A = (i/2)\sigma_A$, and σ_A are the Pauli matrices. Since the trace is vanishing for any matrix of the type (2.12), the only meaningful invariant is

$$\det a = \tfrac{1}{4} \sum_{A=1,3} {b_A}^2 = \tfrac{1}{4} R^2, \tag{2.13}$$

and its level surfaces are ordinary two-spheres. As spheres are simply connected manifolds, they are their own universal covering, so that any homogeneous Hamiltonian SU(2)-space is just a two-sphere. In this case the symplectic structure is proportional to the ordinary Riemannian volume form and the invariance under SU(2) is thus evident.

Indicating by $A_i^A(x)$ the SU(2) gauge potentials and carrying out computations according to the described procedure, we find

$$\Omega_\Theta = dp_i \wedge dx^i + \frac{\partial A_i^A}{\partial x^j} b_A dx^j \wedge dx^i - A_i^A dx^i \wedge db_A - \frac{\epsilon^{ABC}}{2R^2} b_A db_B \wedge db_C, \tag{2.14}$$

where ϵ^{ABC} is the completely antisymmetric tensor of rank three and capital indices are raised and lowered by means of the Euclidean metric δ_A^B.

For a free Hamiltonian $H = H(\eta^{ij}p_ip_j)$, the equations of motion are

$$m\dot{x}^i = \eta^{ij}p_j, \tag{2.15a}$$

$$\dot{p}_i = (\partial A_j^A/\partial x^i - \partial A_i^A/\partial x^j)\dot{x}^j b_A - A_i^A \dot{b}_A, \tag{2.15b}$$

$$\epsilon^{ABC} b_B \dot{b}_C + R^2 A_i^A \dot{x}^i = 0, \tag{2.15c}$$

where, again, $1/m = 2H'$. To these, we add the "constraint" equation deduced from (2.13), i.e.,

$$\dot{b}_A b^A = 0. \tag{2.15d}$$

From (2.15c) and (2.15d) we find

$$\dot{b}_A = -\epsilon_{ABC} A_i^B \dot{x}^i b^C. \tag{2.16}$$

By differentiating (2.15a) and substituting the equation so found, together with (2.16), into (2.15b), we finally get

$$m\ddot{x}^i = \eta^{ik} F_{kj}^A b_A \dot{x}^j, \tag{2.17}$$

where

$$F_{kj}^A = \partial A_j^A/\partial x^k - \partial A_k^A/\partial x^j + \epsilon_{BC}^A A_k^B A_j^C \tag{2.18}$$

is the Yang–Mills field strength.

Let us conclude this section by showing how the usual minimal substitution point of view can be locally recovered in any gauge theory. If we have a local trivialization of the bundle E by means of a section s over $U \subseteq M$— in physical terms a local gauge—then the connection Θ is completely determined by the Lie$G$-valued form $A = s^*\Theta$ on U, the components of A are the gauge potentials and relation (2.7) implies

$$\sigma = d\langle \mu, A\rangle + \Omega. \tag{2.19}$$

We may think of $\langle \mu, A\rangle$ as a one-form on $U \times F$ and perform the minimal substitution $\psi: T^*U \times F \to T^*U \times F$,

$$(x,p,y) \to (x, p + \langle \mu(y), A(x)\rangle, y). \tag{2.20}$$

The equations of motion

$$X_H \lrcorner (\omega + \Omega) = -d(\psi^{-1*}H), \tag{2.21}$$

where H is again a free Hamiltonian and

$$(\psi^{-1*}H)(x,p,y) = H(x, p - \langle \mu(y), A(x)\rangle) \tag{2.22}$$

are equivalent to (2.9) in the sense that integral curves of both (2.9) and (2.21) have identical projections on $U \times F$, while the corresponding momenta are related by means of ψ.

In the Yang–Mills case the minimal substitution is

$$H(p^2) \to H[(p_i - A_i^A b_A)^2], \tag{2.23}$$

and the equations of motion (2.21) read

$$\dot{x}^i = \frac{\partial H}{\partial p_i} = \frac{1}{m}\eta^{ij}(p_j - A_j^A b_A), \tag{2.24a}$$

$$\dot{p}_i = -\frac{\partial H}{\partial x^i} = \frac{1}{m}\eta^{jk}(p_j - A_j^A b_A)\frac{\partial A_k^B}{\partial x^i} b_B, \tag{2.24b}$$

$$\frac{\epsilon^{ABC}}{R^2} b_B \dot{b}_C = -\frac{1}{m}\eta^{ij}(p_i - A_i^D b_D) A_j^A, \tag{2.24c}$$

where $1/m = 2H'[(p_i - A_i^A b_A)^2] = \text{const}$.

It is trivial to verify that the system (2.24) is equivalent to (2.15). However, we want to stress that Eqs. (2.9) have a global meaning even if E is not a trivial bundle, and can be written for an arbitrary symplectic manifold. The minimal substitution, on the other hand, can be given a local meaning only.

We also notice that the system (2.24) has been found by taking the classical limit of the quantum Yang–Mills field equations,[8] while a system analogous to (2.24) has been obtained by the use of anticommuting variables.[9]

III. GRAVITATIONAL FORCE AS A GAUGE INTERACTION OF THE POINCARÉ GROUP

In the present section we want to show how the method of introducing the gauge field–particle interaction by modifying the symplectic form can be adapted to the case of the gravitational coupling. It is therefore necessary to illustrate, in the first place, the geometrical peculiarities of a gauge theory of the Poincaré group and its relationship to an Einstein–Cartan theory.

A gauge theory of the Poincaré group P is given by a principal bundle $\pi: E_P \to M$ over the space–time M with P as structure group. A gauge potential is a connection on E_P, namely a Lie$P$-valued one-form $\tilde{\omega}$ of the adjoint type. To give a physical meaning to E_P the "rotations" must be distinguished by the "translations" in each fiber of E_P by means of a reduction of the structure group P to the Lorentz group L. This means that we have an injection

$$\gamma: E_L \to E_P, \tag{3.1}$$

where $\pi: E_L \to M$ is a principal bundle with structure group L. Such a reduction (always existing for a paracompact M) is described by a tensorial 0-form Φ of type $(\rho, \mathbb{R}^{(1,3)})$, namely by a map

$$\Phi: E_P \to \mathbb{R}^{(1,3)} \tag{3.2}$$

such that

$$r^*_{(\Lambda,a)}\Phi \equiv \Phi \circ r_{(\Lambda,a)} = \rho((\Lambda,a)^{-1})\Phi, \tag{3.3}$$

where $(\Lambda,a) \in P$, r is the right translation in E_P, and ρ is the affine representation of P on $\mathbb{R}^{(1,3)}$ defined by

$$\rho(\Lambda,a)\Phi^i = \Lambda^i_j \Phi^j + a^i. \tag{3.4}$$

Explicitly we have

$$E_L = \{u \in E_P \mid \Phi(u) = 0\}, \tag{3.5}$$

so that E_L is a submanifold of E_P and γ is the canonical injection. We can also define a projection $\beta: E_P \to E_L$ given by

$$\beta(u) = r_{(1,\Phi(u))}(u), \quad u \in E_P, \tag{3.6}$$

where $\mathbb{1}$ is the unity of the Lorentz group. Of course $\beta \circ \gamma$ is the identity mapping of E_L into itself.

Since

$$\mathrm{Lie}P = \mathrm{Lie}L \oplus \mathbb{R}^{(1,3)}, \tag{3.7}$$

the pullback $\gamma^*\tilde{\omega}$—which is the restriction of $\tilde{\omega}$ to $E_L \subseteq E_P$ —can be decomposed as[19]

$$\gamma^*\tilde{\omega} = \omega + \varphi, \tag{3.8}$$

where ω is a connection on E_L and φ is a tensorial one-form of type $(\rho_L, \mathbb{R}^{(1,3)})$, where ρ_L is the obvious representation of L induced by (3.4), so that

$$(r^*_\Lambda \ \varphi)^i = (\rho_L(\Lambda)\varphi)^i = \Lambda^i_{\ j}\varphi^j. \tag{3.9}$$

Indeed, by letting $\tilde{\omega} = \omega_1 + \omega_2$ in the decomposition (3.7), we have

$$r^*_{(\Lambda,0)} \ \tilde{\omega} = \mathrm{ad}_{(\Lambda,0)}\tilde{\omega} = \mathrm{ad}_\Lambda \omega_1 + \rho_L(\Lambda)\omega_2, \tag{3.10}$$

and setting $\omega = \omega_1|_{E_L}$ and $\varphi = \omega_2|_{E_L}$, the above property is verified.

It is very important to notice that the connection $\tilde{\omega}$ can be reconstructed from ω and φ. We find

$$\tilde{\omega} = \beta^*\omega + \beta^*\varphi - d\Phi - \rho'(\beta^*\omega)\Phi, \tag{3.11}$$

where ρ' is the representation of Lie$P$ on $\mathbb{R}^{(1,3)}$ obtained by (3.4), and to which it is formally equal. We give in the Appendix a detailed proof of (3.11), since the same type of computations may be applied to more general cases.

An important geometrical quantity is the covariant differential $D\Phi$ of the reduction function with respect to the connection $\tilde{\omega}$; this can be used to express $\tilde{\omega}$ itself. The explicit form of $D\Phi$ is given by

$$D\Phi = d\Phi + \rho'(\tilde{\omega})\Phi. \tag{3.12}$$

Indeed, let h and v denote the horizontal and vertical projections determined by $\tilde{\omega}$, so that any $X \in TE_P$ is uniquely written as a sum $X = hX + vX$. Then, by definition of D, we have

$$D\Phi(X) = d\Phi(hX) = d\Phi(X) - d\Phi(vX).$$

But

$$d\Phi(vX) = (vX)\Phi = \frac{d}{dt} r^*_{\exp\{t\tilde{\omega}(X)\}} \Phi \big|_{t=0}$$
$$= \frac{d}{dt}\rho(\exp\{-t\tilde{\omega}(X)\})|_{t=0}\Phi = -\rho'(\tilde{\omega}(X))\Phi,$$

from which (3.12) follows.

Since $\tilde{\omega} = \omega_1 + \omega_2$ and

$$\rho'(\tilde{\omega})\Phi = \rho'(\omega_1)\Phi + \omega_2, \tag{3.13}$$

from (3.12) and (3.13) we have

$$\omega_2 = D\Phi - d\Phi - \rho'(\omega_1)\Phi,$$

so that

$$\tilde{\omega} = \omega_1 + D\Phi - d\Phi - \rho'(\omega_1)\Phi. \tag{3.14}$$

A comparison of (3.11) with (3.14) shows that

$$\omega_1 = \beta^*\omega, \tag{3.15a}$$

$$D\Phi = \beta^*\varphi. \tag{3.15b}$$

We are now in position to clarify the relationship of a Poincaré group gauge theory to an Einstein–Cartan theory. This is usually referred to as the "soldering condition"[26] and may be expressed by requiring that the $\mathbb{R}^{(1,3)}$-valued one-form $D\Phi$ has maximal rank,[15] i.e.,

$$D\Phi(X_u) = 0 \text{ if and only if } X_u \in T_u E_P \text{ is vertical.} \tag{3.16}$$

The physical meaning of this condition is clear when we observe that in a local gauge $\mathfrak{s}$ and in a local chart for M, the matrix corresponding to $\mathfrak{s}^* D\Phi$ is just the usual tetrad field; condition (3.16) expresses, therefore, the nondegeneracy of the tetrad field.[15,16]

We want now to show how the geometry of E_P and especially Eqs. (3.14) and (3.15), together with the soldering condition, may be used to describe the interaction of a particle with a gravitational field as a gauge interaction, in the framework developed in Sec. 2. The peculiarity of this case lies in the fact that the Poincaré group is the group of the kinematical transformations of the space–time itself. Thus the P-Hamiltonian space of the gauge variables must be a local model of the phase space, but for a possible enlargement due to the presence of spin. We shall give a treatment of only the spinless particle and simply indicate a possible way to take spin into account.

The configuration space of a free spinless particle is the usual Minkowski space M, so that the phase space is the cotangent bundle $T^*M \simeq M \times \mathbb{R}^{(1,3)}$, and the natural flat metric determines the general form of the Hamiltonian function suitable for describing the particle dynamics.

In this case there exists a concrete principal bundle E_P, realized by the affine frames of TM with orthonormal basis vectors. An element $u \in E_P$ is thus given by a triple

$$u = \{x, \mathscr{V}, X\}, \tag{3.17}$$

where $x \in M$, $\mathscr{V} = (V_i)_{i=1,4}$ is an orthonormal frame for T_xM and $X \in T_xM$ indicates the "displacement" of the origin of the frame. The action of $(\Lambda, a) \in P$ on E_P is given by

$$\{x, (V_i)_{i=1,4}, X\}(\Lambda^i_{\ j}, a^i) = \{x, (V_j \Lambda^j_{\ i})_{i=1,4}, X - V_i a^i\}, \tag{3.18}$$

and the map $\Phi: E_P \to \mathbb{R}^{(1,3)}$ that gives the reduction of the structural group is defined by

$$\Phi^i(u) V_i = X. \tag{3.19}$$

Let us now observe that any coordinate system on M is obtained by choosing an appropriate local gauge, i.e., a local section $\mathfrak{s}$ of E_P. The coordinate system is then given by $\Phi \circ \mathfrak{s}$. Indeed, if $(\psi^i)_{i=1,4}$ are four independent functions on M, then we define

$$\mathfrak{s}_\psi(x) = \{x, \mathscr{V}, X\} = \{x, (V_i)_{i=1,4}, \psi^i V_i\}, \tag{3.20}$$

where $\mathscr{V} = (V_i)_{i=1,4}$ is quite general.

A local gauge provides also a coordinate system for T^*M, namely

$$(x,p) \to (x^i, p_i) \equiv (\psi^i, \langle p, V_i \rangle). \tag{3.21}$$

The action of P on the principal bundle E_P is thus transferred on the range of the coordinate system (x^i, p_i), i.e., $T^*\mathbb{R}^{(1,3)}$, according to

$$(\Lambda^k_{\ j}, a^k)(x^i, p_i) = (\Lambda^i_{\ j} x^j + a^i, p_j (\Lambda^{-1})^j_{\ k}). \tag{3.22}$$

Notice that the relation (3.22) determines a canonical action on $T^*\mathbb{R}^{(1,3)}$ (with respect to the natural symplectic form $dp_i \wedge dx^i$) that endows $T^*\mathbb{R}^{(1,3)}$ with the structure of Hamiltonian P-space. However, the action induced by (3.22) on T^*M by means of the coordinate system is not, in general, a canonical action, as (x^i,p_i) is not in general a Darboux coordinate system.[25]

We shall use $T^*\mathbb{R}^{(1,3)}$ as the domain of the gauge variables and the required identification with T^*M is done by any chart (x^i,p_i) arising from a local gauge. The gauge invariance of the theory implies directly, therefore, its invariance under general coordinate transformations performed in the configuration space M.

Let us now consider the momentum mapping $\mu: T^*\mathbb{R}^{(1,3)} \to (\mathrm{Lie}P)^*$ associated with the action (3.22). We find that

$$\mu(x^i,p_i) = x^i p_j \mathscr{M}^i_j + p_i \mathscr{P}^i, \tag{3.23}$$

where $(\mathscr{M}^j_i, \mathscr{P}^i)$ are the matrices forming the basis of $(\mathrm{Lie}P)^* = (\mathrm{Lie}L)^* \oplus \mathbb{R}^{(1,3)*}$ dual to the basis (M^j_i, P_i) of the standard Lie$P$ representation deduced by (3.4). An explicit proof of (3.23) is obtained by considering the generating function $f_{X_{(\lambda,\alpha)}}$ of the canonical transformation associated with the action of the one-parameter group $\exp(t(\lambda,\alpha))$ generated by an element $(\lambda,\alpha) = \lambda^j_i M^i_j + \alpha^i P_i \in \mathrm{Lie}P$. The function $f_{X_{(\lambda,\alpha)}}$ is obtained[22,25] by contracting the Liouville one-form $p_i dx^i$ of $T^*\mathbb{R}^{(1,3)}$ with the vector field $X_{(\lambda,\alpha)}$ tangent to the flow of the one-parameter group $\exp(t(\lambda,\alpha))$, i.e.,

$$X_{(\lambda,\alpha)}(x^i,p_i) = \frac{d}{dt}\exp(t(\lambda,\alpha))(x^i,p_i)\big|_{t=0}. \tag{3.24}$$

The rhs of equality (3.24) is more simply evaluated by applying the chain rule for derivatives; this reduces the problem to the application of the Jacobian matrix of the group action—evaluated in the identity of P—to the element (λ,α), namely

$$X_{(\lambda,\alpha)}(x^i,p_i) = A^i \frac{\partial}{\partial x^i} + B_i \frac{\partial}{\partial p_i},$$

where the coefficients A^i, B_i are given by

$$\begin{bmatrix} A^i \\ B_i \end{bmatrix} = \begin{bmatrix} \dfrac{\partial}{\partial \Lambda^h_k} (\Lambda^i_j x^j + a^i) & \dfrac{\partial}{\partial a^h} (\Lambda^i_j x^j + a^i) \\ \dfrac{\partial}{\partial \Lambda^h_k} (\Lambda^{-1})^j_i p_j & 0 \end{bmatrix} \begin{bmatrix} \lambda^h_k \\ \alpha^h \end{bmatrix} \Bigg|_{\substack{\Lambda^h_k = \delta^h_k \\ a^h = 0}}$$

$$= \begin{bmatrix} \lambda^i_j x^j + \alpha^i \\ -p_j \lambda^j_i \end{bmatrix}.$$

Therefore,

$$f_{X_{(\lambda,\alpha)}}(x^i,p_i) = x^i p_j \lambda^j_i + p_i \alpha^i, \tag{3.25}$$

and, according to relation (2.5), the required momentum map reads as in (3.23).

Let us now consider the product space

$$R = E_P \times T^*\mathbb{R}^{(1,3)} \tag{3.26}$$

over which we pull back the momentum map μ, the connection form $\widetilde{\omega}$, and the symplectic form $dp_i \wedge dx^i$ of $T^*\mathbb{R}^{(1,3)}$. On R we can define the two-form

$$\Omega_R = d\,\langle \mu, \widetilde{\omega} \rangle + dp_i \wedge dx^i, \tag{3.27}$$

where $\langle\ ,\ \rangle$ denotes the duality pairing in Lie$P$. The gravitational coupling is given by the first term in the expression of Ω_R. Indeed, denoting by $q: T^*M \to M$ the projection of the cotangent bundle and choosing a local gauge $\sigma: M \to E_P$, we may define a map

$$\Delta_\sigma: T^*M \to E_P \times T^*\mathbb{R}^{(1,3)} \tag{3.28}$$

such that

$$\Delta_\sigma(x,p) = (\sigma(x), \Phi^i \circ \sigma(x), \langle p, V_i \rangle) \equiv (\sigma(x), \xi^i, \zeta_i) \tag{3.29}$$

where $u = \sigma(x)$ is given by (3.17). The gauge σ is now quite arbitrary, contrary to what we required for the definition of a coordinate system, as in Eq. (3.21).

The map Δ_σ can be used to pull back the two-form Ω_R from R to T^*M. Taking into account Eq. (3.11) and since

$$\Delta^*_\sigma x^i = \xi^i, \quad \Delta^*_\sigma p_i = \zeta_i, \tag{3.30}$$

we find

$$\begin{aligned} \Delta^*_\sigma \Omega_R &= d\,[\Delta^*_\sigma \langle \widetilde{\omega}, \mu \rangle + \zeta_i\, d\xi^i] \\ &= d\,[(\sigma^*\beta^*\omega)^j_i \xi^i \zeta_j + (\sigma^*\beta^*\varphi)^i \zeta_i - d(\sigma^*\Phi)^i \zeta_i \\ &\quad - (\sigma^*\beta^*\omega)^j_i (\sigma^*\Phi)^i \zeta_j + \zeta_i d\xi^i] \\ &= d[(\beta\circ\sigma)^*\varphi^i \zeta_i]. \end{aligned} \tag{3.31}$$

From this last expression it appears that the two-form $\Delta^*_\sigma \Omega_R$ is actually independent of the choice of the local gauge σ and defines an exact two-form Ω on T^*M. Indeed, let σ' be another local gauge. Then for any $x \in M$ we have

$$\sigma'(x) = \sigma(x)(\Lambda(x), a(x)),$$

where $x \to (\Lambda(x), a(x))$ is the appropriate map of M into the Poincaré group, i.e., a change of affine reference system depending on the position. Labelling by primed letters the variables defined in the gauge σ' and recalling the transformation property (3.9) of φ^i and the definition (3.29) of ζ_i, which implies that $\zeta'_i = \Lambda^j_i \zeta_j$, we find

$$\begin{aligned} \Delta^*_{\sigma'} \Omega_R &= d\,[(\beta\circ\sigma')^*\varphi^i \zeta'_i] = d\,[(\Lambda^{-1})^i_j (\beta\circ\sigma)^*\varphi^j \Lambda^k_i \zeta_k] \\ &= d[(\beta\circ\sigma)^*\varphi^i \zeta_i] = \Delta^*_\sigma \Omega_R. \end{aligned} \tag{3.32}$$

The gauge invariance of Ω is therefore proved. From this, in particular, we also get the invariance of the theory under general coordinate transformations.

It remains now to prove that Ω is actually a symplectic form. In the first place Ω is closed, as it is exact. Secondly, Ω is nondegenerate. Indeed, due to its gauge independence, we can express the two-form Ω in a local gauge σ that provides a coordinate system (x^i,p_i) as in Eq. (3.21). Using the relation (3.15b) we thus have

$$\Omega = d\,[\sigma^* D\Phi^i p_i] = d\,[A^i_j(x) p_i dx^j], \tag{3.33}$$

where $A^i_j(x) dx^j$ is the coordinate expression of $\sigma^* D\Phi^i$ and the matrix $A^i_j(x)$—that depends only on the base point x but is independent of the momentum p—is nonsingular since it is the local expression of the tetrad field, as stressed when we discussed the meaning of the soldering condition. Performing the exterior derivative, we find

$$\Omega = A^i_j(x) dp_i \wedge dx^j + (\partial A^i_j / \partial x^m) p_i dx^m \wedge dx^j, \tag{3.34}$$

which is clearly nondegenerate.

The equations of motion deduced by the symplectic form Ω and a Hamiltonian $H = H(\eta^{ij}p_i p_j)$ read

$$\left(\dot{x}^i \frac{\partial}{\partial x^i} + \dot{p}_i \frac{\partial}{\partial p_i}\right) \lrcorner \left(A^i_{\ j} dp_i \wedge dx^j + \frac{\partial A^i_{\ j}}{\partial x^k} p_i dx^k \wedge dx^j\right)$$
$$= -\frac{1}{m}\eta^{ij} p_i dp_j, \tag{3.35}$$

with $1/m = 2H'$, so that

$$p_j = m\eta_{rj}\dot{x}^k A^r_{\ k},$$
$$A^j_{\ k}\dot{p}_j + m\eta_{rj}A^r_{\ s}(\partial A^j_{\ k}/\partial x^i - \partial A^j_{\ i}/\partial x^k)\dot{x}^i\dot{x}^s = 0. \tag{3.36}$$

Defining

$$g_{kr} = \eta_{ab}A^a_{\ k}A^b_{\ r}, \tag{3.37}$$

and eliminating the momentum in the system (3.36), we find the usual geodesic equation

$$\ddot{x}^r + \Gamma^r_{\ ks}\dot{x}^k\dot{x}^s = 0, \tag{3.38}$$

with

$$\Gamma^r_{\ ks} = \tfrac{1}{2}g^{ur}(\partial g_{uk}/\partial x^s + \partial g_{us}/\partial x^k - \partial g_{sk}/\partial x^u) \tag{3.39}$$

and

$$g^{ur} = \eta^{ij}(A^{-1})^u_i(A^{-1})^r_j. \tag{3.40}$$

We notice that Eqs. (3.39) and (3.40) express the well-known relationship between metric tensor and tetrad field.[16] Moreover, in the spirit of the minimal substitution, it is easily seen that Eq. (3.38) is also obtained by the canonical symplectic form $dp_i \wedge dx^i$ and the Hamiltonian $H = H(g^{ur}p_u p_r)$, according to the procedure explained in Sec. 2. A discussion of the method of getting Eq. (3.38) from a different standpoint has been given also in Ref. (23).

IV. CONCLUSIONS

In this paper we have performed the detailed construction of a mechanical system representing a scalar particle interacting with the gravitational field as a gauge theory of the Poincaré group. The realization has been obtained in geometrical form, so that possible problems of interpretation turn out to be clarified. The results so far achieved naturally suggest the use of the same method for the investigation of other physically interesting situations.

In the first place we face the more general question of the spinning particle. In this case we need a symplectic manifold where the spin variables can be accomodated. The solution can be sought on a group theoretical basis. Starting with the cotangent bundle of the Minkowski space, T^*M, we may observe that the action of the Poincaré group on each fiber reduces to the action of the Lorentz group on the dual space of the translations subalgebra. As suggested in Ref. (27), a possible phase space for the spin variables, attached at any point $(x,p)\in T^*M$, is obtained by considering the orbits of the isotropy subgroup of the momentum p under the coadjoint representation of the subgroup itself. The problems posed by the patching of these spaces can easily be solved for a free particle. However, when considering gravitational gauge interactions, the new form of the soldering condition must take into account the additional structure introduced. One expects to obtain, in this way, an answer to the problem of spin-curvature coupling for particles of any mass.

A different application of the method we have developed, which is now under investigation, concerns supergravity. Indeed it has been shown that dynamical systems involving anticommuting variables may be described by a graded symplectic formalism[28,29] allowing the establishment of a Hamiltonian mechanics. Besides its direct interest in the realm of supersymmetry theories, such an investigation may be compared with the previous approach to give some insight in the relations between anticommuting variables and spin structures.[30]

APPENDIX

The aim of the present Appendix is to prove Eq. (3.11), i.e.,

$$\tilde{\omega} = \beta^*\omega + \beta^*\varphi - d\Phi - \rho'(\beta^*\omega)\Phi. \tag{A1}$$

The interest in a detailed demonstration lies in the possible generalizations of the procedure, namely in the fact that the same type of computations may be applied—but for minor changes—to those cases in which P and L are replaced by a Lie group G and a subgroup H, and the homogeneous space $P/L \simeq \mathbb{R}^{(1,3)}$ by G/H, with only the condition for G/H to be reductive.[19b,31]

Therefore consider the reduction function

$$\Phi: E_P \to \mathbb{R}^{(1,3)} \tag{A2}$$

and the bundle

$$E_L = \{u\in E_P | \Phi(u) = 0\}. \tag{A3}$$

The reduction map

$$\gamma: E_L \to E_P \tag{A4}$$

is the usual subbundle injection.

Let us show that the assignment on E_L of a connection one-form ω and a $\mathbb{R}^{(1,3)}$-valued one-form φ with the transformation property

$$(r^*_\Lambda \cdot \varphi)^i = \Lambda^i_{\ j}\varphi^j, \tag{A5}$$

uniquely determines a connection one-form $\tilde{\omega}$ on E_P such that

$$\gamma^*\tilde{\omega} = \omega + \varphi. \tag{A6}$$

At any point $u\in E_L \subseteq E_P$ a tangent vector $X_u \in T_u E_P$ admits a unique decomposition

$$X_u = Y_u + Z^*_u, \tag{A7}$$

where $Y_u \in T_u E_L$ and Z^*_u is the field tangent to the flow induced by the action of the one-parameter subgroup generated by the element $(0,Z) \in \mathrm{Lie}P$. This is obvious, since E_P is given by the union of the orbits of the translation subgroup of P at all points of E_L.

Let us define the Lie$P$-valued one-form $\tilde{\omega}$ on E_L

$$\tilde{\omega}(X_u) = (\omega + \varphi)(Y_u) + Z = \beta^*(\omega + \varphi)(X_u) + Z. \tag{A8}$$

Now,

$$d\Phi(Z_u) = Z^*_u\Phi = \frac{d}{dt}r^*_{\exp(t(0,Z))}\Phi|_{t=0} = -\rho'(0,Z)\Phi = -Z$$

and

$$(\beta^* d\Phi)(Y_u) = 0,$$

since Φ is constant on E_L. Therefore

$$d\phi(X_u) = -z \tag{A9}$$

and equation (A8) can be written

$$\widetilde{\omega}(X_u) = \beta^*(\omega + \varphi)(X_u) - d\Phi(X_u). \tag{A10}$$

The domain of $\widetilde{\omega}$ can be readily extended from E_L to the whole E_P. Indeed if $\tilde{u} \in E_P$, then

$$u = \tilde{u}(\mathbb{1}, \Phi(\tilde{u})) \in E_L$$

is the unique point of E_L that belongs to the orbit of the translation subgroup of P passing through u. If $X_{\tilde{u}} \in T_{\tilde{u}} E_P$, denoting

$$X_u = dr_{(1,\Phi(\tilde{u}))}(X_{\tilde{u}}) \in T_u E_P,$$

we define

$$\widetilde{\omega}(X_{\tilde{u}}) = \mathrm{ad}_{(1,-\Phi(\tilde{u}))}\,\widetilde{\omega}(X_u), \tag{A11}$$

where $\widetilde{\omega}(X_u)$ is given by (A10).

A straightforward computation shows that the one-form $\widetilde{\omega}$ given by (A11) satisfies the properties of a connection one-form.

Recalling the adjoint representation of the Poincaré group,

$$\mathrm{ad}_{(\Lambda,a)}(\lambda,\alpha) = (\mathrm{ad}_\Lambda \lambda, \Lambda\alpha - (\mathrm{ad}_\Lambda \lambda)a), \tag{A12}$$

we can explicitly evaluate the expression (A11), that gives

$$\widetilde{\omega}(X_{\tilde{u}}) = \beta^*(\omega + \varphi)(X_u) - d\Phi(X_u) - \rho'(\beta^*\omega(X_u))\Phi(\tilde{u}). \tag{A13}$$

Noticing that

$$\begin{aligned} d\Phi(X_u) &= d\Phi(dr_{(\mathbb{1},\Phi(\tilde{u}))}(X_{\tilde{u}})) = r^*_{(\mathbb{1},\Phi(\tilde{u}))} d\Phi(X_{\tilde{u}}) \\ &= d(\Phi + \Phi(\tilde{u}))(X_{\tilde{u}}) = d\Phi(X_{\tilde{u}}), \end{aligned} \tag{A14}$$

and obviously

$$\beta^*(\omega + \varphi)(X_u) = \beta^*(\omega + \varphi)(X_{\tilde{u}}), \tag{A15}$$

(A1) follows immediately from (A13).

[1] M. Daniel and C. M. Viallet, Rev. Mod. Phys. **52**, 175–197 (1980).
[2] S. Sternberg, Proc. Natl. Acad. Sci. USA **74**, 5253–5254 (1977).
[3] A. Weinstein, "Lectures on Symplectic Manifolds," C.B.M.S. Reg. Conf. Series in Math. 29 (Amer. Math. Soc., Providence, 1977) and references therein.
[4] W. Drechsler and M. E. Mayer, "Fiber Bundle Techniques in Gauge Theories," *Lecture Notes in Physics*, Vol. 67 (Springer-Verlag, Berlin, 1977).
[5] For the concept of deformation see M. Flato, A. Lichnerowicz, and D. Sternheimer, J. Math. Phys. **17**, 1754–1762 (1976).
[6] A. Kirillov, *Eléments de la Théorie des Représentations* (Editions MIR, Moscow, 1974).
[7] B. Kostant, "Quantization and Unitary Representations," in "Lectures on Modern Analysis and Applications III," *Lecture Notes in Mathematics* (Springer-Verlag, Berlin, 1970), Vol. 170, pp. 87–207.
[8] S. K. Wong, Nuovo Cimento A **65**, 689–694 (1970).
[9] A. Barducci, R. Casalbuoni, and L. Lusanna, Nucl. Phys. B **124**, 93–108 (1977); A. P. Balachandran, P. Salomonson, B. S. Skagerstam, and J. O. Winnberg, Phys. Rev. D **15**, 2308–2317 (1977).
[10] J. M. Souriau, Ann. Inst. Henri Poincaré, A **20**, 315–364 (1974).
[11] F. W. Hehl, P. von der Heyde, G. D. Kerlick, and J. M. Nester, Rev. Mod. Phys. **48**, 393–416 (1976); F. W. Hehl, "Four Lectures on Poincaré Gauge Field Theory," Lectures given at the 6th Course of the Theoretical School of Cosmology and Gravitation on "Spin, Torsion, Rotation, and Supergravity" (Erice, May 1979).
[12] R. Utiyama, Phys. Rev. **101**, 1597–1607 (1956).
[13] K. Hayashi and T. Nakano, Prog. Theor. Phys. **38**, 491–516 (1967).
[14] Y. M. Cho, Phys. Rev. D **14**, 2521–2525 (1976).
[15] K. A. Pilch, Lett. Math. Phys. **4**, 49–51 (1980).
[16] R. Giachetti, R. Ricci, and E. Sorace, Lett. Math. Phys. **4**, 297–300 (1980); and **5**, 85–91 (1981).
[17] T. W. B. Kibble, J. Math. Phys. **2**, 212–221 (1961).
[18] C. P. Luehr and P. M. Rosenbaum, J. Math. Phys. **21**, 1432–1438 (1980).
[19] (a) S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963 and 1969); Vol. I and Vol. II; (b) R. Giachetti, R. Ricci, and E. Sorace, "Manifold Geometry from Soldered Gauge Theories," preprint (Firenze, March 1981).
[20] S. Sternberg, *Lectures on Differential Geometry*, (Prentice-Hall, Englewood Cliffs, N.J., 1964).
[21] S. Sternberg, "On the Role of Field Theories in Our Physical Conception of Geometry," in "Differential Geometrical Methods in Mathematical Physics II," *Lecture Notes in Mathematics*, (Springer-Verlag, Berlin, 1978), Vol. 676, pp. 1–80.
[22] R. Abraham and J. E. Marsden, *Foundations of Mechanics*, (Benjamin, Reading, Mass., 1978).
[23] R. Giachetti and E. Sorace, Nuovo Cimento B **56**, 263–301 (1980).
[24] J. M. Souriau, *Structure des Systèmes Dynamiques*, (Dunod, Paris, 1970).
[25] V. Arnold, *Méthodes Mathématiques de la Máecanique Classique*, Appendice 5 (Editions MIR, Moscow, 1976).
[26] A. Trautman, Czech. J. Phys. B **29**, 107–116 (1979).
[27] S. Sternberg, Trans. Amer. Math. Soc. **212**, 113–130 (1975).
[28] B. Kostant, "Graded Manifolds, Graded Lie Theory and Prequantization," in "Differential Geometrical Methods in Mathematical Physics," *Lecture Notes in Mathematics*, Vol. 570 (Springer-Verlag, Berlin, 1977), pp. 177–306.
[29] R. Giachetti, R. Ragionieri, and R. Ricci, "Symplectic Structures on Graded Manifolds," to be published in J. Differential Geom.
[30] K. Bugajska, J. Math. Phys. **21**, 2091–2096 and 2097–2101 (1980).
[31] In the case of the affine group $GL(n) \times \mathbb{R}^n$ the analogous formula is given, without proof, by A. Trautman, "On the Structure of the Einstein–Cartan Equations," in *Symposia Mathematica XII, I.N.D.A.M.* (Academic, London, 1973), pp. 139–162.

Orthogonal transitivity, invertibility, and null geodesic separability in type D electrovac solutions of Einstein's field equations with cosmological constant

R. Debever
Département de Mathématique, Université Libre de Bruxelles, Brussels, Belgium
R. G. McLenaghan
Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, Canada

(Received 28 July 1980; accepted for publication 11 March 1981)

It is shown without explicit integration that all Petrov type D electrovac solutions with cosmological constant for an aligned, nonsingular electromagnetic field which satisfy the generalized Goldberg–Sachs theorem, admit at least a two-parameter, abelian, orthogonally transitive group of local isometries. In the case when the group orbits are non-null the group is invertible, and a symmetric null tetrad is shown to exist in which the principle null congruences defined by the type D Weyl tensor are indistinguishable. An explicit example is given of a solution with null group orbits which contains as a subcase a Kinnersley vacuum solution (with the same property). It is also demonstrated that the Hamilton–Jacobi equation for the null geodesics is always solvable by separation of variables in these solutions, a fact which explains the existence of a conformal Killing tensor therein, and which gives rise to a coordinate system in which the field equations may be integrated in terms of polynomial functions.

PACS numbers: 04.20.Jb, 02.30.Jr

I. INTRODUCTION

In his exhaustive integration of Einstein's vacuum field equations for Petrov type D Kinnersley[1] observed that *all* the solutions admit at least a two-parameter abelian isometry group. This fact has been proved without integration by Hughston and Sommers,[2] who extended the result to the class of Einstein-Maxwell solutions with nonsingular, aligned electromagnetic field. Successively simpler forms of Kinnersley's metrics including charge and cosmological constant have been given by Debever,[3] Plebański,[4] Plebański and Demiański,[5] and Weir and Kerr,[6] all these authors assuming implicitly that the two-parameter, abelian group of motions corresponding to the stationarity and axisymmetry of the metric is *invertible*.[7] These solutions, in all but the most general case, correspond to a class of solutions of the electrovac field equations with cosmological constant discovered earlier by Carter[8] under the hypotheses of an invertible, abelian, two-parameter isometry group and the separability of the Hamilton–Jacobi equation and the Schrödinger equation for the charged particle orbits, all of these solutions being necessarily of Petrov type D. Carter's hypotheses have been analysed from a conformal viewpoint by Debever,[9] who showed that Schrödinger separability was superfluous, and weakened the Hamilton–Jacobi separability to that for light-like particles. The separability of the Hamilton–Jacobi equation, as has been shown independently by Matravers[10] and Carter,[11] gives rise to a *fourth constant of the motion* for the particle orbits (the three other constants being the rest mass, the energy, and the angular momentum about the symmetry axis), which implies the existence of a second-rank *Killing tensor* in these solutions. Matravers has also shown the separability of the Hamilton-Jacobi equation for the null geodesics (generalized Carter metrics) gives rise to a nonzero quadratic first integral for the null geodesic equations which implies the existence of a *conformal Killing tensor.* Kinnersley's metrics have likewise been analyzed by Matravers, who finds that for all metrics except Case III the Hamilton-Jacobi equation is solvable by separation of variables, while for Case IIIA (the C metric) it is integrable in this way only for the null geodesics (Case IIIB, the C NUT, was not considered by Matravers). The results of Matravers and Carter are closely related to some earlier theorems of Walker and Penrose,[12] who sought to derive the existence of the fourth constant of motion in the Kerr[13] solution directly from the type D vacuum field equations *without explicit integration*. They found that *all* the vacuum type D solutions admit an irreducible second rank conformal Killing tensor which gives rise to a quadratic first integral for the null geodesics, while an irreducible second rank Killing tensor and its corresponding first integral for all geodesics exists in only a subclass of these solutions (including the Kerr solution). The results of Walker and Penrose have been extended to type D electrovac solutions with aligned nonsingular electromagnetic field by Hughston, Penrose, Sommers, and Walker[14] and by Hughston and Sommers.[15] The latter authors show that the C metric and the C NUT metric and their electrovac generalizations are the only metrics in the class which do not admit the full Killing tensor. It should be pointed out that the mere existence of four independent constants of motion does not imply directly that the Hamilton-Jacobi equation is solvable by separation of variables. Sufficient conditions for this to be true which seem to apply to the type D electrovac case just described have been given by Woodhouse.[16]

The purpose of the present paper[17] is to provide a uni-

[a]This work was supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada.

 © 1981 American Institute of Physics

fied and complete treatment of the results just described, from the perspective of the solutions of the Petrov type D electrovac field equations with cosmological constant for an aligned, nonsingular electromagnetic field which satisfy the generalized Goldberg–Sachs theorem.[18] (This class of solutions will be denoted by $\mathfrak{D}$.) We shall show *without explicit integration* that the field equations imply the existence of at least a two-parameter *orthogonally transitive*[19] abelian isometry group. In the case when the group orbits are non-null the group is necessarily invertible. However, when the orbits are null the group is *not* invertible. We give the general form of the metric in the latter case and exhibit an explicit solution of the electrovac field equations which contains as a subcase a Kinnersley metric admitting an orthogonally transitive isometry group with null orbits. The existence of such a solution is rather surprising since it has been assumed in the literature that all members of the class $\mathfrak{D}$ admit an invertible abelian two-parameter isometry group.

It will also be shown that our hypotheses imply the existence of a system of coordinates in which the Hamilton-Jacobi equation for the null geodesics is *always* solvable by separation of variables. The separation constant provides a quadratic fourth constant of motion for the null geodesics, which in turn gives rise to a conformal Killing tensor identical to the one previously found by Hughston and Sommers[20] from an analysis of the Bianchi identities for the electrovac field equations. It seems, however, that the existence of such a tensor for solutions in $\mathfrak{D}$ is related more fundamentally to the separability of the Hamilton-Jacobi equation for the null geodesics. It should be emphasized that the Hamilton-Jacobi equation for the case of the group with null orbits is included in the above analysis and gives rise to a kind of separability which does not seem to have been previously considered.[21]

Finally, we give a canonical form for the metric (in terms of the separable coordinates) which is essentially the point of departure for the integration of the field equations in terms of polynomial functions previously carried out by Carter,[22] Debever,[23] and Plebański and Demiański.[24] A remarkable feature of our analysis is that from our hypotheses of Petrov type D electrovac field equations, excluding the case of null group orbits, we recover the isometry and separability conditions I, II, III, and IV imposed by Carter[25] for his family of solutions. The only difference that should be mentioned is that we find a weakened version of Condition III, namely separability of the Hamilton–Jacobi equation for only the null geodesics. Even the case of null group orbits parallels closely Carter's situation, the only changes being that invertibility is replaced by orthogonal transitivity in Condition II, and that Condition IV is appropriately modified.

We prove our results by showing that aside from an exceptional case all solutions in the class $\mathfrak{D}$ admit a Riemannian–Maxwellian invertible structure (RMIS) and hence, by a theorem of Debever, McLenaghan, and Tariq[26] (earlier versions of which were given by Debever[27]), possess an invertible abelian two-parameter isometry group. Proving the existence of an RMIS is essentially equivalent to showing the existence of a *symmetric null tetrad* in which the Newman–Penrose[28] (NP) spin coefficients are equal (or opposed) in pairs. Consequently the principal null congruences of the type D Weyl tensor are *indistinguishable* except in the case of the null orbits, which require a separate treatment.

We shall perform our calculations using the NP formalism and the complex vectorial formalism of Cahen, Debever, and Defrise.[29] The relationship between these formalisms is given in DMT.[30]

2. HYPOTHESES AND STATEMENT OF RESULTS

We shall study space-times V_4 which are solutions of the Einstein–Maxwell field equations with cosmological constant

$$R_{ij} - \tfrac{1}{2}Rg_{ij} + \lambda g_{ij} = F_{ik}F_j{}^k - \tfrac{1}{4}g_{ij}F_{kl}F^{kl}, \tag{2.1a}$$

$$F_{ik;}{}^k = 0, \quad F_{[ij;k]} = 0, \tag{2.1b}$$

and which satisfy the following conditions:

H1. The Weyl tensor C_{ijkl} is *Petrov type D*; this is equivalent to the existence of real null vector fields l and n such that at each point

$$l^j l^k C_{ijk[l} l_{m]} = n^j n^k C_{ijk[l} n_{m]} = 0. \tag{2.2}$$

H2. The electromagnetic field tensor F_{ij} is *nonsingular* and its principal null directions are *aligned* with the principal null directions of the Weyl tensor, that is, we have

$$l^i F_{i[j} l_{k]} = n^i F_{i[j} n_{k]} = 0. \tag{2.3}$$

H3. The invariants of the Weyl tensor and the tracefree Ricci tensor S_{ij} satisfy one of the following inequalities:

$$C_{ijkl} {*C}^{ijkl} \neq 0, \tag{2.4}$$

or

$$C_{ijkl}C^{ijkl} \neq \tfrac{4}{3}S_{ij}S^{ij}. \tag{2.5}$$

The last hypothesis is required to insure that the generalized Goldberg-Sachs theorem holds, namely that the principal null congruences defined by l and n are both geodesic and shearfree. The degenerate case when H3 is not satisfied has recently been studied by Plebański and Hacyan.[31] We denote by $\mathfrak{D}$ the class of solutions of (2.1) satisfying the hypotheses H1, H2, and H3. The main results of this paper are contained in the following theorems.

Theorem 1: Every solution in $\mathfrak{D}$ admits at least a two-parameter *orthogonally transitive* abelian isometry group. If the orbits of the group are *non-null* the group is *invertible* and there exists a coordinate system (u,v,w,x) such that the metric and the self-dual Maxwell field have the form

$$\begin{aligned} ds^2 = &-e(Ldu + Mdv)^2 + eR^2dw^2 \\ &+ e(Ndu + Pdv)^2 - eT^2dx^2, \end{aligned} \tag{2.6a}$$

$$\begin{aligned} \overset{+}{F} = B\,[&R(Ldu + Mdv)\wedge dw \\ &- eT(Ndu + Pdv)\wedge dx], \end{aligned} \tag{2.6b}$$

where L, M and R are real valued functions, B is a complex valued function, and N, P, and T are functions satisfying

$$\bar{N} = -eN, \quad \bar{P} = -eP, \quad \bar{T} = eT, \tag{2.6c}$$

where *all* the functions are independent of the coordinates u and v, and where $e = 1$ (spacelike orbits) or $e = -1$ (timelike orbits). If the orbits are *null* the group is *not* invertible and there exists a system of coordinates (u,v,w,x) in which the metric and the self-dual Maxwell field have the form

$$ds^2 = 2Rdw(Ldu + Mdv) - (Ndu + Pdv)^2 - T^2dx^2, \quad (2.7a)$$

$$\overset{+}{F} = B\,[\,-R\,(Ldu + Mdv)\wedge dw + iT(Ndu + Pdv)\wedge dx\,], \quad (2.7b)$$

where L, M, N, P, R and T are real valued functions and B is a complex valued function, all independent of the coordinates u and v.

Since the components of the metrics and the Maxwell fields are independent of u and v it is clear that $\partial/\partial u$ and $\partial/\partial v$ are commuting Killing vector fields for both solutions (2.6) and (2.7). The invertibility of the isometry group for the metric (2.6) is evident from the fact that the transformation

$$(u,v,w,x)\to(-u,-v,w,x) \quad (2.8)$$

is an isometry. However, the transformation (2.8) is not an isometry for the metric (2.7). It should be noted that the Maxwell field tensor is *skew invertible* in the group, since the transformation (2.8) induces the transformation

$$\overset{+}{F}\to -\overset{+}{F}. \quad (2.9)$$

Theorem 1 shows that Carter's[32] isometry Condition I that "the space and the electromagnetic field are invariant under a two-parameter abelian symmetry group" holds for every solution in the class $\mathfrak{D}$, and moreover, that Condition II that "the symmetry group is invertible with non-null surfaces of transitivity" holds when the group orbits are non-null. In the case of null orbits we have to replace "invertible" in Condition II by "orthogonally transitive with null surfaces of transitivity".

Theorem 2: For every solution in $\mathfrak{D}$ there exists a coordinate system (u,v,w,x) in which the Hamilton–Jacobi equation for the null geodesics,

$$g^{ij}\frac{\partial S}{\partial x^i}\frac{\partial S}{\partial x^j} = 0, \quad (2.10)$$

is solvable by separation of variables, in the sence that it possesses a complete integral of the form

$$S = \alpha u + \beta v + S_1(w,\alpha,\beta,\gamma,\delta) + S_2(x,\alpha,\beta,\gamma,\delta), \quad (2.11)$$

where α,β,γ, and δ are constant. With respect to the separable coordinates the metrics have the following canonical forms:

Non-null orbits:

$$ds^2 = e^{2\psi}[\,-e(Ldu + Mdv)^2 + eR^2dw^2 + e(Ndu + Pdv)^2 - eT^2dx^2], \quad (2.12)$$

Null orbits:

$$ds^2 = e^{2\psi}[2R\,dw(Ldu + Mdv) - (Ndu + Pdv)^2 - T^2dx^2], \quad (2.13)$$

where in both cases the metric functions satisfy, in addition to the restrictions of Theorem 1, the conditions

$$\psi_u = \psi_v = 0 \quad (\psi_u = \partial\psi/\partial u, \text{etc.}), \quad (2.14)$$

$$R_x = T_w = 0, \quad (2.15)$$

$$(L/Z)_w = (M/Z)_w = (N/Z)_x = (P/Z)_x = 0, \quad (2.16)$$

where

$$Z = LP - MN. \quad (2.17)$$

If S of the form (2.11) is substituted in Eq. (2.10) one obtains the following equations on taking account of the canonical forms (2.12) and (2.13), respectively:

Non-null orbits:

$$Z^{-2}(\alpha P - \beta N)^2 - R^{-2}(S_1'(w))^2 = Z^{-2}(\beta L - \alpha M)^2 + T^{-2}(S_2'(x))^2, \quad (2.18)$$

Null orbits:

$$R^{-1}S_1'(w)Z^{-1}(\alpha P - \beta N) = \tfrac{1}{2}[Z^{-2}(\beta L + \alpha M)^2 + T^{-2}(S_2'(x))^2]. \quad (2.19)$$

In view of Theorem 2 it is clear that both the above equations separate into ordinary differential equations for the unknown functions S_1 and S_2. Thus Carter's separability Condition III that "the Hamilton–Jacobi equation is soluble by separation of variables in the simplest possible way", weakened to include only the null geodesics, holds for every solution in the class $\mathfrak{D}$. Furthermore, an inspection of Eq. (2.18) shows that Condition IV, that "the separation required by Condition III takes place in such a way that the terms containing derivatives with respect to the ignorable coordinates separate as the sum of two squares each depending on only one of the non-ignorable coordinates", also holds when the group orbits are non-null. An examination of Eq. (2.19) shows that Condtion IV does not hold for the case of the null orbits. Thus it seems that we have an example of an apparently new kind of separability.

It follows directly from Theorem 2 and Eqs. (2.18) and (2.19) that we have the following:

Corollary: Every solution in $\mathfrak{D}$ admits a second-rank conformal Killing tensor B^{ij} defined as follows:

Non-null orbits:

$$B^{ij}p_ip_j = \tfrac{1}{2}[R^{-2}p_w^2 + T^{-2}p_x^2 - Z^{-2}(Pp_u - Np_v)^2 - Z^{-2}(Lp_v - Mp_u)^2], \quad (2.20)$$

Null orbits:

$$B^{ij}p_ip_j = \tfrac{1}{2}[R^{-1}p_wZ^{-1}(Pp_u - Np_v) + \tfrac{1}{2}T^{-2}p_x^2 + \tfrac{1}{2}Z^{-2}(Lp_v - Mp_u)^2], \quad (2.21)$$

where

$$p_i = \partial S/\partial x^i \quad (2.22)$$

denote the components of the canonical momentum covector.

The conformal Killing tensors defined in the corollary are identical with the conformal Killing tensor previously found by Hughston and Sommers[33] in the case $\lambda = 0$ by an analysis of the Bianchi identities for the electrovac field equations. It seems, however, in view of the corollary that its existence for these solutions is more deeply related to the separability of the Hamilton–Jacobi equation for the null geodesics. We note that

$$K = B^{ij}p_ip_j = B_{ij}\dot{x}^i\dot{x}^j, \quad (2.23)$$

where a dot denotes differentiation with respect to an affine parameter, defines a quadratic first integral for the null geodesic equations. Thus by a result of Walker and Penrose[34] the tensor B_{ij} satisfies the conformal Killing equations

$$B_{(ij;k)} - \tfrac{1}{3}g_{(ij}B_{k)l;}{}^{l} = 0. \quad (2.24)$$

Theorem 2 also enables us to find canonical forms for the metric which facilitates the integration of the field equations (2.1). In the case of non-null orbits the canonical forms of the general metric and electromagnetic field are

$$ds^2 = e\exp[2\psi(w,x)]\Big\{-\Big[W(w)\Big(\frac{du+m(x)dv}{p(w)-m(x)}\Big)\Big]^2 + \frac{dw^2}{W^2(w)} + \Big[X(x)\Big(\frac{du+p(w)dv}{p(w)-m(x)}\Big)\Big]^2 + \frac{dx^2}{X^2(x)}\Big\}, \quad (2.25a)$$

$$\overset{+}{F} = \Big(\frac{B(w,x)}{p(w)-m(x)}\Big)[(du+m(x)dv)\wedge dw - ie(du+p(w)dv)\wedge dx], \quad (2.25b)$$

where all functions are real valued except B, which is complex valued and X, which is imaginary valued when $e=1$. The above canonical forms are a little more general than the Hamilton–Jacobi separable form [which requires that $\exp(2\psi)$ be the sum of a function of w and a function x] employed by Carter[35] in his integration procedure. They are the starting point for the explicit integration of the electrovac field equations by one of us[36] in an analysis of Carter's solutions. In order to obtain the metric which is the point of departure for the integration procedure of Plebański and Demiański[37] from the above canonical form, one must integrate most of the remaining type D electrovac field equations.

An example of a solution in $\mathfrak{D}$ possessing a group with null orbits is provided by the following metric and electromagnetic field:

$$ds^2 = 2dw(du + x^2dv) - Z^2\rho\bar{\rho}(du - w^2dv)^2 - dx^2/(Z^2\rho\bar{\rho}), \quad (2.26a)$$

$$\overset{+}{F} = a\rho^2[dw\wedge(du+x^2dv) + i\,dx\wedge(du - w^2dv)], \quad (2.26b)$$

where

$$\rho = -(w+ix)^{-1}, \quad Z^2 = 2lx - 2a\bar{a}, \quad (2.26c)$$

and where a is a complex constant and l a real constant. One has a solution of the vacuum field equations if and only if

$$a = 0, \quad (2.27)$$

in which case one recovers, modulo on obvious coordinate transformation, the solution Case II.E of Kinnersley[38] with $m=b=0$. It should be noted that this solution is a special case of a solution obtained by Leroy[39] under different hypotheses. It is an open question whether or not the above solution can be obtained from the seven-parameter family of Plebański and Demiański by a "limiting transition". The complete integration of the field equation for the solutions in $\mathfrak{D}$ possessing a group with null orbits will be presented elsewhere. Other results on isometry groups with null orbits may be found in Petrov[40] and Bampi and Cianci.[41]

3. NOTATION

We shall employ the complex vectorial formalism[42] with the notation of the Newman and Penrose[43] formalism. A covariant null tetrad of one-forms θ^a ($a = 1,2,3,4$) is defined in which the metric has the form

$$ds^2 = 2\theta^1\theta^2 - 2\theta^3\theta^4. \quad (3.1)$$

In a local coordinate system (x^i) we have

$$\theta^1 = n_i dx^i, \quad \theta^2 = l_i dx^i, \quad \theta^3 = -\bar{m}_i dx^i, \quad \theta^4 = -m_i dx^i, \quad (3.2)$$

where the covariant vectors l_i and n_i are real and null and the complex null vectors m_i and $\bar{m}_i$ are complex conjugate. The basis dual to $\{\theta^a\}$ is denoted by $\{X_a\}$ and the correspondence with the NP operators is given by

$$X_1 = D = l^i\partial_i, \quad X_2 = \Delta = n^i\partial_i, \quad X_3 = \delta = m^i\partial_i, \quad X_4 = \bar{\delta} = \bar{m}^i\partial_i. \quad (3.3)$$

A basis for the space of complex self-dual two-forms is given by

$$Z^1 = \theta^1\wedge\theta^3, \quad Z^2 = \theta^1\wedge\theta^2 - \theta^3\wedge\theta^4, \quad Z^3 = \theta^4\wedge\theta^2. \quad (3.4)$$

The components of the metric in this basis are

$$\gamma^{\alpha\beta} = 4(\delta^{\alpha}_{(1}\delta^{\beta}_{3)} - \delta^{\alpha}_{2}\delta^{\beta}_{2}). \quad (3.5)$$

The complex connection one-forms σ^{α}_{β} are defined by

$$dZ^{\alpha} = \sigma^{\alpha}{}_{\beta}\wedge Z^{\beta}. \quad (3.6)$$

The vectorial connection one-form is defined by

$$\sigma_{\alpha} = \tfrac{1}{8}e_{\alpha\beta\gamma}\gamma^{\gamma\delta}\sigma^{\beta}_{\delta}, \quad (3.7)$$

where $e_{\alpha\beta\gamma}$ is the three-dimensional permutation symbol. The tetrad components $\sigma_{\alpha a}$ defined by $\sigma_{\alpha} = \sigma_{\alpha a}\theta^a$ are 12 complex valued functions which are none other than the NP spin coefficients. The explicit correspondence is

$$\sigma_{\alpha a} = \begin{pmatrix} \kappa & \tau & \sigma & \rho \\ \epsilon & \gamma & \beta & \alpha \\ \pi & \nu & \mu & \lambda \end{pmatrix}. \quad (3.8)$$

The complex curvature two-forms $\Sigma^{\alpha}{}_{\beta}$ are defined by

$$d\sigma^{\alpha}_{\beta} - \sigma^{\alpha}{}_{\gamma}\wedge\sigma^{\gamma}_{\beta} = \Sigma^{\alpha}{}_{\beta}, \quad (3.9)$$

and the vectorial curvature two-form by

$$\Sigma_{\alpha} = \tfrac{1}{8}e_{\alpha\beta\gamma}\gamma^{\gamma\delta}\Sigma^{\beta}{}_{\delta}. \quad (3.10)$$

On expanding Σ_{α} in the basis $[Z^{\alpha},\bar{Z}^{\alpha}]$ one obtains

$$\Sigma_{\alpha} = (C_{\alpha\beta} - \tfrac{1}{6}R\gamma_{\alpha\beta})Z^{\beta} + E_{\alpha\bar{\beta}}\bar{Z}^{\beta}, \quad (3.11)$$

where the quantities $C_{\alpha\beta}$ and $E_{\alpha\bar{\beta}}$ are related to the NP curvature components Ψ_A and Φ_{ab} as follows:

$$C_{\alpha\beta} = \begin{pmatrix} \Psi_0 & \Psi_1 & \Psi_2 \\ \Psi_1 & \Psi_2 & \Psi_3 \\ \Psi_2 & \Psi_3 & \Psi_4 \end{pmatrix}, \quad E_{\alpha\bar{\beta}} = \begin{pmatrix} \Phi_{00} & \Phi_{01} & \Phi_{02} \\ \Phi_{10} & \Phi_{11} & \Phi_{12} \\ \Phi_{20} & \Phi_{21} & \Phi_{22} \end{pmatrix}. \quad (3.12)$$

A self-dual two-form $\overset{+}{F}$ may be expressed as

$$\overset{+}{F} = F_{\alpha}Z^{\alpha}. \quad (3.13)$$

The source-free Maxwell equations have the form

$$d\overset{+}{F} = 0, \quad (3.14)$$

and the Einstein-Maxwell equations may be written

$$E_{\alpha\beta} = A_\alpha \bar{A}_\beta. \tag{3.15}$$

Finally, the relations between the cosmological constant λ, the NP curvature component Λ, and the curvature scalar R are given by

$$\lambda = -6\Lambda = \tfrac{1}{4}R. \tag{3.16}$$

4. BASIC EQUATIONS FOR THE CLASS D

In view of the hypotheses H_1 and H_2 we may choose a local null tetrad whose real null vectors l and n are principal null vectors of the type D Weyl tensor and the nonsingular aligned electromagnetic field. It follows that the only nonvanishing NP components of the curvature are $\Psi_2 = \Psi$, $\Phi_{11} = \Phi$, and Λ, and that the self-dual Maxwell two-form is given by

$$\overset{+}{F} = BZ^2. \tag{4.1}$$

The null tetrad is determined by this choice up to the transformation

$$l' = e^a l, \quad n' = e^{-a} n, \quad m' = e^{ib} m, \tag{4.2}$$

where a and b are real valued functions. This tranformation induces the following transformation of the NP spin coefficients:

$$\kappa' = e^{2a+ib}\kappa, \quad \tau' = e^{ib}\tau, \quad \sigma' = e^{a+2ib}\sigma, \quad \rho' = e^a\rho, \tag{4.3}$$

$$\pi' = e^{-ib}\pi, \quad \nu' = e^{-2a-ib}\nu, \quad \mu' = e^{-a}\mu,$$

$$\lambda' = e^{-a-2ib}\lambda, \tag{4.4}$$

$$\epsilon' = e^a(\epsilon + \tfrac{1}{2}Dp), \quad \gamma' = e^{-a}(\gamma + \tfrac{1}{2}\Delta p), \tag{4.5}$$

$$\beta' = e^{ib}(\beta + \tfrac{1}{2}\delta p), \quad \alpha' = e^{-ib}(\alpha + \tfrac{1}{2}\bar{\delta} p), \tag{4.6}$$

where $p = a + ib$.
It can be shown that

$$dZ^2 = \theta \wedge Z^2, \tag{4.7}$$

where θ is a one-form with complex components defined by

$$\theta = -2(\rho\theta^1 - \mu\theta^2 + \tau\theta^3 - \pi\theta^4). \tag{4.8}$$

It follows from Eqs. (4.1) and (4.7) that Maxwell's equations (3.14) take the form

$$DB = 2\rho B, \quad \Delta B = -2\mu B, \quad \delta B = 2\tau B,$$

$$\bar{\delta} B = -2\pi B, \tag{4.9}$$

and that the integrability condition for these equations can be written as

$$d\theta = 0. \tag{4.10}$$

When this expressed in component form one has

$$\Delta\rho + D\mu = \rho(\gamma + \bar{\gamma}) - \mu(\epsilon + \bar{\epsilon}) + \pi\bar{\pi} - \tau\bar{\tau}, \tag{4.11a}$$

$$\delta\rho - D\tau = \rho(\bar{\alpha} + \beta - \bar{\pi}) - \tau(\bar{\rho} + \epsilon - \bar{\epsilon}) - \kappa\mu + \sigma\pi, \tag{4.11b}$$

$$\bar{\delta}\rho + D\pi = \rho(\alpha + \bar{\beta} - \pi) + \pi(\rho + \bar{\epsilon} - \epsilon) - \bar{\kappa}\mu - \bar{\sigma}\tau, \tag{4.11c}$$

$$\delta\mu + \Delta\tau = \mu(\tau - \bar{\alpha} - \beta) - \tau(\mu - \gamma + \bar{\gamma}) + \bar{\nu}\rho + \bar{\lambda}\pi, \tag{4.11d}$$

$$\Delta\pi - \bar{\delta}\mu = \mu(\alpha + \bar{\beta} - \bar{\tau}) - \pi(\bar{\mu} - \bar{\gamma} + \gamma) - \nu\rho + \lambda\tau, \tag{4.11e}$$

$$\delta\pi + \bar{\delta}\tau = \pi(\bar{\alpha} - \beta) + \tau(\alpha - \bar{\beta}) + \rho\bar{\mu} - \bar{\rho}\mu. \tag{4.11f}$$

In view of Eq. (4.1) the electrovac field equations (3.15) have the form

$$\Phi = B\bar{B}. \tag{4.12}$$

Moreover, the Bianchi identities may be expressed as

$$\kappa(3\Psi - 2\Phi) = \sigma(3\Psi + 2\Phi) = 0, \tag{4.13a}$$

$$\nu(3\Psi - 2\Phi) = \lambda(3\Psi + 2\Phi) = 0, \tag{4.13b}$$

$$D\Psi = \rho(3\Psi + 2\Phi), \quad \Delta\Psi = -\mu(3\Psi + 2\Phi),$$

$$\delta\Psi = \tau(3\Psi - 2\Phi), \quad \bar{\delta}\Psi = -\pi(3\Psi - 2\Phi), \tag{4.14}$$

$$D\Phi = 2(\rho + \bar{\rho})\Phi, \quad \Delta\Phi = -2(\mu + \bar{\mu})\Phi,$$

$$\delta\Phi = 2(\tau - \bar{\pi})\Phi. \tag{4.15}$$

We now invoke the hypothesis H_3 which, in view of the choices already made, becomes

$$9\Psi^2 \neq 4\Phi^2. \tag{4.16}$$

It follows immediately from Eq. (4.13) that

$$\kappa = \sigma = \nu = \lambda = 0. \tag{4.17}$$

This means that *both* the null congruences associated to the type D Weyl tensor are *geodesic* and *shear free*. Thus the generalized Goldberg–Sachs theorem[44] holds for the solutions in the class $\mathfrak{D}$. This result suggests that null congruences corresponding to l and n might be considered as *indistinguishable*. However, to establish this rigorously requires a chain of reasoning which fails in exceptional cases.

The next step is to derive the integrability conditions for Bianchi's identities. This is achieved by evaluation the NP commutators for the quantity Ψ, which yields the following equations on account of Eq. (4.17):

$$(\rho\bar{\mu} - \bar{\rho}\mu + \pi\bar{\pi} - \tau\bar{\tau})\Phi_{11} = 0, \tag{4.18a}$$

$$\delta\rho + D\tau = \rho(\bar{\alpha} + \beta + 2\tau + \bar{\pi}) - \bar{\rho}\tau + \tau(\epsilon - \bar{\epsilon}), \tag{4.18b}$$

$$\bar{\delta}\rho - D\pi = \rho(\alpha + \bar{\beta} - 4\pi - 2\bar{\tau}) + 2\bar{\rho}\pi + \pi(\epsilon - \bar{\epsilon}), \tag{4.18c}$$

$$\Delta\tau - \delta\mu = \mu(\bar{\alpha} + \beta - 4\tau - 2\bar{\pi}) + 2\bar{\mu}\tau + \tau(\gamma - \bar{\gamma}), \tag{4.18d}$$

$$\bar{\delta}\mu + \Delta\pi = -\mu(\alpha + \bar{\beta} + 2\pi + \bar{\tau}) + \bar{\mu}\pi + \pi(\bar{\gamma} - \gamma). \tag{4.18e}$$

Since $\Phi \neq 0$ it follows from Eq. (4.18a) that

$$\rho\bar{\mu} - \bar{\rho}\mu + \pi\bar{\pi} - \tau\bar{\tau} = 0,$$

from which we obtain the *key equations*[45]

$$\rho\bar{\mu} = \bar{\rho}\mu, \tag{4.19}$$

$$\tau\bar{\tau} = \pi\bar{\pi}. \tag{4.20}$$

It should be noted at this point that the conditions (4.17), (4.19), and (4.20) are invariant under the tetrad transformation (4.2).

By combining the Eqs. (4.11) with Eq. (4.18) we obtain the following equations independent of the NP field equations:

$$\Delta\rho + D\mu = \rho(\gamma + \bar{\gamma}) - \mu(\epsilon + \bar{\epsilon}), \tag{4.21a}$$

$$\bar{\delta}\rho = \rho(\alpha + \bar{\beta} - 2\pi - \bar{\tau}) + \bar{\rho}\pi, \tag{4.21b}$$

$$D\pi = \pi(\bar{\epsilon} - \epsilon + 2\rho - \bar{\rho}) + \rho\bar{\tau}, \tag{4.21c}$$

$$\Delta\tau = \tau(\gamma - \bar{\gamma} - 2\mu + \bar{\mu}) - \mu\bar{\pi}, \tag{4.21d}$$

$$\delta\mu = -\mu(\bar{\alpha} + \beta - 2\tau - \bar{\pi}) - \bar{\mu}\tau. \tag{4.21e}$$

Finally the NP curvature equations [Eqs. (4.2a)–(4.2r) in their paper] for the class $\mathfrak{D}$ are

$$D\rho = \rho^2 + (\epsilon + \bar{\epsilon})\rho, \tag{4.22a}$$

$$D\tau = \rho(\tau + \bar{\pi}) + \tau(\epsilon - \bar{\epsilon}), \tag{4.22b}$$

$$D\alpha - \bar{\delta}\epsilon = \alpha(\rho + \bar{\epsilon} - 2\epsilon) - \bar{\beta}\epsilon + \pi(\epsilon + \rho), \tag{4.22c}$$

$$D\beta - \delta\epsilon = \beta(\bar{\rho} - \bar{\epsilon}) - \epsilon(\bar{\alpha} - \bar{\pi}), \tag{4.22d}$$

$$D\gamma - \Delta\epsilon = \alpha(\tau + \bar{\pi}) + \beta(\bar{\tau} + \pi) - \gamma(\epsilon + \bar{\epsilon}) - \epsilon(\gamma + \bar{\gamma}) + \tau\pi + \Psi - \Lambda + \Phi, \tag{4.22e}$$

$$\bar{\delta}\pi = -\pi^2 + \pi(\bar{\beta} - \alpha), \tag{4.22f}$$

$$D\mu - \delta\pi = \bar{\rho}\mu + \pi\bar{\pi} - \mu(\epsilon + \bar{\epsilon}) - \pi(\bar{\alpha} - \beta) + \Psi + 2\Lambda, \tag{4.22g}$$

$$\Delta\pi = -\mu(\bar{\tau} + \pi) + \pi(\bar{\gamma} - \gamma), \tag{4.22h}$$

$$\delta\rho = \rho(\bar{\alpha} + \beta) + \tau(\rho - \bar{\rho}), \tag{4.22i}$$

$$\delta\alpha - \bar{\delta}\beta = \rho\mu + \alpha\bar{\alpha} + \beta\bar{\beta} - 2\alpha\beta + \gamma(\rho - \bar{\rho}) + \epsilon(\mu - \bar{\mu}) - \Psi + \Lambda + \Phi, \tag{4.22j}$$

$$\bar{\delta}\mu = \pi(\bar{\mu} - \mu) - \mu(\alpha + \bar{\beta}), \tag{4.22k}$$

$$\Delta\mu = -\mu^2 - \mu(\gamma + \bar{\gamma}), \tag{4.22l}$$

$$\delta\gamma - \Delta\beta = \gamma(\tau - \bar{\alpha} - \beta) + \beta(\bar{\gamma} - \gamma + \mu) - \mu\tau, \tag{4.22m}$$

$$\delta\tau = \tau(\tau + \beta - \bar{\alpha}), \tag{4.22n}$$

$$\Delta\rho - \bar{\delta}\tau = -\rho\bar{\mu} - \tau\bar{\tau} + \tau(\bar{\beta} - \alpha) + \rho(\gamma + \bar{\gamma}) - \Psi - 2\Lambda, \tag{4.22o}$$

$$\Delta\alpha - \bar{\delta}\gamma = \alpha(\bar{\gamma} - \bar{\mu}) + \gamma(\bar{\beta} - \bar{\tau}). \tag{4.22p}$$

Equations (4.17), (4.19), (4.20), (4.21), and (4.22) are the basic equations required to prove Theorems 1 and 2.

5. PROOF OF THEOREM 1 FOR NON-NULL ORBITS

In order to prove the part of Theorem 1 concerning the non-null orbits it is sufficient to prove the following.

Theorem 3: In every solution belonging to $\mathfrak{D}$ for which $\rho\mu \neq 0$ or $\rho = \mu = 0$, there exists a local null tetrad in which the only nonvanishing NP curvature components are Ψ_2, Φ_{11}, and Λ and in which the NP spin coefficients satisfy the relations

$$\kappa = \nu = \sigma = \lambda = 0, \tag{5.1}$$

$$\mu = -e\rho, \quad \pi = -e\tau,$$

$$\gamma = -e\epsilon, \quad \beta = -e\alpha, \tag{5.2}$$

$$A \equiv 2(\alpha - e\bar{\alpha}) + e\tau - \bar{\tau} = 0, \tag{5.3}$$

$$iE \equiv 2(\epsilon - \bar{\epsilon}) - \rho + \bar{\rho} = 0, \tag{5.4}$$

$$\mathscr{D}\rho = \mathscr{D}\tau = \mathscr{D}\epsilon = \mathscr{D}\alpha = 0, \tag{5.5}$$

$$\mathscr{L}\rho = \mathscr{L}\tau = \mathscr{L}\epsilon = \mathscr{L}\alpha = 0, \tag{5.6}$$

where

$$e^2 = 1, \tag{5.7}$$

and $\mathscr{D}$ and $\mathscr{L}$ are differential operators defined by

$$\mathscr{D} = D - e\Delta, \tag{5.8}$$

$$\mathscr{L} = \delta - e\bar{\delta}. \tag{5.9}$$

In the terminology of DMT the above theorem implies that the solutions in $\mathfrak{D}$ admit a *Riemannian Maxwellian invertible structure*. Hence by the main theorem of DMT they possess locally an invertible two-parameter abelian isometry group and a system of coordinates in which the metric and self-dual Maxwell field have the form (2.6). The fact that B is independent of u and v follows from Eqs. (4.9) and (5.2) and the DMT Eqs. (9.33) and (9.34). Thus once Theorem 3 is proven, the relevant part of Theorem 1 is established. Because of the symmetrical relations (5.1) and (5.2) between the spin coefficients ρ,τ,κ, and σ associated to the tetrad vector l and μ,π,ν, and λ associated to n we say (using another concept from DMT) that the solutions in $\mathfrak{D}$ which satisfy the conditions of Theorem 3 possess a *symmetric null tetrad* which for the metric (2.6) is given by DMT Eqs. (2.14).

In order to prove Theorem 3 we first remark that if we choose the tetrad such that Eq. (4.1) holds we have Ψ_2, Φ_{11}, and Λ as the only nonzero curvature components with Eq. (4.17) holding. Thus it remains to show that a null tetrad can be chosen in the family (4.2) such that Eqs. (5.2)–(5.6) are satisfied. To establish this several cases have to be considered.

Case I: $\rho\mu\tau \neq 0$. On account of Eqs. (4.19) and (4.20) we may use the remaining tetrad freedom of Eq. (4.2) to set

$$\mu = -e\rho, \tag{5.10}$$

$$\pi = -e\tau, \tag{5.11}$$

where $e^2 = 1$. The tetrad is now fixed up to the transformation

$$l' = l, \quad n' = n, \quad m' = -m. \tag{5.12}$$

We now show that the remaining equations of (5.2) and Eqs. (5.3) and (5.4) are satisfied as a consequence of Eqs. (5.10), (5.11), and the basic equations. From Eqs. (4.22f), (4.22l), and (4.22o) one gets

$$\rho^2 - e(\gamma + \bar{\gamma})\rho - \tau^2 + e(\alpha - \bar{\beta})\tau = \rho\bar{\rho} + e(\bar{\beta} - \alpha)\tau - e\tau\bar{\tau} + e(\gamma + \bar{\gamma})\rho - e(\Psi + 2\Lambda). \tag{5.13}$$

On the other hand, Eqs. (4.22b), (4.22g), and (4.22n) yield

$$\tau^2 + (\beta - \bar{\alpha})\tau - \rho^2 - \rho(\epsilon + \bar{\epsilon}) = -\rho\bar{\rho} + e\tau\bar{\tau} + (\epsilon + \bar{\epsilon})\rho + (\bar{\alpha} - \beta)\tau + e(\Psi + 2\Lambda). \tag{5.14}$$

The sum of these equations gives

$$\rho[e(\gamma + \bar{\gamma}) + \epsilon + \bar{\epsilon}] + \tau[\bar{\alpha} - \beta + e(\bar{\beta} - \alpha)] = 0. \tag{5.15}$$

In addition Eqs. (4.21a) and (4.22a), and (4.22l) imply that

$$\rho[\gamma + \bar{\gamma} + e(\epsilon + \bar{\epsilon})] = 0. \tag{5.16}$$

Thus we have

$$\gamma + \bar{\gamma} + e(\epsilon + \bar{\epsilon}) = 0, \tag{5.17}$$

and from (5.15)

$$\tau[\bar{\alpha} - \beta + e(\bar{\beta} - \alpha)] = 0, \tag{5.18}$$

which implies

$$\bar{\alpha} - \beta + e(\bar{\beta} - \alpha) = 0. \tag{5.19}$$

Furthermore, from Eqs. (4.21b) and (4.22k) one obtains

$$\rho[2(\alpha + \bar{\beta}) + e\tau - \bar{\tau}] = 0, \tag{5.20}$$

or

$$2(\alpha + \bar{\beta}) + e\tau - \bar{\tau} = 0. \tag{5.21}$$

If one now adds the complex conjugate of this equation to the original equation multiplied by e, one obtains

$$\bar{\alpha} + \beta + e(\alpha + \bar{\beta}) = 0. \tag{5.22}$$

This equation combined with Eq. (5.19) implies

$$\beta = -e\alpha. \tag{5.23}$$

When the last equation is used to replace β in Eq. (5.21) one obtains Eq. (5.3).

We proceed by equating the right-hand sides of Eqs. (4.21d) and (4.22h), which yields

$$\tau[e(\rho - \bar{\rho}) + 2(\gamma - \bar{\gamma})] = 0, \tag{5.24}$$

or

$$e(\rho - \bar{\rho}) + 2(\gamma - \bar{\gamma}) = 0. \tag{5.25}$$

On the other hand Eqs. (4.21c) and (4.22b) give

$$\tau[\rho - \bar{\rho} + 2(\bar{\epsilon} - \epsilon)] = 0, \tag{5.26}$$

or

$$\rho - \bar{\rho} - 2(\epsilon - \bar{\epsilon}) = 0, \tag{5.27}$$

which is none other than Eq. (5.4). This equation and Eq. (5.25) imply that

$$\gamma - \bar{\gamma} + e(\epsilon - \bar{\epsilon}) = 0, \tag{5.28}$$

which when added to Eq. (5.17) yields

$$\gamma = -e\epsilon. \tag{5.29}$$

Finally it follows from DMT Eqs. (7.26), (7.27), (8.19), and (8.24) that Eqs. (5.5) and (5.6) hold, which completes the proof of Theorem 3 in this case.

Case IIa: $\rho\mu \neq 0$, $\tau = 0$. Here we have

$$\pi = 0 \tag{5.30}$$

by Eq. (4.20). In view of Eq. (4.19) we may choose the function a in the transformation (4.2) such that (on dropping primes)

$$\mu = -e\rho, \tag{5.31}$$

where again $e^2 = 1$.

The remaining liberty in the tetrad is given by the transformation

$$l' = l, \quad n' = n, \quad m' = e^{ib}m. \tag{5.32}$$

By Eqs. (5.16) and (5.20), which still hold in this case, we have

$$\gamma + \bar{\gamma} + e(\epsilon + \bar{\epsilon}) = 0, \tag{5.33}$$

$$\bar{\alpha} + \beta = 0. \tag{5.34}$$

It follows from these equations, and Eqs. (4.21b), (4.21e), (4.22a), and (4.22k) that

$$\delta\rho = \bar{\delta}\rho = 0, \tag{5.35}$$

$$\mathscr{D}\rho = 0. \tag{5.36}$$

On account of Eqs. (4.5) and (5.33) we may use the tetrad transformation (5.32) to set

$$\gamma = -e\epsilon. \tag{5.37}$$

We note that this condition is preserved by transformations (5.32) satisfying

$$(D + e\Delta)b = 0. \tag{5.38}$$

Next we perform a tetrad rotation (5.32) such that (5.38) holds and such that Eq. (5.4) is also valid, that is

$$iE' \equiv 2(\epsilon' - \bar{\epsilon}') - \rho' + \bar{\rho}' = 0. \tag{5.39}$$

In view of Eqs. (4.3) and (4.5) the additional equation to be satisfied by b is

$$Db = i[\tfrac{1}{2}(\bar{\rho} - \rho) + \epsilon - \bar{\epsilon}], \tag{5.40}$$

and since (5.38) must also hold we also have

$$\Delta b = ie[\tfrac{1}{2}(\rho - \bar{\rho}) + \bar{\epsilon} - \epsilon]. \tag{5.41}$$

The integrability condition for these equations,

$$[\Delta, D] = -e(\epsilon + \bar{\epsilon})\mathscr{D}, \tag{5.42}$$

is satisfied on account of Eqs. (4.22l), (4.22o), and (5.36). We now drop the primes on the transformed quantities, noting that Eqs. (5.37) and (5.39) are preserved by transformations (5.32) satisfying

$$Db = \Delta b = 0. \tag{5.43}$$

We now show that it is possible to make a further transformation (4.2) satisfying Eq. (5.43) such that

$$\beta' = -e\alpha' \tag{5.44}$$

which, on account of Eq. (5.34), is equivalent to

$$\bar{\alpha}' = e\alpha'. \tag{5.45}$$

In view of Eq. (4.6) the additional equation to be satisfied by b is

$$e^{ib}(\bar{\alpha} - \tfrac{1}{2}i\,\delta b) = ee^{-ib}(\alpha + \tfrac{1}{2}i\,\bar{\delta} b). \tag{5.46}$$

It remains to verify that Eqs. (5.43) and (5.46) possess a common solution. To show this we define the differential operators

$$F_1(x^i, b, p_i) = Db = l^i p_i, \tag{5.47}$$

$$F_2(x^i, b, p_i) = \Delta b = n^i p_i, \tag{5.48}$$

$$F_3(x^i, b, p_i) = e^{ib}(\bar{\alpha} - \tfrac{1}{2}im^i p_i) - ee^{-ib}(\alpha + \tfrac{1}{2}i\,\bar{m}^i p_i), \tag{5.49}$$

where

$$p_i = \partial b / \partial x^i. \tag{5.50}$$

In order that the system of nonlinear partial differential equations

$$F_\eta(x^i, b, p_i) = 0 \quad (\eta = 1,2,3), \tag{5.51}$$

admit a common solution it is necessary and sufficient[46] (assuming suitable differentiability conditions) that the Poisson brackets

$$(F_\eta, F_\iota) = \frac{\partial F_\eta}{\partial p_i}\frac{dF_\iota}{dx^i} - \frac{\partial F_\iota}{\partial p_i}\frac{dF_\eta}{dx^i}, \tag{5.52}$$

where

$$\frac{dF_\iota}{dx^i} = \frac{\partial F_\iota}{\partial x^i} + \frac{\partial F_\iota}{\partial b}\frac{\partial b_\iota}{\partial x^i} \tag{5.53}$$

vanish for $\eta,\iota = 1,2,3$ as a consequence of Eqs. (5.51). First we have

$$(F_1,F_2) = [D,\Delta]b = 0 \tag{5.54}$$

by Eqs. (5.42) and (5.43). We then calculate

$$(F_1,F_3) = e^{ib}(D\bar{\alpha} + \tfrac{1}{2}i[\delta,D]b) - ee^{-ib}(D\alpha - \tfrac{1}{2}i[\bar{\delta},D]b) + i[e^{ib}(\bar{\alpha} - \tfrac{1}{2}i\delta b) + ee^{-ib}(\alpha + \tfrac{1}{2}i\bar{\delta}b)]Db, \tag{5.55}$$

$$(F_2,F_3) = e^{ib}(\Delta\bar{\alpha} + \tfrac{1}{2}i[\delta,\Delta]b) - ee^{-ib}(\Delta\alpha - \tfrac{1}{2}i[\bar{\delta},\Delta]b) + i[e^{ib}(\bar{\alpha} - \tfrac{1}{2}i\delta b) + ee^{-ib}(\alpha + \tfrac{1}{2}i\bar{\delta}b)]\Delta b. \tag{5.56}$$

On account of Eq. (5.51) for $\eta = 1$ and the commutator

$$[\delta,D] = -\tfrac{1}{2}(\rho + \bar{\rho})\delta, \tag{5.57}$$

Eq. (5.55) becomes

$$(F_1,F_3) = e^{ib}[D\bar{\alpha} - \tfrac{1}{4}i(\rho + \bar{\rho})\delta b] - ee^{-ib}[D\alpha + \tfrac{1}{4}i(\rho + \bar{\rho})\bar{\delta}b]. \tag{5.58}$$

In order to reduce this equation further we need an expression for $D\alpha$. Now Eqs. (4.22c), and (4.22d) yield, respectively,

$$D\alpha - \bar{\delta}\epsilon = \tfrac{1}{2}(\rho + \bar{\rho})\alpha, \tag{5.59}$$

$$D\bar{\alpha} + \delta\epsilon = \tfrac{1}{2}(\rho + \bar{\rho})\bar{\alpha}, \tag{5.60}$$

which imply that

$$\delta(\epsilon + \bar{\epsilon}) = 0. \tag{5.61}$$

On the other hand, by applying δ to both sides of Eq. (5.4) and noting Eq. (5.35), we obtain

$$\delta(\epsilon - \bar{\epsilon}) = 0. \tag{5.62}$$

It thus follows that

$$\delta\epsilon = \delta\bar{\epsilon} = 0 \tag{5.63}$$

and

$$D\alpha = \tfrac{1}{2}(\rho + \bar{\rho})\alpha. \tag{5.64}$$

In view of this equation, Eq. (5.58) becomes

$$(F_1,F_3) = \tfrac{1}{2}(\rho + \bar{\rho})F_3(x^i,b,p_i) = 0, \tag{5.65}$$

on account of Eq. (5.51) for $\eta = 3$. In an entirely similar manner, using Eqs. (4.22m,p) and the commutator

$$[\delta,\Delta] = -(e/2)(\rho + \bar{\rho})\delta, \tag{5.66}$$

it can be shown that

$$(F_2,F_3) = (e/2)(\rho + \bar{\rho})F_3(x^i,b,p_i) = 0. \tag{5.67}$$

Thus we conclude that the system (5.51) has a solution, which implies that there exists a tetrad such that Eqs. (5.1), (5.2), and (5.4) are satisfied. We note that Eq. (5.3) is also satisfied on account of (5.45). We further note that, in view of DMT Eqs. (7.26), (7.27), and (8.19), Eqs. (5.5) and (5.6) are satisfied except that we do not have

$$\mathscr{L}\alpha = 0. \tag{5.68}$$

The remainder of the proof in this case consists in showing that one can make a final tetrad rotation which preserves all the conditions already obtained and such that Eq. (5.68) also holds. In order to preserve conditions (5.2) and (5.4) it is necessary that the function b in Eq. (5.32) satisfy Eqs. (5.43) and the equation

$$e^{ib}(e\alpha - (i/2)\delta b) - ee^{-ib}(\alpha + (i/2)\bar{\delta}b) = 0. \tag{5.69}$$

The transformation law for $\mathscr{L}\alpha$ with b a solution of these equations is

$$\mathscr{L}'\alpha' = \mathscr{L}\alpha + i[\tfrac{1}{2}(\delta\bar{\delta}b + \bar{\delta}\delta b) - \alpha(\delta b + e\bar{\delta}b)]. \tag{5.70}$$

On account of Eq. (5.32) and the commutator

$$[\bar{\delta},\delta] = e(\rho - \bar{\rho})\mathscr{D} + 2\alpha\mathscr{L}, \tag{5.71}$$

Eq. (5.70) reduces to

$$\mathscr{L}'\alpha' = \mathscr{L}\alpha + i(\delta\bar{\delta}b - 2e\alpha\bar{\delta}b). \tag{5.72}$$

Thus in order to set $\mathscr{L}'\alpha' = 0$ the function b must satisfy the second order P. D. E.

$$\delta\bar{\delta}b - 2e\alpha\bar{\delta}b = i\mathscr{L}\alpha, \tag{5.73}$$

in addition to the first-order Equations (5.43) and (5.69). To show that this system of P. D. E.'s possesses a solution we apply in turn the operators D and Δ to both sides of Eq. (5.73). In the first case we have

$$D\delta\bar{\delta}b - 2eD\alpha\bar{\delta}b - 2e\alpha D\bar{\delta}b = iD\mathscr{L}\alpha. \tag{5.74}$$

On account of the Eq. (5.64) and the commutators (5.57) and

$$[D,\mathscr{L}] = \tfrac{1}{2}(\rho + \bar{\rho})\mathscr{L}, \tag{5.75}$$

Eq. (5.74) becomes

$$(\rho + \bar{\rho})(\delta\bar{\delta}b - 2e\alpha\bar{\delta}b) = i(\rho + \bar{\rho})\mathscr{L}\alpha, \tag{5.76}$$

which is *identically satisfied* in view of Eq. (5.73). Similarly by, applying Δ to Eq. (5.73) we obtain

$$\Delta\delta\bar{\delta}b - 2e\Delta\alpha\bar{\delta}b - 2e\alpha\Delta\bar{\delta}b = i\Delta\mathscr{L}\alpha. \tag{5.77}$$

By means of Eqs. (4.22p), (5.63) and the commutators (5.66) and

$$[\Delta,\mathscr{L}] = (e/2)(\rho + \bar{\rho})\mathscr{L}, \tag{5.78}$$

Eq. (5.77) takes the form

$$e(\rho + \bar{\rho})(\delta\bar{\delta}b - 2e\alpha\bar{\delta}b) = ie(\rho + \bar{\rho})\mathscr{L}\alpha, \tag{5.79}$$

which is the same as Eq. (5.76). In view of these results we may conclude that our system of P. D. E.'s for b is *compatible* and hence possesses a solution. Thus we have found a tetrad in which Eqs. (5.1)–(5.6) hold, which completes the proof of Theorem 3 for this case.

Case IIb: $\rho = \mu = 0$, $\tau \neq 0$. In view of Eq. (4.20) we may use the function b in the transformation (4.2) to set

$$\pi = \tau. \tag{5.80}$$

Alternatively we could have set $\pi = -\tau$. However, there is no essential difference between these possibilities. Condition (5.80) is preserved by the transformation

$$l' = e^a l, \quad n' = e^{-a}n, \quad m' = e_1 m, \tag{5.81}$$

where a is an arbitrary function and $e_1^2 = 1$. Equations (5.18), (5.24), and (5.26), which still hold in this case with $e = -1$, imply

$$\beta + \bar{\beta} = \alpha + \bar{\alpha}, \tag{5.82}$$

$$\gamma = \bar{\gamma}, \tag{5.83}$$

$$\epsilon = \bar{\epsilon}. \tag{5.84}$$

It follows from Eqs. (4.22b), (4.22f), (4.22h), and (4.22n) that

$$D\tau = \Delta\tau = \mathscr{L}\tau = 0. \tag{5.85}$$

In view of Eqs. (4.6) and (5.82) we may use the transformation (5.81) with $e_1 = 1$ to set (dropping primes)

$$\beta = \alpha. \tag{5.86}$$

We note that this condition is preserved by transformations satisfying

$$(\delta - \bar{\delta})a = 0. \tag{5.87}$$

The next step consists in using the tranformation (5.81) to set

$$A' \equiv 2(\alpha' + \bar{\alpha}') - \tau' - \bar{\tau}' = 0, \tag{5.88}$$

while preserving Eq. (5.86). By Eqs. (4.3), (4.6), and (5.87) the equations to be satisfied by a are

$$\delta a = \bar{\delta} a = \tfrac{1}{2}(\tau + \bar{\tau}) - \alpha - \bar{\alpha}. \tag{5.89}$$

The integrability condition for these equations,

$$[\bar{\delta},\delta] = (\alpha - \bar{\alpha})\mathscr{L}, \tag{5.90}$$

is satisfied by virtue of Eqs. (4.22g) (4.22j), and (5.85). We drop the primes on the transformed quantities and note that Eqs. (5.86) and (5.88) are preserved by transformations satisfying

$$\delta a = \bar{\delta} a = 0. \tag{5.91}$$

We proceed by using a transformation which preserves Eqs. (5.86) and (5.88) to set

$$\gamma' = \epsilon'. \tag{5.92}$$

The equation to be satisfied by a in addition to Eq. (5.91) is

$$e^{-a}(\gamma + \tfrac{1}{2}\Delta a) = e^{a}(\epsilon + \tfrac{1}{2}Da). \tag{5.93}$$

It must now be verified that these equations possess a common solution. To this end we introduce the partial differential operators (as in Case II a)

$$G_1(x^i,a,p_i) = \delta a = m^i p_i, \tag{5.94}$$

$$G_2(x^i,a,p_i) = \bar{\delta} a = \bar{m}^i p_i, \tag{5.95}$$

$$G_3(x^i,a,p_i) = e^{-a}(\gamma + \tfrac{1}{2}n^i p_i) - e^{a}(\epsilon + \tfrac{1}{2}l^i p_i). \tag{5.96}$$

To establish that the system of nonlinear P. D. E.'s

$$G_\eta(x^i,a,p_i) = 0 \quad (\eta = 1,2,3), \tag{5.97}$$

has a solution we must show that the Poisson brackets (G_η,G_ι) for $\eta,\iota = 1,2,3$ vanish as a consequence of Eqs. (5.97). To begin we note that by Eqs. (5.90) and (5.91)

$$(G_1,G_2) = [\delta,\bar{\delta}]a = (\bar{\alpha} - \alpha)\mathscr{L} a = 0. \tag{5.98}$$

Next we calculate

$$\begin{aligned}(G_1,G_3) &= e^{-a}(\delta\gamma + \tfrac{1}{2}[\delta,\Delta]a) - e^{a}(\delta\epsilon + \tfrac{1}{2}[\delta,D]a)\\ &\quad - [e^{-a}(\gamma + \tfrac{1}{2}\Delta a) + e^{a}(\epsilon + \tfrac{1}{2}Da]\delta a,\end{aligned} \tag{5.99}$$

and note that

$$(G_2,G_3) = \overline{(G_1,G_3)}. \tag{5.100}$$

On account of Eq. (5.91) and the commutators

$$[\delta,D] = \tfrac{1}{2}(\tau - \bar{\tau})D, \tag{5.101}$$

$$[\delta,\Delta] = \tfrac{1}{2}(\tau - \bar{\tau})\Delta, \tag{5.102}$$

Eq. (5.99) reduces to

$$(G_1,G_3) = e^{-a}[\delta\gamma + \tfrac{1}{4}(\tau - \bar{\tau})\Delta a] - e^{a}[\delta\epsilon + \tfrac{1}{4}(\tau - \bar{\tau})Da]. \tag{5.103}$$

It now remains to evaluate $\delta\gamma$ and $\delta\epsilon$. To achieve this we note that Eqs. (4.22c), and (4.22d) have the form

$$D\alpha - \bar{\delta}\epsilon = \tfrac{1}{2}\epsilon(\tau - \bar{\tau}), \tag{5.104}$$

$$D\alpha - \delta\epsilon = \tfrac{1}{2}\epsilon(\bar{\tau} - \tau), \tag{5.105}$$

from which it follows on account of Eq. (5.84) that

$$D(\alpha - \bar{\alpha}) = 0. \tag{5.106}$$

On the other hand, by applying D to both sides of Eq. (5.3) and noting Eq. (5.85) we obtain

$$D(\alpha + \bar{\alpha}) = 0. \tag{5.107}$$

It follows from these equations that

$$D\alpha = 0, \tag{5.108}$$

$$\delta\epsilon = \tfrac{1}{2}\epsilon(\tau - \bar{\tau}). \tag{5.109}$$

In a similar manner it may be shown that

$$\Delta\alpha = 0, \tag{5.110}$$

$$\delta\gamma = \tfrac{1}{2}\gamma(\tau - \bar{\tau}). \tag{5.111}$$

On account of Eqs. (5.109) and (5.111), Eq. (5.103) reads

$$(G_1,G_3) = \tfrac{1}{2}(\tau - \bar{\tau})G_3(x^i,a,p_i). \tag{5.112}$$

Thus $(G_1,G_3) = 0$ by virtue of Eq. (5.97) for $\eta = 3$. From Eq. (5.100) we conclude that (G_2,G_3) also vanishes. Thus we are able to conclude that the system (5.97) has a solution and hence that there exists a null tetrad such that Eqs. (5.1), (5.2), and (5.3) hold. We note that Eq. (5.4) is also satisfied by virtue of Eq. (5.84). Furthermore, Eqs. (5.5) and (5.6) hold on account of DMT Eqs. (7.26), (7.27), and (8.24) except for the equation

$$\mathscr{D}\epsilon = 0. \tag{5.113}$$

It remains to be shown that a tetrad transformation can be made which preserves all the conditions previously imposed and such that Eq. (5.113) is satisfied. We note that conditions (5.3), (5.86), and (5.92) (without the primes) are preserved by transformations (5.81) satisfying Eqs. (5.91) and the equation

$$e^{-a}(\epsilon + \tfrac{1}{2}\Delta a) - e^{a}(\epsilon + \tfrac{1}{2}Da) = 0. \tag{5.114}$$

The transformation law for the quantity $\mathscr{D}\epsilon$ under transformations satisfying Eqs. (5.91) and (5.114) is

$$\mathscr{D}'\epsilon' = \mathscr{D}\epsilon + \tfrac{1}{2}(D\Delta a + \Delta Da) - \epsilon(D - \Delta)a, \tag{5.115}$$

which, on account of the commutator,

$$[\Delta,D] = 2\epsilon\mathscr{D} - (\tau + \bar{\tau})\mathscr{L} \tag{5.116}$$

may be rewritten as

$$\mathscr{D}'\epsilon' = \mathscr{D}\epsilon + D\Delta a + 2\epsilon\Delta a. \tag{5.117}$$

Thus in order for $\mathscr{D}'\epsilon' = 0$ to hold the function a must satisfy the second order P. D. E.

$$D\Delta a + 2\epsilon\Delta a = -\mathscr{D}\epsilon \tag{5.118}$$

in addition to Eqs. (5.91) and (5.114). These equations are compatible and hence admit a common solution since the application of the operator δ (or $\bar{\delta}$) to both sides of Eq. (5.118) yields no new equation to be satisfied by a. In fact, we have

$$\delta D\Delta a + 2\delta\epsilon\Delta a + 2\epsilon\delta\Delta a = -\delta\mathscr{D}\epsilon, \tag{5.119}$$

which may be written as

$$(\tau - \bar{\tau})(D\Delta a + 2\epsilon\Delta a) = -(\tau - \bar{\tau})\mathscr{D}\epsilon, \tag{5.120}$$

on account of the commutators (5.101) and (5.102) and Eq. (5.109). We see that Eq. (5.120) is *identically satisfied* on account of Eq. (5.118). The proof for this case is now complete since the above argument shows that there exists a null tetrad for which Eqs. (5.1)-(5.6) are satisfied.

Case III: $\rho = \mu = \tau = 0$. We first note that as a consequence of Eq. (4.20)

$$\pi = 0. \qquad (5.121)$$

These conditions are preserved by the complete group of transformations (4.2). We begin the construction of the tetrad of Theorem 3 by performing a tetrad transformation such that

$$\alpha' = -\bar{\beta}', \qquad (5.122)$$

$$\epsilon' = \bar{\epsilon}', \qquad (5.123)$$

$$\gamma' = \bar{\gamma}'. \qquad (5.124)$$

In view of Eqs. (4.5) and (4.6) the differential equations to be satisfied by a and b are

$$\delta a = -(\bar{\alpha} + \beta), \qquad (5.125)$$

$$Db = i(\epsilon - \bar{\epsilon}), \qquad (5.126)$$

$$\Delta b = i(\gamma - \bar{\gamma}). \qquad (5.127)$$

In order to verify the integrability conditions for the above equations one needs the commutators

$$[\bar{\delta},\delta] = (\alpha - \bar{\beta})\delta + (\beta - \bar{\alpha})\bar{\delta}, \qquad (5.128)$$

$$[\Delta, D] = (\gamma + \bar{\gamma})D + (\epsilon + \bar{\epsilon})\Delta. \qquad (5.129)$$

It follows from Eqs. (4.22g), and (4.22j) that the commutation relations for Eq. (5.125) are satisfied while the same conclusion holds for Eqs. (5.126) and (5.127) on account of Eqs. (4.22e), and (4.22g). We now suppress the primes on the transformed quantities and note that conditions (5.122), (5.123), and (5.124) are *preserved* by transformations (4.2) satisfying

$$\delta a = 0, \qquad (5.130)$$

$$Db = \Delta b = 0. \qquad (5.131)$$

We next use such transformations to set

$$\gamma' = \epsilon', \qquad (5.132)$$

$$\beta' = \alpha'. \qquad (5.133)$$

The additional equations to be satisfied by a and b are

$$e^{-a}(\gamma + \tfrac{1}{2}\Delta a) = e^{a}(\epsilon + \tfrac{1}{2}Da), \qquad (5.134)$$

$$e^{ib}(\bar{\alpha} - \tfrac{1}{2}i\delta b) = -e^{-ib}(\alpha + \tfrac{1}{2}i\bar{\delta}b), \qquad (5.135)$$

which are identical to Eqs. (5.93) and (5.46) (with $e = -1$) of Case IIb and IIa. The verification that the systems of equations (5.130) and (5.134), and (5.131) and (5.135) each admit a solution is almost the same as that given for the identical systems in the two cases just mentioned. In fact, it is easier in the present case since Eqs. (4.22c), (4.22d), (4.22m), and (4.22p) imply that

$$D\alpha = \Delta\alpha = 0, \qquad (5.136)$$

$$\delta\epsilon = 0, \qquad (5.137)$$

$$\delta\gamma = 0. \qquad (5.138)$$

Again dropping the primes we note that Eqs. (5.122), (5.123), (5.124), (5.132), and (5.135) are preserved by transformations satisfying Eqs. (5.130), (5.131), and the equations

$$e^{-a}(\epsilon + \tfrac{1}{2}\Delta a) = e^{a}(\epsilon + \tfrac{1}{2}Da), \qquad (5.139)$$

$$e^{ib}(-\alpha - \tfrac{1}{2}i\delta b) = -e^{-ib}(\alpha + \tfrac{1}{2}i\bar{\delta}b). \qquad (5.140)$$

We also remark that Eqs. (5.3) and (5.4) hold in view of Eqs. (5.122), (5.123), (5.124), (5.132), and (5.133). Furthermore, Eqs. (5.5) and (5.6) are satisfied on account of the DMT Eqs. (7.26) and (7.27) except for the equations

$$\mathscr{D}\epsilon = 0, \qquad (5.141)$$

$$\mathscr{L}\alpha = 0. \qquad (5.142)$$

The last step in the proof consists in showing that there exists a transformation (4.2) which preserves all the conditions just imposed and such that Eqs. (5.14) and (5.142) are satisfied. The demonstration that such a transformation does exist will be omitted since it is identical to the proof of the same result in Cases IIa and IIb.

This completes the proof of Theorem 3. However, it is worth pointing out at this juncture where some of the known solutions fit into the classification scheme just given. We first mention that Case I contains the most general solutions admitting only two Killing vectors and hence contains the charged versions of the Kerr solution and the C-NUT solution. The solutions in Case II possess four Killing vectors and include for example the Reissner–Nordstrom solution. The solution in Case III is the Robinson–Bertotti solution, which admits a six-parameter isometry group.

6. PROOF OF THEOREM 1 FOR NULL ORBITS

We now conclude the proof of Theorem 1 by considering the cases excluded in Theorem 3, namely, the cases when $\rho \neq 0$ and $\mu = 0$ or vice versa. These cases represent, in fact, the same situation since we can pass from one to the other by interchanging the tetrad vectors l and n. The method of proof will consist in giving an explicit construction of the metric and Maxwell field of Eq. (2.7) and showing that they have the properties stated in the theorem.

We first remark that *the case*

$$\rho \neq 0, \quad \mu = \tau = 0 \qquad (6.1)$$

is impossible. Indeed Eq. (4.20) implies

$$\pi = 0, \qquad (6.2)$$

so that Eq. (4.22g) becomes

$$\Psi = -2\Lambda = \text{const.} \qquad (6.3)$$

Equation (4.14) thus implies

$$\rho(3\Psi + 2\Lambda) = 0, \qquad (6.4)$$

which is impossible since neither ρ nor $3\Psi + 2\Lambda$ may vanish [the latter inequality arising from (4.16)].

In view of the above result we shall assume in the sequel

$$\rho\tau \neq 0 \qquad (6.5)$$

and

$$\mu = 0. \qquad (6.6)$$

As in Cases I and IIa of Sec. 5 we may, on account of Eq. (4.20), use the transformation (4.2) to set

$$\pi = \tau. \qquad (6.7)$$

This condition is preserved by the transformation

$$l' = e^{a}l, \quad n' = e^{-a}n, \quad m' = m. \qquad (6.8)$$

It follows from Eqs. (4.21d) and (4.22h) that

$$\gamma = \bar{\gamma}, \tag{6.9}$$

and from Eqs. (4.21c) and (4.22b) that

$$\epsilon - \bar{\epsilon} = \tfrac{1}{2}(\rho - \bar{\rho}). \tag{6.10}$$

In addition Eqs. (4.21a) and (4.22f,g,n,o) imply

$$\alpha + \bar{\alpha} = \beta + \bar{\beta}. \tag{6.11}$$

On account of Eq. (6.9) we may use the transformation (6.8) to set (dropping the primes)

$$\gamma = 0. \tag{6.12}$$

This condition is preserved by transformations satisfying

$$\Delta a = 0. \tag{6.13}$$

The next step is to use the transformation (6.8) to set

$$2(\alpha' + \bar{\beta}') = \tau' + \bar{\tau}' \tag{6.14}$$

while preserving Eq. (6.12). To achieve this the function a must satisfy in addition to Eq. (6.13) the equations

$$\delta a = \tfrac{1}{2}(\tau + \bar{\tau}) - \bar{\alpha} - \beta, \quad \bar{\delta} a = \tfrac{1}{2}(\tau + \bar{\tau}) - \alpha - \bar{\beta}. \tag{6.15}$$

It remains to verify that this system of equations admits a solution by showing that the commutation relations

$$[\delta,\Delta]a = (\tau - \bar{\alpha} - \beta)\Delta a, \tag{6.16}$$

$$[\bar{\delta},\delta]a = (\bar{\rho} - \rho)\Delta a + (\beta - \bar{\alpha})\bar{\delta} a + (\alpha - \bar{\beta})\delta a, \tag{6.17}$$

are satisfied. The right-hand side of the first of these relations is zero by virtue of Eq. (6.13) while the left-hand side vanishes on account of Eqs. (4.21d), (4.22m), and (4.22p), and (6.12). It follows from Eqs. (4.22g), (4.22j), (6.13), and (6.15) that the second relation is also satisfied. We now suppress the primes on the transformed quantities and note that Eqs. (6.12) and (6.14) are preserved by transformations (6.8) satisfying Eq. (6.13) and

$$\delta a = 0. \tag{6.18}$$

We also note that Eqs. (6.11) and (6.14) imply

$$\alpha = \beta. \tag{6.19}$$

Thus Eq. (6.14) can be rewritten as

$$\alpha + \bar{\alpha} = \tfrac{1}{2}(\tau + \bar{\tau}). \tag{6.20}$$

Consequently, the commutator (6.16) reads

$$[\delta,\Delta] = \tfrac{1}{2}(\tau - \bar{\tau})\Delta, \tag{6.21}$$

from which it follows that

$$[\mathscr{L},\Delta] = 0, \tag{6.22}$$

where we recall

$$\mathscr{L} = \delta + \bar{\delta}. \tag{6.23}$$

Further consequences of Eqs. (4.21) and (4.22) are

$$\Delta\rho = \mathscr{L}\rho = 0, \tag{6.24}$$

$$\Delta\tau = \mathscr{L}\tau = 0, \tag{6.25}$$

$$\Delta\alpha = 0, \tag{6.26}$$

from which it follows by Eqs. (6.10) and (6.20) that one also has

$$\mathscr{L}(\alpha + \bar{\alpha}) = 0, \tag{6.27}$$

$$\Delta(\epsilon - \bar{\epsilon}) = \mathscr{L}(\epsilon - \bar{\epsilon}) = 0. \tag{6.28}$$

Moreover, by applying the operator $\mathscr{L}$ to both sides of Eq. (4.22n) one obtains, on account of the commutator (6.17) and Eqs. (6.25),

$$\mathscr{L}(\alpha - \bar{\alpha}) = 0. \tag{6.29}$$

When this equation is combined with Eq. (6.27) one has

$$\mathscr{L}\alpha = 0. \tag{6.30}$$

Similarly, by applying in turn the operators Δ and $\mathscr{L}$ to both sides of Eq. (4.22a) one obtains with the help of Eq. (6.24) and the commutators

$$[\Delta,D] = (\epsilon + \bar{\epsilon})\Delta - (\tau + \bar{\tau})\mathscr{L}, \tag{6.31}$$

$$[\mathscr{L},D] = -\tfrac{1}{2}(\rho + \bar{\rho})\mathscr{L} \tag{6.32}$$

the equations

$$\Delta(\epsilon + \bar{\epsilon}) = \mathscr{L}(\epsilon + \bar{\epsilon}) = 0. \tag{6.33}$$

It follows from these equations and Eq. (6.28) that

$$\Delta\epsilon = \mathscr{L}\epsilon = 0. \tag{6.34}$$

The commutation relation (6.22) implies that there exist coordinates such that

$$\Delta = \partial/\partial\tilde{u}, \quad \mathscr{L} = \partial/\partial\tilde{v}. \tag{6.35}$$

It then follows from Eqs. (6.24), (6.25), (6.26), (6.30), and (6.34) that the spin coefficients of the problem are *independent* of the variables $\tilde{u}$ and $\tilde{v}$. The rest of the proof consists in showing the stronger result, namely, that there exists a system of coordinated in which the *components of the metric* are independent of two coordinates and such that the other conditions in Theorem 1 are satisfied. The proof is similar to that given in Sec. 9 of DMT in the case of a space-time admitting a general RMIS. We start by writing Cartan's structure equations for the present case, which are

$$d\theta^1 = \tfrac{1}{2}(\bar{\tau} - \tau)(\theta^3 - \theta^4)\wedge\theta^1, \tag{6.36a}$$

$$d\theta^2 = \tfrac{1}{2}(\tau - \bar{\tau})(\theta^4 - \theta^3)\wedge\theta^2 + (\bar{\rho} - \rho)\theta^3\wedge\theta^4 + (\epsilon + \bar{\epsilon})\theta^1\wedge\theta^2, \tag{6.36b}$$

$$d\theta^3 = -\tfrac{1}{2}(\rho + \bar{\rho})\theta^1\wedge\theta^3 - (\tau + \bar{\tau})\theta^1\wedge\theta^2 + (\alpha - \bar{\alpha})\theta^3\wedge\theta^4, \tag{6.36c}$$

$$d\theta^4 = -\tfrac{1}{2}(\rho + \bar{\rho})\theta^1\wedge\theta^4 - (\tau + \bar{\tau})\theta^1\wedge\theta^2 + (\alpha - \bar{\alpha})\theta^3\wedge\theta^4. \tag{6.36d}$$

It follows that

$$d\theta^1\wedge\theta^1 = 0, \tag{6.37a}$$

$$d(\theta^3 - \theta^4)\wedge(\theta^3 - \theta^4) = 0, \tag{6.37b}$$

$$d\theta^2\wedge\theta^2\wedge(\theta^3 + \theta^4) = 0, \tag{6.37c}$$

$$d(\theta^3 + \theta^4)\wedge\theta^2\wedge(\theta^3 + \theta^4) = 0, \tag{6.37d}$$

which on account of Frobenius's theorem implies the existence of a local coordinate system (u,v,w,x) such that

$$\theta^1 = R\,dw, \tag{6.38a}$$

$$\theta^2 = L du + M dv, \tag{6.38b}$$

$$\theta^3 + \theta^4 = \sqrt{2}(N du + P dv), \tag{6.38c}$$

$$\theta^3 - \theta^4 = i\sqrt{2}\,T dx, \tag{6.38d}$$

where $L,M,N,P,R,$ and T are real valued functions of the four coordinates. If one replaces the differential forms in Eqs. (6.36) by their values given by Eqs. (6.38) and equates corresponding coefficients of the differentials, the following equations are obtained:

$$R_u = R_v = 0, \tag{6.39a}$$

$$T_u = T_v = 0, \tag{6.39b}$$

$$LM_w - ML_w = 0, \tag{6.39c}$$

$$PN_x - NP_x = 0, \tag{6.39d}$$

$$L_v = M_u, \tag{6.39e}$$

$$N_v = P_u, \tag{6.39f}$$

$$\rho + \bar{\rho} = -2(RT)^{-1}T_w = 2(RZ)^{-1}(MN_w - LP_w), \tag{6.39g}$$

$$\rho - \bar{\rho} = i(TZ)^{-1}(LM_x - ML_x), \tag{6.39h}$$

$$\tau + \bar{\tau} = (\sqrt{2}RZ)^{-1}(NP_w - PN_w), \tag{6.39i}$$

$$\tau - \bar{\tau} = i\sqrt{2}(RT)^{-1}R_x = i\sqrt{2}(TZ)^{-1}(PL_x - NM_x), \tag{6.39j}$$

$$\alpha - \bar{\alpha} = i(\sqrt{2}TZ)^{-1}(MN_x - LP_x), \tag{6.36k}$$

$$\epsilon + \bar{\epsilon} = (RZ)^{-1}(PL_w - NM_w), \tag{6.39l}$$

where

$$Z = LP - MN. \tag{6.39m}$$

On account of Eqs. (6.39a), and (6.39b) the functions R and T are independent of the coordinates u and v. In order to determine the u,v dependence of the remaining functions we need the following expressions for the partial derivatives in terms of the NP operators and vice-versa:

$$f_u = L\Delta f + (N/\sqrt{2})(\delta + \bar{\delta})f, \tag{6.40a}$$

$$f_v = M\Delta f + (P/\sqrt{2})(\delta + \bar{\delta})f, \tag{6.40b}$$

$$f_w = RDF, \tag{6.40c}$$

$$f_x = (i/\sqrt{2})B(\delta - \bar{\delta})f, \tag{6.40d}$$

$$Df = R^{-1}f_w, \tag{6.41a}$$

$$(\delta - \bar{\delta})f = -(i\sqrt{2}/T)f_x, \tag{6.41b}$$

$$\Delta f = Z^{-1}(Pf_u - Nf_v), \tag{6.42a}$$

$$(\delta + \bar{\delta})f = \sqrt{2}Z^{-1}(Lf_v - Mf_u). \tag{6.42b}$$

The partial derivatives L_w, L_x, N_w, and N_x may be obtained by solving Eqs. (6.39). One obtains

$$L_w = R(\epsilon + \bar{\epsilon})L, \tag{6.43a}$$

$$L_x = (i/\sqrt{2})T(\tau - \bar{\tau})L + iT(\bar{\rho} - \rho)N, \tag{6.43b}$$

$$N_w = -\sqrt{2}R(\tau + \bar{\tau})L - \tfrac{1}{2}R(\rho + \bar{\rho})N, \tag{6.43c}$$

$$N_x = i\sqrt{2}T(\alpha - \bar{\alpha})N. \tag{6.43d}$$

The equations for the corresponding derivatives of M and P may be obtained by replacing L by M and N by P in the above equations. We observe that Eqs. (6.43) are a *linear system of A. Mayer in the unknowns L and N whose coefficients functions are independent of the coordinates u and v*. [The last remark follows from Eqs. (6.24), (6.25), (6.30), (6.34), (6.39a), (6.39b), (6.42c), and (6.42d).] In order to verify that the system is *completely integrable* we rewrite it in terms of the differential operators D and

$$\mathscr{L}^+ = \delta - \bar{\delta}, \tag{6.44}$$

obtaining by means of Eqs. (6.42) the equivalent system

$$DL = (\epsilon + \bar{\epsilon}), \tag{6.45a}$$

$$DN = -\sqrt{2}(\tau + \bar{\tau}) - \tfrac{1}{2}(\rho + \bar{\rho})N, \tag{6.45b}$$

$$\mathscr{L}^+ L = (\bar{\tau} - \tau)L + \sqrt{2}(\bar{\rho} - \rho)N, \tag{6.45c}$$

$$\mathscr{L}^+ N = 2(\alpha - \bar{\alpha})N. \tag{6.45d}$$

The required commutator is

$$[\mathscr{L}^+, D] = (\tau - \bar{\tau})D - \tfrac{1}{2}(\rho + \bar{\rho})\mathscr{L}^+, \tag{6.46}$$

which we first apply to the function L. In view of Eqs. (6.45) the left-hand side is given by

$$(\tau - \bar{\tau})DL - \tfrac{1}{2}(\rho + \bar{\rho})\mathscr{L}^+ L = (\tau - \bar{\tau})[\epsilon + \bar{\epsilon} + \tfrac{1}{2}(\rho + \bar{\rho})]L + (1/\sqrt{2})(\rho + \bar{\rho})(\rho - \bar{\rho})N. \tag{6.47}$$

On the other hand, the right-hand side can be written as

$$[\mathscr{L}^+, D]L = [\mathscr{L}^+(\epsilon + \bar{\epsilon}) + D(\tau - \bar{\tau}) + 2(\bar{\rho} - \rho)(\tau + \bar{\tau})]L + \sqrt{2}[D(\rho - \bar{\rho}) + (\bar{\rho} - \rho)(\epsilon + \bar{\epsilon} + \tfrac{1}{2}(\rho + \bar{\rho}))]N, \tag{6.48}$$

which by virtue of Eqs. (4.22a)-(4.22d) and (6.34) is identical to the right-hand side of Eq. (6.46). In a similar fashion by means of Eqs. (4.21b), (4.22c), (4.22d), (4.22i), (4.22n), (6.24), and (6.25) one may verify that the commutator (6.46) is satisfied when applied to the function N. It follows from this discussion that the system (6.43) is completely integrable and hence possesses a unique solution that assumes a given value at a given point. Since the equations are linear there exist two linearly independent solutions

$$X_1 = \begin{pmatrix} L_1 \\ N_1 \end{pmatrix}, \quad X_2 = \begin{pmatrix} L_2 \\ N_2 \end{pmatrix}, \tag{6.49}$$

where the functions L_1, L_2, N_1, and N_2 are independent of the coordinates u and v, such that every solution of the system (6.43) can be expressed as a linear combination of the vectors X_1 and X_2 with coefficients independent of the coordinates w and x. Thus we have

$$\mathscr{N} = E(u,v)X_1 + F(u,v)X_2, \tag{6.50a}$$

$$\mathscr{P} = G(u,v)X_1 + H(u,v)X_2, \tag{6.50b}$$

where

$$\mathscr{N} = \begin{pmatrix} L \\ N \end{pmatrix}, \tag{6.50c}$$

$$\mathscr{P} = \begin{pmatrix} M \\ P \end{pmatrix}. \tag{6.50d}$$

In terms of the above notation Eqs. (6.39e), and (6.39f) have the form

$$\mathscr{N}_v = \mathscr{P}_u, \tag{6.51}$$

which on account of Eqs. (6.50) implies

$$E_v = G_u, \tag{6.52a}$$

$$F_v = H_u. \tag{6.52b}$$

Thus there exist potential functions I and J such that

$$E = I_u, \tag{6.53a}$$

$$G = I_v, \tag{6.53b}$$

$$F = J_u, \tag{6.53c}$$

$$H = J_v. \tag{6.53d}$$

It follows from Eqs. (6.49), (6.50), and (6.53) that

$$L = I_u L_1 + J_u L_2, \tag{6.54a}$$

$$N = I_u N_1 + J_u N_2, \tag{6.54b}$$

$$M = I_v L_1 + J_v L_2, \tag{6.54c}$$

$$P = I_v N_1 + J_v N_2, \tag{6.54d}$$

and hence by Eqs. (6.38b) and (6.38c) that

$$\theta^2 = L_1 du' + L_2 dv', \tag{6.55a}$$

$$\theta^3 + \theta^4 = \sqrt{2}(N_1 du' + N_2 dv'), \tag{6.55b}$$

where

$$du' = I_u du + I_v dv, \tag{6.56a}$$

$$dv' = J_u du + J_v dv. \tag{6.56b}$$

By redefining $L_1 = L$, $L_2 = M$, $N_1 = N$, and $N_2 = P$ and dropping the primes on u and v we find that the null tetrad has the form (6.38), where now the functions $L,M,N,P,R,$ and T are independent of the coordinates u and v. Thus by Eq. (3.1) the metric in the case of null orbits has the form (2.7a), which, as already indicated in Sec. 2, admits two commuting Killing vectors,

$$k_1 \equiv k^i_1 \frac{\partial}{\partial x^i} = \frac{\partial}{\partial u}, \tag{6.57a}$$

$$k_2 \equiv k^i_2 \frac{\partial}{\partial x^i} = \frac{\partial}{\partial v}. \tag{6.57b}$$

Since

$$(k_1,k_2)^2 - (k_1,k_1)(k_2,k_2) = g^2_{uv} - g_{uu} g_{vv} = 0, \tag{6.58}$$

the orbits of the abelian isometry group defined by Eq. (6.57) are *null* which implies that the group is *not* invertible. In order to show that the group is orthogonally transitive we introduce the Killing forms ω^1 and ω^2 defined by

$$\omega^1 = k^i_1 g_{ij} dx^j, \quad \omega^2 = k^i_2 g_{ij} dx^j. \tag{6.59}$$

On account of Eqs. (2.7a) and (6.57) we have explicitly

$$\omega^1 = RL dw - N^2 du - NP dv, \tag{6.60a}$$

$$\omega^2 = RM dw - NP du - P^2 dv. \tag{6.60b}$$

If one defines

$$\Omega = \omega^1 \wedge \omega^2 = RZ(N du + P dv) \wedge dw, \tag{6.61}$$

it can be shown that

$$d\omega^1 \wedge \Omega = RNZ(PN_x - NP_x) du \wedge dv \wedge dw \wedge dx = 0 \tag{6.62}$$

and

$$d\omega^2 \wedge \Omega = RPZ(PN_x - NP_x) du \wedge dv \wedge dw \wedge du = 0, \tag{6.63}$$

on account of Eq. (6.39d). It thus follows from a lemma of Kundt and Trümper[47] that the group orbits admit orthogonal two-surfaces.

It remains to show that the self-dual Maxwell field has the form (2.7b). This follows from Eqs. (3.4), (4.1), and (6.38) upon noting that the function B satisfies Eqs. (4.9), which in this case reduce to

$$DB = 2\rho B, \tag{6.64a}$$

$$\Delta B = 0, \tag{6.64b}$$

$$\delta B = 2\tau B, \tag{6.64c}$$

$$\bar{\delta} B = -2\tau B. \tag{6.64d}$$

These equations imply that

$$\Delta B = \mathscr{L} B = 0, \tag{6.65}$$

and hence by Eqs. (6.42c), and (6.42d) that B is independent of the coordinates u and v. This completes the proof of Theorem 1.

7. PROOF OF THEOREM 2

In order to prove this theorem it is convenient to make a conformal transformation of the metric

$$\tilde{g}_{ab} = e^{2\psi} g_{ab}, \tag{7.1}$$

where ψ is a real function. This follows from the fact that the Hamilton–Jacobi equation (2.10) is invariant under such transformations, a property which follows from the transformation law for the contravariant metric

$$\tilde{g}^{ab} = e^{-2\psi} g^{ab}. \tag{7.2}$$

The transformation (7.1) is induced by the following transformation law for the basis one-forms:

$$\tilde{\theta}^a = e^{\psi} \theta^a, \quad a = 1,2,3,4, \tag{7.3}$$

which preserves the symmetry between the null vectors l and n. It follows that the transformation laws for the NP operators and spin coefficients are

$$\tilde{D} = e^{-\psi} D, \tag{7.4a}$$

$$\tilde{\Delta} = e^{-\psi} \Delta, \tag{7.4b}$$

$$\tilde{\delta} = e^{-\psi} \delta, \tag{7.4c}$$

$$\tilde{\kappa} = e^{-\psi} \kappa, \tag{7.5a}$$

$$\tilde{\sigma} = e^{-\psi} \sigma, \tag{7.5b}$$

$$\tilde{\rho} = e^{-\psi}(\rho - D\psi), \tag{7.5c}$$

$$\tilde{\tau} = e^{-\psi}(\tau - \delta\psi), \tag{7.5d}$$

$$\tilde{\nu} = e^{-\psi} \nu, \tag{7.5e}$$

$$\tilde{\lambda} = e^{-\psi} \lambda, \tag{7.5f}$$

$$\tilde{\mu} = e^{-\psi}(\mu + \Delta\psi), \tag{7.5g}$$

$$\tilde{\pi} = e^{-\psi}(\pi + \bar{\delta}\psi), \tag{7.5h}$$

$$\tilde{\epsilon} = e^{-\psi}(\epsilon + \tfrac{1}{2} D\psi), \tag{7.5i}$$

$$\tilde{\gamma} = e^{-\psi}(\gamma - \tfrac{1}{2}\Delta\psi), \tag{7.5j}$$

$$\tilde{\alpha} = e^{-\psi}(\alpha - \tfrac{1}{2}\bar{\delta}\psi), \tag{7.5k}$$

$$\tilde{\beta} = e^{-\psi}(\beta + \tfrac{1}{2}\delta\psi). \tag{7.5l}$$

These equations imply that the complex one-form θ defined by Eq. (4.8) transforms as

$$\tilde{\theta} = \theta + 2d\psi. \tag{7.6}$$

Thus we can choose the function ψ such that

$$\tilde{\theta} + \bar{\tilde{\theta}} = 0, \tag{7.7}$$

since the integrability condition for such a choice,

$$d(\theta + \bar{\theta}) = 0, \tag{7.8}$$

is satisfied on account of the integrability condition (4.10) for Maxwell's equations. The resulting differential equations to be satisfied by ψ are

$$D\psi = \tfrac{1}{2}(\rho + \bar{\rho}), \tag{7.9a}$$

$$\Delta\psi = -\tfrac{1}{2}(\mu + \bar{\mu}), \tag{7.9b}$$

$$\delta\psi = \tfrac{1}{2}(\tau - \bar{\pi}). \tag{7.9c}$$

We now have to consider two separate cases according to whether the group orbits are non-null or null.

Case A: *non-null orbits.* In this case we have Theorem 3 holding, which implies by Eq. (5.2) that Eqs. (7.9) have the form

$$D\psi = \tfrac{1}{2}(\rho + \bar{\rho}), \tag{7.10a}$$

$$\Delta\psi = \tfrac{1}{2}e(\rho + \bar{\rho}), \tag{7.10b}$$

$$\delta\psi = \tfrac{1}{2}(\tau + e\bar{\tau}). \tag{7.10c}$$

It thus follows that

$$\mathscr{D}\psi = \mathscr{L}\psi = 0. \tag{7.11}$$

We shall now show that *all the conditions of the DMT theorem are invariant under a conformal transformation satisfying Eq.* (7.11). This is immediate for conditions on the spin coefficients of the type $\mu = -e\rho$, etc., given in DMT Eq. (5.31), in view of the transformation laws (7.5). We also see that $\tilde{A} = \tilde{E} = 0$ if and only if $A = E = 0$ [DMT Equation (6.27)] on account of the transformation laws

$$\tilde{A} = e^{-\psi}A, \tag{7.12a}$$

$$\tilde{E} = e^{-\psi}E, \tag{7.12b}$$

which follow from Eqs. (7.5) and DMT Eqs. (6.22) and (6.23). The differential conditions DMT Eqs. (7.26) and (7.27) [our Eqs. (5.5) and (5.6)] are also preserved. For example,

$$\begin{aligned}\tilde{\mathscr{D}}\tilde{\rho} &= e^{-\psi}\mathscr{D}(e^{-\psi}(\rho - D\psi)) \\ &= e^{-2\psi}(-\mathscr{D}\psi(\rho - D\psi) + \mathscr{D}\rho - \mathscr{D}D\psi) = 0\end{aligned} \tag{7.13}$$

on account of Eqs. (7.4a), (7.4b), (7.5c), and the commutator

$$[\mathscr{D}, D] = (\epsilon + \bar{\epsilon})\mathscr{D} + (e\bar{\tau} - \tau)\mathscr{L}. \tag{7.14}$$

It remains to verify the invariance of the symmetry conditions on the curvature components given in DMT Eqs. (7.28) and (7.29), which are

$$\Phi_{22} = \Phi_{00}, \tag{7.15a}$$

$$\Phi_{12} = \Phi_{10}, \tag{7.15b}$$

$$\Phi_{20} = \Phi_{02}, \tag{7.15c}$$

$$\Psi_4 = \Psi_0, \tag{7.16a}$$

$$\Psi_3 = \Psi_1. \tag{7.16b}$$

As an example we establish the invariance of Eq. (7.15a). The required transformation laws are

$$\begin{aligned}\tilde{\Phi}_{00} = e^{-2\psi}[&\Phi_{00} + (D\psi)^2 - D^2\psi \\ &+ (\epsilon + \bar{\epsilon})D\psi - \kappa\bar{\delta}\psi - \bar{\kappa}\delta\psi],\end{aligned} \tag{7.17}$$

$$\begin{aligned}\tilde{\Phi}_{22} = e^{-2\psi}[&\Phi_{22} + (\Delta\psi)^2 - \Delta^2\psi \\ &- (\gamma + \bar{\gamma})\Delta\psi + \nu\delta\psi + \bar{\nu}\bar{\delta}\psi].\end{aligned} \tag{7.18}$$

In view of the fact that $\nu = -e\kappa$, $\gamma = -e\epsilon$, $\Delta\psi = eD\psi$, $\bar{\delta}\psi = e\delta\psi$, and the commutator

$$[\Delta, D] = -e(\epsilon + \bar{\epsilon})\mathscr{D} - (\bar{\tau} - e\tau)\mathscr{L},$$

it follows that

$$\tilde{\Phi}_{22} = \tilde{\Phi}_{00}. \tag{7.19}$$

The verification of Eqs. (7.15b), and (7.15c) is similar. Finally, the invariance of Eqs. (7.16) is a consequence of the transformation laws

$$\tilde{\Psi}_a = e^{-2\psi}\Psi_a, \quad a = 1,2,3,4, \tag{7.20}$$

for the Weyl tensor components.

In view of this general result and Eqs. (7.5) all the conditions of Theorem 3 are invariant under the conformal transformation defined by Eqs. (7.10) except that Φ_{11} is no longer the only nonzero tetrad component of the trace-free Ricci tensor. However, the relations (7.15) will hold between the nonzero components of $\tilde{\Phi}_{ab}$. It thus follows that the conditions of the DMT theorem still hold for the conformally related space. This implies that there exists a coordinate system (u,v,w,x) such that the metric of this space given by

$$d\tilde{s}^2 = 2\tilde{\theta}^1\tilde{\theta}^2 - 2\tilde{\theta}^3\tilde{\theta}^4 \tag{7.21}$$

has the form of Eq. (2.6). However, the metric functions must satisfy some additional restrictions to those required in Theorem 1. These arise from condition (7.7), which in terms of the spin coefficients reads

$$\tilde{\rho} + \bar{\tilde{\rho}} = 0, \tag{7.22a}$$

$$\tilde{\tau} + e\bar{\tilde{\rho}} = 0, \tag{7.22b}$$

and Eqs. (5.1), which are still valid in the conformally related space. On account of DMT Eqs. (9.27)–(9.30) these conditions imply that the metric functions R and T must satisfy Eqs. (2.15) while the functions $L,M,N,$ and P must satisfy the equations

$$ML_w - LM_w = PL_x - NM_x = 0 \tag{7.23a}$$

$$NP_x - PN_x = LP_w - MN_w = 0 \tag{7.23b}$$

which, when Eqs. (2.15) hold, are equivalent to the statement that the principal null congruences defined by l and n are geodesic and shearfree.[48] It is not difficult to show that the Eqs. (7.23) are equivalent to Eqs. (2.16). In order to establish Eqs. (2.12) and (2.14) we note that the inverse conformal transformation, which transforms the conformally related metric (7.21) back to the original solution in $\mathfrak{D}$, is given (for the covariant metric) by $\exp(-2\psi)$, where ψ is a solution of Eqs. (7.10). Furthermore, by Eqs. (7.11) we have

$$\tilde{\mathscr{D}}\psi = e^{-\psi}\mathscr{D}\psi = 0, \tag{7.24a}$$

$$\tilde{\mathscr{L}}\psi = e^{-\psi}\mathscr{L}\psi = 0, \tag{7.24b}$$

which by DMT Eqs. (9.33) and (9.34) imply that ψ satisfies

$$\psi_u = \psi_v = 0 \tag{7.25}$$

where u and v are coordinates in the metric (2.12). On making the substitution $\psi \rightarrow -\psi$ we find the conformal factor of Theorem 2 which satisfies Eq. (2.14). The last step in the proof is to note that the required complete integral should have the form of Eq. (2.11) in view of the fact that u and v are ignorable coordinates in the metric (2.12).

Case B: *null orbits.* Equations (6.6) and (6.7) imply in this case that Eqs. (7.9) become

$$D\psi = \tfrac{1}{2}(\rho + \bar{\rho}), \tag{7.26a}$$

$$\Delta\psi = 0, \tag{7.26b}$$

$$\delta\psi = \tfrac{1}{2}(\tau - \bar{\tau}). \tag{7.26c}$$

From Eq. (7.26c) and its complex conjugate we deduce that

$$\mathscr{L}\psi = 0. \tag{7.27}$$

By means of the transformation laws (7.5) it can be shown that all the conditions on the spin coefficients used to derive the metric form (2.7a) of Theorem 1 in Sec. 6 are invariant under a conformal transformation satisfying Eqs. (7.26b) and (7.27). The demonstration is similar to that used to prove the invariance of the DMT theorem in Case A. Thus there exists a coordinate system (u,v,w,x) such that the conformally related metric

$$d\tilde{s}^2 = 2\tilde{\theta}^1\tilde{\theta}^2 - 2\tilde{\theta}^3\tilde{\theta}^4 \tag{7.28}$$

has the form (2.7a). However, as in the previous case the metric functions must satisfy some additional restrictions which arise from Eqs. (4.17) (preserved under the conformal transformation) and Eq. (7.7), which has the form [of Eq. (7.22) with $e = -1$]

$$\tilde{\rho} + \bar{\tilde{\rho}} = 0, \tag{7.29a}$$

$$\tilde{\tau} - \bar{\tilde{\tau}} = 0. \tag{7.29b}$$

It follows from these equations and from Eqs. (6.39c), (6.39d), (6.39g), and (6.39j) that the metric functions satisfy Eqs. (2.15) and (2.16). We also note that by an argument identical to that in Case A it may be shown that the inverse conformal transformation required to transform the metric (7.29) back to the starting metric (a solution in $\mathfrak{D}$) is given by $\exp(2\phi)$, where $\phi = -\psi$ and the function ψ satisfies the equations

$$\psi_u = \psi_v = 0, \tag{7.30}$$

where u and v are coordinates in the metric (2.13). We get the final form (2.13) of the metric by the substitution $\phi \rightarrow \psi$. For the same reasons given in Case A the complete integral of Eq. (2.10) should have the form (2.11). This completes the proof of Case B.

A remark applicable to both cases is that the key to obtaining the canonical separable coordinate system for Eq. (2.10) is the possibility of performing a conformal transformation such that Eq. (7.9) or equivalently Eq. (7.23) holds. Finally, we note that the Hamilton–Jacobi equation for *the non-null geodesics*,

$$g^{ij}\frac{\partial S}{\partial x^i}\frac{\partial S}{\partial x^j} = -m^2,$$

where m is constant, is solvable by separation of variables in the manner of Theorem 2 provided that the conformal factor $\exp(2\psi)$ in the metrics (2.12) and (2.13) is expressable as the *sum* of a function of w and a function of x.

ACKNOWLEDGMENTS

We would like to express our appreciation to Brandon Carter and Jules Leroy for their continued interest in this work and for many enlightening discussions. In particular we would like to thank the former for his helpful comments on an early version of the manuscript. We would also like to thank Stephen Czapor for checking some of the calculations and for his careful proofreading of the manuscript. One of the authors (R.D.) would like to express his thanks to the members of the Department of Applied Mathematics of the University of Waterloo for their hospitality during a recent visit to Waterloo.

[1] W. Kinnersley, J. Math. Phys. **10**, 1195 (1969).
[2] L. P. Hughston and P. Sommers, Commun. Math. Phys. **33**, 129 (1973).
[3] R. Debever, Bull. Soc. Math. Belg. **XXIII**, 360 (1971).
[4] J. F. Plebański, Ann. Phys. (N.Y.) **90**, 196 (1975).
[5] J. F. Plebański and M. Demiański, Ann. Phys. (N.Y.) **98**, 9 (1976).
[6] G. J. Weir and R. P. Kerr, Proc. R. Soc. London Ser. A **355**, 31 (1977); these authors study the expanding type D vacuum solutions in the context of algebraically special solutions admitting two commuting Killing vectors.
[7] An isometry group is said to be *invertible* if the isotropy subgroup contains an element which inverts the sense of p independent directions in the surface of transitivity at each point P but leaves unaltered the sense of the directions orthogonal to the surface of transitivity at P. This concept was introduced by B. Carter, J. Math. Phys. **10**, 70 (1969).
[8] B. Carter, Commun. Math. Phys. **10**, 280 (1968).
[9] R. Debever, Bull. Cl. Sci. Acad. R. Belg. **LV**, 8 (1969).
[10] D. Matravers, Util. Math. **2**, 55 (1972).
[11] B. Carter, Phys. Rev. **174**, 1559 (1968), and in *Black Holes*, edited by B. DeWitt and C. DeWitt (Gordon and Breach, New York, 1973), p. 118.
[12] M. Walker and R. Penrose, Commun. Math. Phys. **18**, 265 (1970).
[13] R. P. Kerr, Phys. Rev. Lett. **11**, 237 (1963).
[14] L. P. Hughston, R. Penrose, P. Sommers, and M. Walker, Commun. Math. Phys. **27**, 303 (1972).
[15] L. P. Hughston and P. Sommers, Commun. Math. Phys. **32**, 147 (1973).
[16] N. M. J. Woodhouse, Commun. Math. Phys. **44**, 9 (1975).
[17] A preliminary version of which appears in Bull. Cl. Sci. Acad. R. Belg. **LXVI**, 204 (1980).
[18] I. Robinson and A. Schild, J. Math. Phys. **4**, 484 (1962); W. Kundt and A. H. Thompson, C. R. Acad. Sci. (Paris) **254**, 4257 (1962).
[19] An isometry group is said to be *orthogonally transitive* if the p-dimensional orbits of the group are orthogonal to a family of $(n$-$p)$-dimensional surfaces. See Ref. 7.
[20] Reference 15.
[21] See, for example, S. Benenti and M. Francaviglia, Gen. Relativ. Gravit. **10**, 79 (1979).
[22] Reference 8.
[23] Reference 9.
[24] References 4 and 5.
[25] Reference 8.
[26] R. Debever, R. G. McLenaghan, and N. Tariq, Gen. Relativ. Gravit. **10**, 853 (1979).
[27] R. Debever, in *Proceedings of the Conference on Differential Geometrical Methods in Mathematical Physics*, Lecture Notes in Mathematics No. 570 (Springer, Berlin, 1977), p. 518; Bull. Cl. Sci. Acad. R. Belg. **LXII**, 662 (1976).
[28] E. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
[29] R. Debever, Cah. Phys. **168–169**, 303 (1964); M. Cahen, R. Debever, and L. Defrise, J. Math. Mech. **16**, 761 (1967).
[30] Reference 26.
[31] J. F. Plebański and S. Hacyan, J. Math. Phys. **20**, 1004 (1979).
[32] Reference 8.
[33] Reference 15.
[34] Reference 12.
[35] Reference 8.
[36] Reference 9.
[37] Reference 5.
[38] Reference 1.
[39] J. Leroy, Bull. Cl. Sci. Acad. R. Belg. **LXIV**, 130 (1978); see Eq. (3.36) with

$b = 0$; the solution is also given in Ref. 3 by Eq. (2.12), when $b = 0$.

[40]A. Z. Petrov, *Einstein Spaces* (Pergamon, Oxford, 1964), p. 201.

[41]F. Bampi and R. Cianci, Commun. Math. Phys. **70**, 69 (1979).

[42]Reference 29.

[43]Reference 28.

[44]Reference 18.

[45]These equations are also verified in the vacuum case ($\Phi = \Lambda = 0$), a result which follows from the calculations in Ref. 1; see in particular Eqs. (3.25), (3.27), and (3.36).

[46]See, for example, G. F. D. Duff, *Partial Differential Equations* (University of Toronto press, Toronto, 1956), p. 62.

[47]W. Kundt and M. Trümper, Z. Phys. **192**, 419 (1966).

[48]By Theorem 1 of Ref. 9 this is a sufficient condition for the separability of Eq. (2.10) for a metric of the form (2.6a) for which $R_x = T_w = 0$.

On the application of energy conditions and algebraic classification in general relativity [a)]

Michael C. Moody and John R. Ray
Department of Physics and Astronomy, Clemson University, Clemson, South Carolina 29631

(Received 30 September 1980; accepted for publication 11 March 1981)

The application of the ideas of the symmetry classification of space-time, the algebraic classification of the energy-momentum tensor, and the dominant energy conditions for macroscopic matter to the problem of searching for solutions to the field equations of general relativity is discussed. These ideas are then specifically applied to certain space-times of G_4 symmetry to demonstrate the ability of this approach to obtain both new and general solutions. For macroscopic matter, $G_4\mathrm{I}_1$ symmetry is shown to have only one general solution; symmetries $G_4\mathrm{V}$, $G_4\mathrm{VII}_1$, and $G_4\mathrm{VII}_2$ are shown to allow no solutions whatever; and the method is used to discover a new solution for $G_4\mathrm{VIII}_1$ symmetry. A similar solution would be expected for $G_4\mathrm{VIII}_2$ symmetry.

PACS numbers: 04.20.Jb, 04.20.Cv

I. INTRODUCTION

In Einstein's theory of general relativity, the geometry of space-time is assumed related to the matter present in the universe through the field equations (given here with the cosmological term)[1,2]

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi T_{\mu\nu}, \tag{1}$$

where $G_{\mu\nu}$ is the Einstein tensor of space-time, Λ the cosmological constant, $g_{\mu\nu}$ the metric tensor defining the space-time, and $T_{\mu\nu}$ the energy-momentum tensor of matter. [Throughout this work Greek indices assume the values 1,2,3,4, and the summation convention is used. A metric signature of $(+\ +\ +\ -)$ has been chosen.] The Einstein tensor is defined in terms of the metric tensor and its derivatives to second order.

Since their formulation much theoretical work has been involved with finding and studying solutions to the field equations for the gravitational field, determined by $g_{\mu\nu}$, and the matter field, determined by $T_{\mu\nu}$. As the field equations are nonlinear, second order, partial differential equations and the constitution of matter throughout the universe is unknown, this is in general an intractable problem.

Some simplification can be introduced by choosing a specific model for the matter, for instance the perfect fluid model or the scalar field. More general energy-momentum tensors may be obtained by adding (superposing) specific models. The space-time may be sectioned into regions such that in each some specific model is appropriate. These regions can then be joined using proper junction conditions. With such simplification there is still the formidable difficulty of solving nonlinear partial differential equations. This approach is referred to as the T-method by Synge.[3,4] In short, given the energy-momentum tensor solve for the geometry.

Solutions to the field equations may often be found via this method by assuming, in addition to a specific matter field, certain symmetries for the space-time. This simplifies the Einstein tensor and reduces the number of unknowns. A complete classification of space-times (based on the number and types of symmetry transformations and the types of generated subspaces) together with the metrics has been given by Petrov.[5] Thus the search for solutions to the field equations via this approach requires investigating the compatibility of each possible metric with the various specific energy-momentum tensors.

This method of searching for solutions has been applied to space-times admitting a four-parameter group of motions G_4, acting on null three-dimensional hypersurfaces V_3^*, by Lauten and Ray.[6,7] As the Bondi-Pirani-Robinson plane gravitational wave metric[8] is among this class ($G_4\,\mathrm{VI}_2$), solutions corresponding to gravitational radiation were expected as suggested by Pirani.[9] A total of 54 cases were investigated and in only one case was the metric and matter compatible with the field equations. All other cases studied were shown to be incompatible with the field equations; there are no solutions in these cases! In this application a major consequence of the T-method approach is the question of why there are so few solutions.

There is, however, another point of view leading to an alternative approach. Consider the field equations rewritten as

$$T_{\mu\nu} = (1/8\pi)(G_{\mu\nu} + \Lambda g_{\mu\nu}). \tag{2}$$

In this way the field equations can be interpreted as defining a geometrical object $T_{\mu\nu}$ derivable from the metric tensor of space-time. Thus from this viewpoint there is no longer a set of nonlinear, partial differential equations to be solved, but merely a definition. For any ten reasonably well-behaved functions $g_{\mu\nu}$ such that a metric of proper signature can be formed, a space-time can be constructed in which the matter is described by an energy-momentum tensor given by (2). This method (Synge's g-method)[3,4] provides an endless supply of space-times unrestricted by symmetry. Of course, not all of these space-times will be analogous to the universe we live in and there remains the nontrivial task of interpreting the matter associated with the energy-momentum tensor.

Since it is the nature of our own universe we wish to describe, some method is needed to guide the search; to re-

[a)] Based in part on the M. S. Thesis of M. C. Moody.

 0022-2488/81/081727-09$01.00 © 1981 American Institute of Physics

duce the number of possibilities and, hopefully, isolate space-times like our own. To isolate space-times with matter having properties common to our own, mathematical conditions describing these properties are needed. With the addition of these conditions, (2) is no longer merely a definition, but rather part of a model with a limited range of allowable solutions. Some physically reasonable restrictions (at least for macroscopic matter) are the energy conditions[10,11]: (i) the energy density for all observers is non-negative, and (ii) energy cannot be transferred at speeds greater than the speed of light. With these restrictions there is the model

$$\begin{aligned} &T_{\mu\nu} = (1/8\pi)(G_{\mu\nu} + \Lambda g_{\mu\nu}) \\ &\text{(i)}\, T_{\mu\nu} v^\mu v^\nu \geqslant 0 \quad \text{for all } v_\mu v^\mu \leqslant 0. \\ &\text{(ii)}\, T_{\mu\nu} v^\nu T_{\mu\rho} v^\rho \leqslant 0 \end{aligned} \tag{3}$$

The search for allowable solutions to the model (3) can be simplified by an algebraic classification of the energy-momentum tensor[10–18] in terms of its eigenvalues, eigenvectors, and Segrè type. Any symmetric second-rank tensor will have (excluding degeneracies) one of six possible Segrè types. An energy-momentum tensor in general relativity satisfying the above energy conditions must be in only certain subsets of only two of these Segrè types.[10,11,18] By the field equations this is also true for the Einstein tensor.

To inspect a given space-time for agreement with the energy conditions, the energy-momentum tensor is calculated and its Segrè type determined. (It is actually more convenient to do this for the Einstein tensor.) The Segrè type depends on the metric. If the general metric is not of an allowed Segrè type, it may reduce to an allowed type under certain restrictions on the metric. Within the allowed Segrè types, still only certain subsets can satisfy the energy conditions. Inclusion in these subsets depends on conditions on the eigenvalues for the energy-momentum tensor. (And also on the eigenvalues of the Einstein tensor. But it must be remembered that with the inclusion of the cosmological term in the field equations these are no longer proportional.) The eigenvalues also depend on the metric so it is determined that, if at all, the space-time can satisfy the energy conditions only for certain restrictions on the metric. Of course, finally, there remains the problem of interpreting the corresponding energy-momentum tensor.

When the geometry does agree with the energy conditions, the energy-momentum tensor must assume certain allowed Segrè types. The type may depend on restrictions on the metric. If a simple model of matter is to be suitable as a source, its energy-momentum tensor must have the same Segrè type. It is only for those geometries and matter fields with energy-momentum tensors of mutual Segrè type that solutions to the field equations might be obtained. Unfortunately, an algebraic classification has been given only for the simplest models of matter.[19,20] These results may be used in some cases to give an interpretation of the energy-momentum tensor obtained in the g-method approach.

In this work, the g-method approach is applied to the space-times admitting a four-parameter group of motions G_4, acting on null three-dimensional hypersurfaces V_3^*, previously studied by Lauten and Ray. It is found that for $G_4\mathrm{I}_1$ symmetry the one particular solution found previously is the general solution to the field equations. There can be no other solution for geometries having this particular type of symmetry that satisfies the energy conditions! In this solution the matter can be interpreted as a perfect fluid whose four velocity is null (photons for example). For the symmetries $G_4\mathrm{V}$, $G_4\mathrm{VII}_1$, and $G_4\mathrm{VII}_2$ it is shown that there can be no solutions. The symmetries $G_4\mathrm{VIII}_1$ and $G_4\mathrm{VIII}_2$ are found to satisfy the energy conditions, but the energy-momentum tensor obtained is not obviously of the form of some simple specific model of matter. By comparison of Segrè types, it is discovered that the energy-momentum tensor may be interpreted as a superposition of a null perfect fluid and a non-null electromagnetic field. Using this model of matter, a solution is found for $G_4\mathrm{VIII}_1$ symmetry. A similar solution would be expected for $G_4\mathrm{VIII}_2$ symmetry.

II. THEORY

The procedure employed here relies primarily on three major concepts: (i) the classification of the geometry of general relativity according to symmetries, (ii) the algebraic classification of the energy-momentum tensor of the matter in terms of its canonical forms, and (iii) the restrictions imposed on the classifications by the energy conditions. In this section the fundamental ideas of these concepts will be summarized.

A. Symmetry classification

A complete classification of all gravitational fields can be made in terms of the groups of motions they admit. A motion in a space-time is a transformation that maps the space-time into itself while preserving the metric. This is, then, an invariant property of the group. In order that a group be admitted, it is necessary and sufficient that each of r generators of the group satisfies Killing's equations. The solutions are called Killing vectors. The generators are associated with the transformations

$$x'^\mu = f^\mu(x, a^i), \quad i = 1, \ldots, r,$$

where the a^i are r essential parameters labeling the transformations. Each generator is defined by

$$\xi^\mu_i(x) = \partial f^\mu(x, a^i)/\partial a^i, \quad \text{all } a^i = 0.$$

If the transformations form a group, the generators satisfy the fundamental commutators and determine the structure constants of the group. The equations of structure for $r = 2,3$ and a classification in terms of them has been given by Bianchi.[21] This has also been done for $r = 4$ by Lie[22] and Kruchkovich.[23] The classification is summarized by Petrov.[5]

For each structure the Killing vectors (generators) may be determined in a simple system of coordinates. The Killing equations can then be intergrated to determine the metric. Further specification of the gravitational field arises from also considering the types of subspaces the group of transformations generates. Fortunately, this problem has already been solved and is presented in Chap. 5 of Petrov which may be referred to for the results.

B. Algebraic classification

A complete algebraic classification of a general second-order symmetric tensor $S_{\mu\nu}$ can be given in terms of (i) its eigenvalues, (ii) its eigenvectors, and (iii) its Segrè characteristics determined by the elementary divisors. In general relativity this is a linear algebra problem with an underlying four-dimensional space having an indefinite norm. Such a classification has been worked out by Plebañski and others.[10–18] In a general four-dimensional space the tensor is represented by a 4×4 matrix which, when the proper orthonormal tetrad is chosen as basis, can assume (excluding degeneracies) one of six canonical forms. Each form is conveniently described by its corresponding Segrè characteristics. The following types of Segrè characteristics are possible:

(i)[1,1,1,1], (ii) [1,1,2], (iii) [2,2],

(iv) [1,3], (v) [4], (vi) $[1,1,\bar{z},z]$.

Each digit corresponds with an eigenvalue of the matrix and equals the elementary divisor corresponding to that eigenvalue. The convention in which the last eigenvalue corresponds with a timelike or null eigenvector is chosen. The number of digits in each Segrè type gives the number of eigenvectors. The symbol z denotes a complex eigenvalue. The possibility of the eigenvalues being equal exists, and such degeneracy is denoted by enclosing the appropriate characteristics in parenthesis. Taking degeneracies into account yields 18 distinct Segrè types. From the condition of symmetry it follows that Segrè types [4] and [2,2] (including its degeneracy) are empty in general relativity.[18]

For each of the remaining types there exists an orthonormal tetrad in which frame $S_{\mu\nu}$ takes one of the following canonical forms:

$$\text{(i) } [1,1,1,1] \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 \\ 0 & 0 & 0 & -\lambda_0 \end{bmatrix},$$

$$\text{(ii) } [1,1,2] \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_0+\epsilon & \epsilon \\ 0 & 0 & \epsilon & -\lambda_0+\epsilon \end{bmatrix}, \quad \text{with } \epsilon \neq 0,$$

$$\text{(iii) } [1,3] \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_0 & \epsilon & 0 \\ 0 & \epsilon & \lambda_0 & -\epsilon \\ 0 & 0 & -\epsilon & -\lambda_0 \end{bmatrix}, \quad \text{with } \epsilon > 0,$$

$$\text{(iv) } [1,1,\bar{z},z] \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda' & \lambda'' \\ 0 & 0 & \lambda'' & -\lambda' \end{bmatrix},$$

$$\text{with } \lambda_0 = \lambda' + i\lambda'', \lambda_3 = \lambda' - i\lambda'', \tag{4}$$

where λ_i's are the eigenvalues of $S_{\mu\nu}$: i.e., are the roots of the corresponding characteristic equation

$$\det(S_{\mu\nu} - \lambda g_{\mu\nu}) = 0.$$

A vector V^ν satisfying

$$(S_{\mu\nu} - \lambda_i g_{\mu\nu}) V^\nu = 0$$

is an eigenvector corresponding to the eigenvalue λ_i. Due to the arbitrariness that exist in naming the spacelike coordinates, different choices have correspondingly altered canonical forms.

With consideration of all possible degeneracies there exist 15 nonempty algebraic types for a general second-order symmetric tensor in general relativity. These types can be divided into three more general classes according to the number of admitted eigenvectors. These may be conveniently tabulated (originally done by Plabañski)[10] as shown in Table I.

Class I is separated into two subclasses, Z and R. Class I_R is diagonalizable and Class I_Z has members with complex eigenvalues. Members of Classes I_R, II, and III have all real eigenvalues. The graphic symbols in Table I are intended to indicate kinds of eigenvectors admitted by each class. Class I_Z admits two complex eigenvectors. Class I_R admits a timelike eigenvector. The eigenvectors corresponding to the multiple eigenvalues of Classes II and III are null. All other eigenvectors are real and spacelike. In cases of degeneracy, rotations in the subspaces spanned by the degenerate eigenvalues are allowed. For example, in type [1,1,(1,1)], instead of choosing one timelike and one spacelike eigenvector, two null principal directions could be chosen.

Types with the same multiplicity of eigenvalues are along horizontal lines. The diagonal arrows connect types with the same order of minimal polynomial. Note that these two quantities determine all types uniquely except for the two pairs [1,1(1,1)], [1,(1,1),1] and [1,(1,1,1)] [(1,1,1),1]. These may be easily distiguished by diagonalizing the matrix directly.

Thus to determine the Segrè type, the characteristic equation is first solved for the eigenvalues and their multipli-

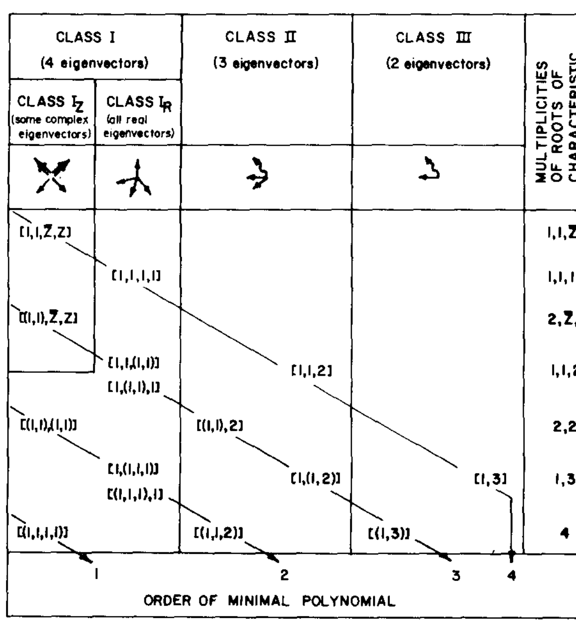


TABLE I. Summary of algebraic classification of the Einstein tensor.

cites determined. Next the order of the minimal polynomial must be determined. The minimal polynomial of a matrix is that polynomial of least degree for which the matrix itself is a root. The Cayley–Hamilton theorem states any matrix is a root of its characteristic polynomial. Also, the characteristic polynomial can be factored in terms of its roots, the eigenvalues. It follows as a theorem in linear algebra that the minimal polynomial is a divisor of the characteristic polynomial and has the same roots. Thus the minimal polynomial is some product of distinct factors of the characteristic polynomial whose factors are perhaps raised to lower powers. Due to the limited number of possiblities in four dimensions it is easy to simply search for the minimal polynomial by direct substitution. This procedure will be illustrated in the next section. This method also yields the restrictions necessary in order that the minimal polynomial be of lower order. Once the multiplicity of the eigenvalues and the order of the minimal polynomial are determined, the Segrè type can be read from Table I.

It should also be noted that a complete algebraic classification of second-order symmetric tensors in general relativity can also be given in terms of their invariant two-space structure. The relationship between this classification and Segrè types has been given by Cormack and Hall.[24]

Of the particular second-order symmetric tensors known as energy-momentum tensors in general relativity, only a few of those corresponding to simple types of matter have been classified.[19,20] The results for four of the simplest models follow:

(i) A dust of noninteracting uncharged particles

$$T_{\mu\nu} = \rho U_\mu U_\nu,$$

where ρ is the rest energy density and U_μ is the four velocity. The Segrè type is $[(1,\overset{0}{1},1)1]$. The raised zero indicates that the corresponding eigenvalues are zero.

(ii) The perfect fluid model

$$T_{\mu\nu} = (\rho + P)U_\mu U_\nu + P g_{\mu\nu},$$

where ρ is again the rest energy density and P is the pressure. The Segrè type is [(1,1,1),1]. Case (i) is the special case of (ii) for which the pressure is zero.

(iii) The null fluid model—

This is the special case of (i) and (ii) for which the four-velocity is null (for example, neutrinos). The Segrè type is [(1,1,2)]. If the pressure vanishes the eigenvalues are zero.

(iv) The electromagnetic field

$$T_{\mu\nu} = g_{\mu\sigma} F_{\nu\rho} F^{\sigma\rho} - \tfrac{1}{4} g_{\mu\nu} F_{\sigma\rho} F^{\sigma\rho},$$

where $F_{\mu\nu}$ is the electromagnetic field tensor. Two types are possible: (a) [(1,1),(1,1)], a non-null field, and (b) $[(1,\overset{0}{1},2)]$, a null field (pure radiation).

C. The energy conditions

In order to focus attention on those space-times defined in (2) having matter with properties similar to that of our own, the addition of mathematical conditions describing some reasonable properties was needed. The conditions chosen were those denoted by Hawking and Ellis[11] as the dominant energy conditions:

For all timelike, and by continuity, all null vectors v^μ,

(i) $T_{\mu\nu} v^\mu v^\nu \geqslant 0$,

(ii) $T_{\mu\nu} v^\mu$ is non-space-like.

The first condition is an assumption that the energy density as measured by any observer is positive definite. The second condition states the assumption that the local energy flux vector is non-space-like, i.e., energy cannot be transferred at speeds greater than the speed of light. These conditions seem physically reasonable for macroscopic matter.

With the assumption of energy conditions the system (3) is obtained. Now the energy-momentum tensor which must satisfy the restricting energy conditions must be of one of the forms (4) obtained in the algebraic classification. Application of the energy conditions to each of these classes shows that not all classes are able to satisfy them. The results can be summarized by the following statements:

(i) Neither energy-momentum tensors of class I_Z nor Class III can satisfy either of the energy conditions.

(ii) Energy-momentum tensors of either Class I_R or Class II satisfy the energy conditions if and only if

$$\lambda_0 \leqslant 0 \text{ and } \lambda_0 \leqslant \lambda_i \leqslant -\lambda_0. \tag{5}$$

For Class II the condition that ϵ be positive must be added. A proof of these statements has been given by Plebański.[10]

The explicit canonical forms of the energy-momentum tensors must therefore correspond to one of the 11 allowed Segrè types given below.

$$\begin{aligned}
&[1,1,1,1],\\
&[1,1,(1,1)],[1,(1,1),1],[(1,1),(1,1)],\\
&[1,(1,1,1)],[(1,1,1),1],\\
&[(1,1,1,1)],\\
&[1,1,2],\\
&[1,(1,2)],[(1,1),2],\\
&[(1,1,2)].
\end{aligned} \tag{6}$$

For a given space-time the energy-momentum tensor must satisfy (2). To be an allowable source for the model (3), this energy-momentum tensor must be found to be of one of the allowed classes given in (6). If this is true, the eigenvalues must then still satisfy (5). Under these conditions, the space-time corresponds to a matter field that satisfies the energy conditions.

The problem of obtaining specific solutions to the field equations is now the problem of interpreting the type of matter described by the energy-momentum tensor. The relationship between the energy-momentum tensor and the corresponding matter field is not necessarily unique.[25,26] For example, the type $[(1,\overset{0}{1},2)]$ may be associated with either a neutrino field (null fluid) or a null electromagnetic field, or a superposition of the two fields. In order that a given energy-momentum tensor have as its source some simple model of matter, they must have an identical Segrè types. Once it is established that the energy conditions may be satisfied, only those models of matter yielding an energy-momentum tensor of the same Segrè type as the space-time need be investigated for solutions.

Concerning the interpretation of the energy-momentum tensor, it has been shown by Williams[27] that any second-rank symmetric tensor can be decomposed into the difference of a perfect fluid energy-momentum tensor and one of the types of the electromagnetic energy-momentum tensor. He gives the explicit forms of the decomposition for each of the possibilities in (6). (They are not all real.) Williams does not, however,consider the field equations of the matter (for example, Maxwell's equations). His work raises the question of the form of a decomposition into other models of matter.

The solution to the problem of interpreting the energy-momentum tensor (after establishing that it satisfies the energy conditions) is thus incomplete and not unique. Still these ideas can be useful in searching for solutions to the field equations as will be demonstrated in the next section.

III. APPLICATION TO G_4 SYMMMETRY

Again, a summary of the procedure. First, using a geometrical approach, the space-time metric can be obtained for a particular symmetry. The Einstein tensor can then be calculated and algebraically classified via investigation of the multiplicity of its eigenvalues and the order of its minimal polynomial. (This will also be the classification of the energy-momentum tensor.) Allowed candidates for solutions are then found by checking the eigenvalues of the energy-momentum tensor (not the Einstein tensor!), using (5), for agreement with the energy conditions. The eigenvalues of $T_{\mu\nu}$ and $G_{\mu\nu}$ are related by

$$\underset{T}{\lambda} = (1/8\pi)\Big(\underset{G}{\lambda} + \Lambda\Big),$$

If then the energy conditions are satisfied, those simple models of matter whose energy-momentum tensor is of the proper Segrè type may be investigated for solutions to the field equations.

In these applications certain space-times admitting a G_4 group of motions acting on null three-dimensional hypersurfaces are considered.

A. G_4I_1 symmetry

Petrov gives the metric as

$$ds^2 = \alpha(x^4)\exp(-2x^3)[2\,dx^1\,dx^4 + (dx^2)^2] + \beta(x^4)(dx^3)^2,$$

where α and β are arbitrary functions. The Einstein tensor in an orthonormal frame is

$$G_{\mu\nu} = \begin{bmatrix} p + 3/B^2 & 0 & -q & -p \\ 0 & 3/B^2 & 0 & 0 \\ -q & 0 & 3/B^2 & q \\ -p & 0 & q & p - 3/B^2 \end{bmatrix},$$

where

$$p = \frac{1}{A^2\exp(-2x^3)}\Big[2\Big(\frac{A'}{A}\Big)^2 - \frac{A''}{A} - \frac{B''}{B} + \frac{2A'B'}{AB}\Big],$$

$$q = \frac{1}{AB\exp(-2x^3)}\Big(2\frac{B'}{B}\Big),\ A = \alpha^{1/2},\ B = \beta^{1/2}.$$

Here the primes refer to differentiation with respect to $(x^1 - x^4)$.

The characteristic deteminant gives the characteristic equation

$$(\lambda - 3/B^2)^4 = 0,$$

which has one root, $\lambda = 3/B^2$, of multiplicity 4. The candidates for the minimal polynomial are

(i) $(\lambda - 3/B^2)$, (ii) $(\lambda - 3/B^2)^2$,
(iii) $(\lambda - 3/B^2)^3$, (iv) $(\lambda - 3/B^2)^4$,

with the corresponding Segrè types from Table I,

(i) [(1,1,1,1)], (ii) [(1,1,2)],
(iii) [(1,3)], (iv) [4].

The last of these cannot occur since the signature of the metric is Minkowskian. Since a matrix must be a root of its minimal polynomial the requirements that must be investigated are

(i) $M = 0$, (ii) $M^2 = 0$, (iii) $M^3 = 0$,

where

$$M^\mu{}_\nu = (G^\mu{}_\nu - 3/B^2\delta^\mu{}_\nu),$$

$$M = \begin{bmatrix} p & 0 & -q & p \\ 0 & 0 & 0 & 0 \\ -q & 0 & 0 & q \\ p & 0 & -q & -p \end{bmatrix}.$$

Then

$$M^2 \begin{bmatrix} q^2 & 0 & 0 & -q^2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ q^2 & 0 & 0 & -q^2 \end{bmatrix},\ M^3 = 0.$$

Thus the minimal polynomial is $(\lambda - 3/B^2)^3$ and the Segrè type is [(1,3)], for which the energy conditions cannot be satisfied. Here this is due to the fact that the $G_{34} = q$ term allows, for some coordinate systems, the enery flux vector to become spacelike for large enough q, and/or the energy density to become negative.

If q vanishes then $M^2 = 0$ and the minimal polynomial reduces to $(\lambda - 3/B^2)^2$ with corresponding Segrè type [(1,1,2)], a type allowed by the energy conditions. But only a subset of this type satisfying (5) is allowed. The restriction $q = 0$ implies B is constant ($B' = 0$) and $G_{\mu\nu}$ reduces to the canonical form (ii) of Eq. (4). The restrictions (5) that must be satisfied are

$$\epsilon = p > 0,\ \underset{T}{\lambda}{}_0 \leqslant 0,\ \underset{T}{\lambda}{}_0 \leqslant \underset{T}{\lambda}{}_i \leqslant \underset{T}{\lambda}{}_0,$$

or 2

$$2(A'/A)^2 - A''/A > 0,\ 3/B^2 + \Lambda \leqslant 0,$$
$$3/B^2 + \Lambda \leqslant 3/B^2 + \Lambda \leqslant -3/B^2 - \Lambda.$$

The last condition is equivalent to the second one.

It was only for this symmetry and in agreement with these conditions that the one solution of Lauten and Ray[7] was found. The solution was to the field equations without the cosmological term and was for a perfect fluid model of matter. The four-velocity of the fluid was found to be null and the constant pressure found was interpreted as a cosmological constant.

When the cosmological term is included in the field equations the solution goes through in the same manner. The energy-momentum tensor of a perfect fluid is given by

$$T_{\mu\nu} = (\rho + P)U_\mu U_\nu + Pg_{\mu\nu},$$

where ρ is the energy density, P is the pressure, and U_μ is the four-velocity of the fluid. The field equations (1) are

$$\begin{aligned}
&(1,1) \quad p + 3/B^2 + \Lambda = 8\pi[(\rho + P)U_1^2 + P],\\
&(1,3) \quad -q = 8\pi(\rho + P)U_1U_3,\\
&(1,4) \quad -p = 8\pi(\rho + P)U_1U_4\\
&(2,2) \quad 3/B^2 + \Lambda = 8\pi[(\rho + P)U_2^2 + P],\\
&(3,3) \quad 3/B^2 + \Lambda = 8\pi[(\rho + P)U_3^2 + P],\\
&(3,4) \quad q = 8\pi(\rho + P)U_3U_4,\\
&(4,4) \quad p - 3/B^2 - \Lambda = 8\pi[(\rho + P)U_4^2 - P].
\end{aligned} \tag{7}$$

From these it follows that $U_1 = -U_4$ and $U_2 = U_3 = 0$, so that the four-velocity is null. Also resulting is $q = 0$, which implies B is constant ($B' = 0$). The pressure is found to be

$$P = (1/8\pi)(3/B^2 + \Lambda) = \text{const}.$$

Thus the field equations reduce to one equation

$$p = 2(A'/A)^2 - A''/A = 8\pi(\rho + P)U_1^2, \tag{8}$$

which may be solved one the right-hand side is specified.

Since the pressure is constant it can be interpreted as part of the effective cosmological term and transposed to the left-hand side of the field equations (1)

$$G_{\mu\nu} + \Lambda_{\text{eff}} g_{\mu\nu} = 8\pi(\rho + P)U_\mu U_\nu,$$

where

$$\Lambda_{\text{eff}} = \Lambda - 8\pi P = -3/B^2.$$

The energy-momentum tensor is then reduced to

$$T_{\mu\nu} = \begin{pmatrix} p & 0 & 0 & -p \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -p & 0 & 0 & p \end{pmatrix},$$

and is of the cannonical form (ii) of Eq. (4) and Segrè type [(1,1,2)]. The energy conditions are satisfied provided

$$p = 2(A'/A)^2 - A''/A > 0.$$

The inclusion of the cosmological term (not dealt with in the previous study) now allows several interpretations of the above solution. With the restriction $\Lambda = -3/B^2$, the source of the energy-momentum tensor may be taken as, besides the null fluid, a null electromagnetic field, a massless scalar field, or some combination of these. The right-hand side of (8) will depend on the interpretation, and once specified allows a solution.

Further reduction in the Segrè type requires the additional restriction

$$2(A'/A)^2 - A''/A = 0 \quad (p = 0),$$

which can be solved for A. Then $M = 0$, the minimal polynomial is $(\Lambda - 3/B^2)$ and the Segrè type is [(1,1,1,1)]. This is just the special case of the solution (8) for vacuum.

The possible Segrè types for the space-time are now exhausted, therefore completing this application of the procedure. The interesting conclusion reached, then, is that the only solution allowed by the energy conditions is of the form

$$2(A'/A)^2 - A''/A = f(x^1 - x^4),$$

where the function f depends on the interpretation of the matter. The substitution

$$y = 1/A$$

simplifies the equation to

$$y'' - fy = 0, \tag{9}$$

which may be solved once the function f is specified. No other solution is allowed. The procedure thus furnishes a proof of the startling generality of the Lauten-Ray solution (9). That is, the application of the energy conditions leads to a generalized Birkhoff-type theorem for this symmetry. No other method of establishing these general results is known.

B. G_4V and G_4VII symmetry

The metric for G_4V is, according to Petrov,

$$ds^2 = 2\,dx^1\,dx^4 + \alpha(x^4)\exp(-x^1)[(dx^2)^2 + (dx^3)^2],$$

where α is an arbitrary function. Tranformed to an orthonormal frame the Einstein tensor becomes

$$G_{\mu\nu} = \begin{bmatrix} -a^2 + 2b - \frac14 & 0 & 0 & a^2 - \frac14 \\ 0 & b & 0 & 0 \\ 0 & 0 & b & 0 \\ a^2 - \frac14 & 0 & 0 & -a^2 - 2b - \frac14 \end{bmatrix},$$

where

$$a^2 = 2A''/A,\quad b = -2^{1/2}A'/A,\quad A = \alpha^{1/2}.$$

Again primes refer to differentiation with respect to $(x^1 - x^4)$.

The characteristic equation has the simple roots $2b + a$, $2b - a$ and the root b of multiplicity 2. A direct investigation shows the eigenvector corresponding to the eigenvalue $2b + a$ is timelike where a is taken as the positive square root. Also there are three spacelike eigenvectors corresponding to the other three eigenvalues. Thus the most general Segrè type is [1,(1,1),1]. This type is allowed only if the eigenvalues of the energy-momentum tensor satisfy (5). This condition gives $a \leqslant 0$, which contradicts the definition of a unless $a = 0$. With $a \neq 0$ the Segrè type will degenerate when the eigenvalue b equals either of the others. When $b = a$ ($b = 2b - a$) the Segrè type is [(1,1,1),1], and when $b = -a$ ($b = 2b + a$) the Segrè type is [1,(1,1,1)]. In these cases the condition yielding the contradiction above is still required.

If $a = 0$ the eigenvalues are $2b$ and b, each with multiplicity 2. The Einstein tensor reduces to the canonical form (ii) of Eq. (4) corresponding to Segrè type [(1,1),2] with $\epsilon = -\frac14$. To satisfy conditions (5), ϵ must be positive. If $b = 0$ the Segrè type is [(1,1,2)] but ϵ is unaltered. Thus in all cases the G_4V symmetry cannot satisfy the energy conditions. There are always frames in which the energy density becomes negative.

Similar contradictions arise in the application to $G_4\text{VII}_1$ and $G\,\text{VII}_2$ symmetries.

C. G_4VIII symmetry

Petrov gives the metric for $G_4\mathrm{VIII}_1$ as

$$ds^2 = 2\,dx^1\,dx^4 + \beta(x^4)[\cos^2(x^3)(dx^2)^2 + (dx^3)^2].$$

In an orthonormal frame the Einstein tensor is

$$G_{\mu\nu} = \begin{bmatrix} a+b & 0 & 0 & -b \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -b & 0 & 0 & -a+b \end{bmatrix},$$

where

$$a = -1/A^2,\ b = -A''/A,\ A = \beta^{\frac{1}{2}}.$$

The primes refer to differentiation with respect to $(x^1 - x^4)$.

The characteristic equation has the roots 0 and a, each of multiplicity two. The Einstein tensor is of the canonical form (ii) of Eq. (4) and the Segrè type is [(1,1),2]. The conditions (5) required by imposing the energy conditions are

$$-A''/A > 0,$$
$$-1/A^2 + \Lambda \leqslant 0,$$
$$-1/A^2 + \Lambda \leqslant \Lambda \leqslant 1/A^2 - \Lambda,$$

which are satisfied provided

$$\Lambda \leqslant 1/(2A^2) \text{ and } A''/A < 0. \tag{10}$$

The energy-momentum tensor is

$$T_{\mu\nu} = \frac{1}{8\pi}\begin{bmatrix} a+b+\Lambda & 0 & 0 & -b \\ 0 & \Lambda & 0 & 0 \\ 0 & 0 & \Lambda & 0 \\ -b & 0 & 0 & -a+b+\Lambda \end{bmatrix},$$

and is apparently not one of the simple models investigated by Lauten and Ray, since they did not find solutions for this symmetry.

Since the energy-momentum tensor obtained by calculation from the metric is of Segrè type [(1,1),2], any simple model of matter acting as a source in this space-time must yield an energy-momentum tensor of the same Segrè type. The results of the algebraic classification of the simple models of matter given in Sec. II do not include this type. However (as indicated by the decompositon given by Williams)[27] the desired Segè type in this case can be obtained by adding a null fluid model with Segrè type [(1,1,2)] and a non-null electromagnetic field with Segrè type [(1,1),(1,1)]. This may be seen directly by observing the resulting sum of the corresponding canonical forms. Since the energy conditions can be satisfied, a solution for a coupled electromagnetic field and null fluid would be expected.

To investigate this expectation the field equations must be solved for the energy-momentum tensor

$$T_{\mu\nu} = (\rho + P)U_\mu U_\nu + Pg_{\mu\nu} + g_{\mu\sigma}F_{\nu\rho}F^{\sigma\rho} - \tfrac{1}{4}g_{\mu\nu}F_{\sigma\rho}F^{\sigma\rho}.$$

The part of the energy-momentum tensor related to the electromagnetic field must have the canonical form [(1,1),(1,1)]. This is precisely the form obtained by assuming that the electromagnetic field tensor $F_{\mu\nu}$ has the same symmetry as the metric, that is

$$L_\xi F_{\mu\nu} = 0.$$

Solving for $F_{\mu\nu}$,

$$F_{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & -\alpha \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \alpha & 0 & 0 & 0 \end{pmatrix},$$

where α is a function of $(x^1 - x^4)$. Maxwell's equations are satisfied by this field tensor for a null current[28]

$$j^\mu = (-1/4\pi)[\alpha' + 2(A'/A)\alpha, 0, 0, \alpha' + 2(A'/A)\alpha].$$

Thus,

$$\underset{\mathrm{EM}}{T_{\mu\nu}} = \begin{bmatrix} \frac{-\alpha^2}{2} & & & 0 \\ & \frac{\alpha^2}{2} & & \\ & & \frac{\alpha^2}{2} & \\ 0 & & & \frac{\alpha^2}{2} \end{bmatrix}.$$

The field equations (1) are

$$\begin{aligned}
&(1,1) & a + b + \Lambda &= 8\pi[(\rho + P)U_1^2 + P - \alpha^2/2],\\
&(1,4) & -b &= 8\pi[(\rho + P)U_1U_4\\
&(2,2) & \Lambda &= 8\pi[(\rho + P)U_2^2 + P + \alpha^2/2],\\
&(3,3) & \Lambda &= 8\pi[(\rho + P)U_3^2 + P + \alpha^2/2],\\
&(4,4) & -a + b - \Lambda &= 8\pi[(\rho + P)U_4^2 - P + \alpha^2/2].
\end{aligned} \tag{11}$$

From these it follows that $U_1 = -U_4$ and $U_2 = U_3 = 0$, so that the four-velocity of the matter is null (a needed result for agreement of Segrè types). The pressure is found to be

$$P = \Lambda/(8\pi) - \alpha^2/2, \tag{12}$$

and the field equations reduce to

$$1/A^2 = 8\pi\alpha^2 \text{ and } A''/A = -8\pi(\rho + P)U_1^2,$$

which may be solved once the right-hand sides, satisfying condition (12), are specified. The energy conditions require that the pressure be negative in this case. The realization of negative pressure in the expectation value of the energy-momentum tensor has been discussed by Parker and Fulling[29] and Ford.[30]

In this application the energy conditions were found to be satisfied under conditions (10), indicating the possibility of obtaining solutions to the field equations. The necessity of agreement between the algebraic clasification of the Einstein tensor of the space-time and the energy-momentum tensor of the matter was then employed to determine a likely simple model of matter for obtaining a solution to the field equations. Finally, the field equations were investigated for this form of matter and a solution obtained.

When $A'' = 0$ $(b = 0)$ the Segrè type reduces to [(1,1),(1,1)]. The requirement of the energy conditon is

$$\Lambda \leqslant 1/(2A^2).$$

Since the non-null electromagnetic field has the proper type of energy-momentum tensor, it might be expected to be a satisfactory source. The field equations (11) with $b = 0$ give

$$\alpha^2 = \Lambda/(4\pi) = \text{const}., \text{ and } 1/A^2 = 8\pi\alpha^2 = \text{const}.$$

The Riemann tensor does not vanish, so the space-time cannot be flat. This is a special case of the previous solution with

$$\rho = -P = 0.$$

For $G_4\mathrm{III}_2$ symmetry, the Segré type is [(1,1),2] and the energy conditions are satisfied if

$$\Lambda \leqslant 1/(2A)' \quad \text{and} \quad 2A''/A \leqslant 1/(2A^4).$$

When $2A/A = 1/(2A^4)$ the Segrè type reduces to [(1,1),(1,1)] with the same energy condition requirement. Analogous to the $G_4\mathrm{VIII}_1$ symmetry, solutions for combinations of null fluid and non-null electromagnetic fields would be expected.

IV. CONCLUSION

To summarize the results obtained by applying the energy conditions to those space-times with G_4 symmetry studied, it was found that for $G_4\mathrm{I}_1$ symmetry the only solution allowed by the energy conditions (3) corresponds to a null fluid model of matter. Also the energy conditions allow no solutions for the symmetries $G_4\mathrm{V}$, $G_4\mathrm{VII}_1$, and $G_4\mathrm{VII}_2$. The $G_4\mathrm{VIII}_1$ and $G_4\mathrm{VIII}_2$ symmetries correspond with an energy-momentum tensor that agrees with the energy conditions, but the corresponding matter is not described by one of the simple models studied previously. Through consideration of the algebraic classification of the energy-momentum tensor a combination of these simple models was determined to be suitable and the solution found.

It should be emphasized that the T-method approach as applied by Lauten and Ray, though giving a particular solution to the field equations, offered little explanation for the scarcity of solutions and could not yield conclusive information concerning other possible solutions. In contrast, the g-method approach used here obtains general results and explains the lack of solutions. The energy conditions cannot be satisfied in general for certain symmetries. However, the simple models of matter fields can only be found when the geometry agrees with the energy conditions.

Further, there can be solutions only for simple models of matter whose corresponding energy-momentum tensor allows the same algebraic classification (Segrè type) as that obtained from the geometry. This considerably restricts the matter fields that need be investigated for solutions, thus giving a more direct procedure for searching.

It should be emphasized even more that the results obtained here depend on the validity of the energy conditions which is assumed throughout this work. The possibility of obtaining more general results, as illustrated here, supports the need for evaluating their validity.[10,11,31–33]

The dominant energy conditions are not the only conditions that may be chosen. Hawking and Ellis[11] also discuss the less stringent weak energy condition (positive definite energy only) and the more stringent strong energy conditions ($T_{\mu\nu}v^\mu v^\nu \geqslant \frac{1}{2}Tv^\mu v_\mu$, $\Lambda = 0$,) from which follow the singularity theorems. Either of these conditions allows the same Segrè types (6) as the dominant energy conditions, but the restrictions analogous to (5) are, respectively, less and more stringent. The general results obtained in this work would not be altered except for slight changes in allowed values for the parameters in the solutions.

Besides the question of the validity of the energy conditions, just how applicable the g-method type approach taken here may be is not apparent. In initial investigations of space-times with less restrictive symmetries (spherical symmetry, the Farnsworth metric, the Kundt metric)[34,35] the energy conditions exhibit themselves through nonlinear partial differential inequalities, the value of which is not transparent. This is the result of the more complicated form of the eigenvalues. The Segrè types of the energy-momentum tensor and the conditions on the metric for which each of these types is obtained is easily determined. Only those matter fields with an energy-momentum tensor of the same Segrè type may yield a solution to the field equations. Other workers have considered the Segrè classification of various types of gravitational fields.[36–40] These results would be useful in a further study of the types of matter allowed by these gravitational fields following and extending the methods and results of this paper. If the algebraic classification of all macroscopic matter fields was known, those (and combinations thereof) that were suitable could be investigated for solutions. By comparing the canonical forms of both sides of the field equations, the simplest form of the equations for all allowed matter fields would be found, though their solution might be untenable. The energy conditions would then place further restrictions on the allowed values of the parameters in the equations.

Of course, the energy conditions could be imposed using the T-method also. Then, at first ignoring the question of the type of matter present, the energy-momentum tensor can only assume one of the canonical forms given in (6), with additional restrictions on the elements given by (5). In this way the number of unknowns in the field equations is considerable reduced. The problem of solving these nonlinear partial differential equations is still to be faced.

As the subject considered here is that of finding solutions to the field equations, the strength of the equations themselves should also be considered. In closing, we present a reminder of some ideas of the equations' author.[41]

"Not for a moment, of course, did I doubt that this formulation (1) was merely a makeshift in order to give the general principle of relativity a preliminary closed expression. For it was essentially not anything more than a theory of the gravitational field, which was somewhat artificially isolated from a total field....

"If anthing in the theory...can possibly make the claim to final significance, then it is the theory of the limiting case of the pure gravitational field....

"If one had the field-equation of the total field,...only then would the general theory of relativity be a complete theory."

ACKNOWLEDGMENTS

The authors wish to thank Professor J. Safko for providing an important reference, and Professor M. MacCallum and Professor G. Hall for providing preprints of their work.

[1]B. M. Tinsley, Phys. Today **30** (6), 32 (1977).
[2]Y. B. Zeldovich and I. D. Novikov, *Relativistic Astrophysics*, Vol. I: *Stars and Relativity*, (University of Chicago,Chicago, 1971), p. 28.
[3]J. L. Synge, *Relativity: The General Theory* (North-Holland, Amsterdam, 1960), p. 184.
[4]J. L. Synge, in *Perspectives in Geometry and Relativity*, edited by B. Hoffmann (Indiana U. P., Bloomington, 1966), p. 7.
[5]A. Z. Petrov, *Einstein Spaces* (Pergamon, New York, 1969).
[6]W. T. Lauten, III and J. R. Ray, Lett. Nuovo Cimento **14**, 63, (1975).
[7]W. T. Lauten, III and J. R. Ray, J. Math. Phys. **18**, 885 (1977).
[8]H. Bondi, F. A. E. Pirani, and I. Robinson, Proc. R. Soc. London, Ser. A **251**, 519 (1959).
[9]F. A. E. Pirami, in *Recent Developments in General Relativity*, edited by L. Infeld (Pergamon, New York, 1962) p. 103.
[10]J. Plebański, Acta Phys. Pol. **26**, 963 (1964).
[11]S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-time*, (Cambridge U. P., Cambridge, 1973), p. 88.
[12]L. P. Eisenhart, *Riemannian Geometry* (Princeton U. P., 1926).
[13]R. V. Churchill, Trans. Amer. Math. Soc. **34**, 347 (1932).
[14]L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, rev. 2nd ed. (Pergamon, London, 1956), p. 269.
[15]See Ref. 5, pp. 323-326.
[16]C. D. Collinson and R. Shaw, Int. J. Theor. Phys. **6**, 347 (1972).
[17]G. S. Hall, J. Phys. A **9**, 541 (1976).
[18]G. S. Hall, "The Classification of Second Order Symmetric Tensors in General Relativity Theory," paper delivered at the Differential Geometry Semester in the Stefan Banach International Mathematical Centre, Warsaw, Poland, Autumn, 1979.
[19]See Ref. 5, pp. 322-326.
[20]A. Z. Petov, in *Perspectives in Geometry and Relativity*, edited by B. Hoffmann (Indiana U. P., Bloomington, 1966), p. 278.
[21]L. Bianchi, Soc. Ital. Sci. mem. mat. (3), **11**, 267 (1897).
[22]S. Lie and F. Engel, *Theorie der Transformationgruppen* (Teubner, Leipzig, 1890).
[23]G. I. Kruchkovich, Math. Sbor. **41**, 195 (1957).
[24]W. J. Cormack and G. S. Hall, J. Phys. A **12**, 55 (1979).
[25]J. B. Griffiths, Gen. Relativ. Gravit. **5**, 453 (1974).
[26]M. Novello, "A Note Concerning the Undistinguishability of Neutrino Field and Stokesian Fluids," preprint.
[27]E. Williams, "An Invariant Decomposition of a Symmetric Tensor of Rank Two in Pseudo-Euclidean Space," Ph. D. Thesis, University of Michigan, March, 1946.
[28]G. M. O'Brien, Phys. Lett. A **77**, 105 (1980).
[29]L. Parker and S. A. Fulling, Phys. Rev. D **7**, 2357 (1973).
[30]L. H. Ford, Phys. Rev. D **11**, 3370 (1975).
[31]See Ref. 2, pp. 194–200.
[32]J. D. Bekenstein, Int. J. Theor. Phys. **13**, 317 (1975).
[33]F. J. Tipler, Phys. Rev. D **17**, 2521 (1978).
[34]W. Kundt, Z. Phys. **163**, 77 (1961).
[35]D. Kramer, J. Phys. A **13**, *L* 43 (1980).
[36]H. Goenner and J. Stachel, J. Math. Phys. **11**, 3358 (1970).
[37]M. A. H. MacCallum, "Locally Isotropic Space-times with Non-null Homogeneous Hypersurfaces,"preprint.
[38]R. F. Crade and G. S. Hall, Phys. Lett. A **75**, 17 (1979).
[39]M. Abdel-Megied, J. Phys. A **13**, 1709 (1980).
[40]M. Abdel-Megied, J. Phys. A **13**, 1721 (1980).
[41]*Albert Einstien: Philosopher-Scientist*, edited by P. A. Schlipp (Library of Living Philosophers, Evanston, Ill., 1949), p. 75.

An exact stationary solution of the combined Einstein–Maxwell–Klein–Gordon equations

J. D. Gegenberg and A. Das
Department of Mathematics, Simon Fraser University, Burnaby, B.C., V5A 1S6, Canada

(Received 22 July 1980; accepted for publication 21 January 1981)

The Einstein–Maxwell–Klein–Gordon equations are simplified by imposing stationarity, isometric motion, the Weyl–Majumdar–Papapetrou condition, and axial symmetry. An exact (nonstatic) stationary solution is found such that the electric field vanishes, the magnetic field is constant and parallel to the polar axis, and the wavefunction of the matter field is of the form of a "pure phase." The energy–momentum tensor satisfies the strong energy condition of Hawking and Ellis. The metric tensor resembles that of the Gödel solution and has similar causal properties.

PACS numbers: 04.40. + c, 04.20.Jb, 03.50.Kk

I. INTRODUCTION

In anticipation of the construction of a quantum theory of gravitation, which will unify gravity and elementary particle phenomena, it seems that a useful purpose will be served in finding exact solutions of the classical Einstein–Maxwell–Klein–Gordon equations. Such solutions, notably if they are stationary, could provide examples of self-consistent classical models of massive spin-zero bosons. Although these models are probably not realistic (because of their nonquantal nature), nevertheless some of their properties hint at what a fully second quantized theory may yield. For example, the static, spherically symmetric solitonlike solutions of Das and Coffman provide a theoretical estimate of the fine-structure constant in terms of the gravitational constant, the speed of light, and Planck's constant.[1] Furthermore, one should keep in mind the recent results of Jackiw *et al.*,[2] wherein certain exact solutions of classical nonlinear field equations provide approximations of the vacuum expectation values of the quantum fields and reveal a richness in the structure of the corresponding quantum field theories that could not have been uncovered by standard perturbation methods. It is likely that there are solutions of the classical Einstein–Maxwell–Klein–Gordon equations which play a similar role in quantum gravity.

In Sec. II, the Einstein–Maxwell–Klein–Gordon (EMKG) equations will be displayed in their full generality, then simplified by imposing the conditions of stationarity of the fields and that the matter field current is parallel to the timelike Killing vector field (the condition of *isometric motion*[3]). The field equations are further simplified by imposing a condition of the "Weyl–Majumdar–Papapetrou" (WMP) form.[4] The latter implies, as in the static case,[1,5–7] a "balance condition" on the bare charge and mass of the matter field. This section concludes by displaying a class of stationary, but nonstatic axially symmetric solutions of the EMKG equations, the first examples of which, the authors believe, to appear in the literature.[8] In Sec. III, some of the properties of these solutions are explicated, and, in particular, it is shown that the solutions satisfy the so-called strong energy condition of Hawking and Ellis,[9] and that the physical components of the Riemann tensor are constants. It is also shown that for each member of a subclass of these solutions, the curves of the angular (periodic) coordinate are timelike everywhere except in a cylindrical shell of finite thickness centered on the polar axis. Thus, the solutions have the same causal pathologies as the Gödel solution,[7,9] the metric of which resembles the ones found here.

II. THE FIELD EQUATIONS

The notation and conventions used here are as follows. Latin and Greek indices have, respectively, the ranges 1,2,3,4 and 1,2,3. Space–time, M_4, is a C^3_p-differentiable four dimensional semi-Riemannian manifold with metric tensor field $\hat{g}_{ij}$ having signature -2. All "hatted" symbols, e.g., $\hat{R}_i{}^j$, have had their indices raised by $\hat{g}^{ij}$. The Ricci tensor $\hat{R}_{ij} \equiv \hat{g}^{km}\hat{R}_{kijm}$ and the curvature scalar $\hat{R}$ are constructed from the curvature tensor of the metric $\hat{g}_{ij}$. The electromagnetic four-potential is denoted A_i and the field strength is defined as $F_{ij} \equiv A_{i,j} - A_{j,i}$. A comma followed by an index denotes partial differentiation with respect to the coordinates. $\hat{\nabla}_i$ is the covariant derivative on M_4. The complex Klein–Gordon matter field ψ has bare mass m and bare charge e. Finally, natural Gaussian units are used so that $G = c = \hbar = 1$.

The Einstein–Maxwell–Klein–Gordon equations are

$$\hat{R}_{ij} - \tfrac{1}{2}\hat{g}_{ij}\hat{R} = -8\pi T_{ij}, \tag{2.1}$$

$$\hat{\nabla}_j \hat{F}^{ij} = \hat{J}^i, \tag{2.2}$$

$$(\hat{D}^j D_j + m^2)\psi = 0. \tag{2.3}$$

The remaining symbols are defined as

$$D_i \equiv \hat{\nabla}_i + ieA_i, \tag{2.4a}$$

$$D^*_i \equiv \hat{\nabla}_i - ieA_i, \tag{2.4b}$$

$$T_{ij} \equiv M_{ij} + \epsilon_{ij}, \tag{2.5}$$

$$M_{ij} \equiv (D^*_i\psi^*)(D_j\psi) + (D^*_j\psi^*)(D_i\psi) - \hat{g}_{ij}\left[(\hat{D}^{*k}\psi^*)(D_k\psi) - m^2\psi^*\psi\right], \tag{2.6}$$

$$\epsilon_{ij} \equiv -F_{ik}\hat{F}_j{}^k + \tfrac{1}{4}\hat{g}_{ij}F_{km}\hat{F}^{km}, \tag{2.7}$$

$$J_i \equiv -ie\left[(D^*_i\psi^*)\psi - \psi^*(D_i\psi)\right]. \tag{2.8}$$

[a] This work forms part of the author's Ph.D. thesis.
[b] The author is also a member of the Theoretical Science Institute, Simon Fraser University, Burnaby, B.C., V5A 1S6, Canada.

 0022-2488/81/081736-04$01.00 © 1981 American Institute of Physics

A set of four conditions will now be imposed on the field equations. This procedure will simplify the field equations to the point where nontrivial solutions can be found. The four conditions are (i) stationarity of the fields, (ii) isometric motion, (iii) the WMP condition, and (iv) axial symmetry.

(i) Stationarity of the metric, electromagnetic field, and the matter field shall be taken to mean that there exists a timelike Killing vector field ξ^i on M_4 such that

$$\mathscr{L}_\xi \hat{g}_{ij} = \mathscr{L}_\xi A_i = 0,$$
$$\mathscr{L}_\xi \psi = iE\psi,$$

where $\mathscr{L}_\xi$ is the Lie derivative and E is a constant. One can find[10] a class of coordinate systems $x^i = (x^\alpha, t)$ in which $\xi^i = (0,0,0,1)$, and hence in which the metric has the form

$$\hat{\Phi} \equiv \hat{g}_{ij} dx^i dx^j = -e^{-\omega} g_{\alpha\beta} dx^\alpha dx^\beta + e^\omega [a_\alpha dx^\alpha + dt]^2, \quad (2.9)$$

the matter field has the form $\psi = \chi(x^\alpha) e^{iEt}$, and in which A_i is a function of the x^α only. The quantities $g_{\alpha\beta}, a_\alpha$, and ω depend only on the x^α. The $g_{\alpha\beta}$ are the components of the metric tensor on a three-dimensional Riemannian manifold M_3, called the *associated space*. The a_α and ω transform, respectively, as a covariant vector and a scalar on M_3. The twist vector on M_3 is defined as

$$\tau^\mu \equiv \tfrac{1}{2} e^{2\omega} \eta^{\mu\alpha\beta} f_{\alpha\beta}, \quad (2.10)$$

where $\eta^{\mu\alpha\beta} \equiv g^{-1/2} \epsilon_{\mu\alpha\beta}$ is the alternating tensor on M_3, and $f_{\alpha\beta} \equiv a_{\alpha,\beta} - a_{\beta,\alpha}$. The necessary and sufficient condition[10] that the metric $\hat{g}_{ij}$ is static is that $\tau^\mu = 0$.

(ii) From the Maxwell equations (2.2) it follows that there exists a scalar field $A(x^i)$ such that $\xi^j F_{ij} = A_{,i}$. That there exists another scalar field $B(x^i)$ such that $\xi^{j*} F_{ij} = -B_{,i}$, where $^*F_{ij}$ is the dual of F_{ij}, follows if and only if the current $\hat{J}^i$ is parallel to the Killing vector ξ^i ("isometric motion").[3,10] Henceforth, it will be assumed that the coordinates (x^α, t) are those for which $\xi^i = (0,0,0,1)$ and for the remainder of this section all tensor analysis will be done in M_3.

Thus $\hat{J}^\alpha = 0$ and hence

$$\chi^*_{,\alpha}\chi - \chi^*\chi_{,\alpha} = 2ie(A_\alpha + A a_\alpha)\chi^*\chi,$$

where we have chosen, without loss of generality, $A \equiv -(A_4 + E/e)$. From this one easily derives, after differentiating both sides of the above by x^β, antisymmetrizing in α and β, and using (2.10), that

$$B_{,\mu} = A e^{-\omega} \tau_\mu, \quad (2.11)$$

i.e., *the twist vector is parallel to the magnetic field.*

By procedures analogous to those used in Refs. 3, 10, and 11, and using (2.11), the field Eqs. (2.1)–(2.3) are cast into the following useful and elegant form:

$$\begin{aligned} R_{\alpha\beta} &+ \tfrac{1}{2} e^{-2\omega} \mathrm{Re}[(\Gamma_{,\alpha} + 8\pi\phi\, ^*\phi_{,\alpha}) \\ &\times (\Gamma^*_{,\beta} + 8\pi\phi^*_{,\beta}\phi) - 16\pi e^\omega \phi^*_{,\alpha}\phi_{,\beta}] \\ &+ 16\pi\eta_{,\alpha}\eta_{,\beta} - 16\pi g_{\alpha\beta} e^{-2\omega}(e^2A^2 - m^2e^\omega)\eta^2 \\ &= 0, \end{aligned} \quad (2.12a)$$

$$\Delta_2\phi - e^{-\omega}(\Gamma^{,\alpha} + 8\pi\phi\, ^*\phi^{,\alpha})\phi_{,\alpha} = 2e^2 e^{-\omega} A\eta^2, \quad (2.12b)$$

$$\begin{aligned} \Delta_2\Gamma &- e^{-\omega}(\Gamma^{,\alpha} + 8\pi\phi\, ^*\phi^{,\alpha})\Gamma_{,\alpha} \\ &= 16\pi e^{-\omega}[(e^2A^2 - m^2e^\omega) + ie^2AB]\eta^2, \end{aligned} \quad (2.12c)$$

$$\Delta_2\eta + e^{-2\omega}(e^2A^2 - m^2e^\omega)\eta = 0. \quad (2.12d)$$

In these equations $R_{\alpha\beta}$ is the Ricci tensor on M_3 and the remaining symbols are defined as

$$\phi \equiv A + iB, \quad (2.13)$$

$$\Gamma \equiv e^\omega - 4\pi\phi\, ^*\phi + i\Omega, \quad (2.14)$$

$$\eta \equiv (\psi^*\psi)^{1/2} = (\chi^*\chi)^{1/2}, \quad \Delta_2\phi \equiv g^{\alpha\beta}\nabla_\beta\phi_{,\alpha},$$

where Ω, the "twist potential" is determined by

$$\Omega_{,\alpha} = \tau_\alpha + 4\pi i(\phi\, ^*\phi_{,\alpha} - \phi^*_{,\alpha}\phi), \quad (2.15)$$

and where ∇_α is the covariant derivative on M_3.

(iii) the quantity $(e^2A^2 - m^2e^\omega)$ occurs frequently in the field Eqs. (2.12), and thus, a great simplification would occur if that quantity were to vanish. Now $e^\omega = e^2A^2/m^2$ is of the form of the WMP condition: $\hat{g}_{44} = [c_1(c_2 \pm A_4)]^2$, where c_1 and c_2 are constants. The WMP condition has formerly been associated with static solutions of the Einstein–Maxwell equations without sources[7] and with sources[1,5,6,11] when one requires that in some sense, the gravitational and electrostatic forces are balanced so that the solution is stable. Characteristically, the WMP condition is accompanied by a relation of the form (charge density)/(mass density) = const. Such, in fact, is the case here, because of the following result: If one imposes $e^\omega = e^2A^2/m^2$ on Eqs. (2.12), then the contracted Bianchi identities $\nabla_\beta R^{\alpha\beta} - \tfrac{1}{2} g^{\alpha\beta} R_{,\beta} = 0$ imply that either $e^2 = 16\pi m^2$ or $B(x^\alpha) = f(A(x^\alpha))$, where f is an arbitrary C^2 function. Henceforth it will be assumed that $e^\omega = e^2A^2/m^2$ and $e^2 = 16\pi m^2$. It is worth mentioning here that in the static case $e^\omega = e^2A^2/m^2$ if and only if $e^2 = 4\pi\ m^2$ holds.[6]

(iv) It shall now be assumed that there exists another Killing vector field in M_4 whose orbits are closed curves and which is independent of the previously introduced timelike Killing vector field. Coordinates (r,z,φ,t) adapted to the Killing motion will henceforth be used so that the φ lines coincide with the orbits of the periodic Killing vector field and the z axis is the polar axis. In these coordinates the metric of M_4 will take the form

$$\hat{\Phi} = -e^{-\omega}[e^\nu(dr^2 + dz^2) + e^{2\lambda}d\varphi^2] + e^\omega(a\,d\varphi + dt)^2, \quad (2.16)$$

where ω, ν, λ, and a are functions of (r,z) only. Furthermore, it is assumed that the potentials A and B are functions of (r,z) only and that the matter field is of the form $\psi = \eta(r,z)e^{i(L\varphi + Et)}$, where L is a constant and $\eta(r,z)$ is real.[12]

Consider the Eq. (2.12a) with indices $\alpha = \beta = 3$ and with the assumption of axial symmetry imposed on the metric tensor and the other fields, but without assuming, for the moment, that the WMP condition holds. A straightforward computation yields[10]

$$e^{-\lambda}\Delta(e^\lambda) = 16\pi e^{-2\omega + \nu}(e^2A^2 - m^2e^\omega)\eta^2,$$

where $\Delta \equiv \partial^2/\partial r^2 + \partial^2/\partial z^2$ is the Laplacian on $\mathbb{R}^2$. Hence e^λ is a harmonic function of (r,z) if and only if $e^\omega = e^2A^2/m^2$. Thus the metric (2.16) can be written in the Weyl–Lewis–Papapetrou form[13]

$$\hat{\Phi} = -e^{-\omega}[e^{\nu}(dr^2 + dz^2) + r^2 d\varphi^2] + e^{\omega}(a d\varphi + dt)^2, \qquad (2.17)$$

if and only if the WMP condition holds.

All the field Eqs. (2.12) can now be written as partial differential equations on an auxilliary Euclidean $\mathbb{R}^3$. The equations below are linear combinations of Eqs. (2.12) with the WMP condition, $e^2 = 16\pi m^2$, and axial symmetry conditions imposed:

$$\Delta \nu = -\tfrac{3}{2}A^{-2}|\nabla A|^2 - 16\pi|\nabla\eta|^2, \qquad (2.18a)$$

$$\nu_{,r} = r\left[\tfrac{3}{2}A^{-2}(A_{,r}^2 - A_{,z}^2) + 16\pi(\eta_{,r}^2 - \eta_{,z}^2)\right], \qquad (2.18b)$$

$$\nu_{,z} = r(3A^{-2}A_{,r}A_{,z} + 32\pi\eta_{,r}\eta_{,z}), \qquad (2.18c)$$

$$\nabla^2 B - 3A^{-1}\nabla A\cdot\nabla B = 0, \qquad (2.18d)$$

$$\nabla^2 A - A^{-1}|\nabla A|^2 = 0, \qquad (2.18e)$$

$$|\nabla B|^2 - |\nabla A|^2 = 2m^2 e^{\nu}\eta^2, \qquad (2.18f)$$

$$\nabla^2\eta = 0, \qquad (2.18g)$$

where $A_{,r} \equiv \partial A/\partial r$, $A_{,z} \equiv \partial A/\partial z$, ∇ is the usual gradient operator in cylindrical polar coordinates on Euclidean $\mathbb{R}^3$, ∇^2 is the Laplacian, $\nabla A\cdot\nabla B$ is the scalar product, and $|\nabla A| \equiv (\nabla A\cdot\nabla A)^{1/2}$.

Given solutions of (2.18e) and (2.18g), it is easy to integrate (2.18b) and (2.18c) to obtain $\nu(r,z)$, and (2.18a) is identically satisfied. Hence a complete solution of the field equations hinges on finding a function $B(r,z)$ satisfying (2.18d) and (2.18f).

If one further requires that $A = \text{const} \neq 0$, then there is just one class of solutions of (2.18). The proof of this is relegated to the appendix. These solutions are characterized by $\eta = \text{const}$, $\nu = \text{const}$, and $B = \pm pz + B_0$, where B_0 is an arbitrary constant and $p^2 \equiv 2m^2 e^{\nu}\eta^2$. There is clearly no loss of generality in choosing $\eta = 1$ and $\omega = \nu = 0$. The function $a(r,z)$ can be determined from Eqs. (2.11). Hence a class of exact solutions of the Einstein–Maxwell–Klein–Gordon equations is given by

$$\hat{\Phi} = -(dr^2 + dz^2 + r^2 d\varphi^2) + [a(r)d\varphi + dt]^2, \qquad (2.19)$$

$$a(r) = \pm(8\pi)^{1/2} m r^2 + a_0,$$

$$A = \pm(16\pi)^{-1/2}, \qquad (2.20)$$

$$B = \pm 2^{1/2} m z + B_0, \qquad (2.21)$$

$$\psi = e^{i(L\varphi + Et)}. \qquad (2.22)$$

The quantities a_0, B_0, and L and E are arbitrary parameters.

III. SOME PROPERTIES OF THE SOLUTION

Some geometrical and physical properties of the solutions (2.19)–(2.22) are:

(i) The metric (2.19) is flat only if $m = 0$. The nonvanishing invariant components of the curvature tensor are

$$\hat{R}_{(1441)} = 8\pi m^2,$$
$$\hat{R}_{(1331)} = 24\pi m^2, \qquad (3.1)$$
$$\hat{R}_{(3443)} = 8\pi m^2,$$

where $\hat{R}_{(ijkm)}$ denote the components of the curvature tensor with respect to the orthonormal basis of 1-form fields $\{dr,\ dz,\ r d\varphi, [a(r)d\varphi + dt]\}$. The metric $g_{\alpha\beta}$ on M_3 is, of course, flat.

(ii) The metric (2.19) has four linearly independent Killing vector fields. In the coordinates (r,z,φ,t), these are $\partial/\partial t$, $\partial/\partial\varphi$, $\partial/\partial z$, and $\cos\varphi\,\partial/\partial r - r^{-1}\sin\varphi\,\partial/\partial\varphi + [a_0 r^{-1} \mp (8\pi)^{1/2} m r]\sin\varphi\,\partial/\partial t$. The existence of the last Killing vector field above is not obvious, and it was found by the procedure outlined in a recent article by Raychaudhuri and Thakurta.[14] Hence the space–time M_4 with metric $\hat{g}_{ij}$ is homogeneous.

(iii) The twist vector on M_3 is

$$\vec{\tau} = (0, \pm(32\pi)^{1/2} m, 0). \qquad (3.2)$$

Hence, $\hat{g}_{ij}$ is static only if $m = 0$.

(iv) Since $(-\hat{g})^{1/2} = r$, the polar axis cannot be included in the domain of the coordinates (r,z,φ,t). Since $\hat{g}_{33}(r) = -r^2 + [\pm(8\pi)^{1/2} m r^2 + a_0]^2$, the closed φ lines are nonspacelike for $0 < r \leqslant r_-$ and for $r \geqslant r_+$, where r_+ and r_- are the two nonnegative solutions of $\hat{g}_{33}(r) = 0$. If it is required that $a_0^2 < (128\pi m^2)^{-1}$, then in the region defined by $0 \leqslant r_- < r < r_+$, the metric is regular everywhere and the φ lines are spacelike. Since $\hat{g}^{44} > 0$ in the above region, the metric there does not admit smooth, closed, timelike curves.[15] The constancy of $\hat{R}$ precludes the existence of any asymptotically flat regions in M_4.

It would seem that one could eliminate the constant a_0 in the metric (2.19) by the replacement $t' = t + a_0\varphi$. However, the periodicity of φ implies that t' is *not* a proper coordinate, as the following argument[16] demonstrates. Let p and q be two events whose t coordinates are identical, and whose φ coordinates differ by slightly less than 2π. Then their t' coordinates differ by a "jump" of nearly $2\pi a_0$, in spite of the fact that the points are nearly coincident.

Furthermore, the condition of elementary flatness,[7] which in the present case amounts to

$$\lim_{r\to 0} \{2\pi|1 - [\pm(8\pi)^{1/2} m r + a_0 r^{-1}]^2|^{1/2}\} = 2\pi,$$

holds only if $a_0 = 0$.

For these reasons it must be concluded that metrics with differing values of a_0 must be physically and geometrically distinct.

(v) From (2.1) and (3.1), the nonvanishing components of T_{ij} are easily computed:

$$T_{11} = -T_{22} = m^2,\ T_{33} = m^2[r^2 + 3(a(r))^2],\ T_{34} = 3m^2 a(r),\ T_{44} = 3m^2. \qquad (3.3)$$

The eigenvalues of T_{ij} are $-m^2$, $-m^2$, m^2, $3m^2$ and the eigenvectors corresponding to the first three eigenvalues are all spacelike, while that corresponding to the last eigenvalue is timelike. Hence, the energy momentum tensor satisfies the strong energy condition of Hawking and Ellis[9]: $T_{ij}W^iW^j \geqslant \frac{1}{2}W^iW_iT^j{}_j$ for any timelike vector W^i.

(vi) The physical relevance of the solution (2.19)–(2.22) is problematic. On the one hand, the electromagnetic and scalar field "sources" are not obviously unphysical, given the fact that the Hawking–Ellis strong energy condition is satisfied. On the other hand, the metric has causal pathologies of the sort which violate our intuition, and, in any case, have never been observed. The metric is not asymptotically flat, but the constancy of the physical components of the curvature would seem to preclude singularities at radial infinity. In fact, the situation posed here has a precedent in the Gödel solution,[7,9] which has similar causal pathologies, but also has well-behaved sources. If there is one lesson to be learned from the history of general relativity, it is that one should be hesitant about dismissing solutions as "nonphysical" because of their counter-intuitive properties. Hence, it is predicted that if there exist elementary charged massive spin-zero bosons, the configuration (2.19)–(2.22) will be realized. On a less fundamental level, it is possible that the solutions found here could be useful in an attempt to construct a model of a galaxy consisting of a gas of protons (whose spin can be approximately ignored) with a net macroscopically constant magnetic field.

ACKNOWLEDGMENT

One of us (J.D.G.) would like to thank Dr. S. Kloster for many helpful discussions. The work of the other author (A.D.) was supported in part by the NSERC Grant No. 3993.

APPENDIX

Requiring $A=$ const reduces the field Eqs. (2.18) to

$$\nu_{,r} = 16\pi r[\eta_{,r}^2 - \eta_{,z}^2], \tag{A1a}$$

$$\nu_{,z} = 32\pi r \eta_{,r}\eta_{,z}, \tag{A1b}$$

$$\nabla^2 B = 0, \tag{A1c}$$

$$|\nabla B|^2 = 2m^2 e^{\nu}\eta^2, \tag{A1d}$$

$$\nabla^2\eta = 0. \tag{A1e}$$

It will now be shown that (i) if it is required that all the fields are C^2 in (r,z), then Eqs. (A1) have a solution only if $\eta=$ const, and (ii) the solution is unique and is that given by Eqs. (2.19)–(2.22).

(i) If $\eta(r,z)$ is a solution of (A1e), then there is a solution $\nu(r,z)$, unique up to a constant of integration, of (A1a) and (A1b). Write $f(r,z)\equiv 2^{1/2}me^{\nu/2}\eta$. If $B(r,z)$ is C^2 and satisfies (A1c) and (A1d), then there exists a function $\alpha(r,z)$ such that

$$B_{,r} = f\cos\alpha, \tag{A2a}$$

$$B_{,z} = f\sin\alpha, \tag{A2b}$$

if and only if

$$f^{-1}f_{,r} + \alpha_{,z} = -r^{-1}\cos^2\alpha, \tag{A3a}$$

$$f^{-1}f_{,z} - \alpha_{,r} = -r^{-1}\cos\alpha\sin\alpha. \tag{A3b}$$

Now define $\theta\equiv\ln|f|$, differentiate (A3a) with respect to r and (A3b) with respect to z, and add the resulting equations to get

$$\Delta\theta - r^{-1}(\cos 2\alpha\,\theta_{,r} + \sin 2\alpha\,\theta_{,z}) - 2r^{-2}\cos^2\alpha = 0. \tag{A4}$$

Use (A1a), (A1b), and (A1e) to express $\Delta\theta, \theta_{,r}$, and $\theta_{,z}$ in terms of $\eta, \eta_{,r}$ and $\eta_{,z}$ only. Then (A4) is a quadratic equation in r^{-1}:

$$C_1 r^{-2} + C_2 r^{-1} + C_3 = 0, \tag{A5}$$

where

$$C_1 \equiv 2\cos^2\alpha,$$

$$C_2 \equiv 2\eta^{-1}\cos\alpha(\cos\alpha\,\eta_{,r} + \sin\alpha\,\eta_{,z}),$$

$$C_3 \equiv \eta^{-2}[16\pi(\cos\alpha\,\eta_{,r} + \sin\alpha\,\eta_{,z})^2 + |\nabla\eta|^2].$$

Thus

$C_2 - 4C_1C_3 = 4\eta^{-2}[(\vec{n}\cdot\nabla\eta)^2 - 2|\nabla\eta|^2 - 32\pi\eta^2(\vec{n}\cdot\nabla\eta)^2]$, where $\vec{n}\equiv(\cos\alpha, \sin\alpha)$. Since $|\vec{n}|^2 = 1$, it follows that $(\vec{n}\cdot\nabla\eta)^2 - 2|\nabla\eta|^2 < 0$ for $\nabla\eta\neq 0$. Thus if $|\nabla\eta|\neq 0$, $C_1 - 4C_2C_3 < 0$, so r^{-1} must be complex. Hence, in order to get solutions depending on real values of r, it must be that $|\nabla\eta| = 0$, i.e., $\eta=$ const.

(ii) Clearly if $\eta=$ const, then $\nu=$ const. Thus the solution of Eqs. (A1) reduces to finding a C^2 function $B(r,z)$ such that $\nabla^2 B = 0$ and $|\nabla B|^2 = p^2 = 2m^2e^{\nu}\eta^2 =$ const. Now the Eqs. (A3a) and (A3b) become, respectively,

$$\alpha_{,z} = -r^{-1}\cos^2\alpha,$$

$$\alpha_{,r} = r^{-1}\cos\alpha\sin\alpha.$$

The only C^1 functions $\alpha(r,z)$ satisfying these last two equations are constant functions $\alpha = \pm\pi/2, \pm 3\pi/2, \ldots$. Hence, by (A2), $B_{,r} = 0$ and $B_{,z} = \pm p$. In conclusion, $B = \pm pz + B_0$, where B_0 is an arbitrary constant of integration, $\eta=$ const, and $\nu=$ const is the unique class of solutions of Eqs. (A1).

[1] A. Das and C. V. Coffman, J. Math. Phys. 8, 1720 (1967). The value of the fine-structure constant derived here is about two orders of magnitude larger than the experimentally accepted value.
[2] R. Jackiw, Rev. Mod. Phys. 49, 681 (1977).
[3] D. Kramer, G. Neugebauer, and H. Stephani, Fortschr. Phys. 20, 1 (1972).
[4] S. D. Majumdar, Phys. Rev. 72, 390 (1947); A. Papapetrou, Proc. R. Ir. Acad. Sect. A 51, 191 (1947).
[5] A. Das, Proc. R. Soc. London Ser. A 267, 1 (1962).
[6] A. Das, J. Math. Phys. 4, 45 (1963).
[7] J. L. Synge, *Relativity: The General Theory* (North-Holland, Amsterdam, 1960).
[8] M. M. Som and A. K. Raychaudhuri [Proc. R. Soc. London Ser A 304, 81 (1968)], have found a solution of the Einstein–Maxwell equations with a charged dust source in which the metric, electromagnetic field, and charge/mass ratio are of the same form as that of the solution (2.19)–(2.21) with $a_0 = 0$. This correspondence between EMKG and charged dust solutions is not surprising in view of the fact that in the case where $\eta = (\psi^*\psi)^{1/2} =$ const, the EMKG equations reduce to a set of equations which formally resemble the Einstein–Maxwell charged dust equations. The details of this are in J. D. Gegenberg, *On the Stationary Einstein–Maxwell–Klein–Gordon Equations*, Ph.D thesis, Simon Fraser University, Burnaby, B.C., Canada.
[9] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Spacetime* (Cambridge U. P., Cambridge, England, 1973).
[10] S. Kloster, M. M. Som, and A. Das, J. Math. Phys. 15, 1096 (1974); W. Israel and G. A. Wilson, J. Math. Phys. 13, 865 (1972).
[11] S. Kloster and A. Das, J. Math. Phys. 18, 586 (1977).
[12] This expression for the matter field ψ follows from $\mathscr{L}_{\theta}\psi = iL\psi$, where θ' is the periodic Killing vector field.
[13] T. Lewis, Proc. R. Soc. London Ser. A 136, 176 (1932); A. Papapetrou, Ann. Physik 12, 309 (1953).
[14] A. K. Raychaudhuri and S. N. Guha Thakurta, Phys. Rev. D 22, 802 (1980).
[15] R. Geroch, "The Global Properties of Spacetime," in *General Relativity: An Einstein Centenary Volume*, edited by S. W. Hawking and W. Israel (Cambridge U. P., Cambridge, England, 1979).
[16] A. Krasínski, Ann. Phys. (N.Y.) 112, 22 (1978).

On the dynamics of infinite anharmonic systems

C. Marchioro [a)]
Dipartimento di Matematica, libera Università di Trento, Trento, Italy
A. Pellegrinotti [a)]
Istituto Matematico, Università di Camerino, Camerino (MC), Italy
M. Pulvirenti [a)]
Istituto Matematico, Università di Roma, Roma, Italy

(Received 25 February 1980; accepted for publication 6 June 1980)

The nonequilibrium dynamics of a large class of classical systems of anharmonic oscillators is studied. An existence theorem for the solution of the hierarchical equations describing the evolution of the states is given.

PACS numbers: 05.20. – y

1. INTRODUCTION

One of the goals in studying the nonequilibrium statistical mechanics of a physical system is to investigate their dynamical properties when the size of the system becomes very large. A convenient mathematical way to state the problem is to study the time evolution of infinitely extended systems. (For a deeper discussion on this point see Refs. 1 and 2 for more recent results).

In this paper we shall study the time evolution of an infinite system of anharmonic oscillators interacting via a two body polynomial potential with range one. An additional one-body restoring force that is a polynomial of degree greater or equal to that describing the interaction is also assumed. (The methods used in this paper can be easily generalized to more general potential. We avoid such trivial generalization for the sake of clarity.)

The dynamical problem may be stated in the following way. One has to find a subset of the phase space of the infinitely extended system, in which it is possible to define a one parameter group of transformations satisfying the equations of motion. Such a subset of the phase space has to be chosen as large as possible, to be the support of interesting probability measures (physical states) in such a way that the time evolved expectations of suitable functions on the phase space (physical observables), make sense. On the contrary, the subset in which the motion takes place cannot be too large because there are initial conditions giving solutions that become singular in a finite time. This, of course, is a feature occurring also in the case of continuous systems.[1,2]

This program may be performed in the considered case, under an additional condition linking the dimension d of the space with the degree $2k$ of the interaction. More precisely an k is allowed for $d = 1,2$; if $d = 3$ then k has to be at most 2, so covering only the case of the first nonharmonic term in the perturbative development of the potential.[3] This will be done in Sec. 4. In previous papers the same problem for the harmonic case $k = 1$ has been treated.[4] The case in which the restoring one-body force dominates the interaction was also treated in Ref. 5 (in the polynomial case, for example, one has to assume that the first term has double the degree of the second one). The method used in Sec. 4 is essentially based on the energy conservation law.

The main interest of pointwise dynamics is that it induces an evolution of the states and this is the aspect relevant from a physical point of view. Moreover, in considering the time evolution of the states, one expects that it has to be governed by evolution equations of hierarchical type, as the BBGKY equations in the continuous case. These equations have been studied in a strong form in Ref. 6 in the case of the model considered in Ref. 5. In this paper we study such equations in a weak form (this means that we consider an evolution equation for the expectation values and not directly for the measures). In Sec. 3 we give an approach to obtain a theorem of existence of such equations without making use of the pointwise dynamics. We observe that the evolution of the states is easily obtained by the pointwise dynamics when it exists. Since the evolution of states is, after all, the relevant feature of dynamics from the physical point of view, we think that it is interesting to obtain it without any reference to the pointwise dynamics. We remark finally that the two methods used in Sec. 3 and 4, although different, hold under the same limitations on the degree of the potentials and the dimension of lattice. The class of initial states we are considering is large enough to contain interesting states from a physical point of view that are singular w.r.t. any equilibrium state, e.g., states that are products of equilibrium states at different temperatures in different regions.

For a particular class of states that is absolutely continuous w.r.t. some equilibrium state the time evolution can be derived by means of nonconstructive arguments.[5]

The basic idea used in the proof is the conservation law for a Gibbs state "comparable" with the nonequilibrium state we are considering.

2. NOTATIONS AND STATEMENT OF THE PROBLEM

We consider a system of anharmonic oscillators in a lattice $\mathbb{Z}^d$. The phase space of the system is $\mathscr{X} = (\mathbb{R}^2)^{\mathbb{Z}^d}$,

[a)]Research partially supported by Consiglio Nazionale delle Ricerche, G.N.F.M.

 © 1981 American Institute of Physics

whose points are denoted by X, Y, where $X=\{x_i\}_{i\in\mathbb{Z}^d}$, $x_i=(q_i,p_i)$ and q_i and p_i are the position and momentum of the oscillator at the point i of the lattice. For any $\Lambda\subset\mathbb{Z}^d$, Λ bounded, we denote by $\mathscr{X}(\Lambda)$ the phase space associated to the region Λ. Explicitly, $\mathscr{X}(\Lambda)=(\mathbb{R}^2)^\Lambda$. The points of $\mathscr{X}(\Lambda)$ are denoted by X_Λ, Y_Λ etc., where $X_\Lambda=\{x_i\}_{i\in\Lambda}$. The oscillators interact via a family of Hamiltonians $\{H_\Lambda\}$; $H_\Lambda:\mathscr{X}(\Lambda)\to\mathbb{R}$ defined by

$$H_\Lambda(X)=\sum_{i\in\Lambda}\left\{\frac{p_i^2}{2}+\left(\sum_{j\in U_i\cap\Lambda}P(q_i-q_j)\right)+Q(q_i)\right\},\tag{2.1}$$

where Q and P are sums of even positive monomials with maximum degree respectively $2k$ and $\leqslant 2k$. The mass of oscillators is chosen to be one and

$$U_i=\{j\in\mathbb{Z}^d\,|\,|i-j|\equiv\Sigma_{\alpha=1}^d|i_\alpha-j_\alpha|=1\}.$$

A state ρ of the system is a Borel probability measure on $\mathscr{X}$ considered as a topological space w.r.t. the product topology. Alternatively a state may be described in terms of its joint distributions on $\mathscr{X}(\Lambda)$, i.e., a compatible family $\{\rho_\Lambda\}_{\Lambda\subset\mathbb{Z}^d}$ (Λ bounded) of Borel probability measures on $\{\mathscr{X}(\Lambda)\}$.

We shall consider the state described by a family of joint distributions on $\mathscr{X}(\Lambda)$ with the following properties:

$\rho_\Lambda(dx_\Lambda)$ are absolutely continuous w.r.t. the Lebesgue measure dx_Λ on $\mathscr{X}(\Lambda)$ with derivatives

$$\rho_\Lambda(x_\Lambda).\tag{2.2}$$

There exist two constants a, $b>0$ such that the following estimate holds. For any bounded $\Lambda\subset\mathbb{Z}^d$ we have the result

$$\rho_\Lambda(x_\Lambda)\leqslant e^{b|\Lambda|}e^{-a\sum_{i\in\Lambda}(p_i^2+q_i^{2k})},\tag{2.3}$$

where $|\Lambda|$ denotes the number of points in Λ.

The only consequence of (2.3) that we shall use in the sequel is the following condition. Let $f:\mathscr{X}(\Lambda)\to\mathbb{R}^+$ be any bounded measurable positive function. Clearly f may be thought of as a function $\mathscr{X}(\Omega)\to\mathbb{R}$ $\Omega\supset\Lambda$, still denoted by f, by putting $f(x_\Omega)=f(x_{\Omega|\Lambda})$ where $x_{\Omega|\Lambda}$ denotes the restriction of x_Ω to Λ. Putting

$$\omega_\Omega^\beta(f)=Z_\Omega^{-1}\textstyle\int\exp[-\beta H_\Omega(x_\Omega)]f(x_\Omega)\,dx_\Omega,$$

$Z_\Omega=\int\exp[-\beta H_\Omega(x_\Omega)]dx_\Omega$, the Gibbs measure at temperature β^{-1} associated to the region Ω, with zero boundary conditions (2.3) implies the existence of some β for which

$$\begin{aligned}\rho(f)&=\int\rho_\Omega(x_\Omega)f(x_\Omega)\,dx_\Omega\\&\leqslant\int e^{b'|\Omega|}e^{-\beta H_\Omega(x_\Omega)}f(x_\Omega)\,dx_\Omega\\&\leqslant e^{c|\Omega|}\omega_\Omega^\beta(f),\end{aligned}\tag{2.4}$$

where $\Omega\supset\Lambda$ and b' and c are suitable constants. A useful estimate that we shall use in the sequel is the following (see Ref. 8): there exist positive A and B such that

$$\omega_\Omega^\beta(x_\Omega)\leqslant e^{B|\Lambda|}\exp\left[-A\sum_{i\in\Lambda}(p_i^2+q_i^{2k})\right]\tag{2.5}$$

for all $\Omega\supset\Lambda$. Here $\omega_\Omega^\beta(x_\Omega)$ denotes the density of the joint distribution of the phase variables in Λ. In particular (2.5) ensures the existence of the infinite volume Gibbs state.

We want to briefly discuss our hypothesis on the $\{\rho_\Lambda\}$. Condition (2.2) is not particularly deep.

Condition (2.3) means that the joint distributions have to satisfy a superstable estimate like the Gibbs measure generated by H [i.e., inequality (2.5)]. Following Ref. 8 we may conclude that inequality (2.3) is satisfied if ρ is a Gibbs state generated by a family of Hamiltonians $\{h_\Lambda\}$ sufficiently well behaved with respect to $p_i^2+q_i^{2k}$. We do not make more precise conditions on $\{h_\Lambda\}$ since we are more interested in the states than in the Hamiltonians generating them via the Gibbs prescription. For some results on the evolution of states generated by Hamiltonians in the case treated in Ref. 5 (see Ref. 9).

Let us put $\Lambda_n=[-n,n]^d$. The map $x_{\Lambda_n}\to x_{\Lambda_n}^n(t)$, $t\in\mathbb{R}$ of $\mathscr{X}(\Lambda_n)\to\mathscr{X}(\Lambda_n)$ is defined as a solution of the Newton's law of the motion given by the Hamiltonian H_{Λ_n}. This map induces a new map $x\to x^n(t)$ of $\mathscr{X}\to\mathscr{X}$ defining $x^n(t)=x_{\Lambda_n}^n(t)\cup x_{\Lambda_n^c}$, that with an obvious meaning of the symbols means that we study the evolution of the oscillators in Λ_n under the action of H_{Λ_n} and take the others fixed. The Liouville theorem and the invariance of H_{Λ_n}, imply that the measure $\omega_{\Lambda_n}^\beta$ is invariant under the flow $x_{\Lambda_n}\to x_{\Lambda_n}^n(t)$.

Defining

$$\rho_t^n(x_{\Lambda_n})=\rho_{\Lambda_n}(x_{\Lambda_n}^n(-t))$$

and for

$$j\leqslant n\quad\rho_{j,t}^n(x_{\Lambda_j})=\int\rho_{\Lambda_n}(x_{\Lambda_n}^n(-t))\,dx_{\Lambda_n-\Lambda_j}^n,\tag{2.6}$$

we have for $j<n$,

$$\frac{d}{dt}(\rho_{j,t}^n,f_j)=-(\rho_{j,t}^n,\mathscr{L}_jf_j)-(\rho_{j+1,t}^n,C_{j+1,j}f_j),\tag{2.7}$$

where $\{f_j\}_{j\in\mathbb{N}}$ is any sequence of functions with the following properties: $f_j\in C^1[\mathscr{X}(\Lambda_j)]$, f_j is bounded, and putting

$$\nabla f_j\equiv\sum_{i\in\Lambda_j}\left(\left|\frac{\partial f_j}{\partial q_i}\right|+\left|\frac{\partial f_j}{\partial p_i}\right|\right)\tag{2.8}$$

one has $\|\nabla f_j\|_\infty<\infty$.

The operator $\mathscr{L}_j$, the Liouville operator associated to the Hamiltonian H_Λ, is defined as

$$(\mathscr{L}_jf_j)(x_{\Lambda_j})=\{f_j,H_{\Lambda_j}\}(x_{\Lambda_j}),\tag{2.9}$$

where

$$\{f_j,g_j\}=\sum_{i\in\Lambda_j}\left(\frac{\partial f_j}{\partial q_i}\frac{\partial g_j}{\partial p_i}-\frac{\partial f_j}{\partial p_i}\frac{\partial g_j}{\partial q_i}\right)$$

and

$$C_{j+1,j}:C'[\mathscr{X}(\Lambda_j)]\to C[(\mathscr{X}(\Lambda_{j+1})]$$

is defined as

$$(C_{j+1,j}f_j)(x_{\Lambda_{j+1}})=-\sum_{i\in\Lambda_j}\frac{\partial f_j}{\partial p_i}\sum_{k\in\Lambda_{j+1}-\Lambda_j}\frac{\partial P}{\partial q_i}(q_i-q_k),\tag{2.10}$$

and finally

$$(f_j,g_j)\equiv\int dx_{\Lambda_j}f_j(x_{\Lambda_j})\,g_j(x_{\Lambda_j})\tag{2.11}$$

for all couples of f_j and g_j for which the above integral makes sense.

The Eq. (2.7) has a very natural limiting equation:

$$\frac{d}{dt}\rho_t(f_j) = -\rho_t(\mathscr{L}_j f_j) - \rho_t(C_{j+1,j} f_j) \tag{2.12}$$

so that the problem we state is to find a one parameter group of maps $\rho \to \rho_t$ on the Borel probability measures on $\mathscr{X}$ such that (2.12) hold. Moreover we would give a control on the deviation of the solution of (2.12) from the solutions $\rho_t^n(f)$ of truncated Eq. (2.7) in terms of f_j and n. This problem will be approached in Sec. 3.

3. THE HIERARCHICAL EQUATIONS

In this section we investigate the existence and the propeties of the evolved measure ρ_t under the action of the infinite dynamics. In particular we start by proving $L_p(\mathscr{X},\rho)$–convergence of the sequence $\{q_i^n(t)\}$.

Let $x \in \mathscr{X}$ and $(q_i^n(t), p_i^n(t))$ be the coordinate and momentum of the i-oscillator in the configuration $x^n(t)$, with initial condition $x \equiv (q_j, p_j)_{j\in\mathbb{Z}^d}$.

Then $(n>1)$

$$q_i^n(t) = q_i + \theta_{i,n}(p_i t - \int_0^t ds \int_0^s dr \times \Big[\sum_{j\in U_i\cap\Lambda_n} [P'(q_i^n(r) - q_j^n(r))] + Q'(q_1^n(r)) \Big], \tag{3.1}$$

$$p_i^n(t) = q_i^n(t) + (1-\theta_{i,n})p_i,$$

where

$$\theta_{i,n} = \begin{cases} 1 & \text{if } i\in\Lambda_n, \\ 0 & \text{otherwise.} \end{cases}$$

Then putting

$$\delta_i^n(t) = |q_i^n(t) - q_i^{n-1}(t)| \tag{3.2}$$

if $\{a_\alpha\}$ and $\{b_\alpha\}$ are such that $P'(y) = \sum_{\alpha=1}^{2k-1} a_\alpha y^\alpha$ and $Q'(y) = \sum_{\alpha=1}^{2k-1} b_\alpha y^\alpha$ and $i\in\Lambda_j\subset\Lambda_n$ we have

$$\delta_i^n(t) \leqslant \int_0^t dt_1 \int_0^{t_1} dt_2 \Big\{ \sum_{l\in U_i\cap\Lambda_n} \sum_{\alpha=1}^{2k-1} a_\alpha \Big[\sum_{\beta<\alpha} |q_i^n(t_2) - q_l^n(t_2)|^{\alpha-\beta-1} |q_i^{n-1}(t_2) - q_l^{n-1}(t_2)|^\beta \Big]$$
$$\times [\delta_i^n(t_2) + \delta_l^n(t_2)] + \sum_{\alpha=1}^{2k-1} b_\alpha \Big[\sum_{\beta<\alpha} |q_i^n(t_2)|^{\alpha-\beta-1} |q_i^{n-1}(t_2)|^\beta \Big] \delta_i^n(t_2) \tag{3.3}$$
$$\leqslant \int_0^t dt_1 \int_0^{t_2} dt_2 \sum_{l\in V_i} D_{i,l}^n(t_2)\, \delta_l^n(t_2),$$

where

$$V_i = U_i \cup \{i\}$$

and

$$D_{i,l}^n(t) = \begin{cases} A_{i,l}^n(t) & \text{if } i\neq l, \\ A_{i,i}^n(t) + B_i^n(t) & \text{if } i = l, \end{cases} \tag{3.4}$$

where

$$B_i^n(t) = \sum_{\alpha=1}^{2k-1} b_\alpha \sum_{\beta<\alpha} |q_i^n(t)|^{\alpha-\beta-1} |q_i^{n-1}(t)|^\beta,$$

$$A_{i,l}^n(t) = \begin{cases} \sum_{\alpha=1}^{2k-1} a_\alpha \Big(\sum_{\beta<\alpha} |q_i^n(t_1) - q_l^n(t_1)|^{\alpha-\beta-1} |q_i^{n-1}(t_1) - q_l^{n-1}(t_1)|^\beta \Big) & \text{if } i\neq l, |i-l| = 1, \\ 0 & \text{if } i\neq l,\ |i-l|>1, \end{cases} \tag{3.5}$$

$$A_{i,i}^n(t) = \sum_{l\in U_i} A_{i,l}^n(t).$$

Then (3.3) may be iterated at least $n(i) = |n-j|$ times to obtain:

$$\delta_i^n(t) \leqslant \int_0^t \int_0^{t_1} \cdots \int_0^{t_{2n(i)-1}} dt_1 \cdots dt_{2n(i)} \times \sum_{l_n,\ldots,l_{n(i)}} D_{i,l_1}^n(t_2) \cdots D_{l_{n(i)-1},l_{n(i)}}^n(t_{2n(i)})\, \delta_{l_{n(i)}}^n(t_{2n(i)}). \tag{3.6}$$

The main point of this section is the following estimate. Let $O_h(y)$ be any function such that $\lim_{y\to 0} O_h(y)/y = 0$, $h\in\mathbb{Z}^+$. Then

Proposition 3.1: Let $d(1-1/k) < 2, p \geqslant 1$ and ρ^0 be a state satisfying (2.2) and (2.3). Then for all $i\in\Lambda_j$ there exists a $\gamma>0$ (not depending on j) such that

$$\rho(|\delta_i^n(t)|^p) \leqslant O_1(n^{-\gamma n}) t^{2|n-j|}. \tag{3.7}$$

Proof: Let us first put $p = 1$.
Defining

$$\bar{q}_i^l(\tau) = |q_i^l(\tau)| + 1 \quad \tau\in\mathbb{R} \tag{3.8}$$

our first aim to find a bound for $l_i = n$ or $n-1$ and $j_i \in \Lambda_n$:

$$\rho(\bar{q}_{j_1}^{l_1}(\tau_1) \cdots \bar{q}_{j_m}^{l_m}(\tau_m)) \leqslant C_4(k,n), \tag{3.9}$$

where $C_4(k,n)$ depends on k and n, and $m = 2n(k-1)+1$.

In fact with (3.9), using (3.4) and (3.5) we obtain

$$\sum_{j_1 \cdots j_s} \rho(D^n_{i j_1}(t_2) \cdots D^n_{j_s j_s}(t_{2s})\, \delta^n_{j_s}(t_{2s})$$

$$\leqslant 2C_5^s N(d)^{2s}(2k-1)^{2s} 2^{(2k-1)s} C_4(k,n), \tag{3.10}$$

and hence from (3.6)

$$\rho(q_i^n)(t)) \leqslant \frac{t^{2n(i)}}{[2n(i)]!} C_5^{n(i)} N(d)^{2n(i)} C_4(k,n), \tag{3.10$'$}$$

where C_5 and C_6 are constants depending only on the interaction and $N(d)$ is the number of the first neighborhoods at any point of $\mathbb{Z}^d$. We now determine $C_4(k,n)$.

Let $\alpha < 2$ and $X_n : \mathscr{X} \to \mathbb{R}$ be the characteristic function of the set $\{X \in \mathscr{X} | \bar{q}^{l_1}_{j_1}(\tau_1) \cdots \bar{q}^{l_m}_{j_m}(\tau_m) > n^{\alpha n}\}$. Then

$$\rho(\bar{q}^{l_1}_{j_1}(\tau_1) \cdots \bar{q}^{l_m}_{j_m}(\tau_m)) \leqslant n^{\alpha n} + \rho\left[X_n\, \bar{q}^{l_1}_{j_1} \cdots \bar{q}^{l_m}_{j_m}(\tau_m)\right]. \tag{3.11}$$

By virtue of (2.4) and Hölder inequality and time invariance of $\omega^\beta_{\Lambda_n}$,

$$\rho\left[X_n\, \bar{q}^{l_1}_{j_1}(\tau_1) \cdots \bar{q}^{l_m}_{j_m}(\tau_m)\right] \leqslant e^{cn^d} \omega^\beta_{\Lambda_n}(\bar{q}^{p_n}_{j_1})^{1/p_n} \cdots \omega^\beta_{\Lambda_n}(\bar{q}^{p_n}_{j_m})^{1/p_n} \omega^\beta_{\Lambda_n}(X_n)^{1/q_n}, \tag{3.12}$$

$$p_n = \frac{q_n m}{(q_n - 1)}. \tag{3.13}$$

By the superstable estimate (2.5) there exists a constant C_7 such that

$$\omega^\beta_{\Lambda_n}(\bar{q}^{p_n}_j) \leqslant C_7^{p_n/2K} \left(\frac{p_n}{2k}\right)! \tag{3.14}$$

and hence

$$\text{The lhs of (3.12)} \leqslant e^{cn^d} C_7^{m/2K} \left(\frac{p_n}{2k}\right)^{m/2k} \omega^\beta_{\Lambda_n}(X_n)^{1/q_n}. \tag{3.15}$$

On the other hand, let $\bar{\beta} > 0$; then

$$\omega^\beta_{\Lambda_n}(X_n) = \omega^\beta_{\Lambda_n}\left(\left\{X \in \mathscr{X} \middle| \frac{\bar{\beta}}{m} \sum_{r=1}^m \ln \bar{q}^{l_r}_{j_r}(\tau_r) > \frac{\alpha n \bar{\beta}}{m} \ln n\right\}\right) \leqslant e^{-n^{\bar\beta}} \frac{1}{m} \sum_{r=1}^m \omega^\beta_{\Lambda_n} \times \left[\exp\left(e^{(\bar{\beta} m/\alpha n) \ln \bar{q}^{l_r}_{j_r}(\tau_r)}\right)\right], \tag{3.16}$$

where the last step is obtained by the convexity inequality. By the time invariance of the $\omega^\beta_{\Lambda_n}$ and the estimate (2.5) we obtain the existence of a constant C_8 such that

$$\omega^\beta_{\Lambda_n}(\exp[\bar{q}^{\bar{\beta} m/\alpha n}]) \leqslant C_8, \tag{3.17}$$

if $\bar{\beta} m/\alpha n < 2k$ i.e. $\beta < 2k\alpha/[2(k-1) + 1/n]$, (3.18)

and finally

$$\omega^\beta_{\Lambda_n}(X_n) < C_8 e^{-n^{\bar\beta}}. \tag{3.19}$$

Choosing $q_n = n^\epsilon$ with $\epsilon > 0$, we obtain

$$p_n = \frac{n^\epsilon}{n^\epsilon - 1} m \quad\text{and}\quad \frac{m}{2k} \leqslant \bar{\alpha} n, \quad \frac{p_n}{2k} \leqslant \tilde{\alpha} n,$$

with $\bar{\alpha} < 1$ and $\tilde{\alpha} < 1$ for sufficiently large n. Combining (3.19) and (3.15) we obtain

$$\text{lhs of (3.12)} \leqslant C_7^{\bar{\alpha} n} C_8^{1/n^\epsilon} (\tilde{\alpha} n)^{\tilde{\alpha} n} \exp[cn^d - n^{\bar{\beta} - \epsilon}]. \tag{3.20}$$

So that, if $d(1 - 1/K) < 2$, choosing a sufficiently small ϵ and large $\bar{\beta}$, (3.9) is proved with

$$C_4(K, n) = n^{\alpha n} + O_1(e^{-n}). \tag{3.21}$$

By (3.10)$'$ we have the proposition with $p = 1$. Now let p be an integer larger than one. Then

$$\rho(|q_i^n(t) - q_i^{n-1}(t)|^p) \leqslant \sum_{r=1}^{p-1} e_r\, \rho(|\bar{q}_i^n(t)|^r |\bar{q}_i^{n-1}(t)|^{p-r-1} \delta_i^n(t)) \tag{3.22}$$

with some positive coefficients e_r.

Let $\zeta > 1$ then

$$\rho(|\bar{q}_i^n(t)|^r |q_i^{n-1}(t)|^{p-r-1} \delta_i^n(t)) \leqslant \zeta^n \rho(\delta_i^n(t)) + \rho(\bar{X}_n |\bar{q}_i^n(t)|^r |\bar{q}_i^{n-1}(t)|^{p-r-1} \delta_i^n(t)), \tag{3.23}$$

where $\bar{X}_n$ is the characteristic function of the set $\{x \in \mathscr{X} | \bar{q}_i^n(t)|^r |\bar{q}_i^{n-1}(t)|^{p-r-1} > \zeta^n\}$. But

$$\rho(\bar{X}_n |\bar{q}_i^n(t)|^r |\bar{q}_i^{n-1}(t)|^{p-r-1}) \leqslant e^{cn^d} \omega^\beta_{\Lambda_n}(\bar{X}_n)^{1/2} \omega^\beta_{\Lambda_n}(|\bar{q}_i^n(t)|^{2r} |\bar{q}_i^{n-1}(t)|^{2(p-r-1)})^{1/2}, \tag{3.24}$$

where the last term in rhs of (3.24) is uniformly bounded by a constant C_9 because of Swartz's inequality, time invariance of $\omega^\beta_{\Lambda_n}$, and estimate (2.5).

Furthermore, if $\gamma > 0$ is chosen such that $\gamma p < 2k$ then

$$\omega^\beta_{\Lambda_n}(\bar{X}_n) = \omega^\beta_{\Lambda_n}\left(\left\{X \in \mathscr{X} \middle| \gamma\left(\frac{\lg \bar{q}_i^n(t)^r + \ln \bar{q}_i^{n-1}(t)^{p-r-1}}{2}\right) > r\frac{n}{2} \ln \zeta\right\}\right) \leqslant e^{-\zeta^{\gamma n/2}} \left[\tfrac{1}{2} \omega^\beta_{\Lambda_n}(e^{\bar{q}_i^n(t)^{\gamma r}}) + \tfrac{1}{2} \omega^\beta_{\Lambda_n}(e^{\bar{q}_i^n(t)^{\gamma(p-r-1)}})\right]. \tag{3.25}$$

Hence proceeding as above there exists constants C_{10} such that

$$\text{rhs of (3.23)} \leqslant \zeta^n \rho[\delta_i^n(t)]\, e^{cn^d} C_9 C_{10} e^{-\frac{1}{2}(\zeta^{\gamma n/2})}. \tag{3.26}$$

Finally, combining (3.26) with (3.22) we conclude the proof of Proposition 3.1.

Theorem 3.2: Let $d(1 - 1/k) < 2$, $\mathscr{f}_j \in \{g_j \in C^1[\mathscr{X}(\Lambda_j)] | \|\nabla g_j\| < \infty\}$ and ρ a state satisfying (2.2) and (2.3). Then there exists $\gamma > 0$ such that:

$$\text{(i)}\ |\rho_t^n(\mathscr{f}_j) - \rho_t^{n-1}(\mathscr{f}_j)| \leqslant \|\nabla \mathscr{f}_j\|_\infty (2j+1)^d |t|^{2n+1} O_3(n^{-\gamma n}).$$

Moreover there exists a limiting measure ρ_t such that $\rho_t(\mathscr{f}_j)[= \lim_{n\to\infty} \rho_t^n(\mathscr{f}_j)]$ is differentiable and satisfies:

$$\text{(ii)}\ \frac{d\rho_t(\mathscr{f}_j)}{dt} = -\rho_t(\mathscr{L}_j \mathscr{f}_j) - \rho_t(C_{j+1,j} \mathscr{f}_j).$$

Proof: Let $r \geqslant 1$. Then for fixed $t \in \mathbb{R}$ we have

$$\int \rho(dx)\, \left| [q_i^n(t) - q_j^n(t)]^r - [q_i^m(t) - q_j^m(t)]^r \right|$$

$$\leqslant r\int \rho\,(dx)\,\{[\,|q_i^n(t)-q_j^n(t)|^{r-1} + |q_i^m(t)-q_j^m(t)|^{r-1}] \times [\,|(q_i^n(t)-q_j^n(t))-(q_i^m(t)-q_j^m(t))|\,]\}$$

$$\leqslant r\Big[\Big(\int \rho\,(dx)\,|q_i^n(t)-q_i^m(t)|^r\Big)^{1/r} + \Big(\int \rho\,(dx)\,|q_j^m(t)-q_j^n(t)|^r\Big)^{1/r}\Big] \times\Big\{\Big(\int \rho\,(dx)\,|q_i^n(t)-q_j^n(t)|^r\Big)^{(r-1)/r} + \Big(\int \rho\,(dx)\,|q_i^m(t)-q_j^m(t)|^r\Big)^{r-1/r}\Big\}. \tag{3.27}$$

So the rhs of (3.27) goes to 0 as $n\to\infty$ and $m>n$ in virtue of Proposition 3.1.

In particular, putting

$$F_i(x) = -\sum_{j\in U_i} P'(q_i-q_j) - Q'(q_i) \tag{3.28}$$

the force acting on the i-oscillator, the above argument proves

$$\rho\,[\,|F_i(x^n(t))-F_i(x^{n-1}(t))|\,] \leqslant |(t)|^{2n+1}O_2(n^{-\gamma n}) \tag{3.29}$$

and hence

$$\rho[\,|\bar\delta_i^n(t)|\,] \leqslant t^{2(n+1)}O_2(n^{-\gamma n}), \tag{3.30}$$

where $\bar\delta_i^n(t) = |p_i^n(t)-p_i^{n-1}(t)|$. Finally,

$$\int \rho\,(dx)\,|f_j(x^n(t)) - f_j((x^{n-1}(t))| \leqslant \|\nabla f_j\|_\infty \sum_{i\in\Lambda_j}[\,\rho(\delta_i^n(t))] + \rho[(\bar\delta_i^n(t))] \leqslant \|\nabla f_j\|_\infty (2j+1)^d t^{2(n+1)}O_3(n^{-\gamma n}). \tag{3.31}$$

That proves (i).

The measure ρ_t may be built by observing that since the $L_1(\mathscr{X},\rho)$-convergence of $p_i^n(t)$, $q_i^n(t)$, $F_i(x^n(t))$ is integrable it also implies the ρ-almost everywhere convergence and induces a ρ–almost everywhere defined map $x\to x(t)$ and hence an evolved measure.

Let us prove part (ii). We must prove that

$$|(\rho_{j,t}^n, \mathscr{L}_j f_j) - (\rho_{j,t}^m, \mathscr{L}_j f_j)| \underset{n\to\infty}{\overset{m>n}{\longrightarrow}} 0, \tag{3.32}$$

$$|(\rho_{j+1,t}^n, C_{j+1,j} f_j) - (\rho_{j,t}^m, C_{j+1,j} f_j)| \underset{n\to\infty}{\overset{m>n}{\longrightarrow}} 0. \tag{3.33}$$

We prove (3.33).

$$|(\rho_{j+1,t}^n, C_{j+1,j} f_j) - (\rho_{j+1,t}^m, C_{j+1,j} f_j)| \leqslant \sum_{i\in\Lambda_j}\sum_{k\in\Lambda_{j+1}\setminus\Lambda_j}\Big\{\int d\rho\,\Big|\Big[\frac{\partial f_j}{\partial p_i}\frac{\partial P}{\partial q_i}(q_i-q_k)\Big] \times[(x^n(t))] - \Big[\frac{\partial f_j}{\partial p_i}\frac{\partial P}{\partial q_i}(q_i-q_k)\Big][x^m(t)]\Big|\Big\}.$$

Now we have

$$\int d\rho\,\Big|\frac{\partial f_j}{\partial p_i}[x^n(t)]\frac{\partial P}{\partial q_i}[q_i^n(t)-q_k^n(t)] - \frac{\partial f_i}{\partial p_i}(x^m(t))\frac{\partial P}{\partial q_i}[q_i^m(t)-q_k^m(t)]\Big|$$

$$\leqslant \int d\rho\,\Big|\frac{\partial f_j}{\partial p_i}(x^n(t))\Big|\,\Big|\frac{\partial P}{\partial q_i}[q_i^n(t)-q_k^n(t)] - \frac{\partial P}{\partial q_i}[q_i^m(t)-q_k^m(t)]\Big| + \int d\rho\,\Big|\frac{\partial P}{\partial q_i}q_i^m(t)-q_k^m(t)\Big|\,\Big|\frac{\partial f_j}{\partial p_i}(x^n(t)) - \frac{\partial f_j}{\partial p_i}[x^m(t)]\Big|.$$

Using Theorem 3.2 we have the result.

Comment on the proof: Unfortunately we are not able to prove that an evolved state still satisfy the inequality (2.3) (of course with time depending coefficients). It seems that the conservation in time of superstable estimate may be possible only when a linear estimate on a global quantity holds as in Ref. 5. The estimates given in Ref. 9 for the model porposed in Ref. 5 allows us to prove that the time evolved state still satisfies the superstable estimate.

4. REMARKS ON POINTWISE DYNAMICS

In the previous section we introduced the map $x\to x(t)$ ρ-almost everywhere defined, satisfying the Newton's laws of motion. A natural problem arising is to characterize a full measure initial set for which the motion takes place. This can be done in the following way.

We define $H:\mathscr{X}\to\mathbb{R}\cup\{\infty\}$:

$$H(x) = \sup_n\Big(\frac{H_{\Lambda_n}(x_{\Lambda_n})}{|\Lambda_m|}+1\Big). \tag{4.1}$$

H is a measurable function. Putting

$$\mathscr{X}_0 = \{x\in\mathscr{X}\,|H(x) < +\infty\}, \tag{4.2}$$

we have as a consequence of inequaltiy (2.3) an Chebyshev inequality,

$$\rho(\mathscr{X}_0) = 1. \tag{4.3}$$

We sketch the proof that $\mathscr{X}_0$ is a good set of initial conditions and that $H[x(t)]$ is finite for all $t\in\mathbb{R}$ if $x\in\mathscr{X}_0$ and then $x(t)\in\mathscr{X}_0$. From (3.6) we easily get $(i\in\Lambda_j)$

$$\delta_i^n(t) \leqslant \frac{t^{2|n-j|}}{(2(n-j))!}C(k,d)^{(n-j)}[H(x)]^{(2k-2/2k)|n-j|} \times n^{d(1-1/k)(n-j)}, \tag{4.4}$$

where $C(k,d)$ depends only on k and d.

An analog of estimate (4.4) can be proved for $\bar\delta_i^n(t)$ and this is enough in order to prove the existence of the flow $\mathscr{X}_0 \ni x\to x(t)$ satisfying the motion laws if $d(1-1/k)<2$.

We now prove that $H[x(t)] < +\infty$. We have

$$H_\Omega[x_\Omega^m(t)] \leqslant H_\Omega[x_\Omega^s(t)] + \sum_{e=s}^m |H_\Omega[x_\Omega^{l+1}(t)] - H_\Omega[x_\Omega^l(t)]|, \tag{4.5}$$

where $\Omega = \Lambda_j$, $s>j$. Then for some constant C

$$|H_\Omega[x_\Omega^{l+1}(t)] - H_\Omega[x_\Omega^l(t)]| \leqslant CH(x)(l+1)^d$$

$$\times \sup_{i\in\Omega} \delta_i^{l+1}(t), \tag{4.6}$$

so that (4.5), (4.4), and the energy conservation imply

$$H_\Omega(x_\Omega^m(t)) \leqslant H_{\Lambda_s}(x) + 1, \tag{4.7}$$

$$\text{if } s = \text{Integer part of}[\gamma_1 H(x)^{\gamma_2}] + 1 + 2j, \tag{4.8}$$

where γ_1 and γ_2 are suitable constants depending on the parameters of the system. Finally

$$\frac{H_\Omega(x_\Omega^m(t))}{|\Omega|} \leqslant H(x)\frac{|\Lambda_s|}{|\Omega|} + 1 \leqslant \gamma_3 H(x)^{\gamma_4}, \tag{4.9}$$

where γ_3 and γ_4 are constants.

We conclude by remarking that the set

$$\mathscr{X}_a \equiv \left\{ x\in\mathscr{X} \,|\, \bar{H}(x) \equiv \lim_{j\to\infty} \frac{H_{\Lambda_j}(x)}{|\Lambda_j|} = a \right\} \tag{4.10}$$

is time invariant.

This means that $\bar{H}$ is a first integral of the motion. (See Ref. 10) for a discussion on the interest on this point. This may be seen by the inequality

$$H_\Omega[x_\Omega^s(t)] - \sum_{e=s}^{\infty} |H_\Omega[x_\Omega^{l+1}(t)] - H_\Omega[x_\Omega^l(t)]|$$
$$\leqslant H[x_\Omega(t)] \leqslant H_\Omega[x_\Omega^s(t)] + \sum_{e=s}^{\infty} |H_\Omega[x_\Omega^{l+1}(t)]|$$
$$- H_\Omega[x_\Omega^l(t)]|. \tag{4.11}$$

Choosing s such that $(s-j)\to\infty, s^\alpha\cdot(s-j)^{-2}\to 0$, with $\alpha > d(1-1/k)$ and $(s-j)\cdot j^{-1}\to 0$, where $s\to\infty$ and $j\to\infty$ the statement is easily proved in virtue of the above estimates.

Remarks: We observe that from the point of view of the pointwise dynamics the degreee of Q may also be chosen smaller (possibly zero) than the degree of P. In fact, the relation between the degrees of P and Q of Sec. 2 are used only in constructing the equilibrium states.

ACKNOWLEDGMENTS

We would like to thank L. Triolo for his interesting contributions to the preliminary version of this work.

[1]O. E. Lanford, III, *Lecture Notes in Physics*, No. 38, edited by J. Moser (Springer, Berlin, 1975).

[2]R. L. Dobrushin and J. Fritz, Commun. Math. Phys. **55**, 275 (1977); Commun. Math. Phys. **57**, 67 (1977).

[3]N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Holt, Rinehart and Winston, New York, 1976).

[4]O. E. Lanford, III and J. L. Lebowitz, *Lectures Notes in Physics*, No. 38 (Springer, Berlin, 1975); J. L. Van Hemmen, Thesis, University of Groningen (1976).

[5]O. E. Lanford, III, J. L. Lebowitz, and E. H. Lieb, J. Stat. Phys. **16**, 453 (1977).

[6]R. Esposito and M. Pulvirenti, J. Math. Phys. **21**, 1194 (1980).

[7]D. Ruelle, *Statistical Mechanics* (Benjamin, New York, 1969).

[8]D. Ruelle, Commun. Math. Phys. **50**, 189 (1976).

[9]C. Marchioro, A. Pellegrinotti, M. Pulvirenti, Yu. Suhov, Commun. Math. Phys. **66**, 131 (1979).

[10]R. L. Dobrushin and Yu. Suhov, Lect. Notes Phys. **80**, 325 (1978).

Pair site occupation probability for one-dimensional array of dumbbells

A. Maltz and E. E. Mola
Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas, Sección Química Teorica Sucursal 4, Casilla de Correo 16, 1900 La Plata, Argentina

(Received 11 September 1980; accepted for publication 20 February 1981)

We present two different approaches to evaluating the probability of finding in a particular state two neighboring lattice sites on a one-dimensional array of dumbbells. The results we find are exact and the same with either the first or the second procedure.

PACS numbers: 05.50. + q, 68.20. + t

I. INTRODUCTION

Many physical and chemical systems can be represented by the distribution of particles on a lattice space. Among the systems that can be conceptualized in this manner are the magnetic materials, binary alloys, the eleasticity of muscle and some textiles, the helix-random coil transition in solutions of proteins, and the adsorption of gas on a crystalline surface.

An adsorbed film formed by the adsorption on two adjacent vacant sites of the two atoms of a diatomic molecule must be regarded as a geometrically random distribution of pairs of occupied sites. Therefore, when we come to consider the kinetic aspects of adsorption of diatomic gases on metal surfaces, such as sticking probabilities or thermal desorption[1,2] we are immediately faced with a kind of problem that we are going to call pair-site-occupation probability, that is to say, the probability of finding in a particular state not one but two nearest-neighboring lattice sites. This lattice pair site can be in four different states:(a) both sites occupied by the same diatomic molecule (o-o)s; (b) both sites unoccupied u-u; (c) one site occupied by part of a diatomic molecule and the other vacant o-u; (d) both sites occupied by parts of different diatomic molecules (o-o)d [see Fig. 1(c)].

The purpose of this paper is to discuss the occupation probability aspects of a one-dimensional array of dumbbells (particles that occupy two contiguous lattice sites). There exists in the literature a large number of papers[3–8] concerned with the statistical analysis of one-dimensional systems. The rationale for treating such systems is that it is often possible to perform a relatively thorough investigation of their statistical-mechanical properties, and knowledge thus gained may be of value when considering systems of higher dimensionality.

Problems dealing with particles that occupy more than one lattice site have always been troublesome. Unlike simple particles, there is no reciprocity between particles and vacancies.[8–9]

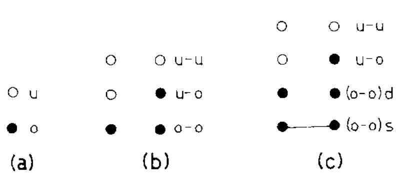


FIG. 1. Different states in which we can find a single site (a) or a pair of neighboring sites (b) and (c) in a lattice space occupied by single particles or by dumbbells.

In Sec. II, as an introduction to the real problem, we consider the case of one site occupation probability.

In Sec. III we show that by applying probability theory we can find the exact pair site occupation probabilities as in Sec. III 2 where we come to the same results combinatorially.

II. ONE-SITE AND PAIR-SITE OCCUPATION PROBABILITY

A. One-site

A particular site can be in only two different states, either occupied or unoccupied [see Fig. 1(a)].

Let $\theta = k/N$ be the occupation probability of a given site by simple particles (particles that occupy a single lattice site), either on a one- or a two-dimensional lattice, where k is the number of simple particles and N is the number of lattice sites. If we are dealing with dumbbells the occupation probability of a given lattice site is also θ ($= 2k/N$) because any site taken at random is again a typical site, either on a one- or a two-dimensional lattice.

B. Pair-site

Let us now ask for the pair-site occupation probability. We are only going to consider the nearest-neighbor pair of sites either on a one- or on a two-dimensional lattice. The number of different states in which we find a given pair of sites will depend now on the kind of particles we are considering

(a) *Single particles*: The pair-site can be in three different states (Fig. 1b), both sites either unoccupied (u-u) or occupied (o-o) or one unoccupied and the other occupied (u-o).

The pair-site probability of states u-u; o-u; and o-o is $(1-\theta)^2, 2\theta(1-\theta)$, and θ^2, respectively, if we are considering noninteracting particles.

(b) *Dumbbells*: the previous result can be taken as only a first approximation to the true value when we are dealing with particles that occupy more than one lattice site, and this is so because a pair site can be now in four different states [Fig. 1(c)]: u-u; u-o; (u-o)d; or (o-o)s. The fourth state is now one where the same particle is occupying both lattice sites. In the present paper we are going to determine the exact pair site occupation probability for a one-dimensional array of dumbbells.

 0022-2488/81/081746-03$01.00 © 1981 American Institute of Physics

III. THE MODEL

The essential motive of this work is the application of probability theory to solving the infinite linear lattice. The aim of the authors is to show with future outlooks more complexes problems, which sometimes can avoid combinatorial calculus, and we only carry them out in Sec. III-2 to show that we can obtain identical results.

The essential condition that the *infinite* linear model must fulfill [see Sec. III A and Fig. 2(a)] is that the probability of a site taken at random being occupied is $\theta\,(=2k/N)$.

This condition is not fulfilled by a finite linear lattice (with extremes) [see Fig. 2(c)] and if x is an end site the probability of its being occupied is

$$\frac{\binom{N-k-1}{k-1}}{\binom{N-k}{k}}=\frac{\theta}{2-\theta}\neq\theta.$$

The only way to simulate in finite form an infinite linear model without end sites would be with the finite circular model, and in this case the probability that a given site taken at random will be occupied is

$$\frac{\binom{N-k-1}{n-k}+\binom{N-k-1}{N-k}}{\binom{N-k}{k}+\binom{N-k-1}{k-1}}=\theta.$$

A. The one dimensional model

1. First procedure

Let us take a one-dimensional lattice occupied by dumbbells where the one-site occupation probability is θ, such as the one shown in Fig. 2(a). Our purpose is to find out what is the probability of finding two contiguous lattice sites, taken at random, in one of those states shown in Fig. 1(c).

Let γ_1, γ_2, γ_3, and γ_4 be the probabilities of finding the states (o — o)s; u — u; o — u; and (o — o)d, respectively. We are going to call **r**, **s**, **t**, and **u** four consecutive lattice sites [Fig. 2(b)].

Let A, B, and C be the following events:

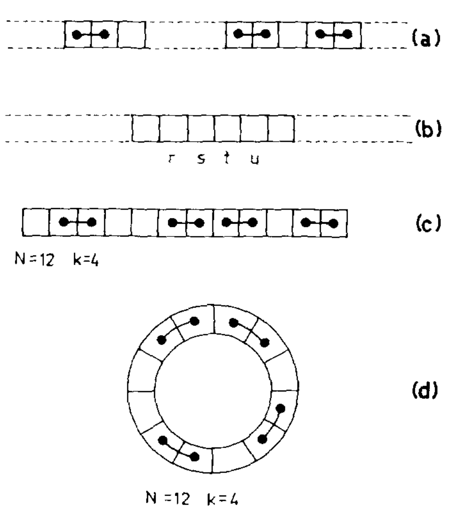


FIG. 2. A one-dimensional lattice space, (a), (b), (c) linear; (d) circular.

A: The event of the occupation of site **s**,

B: The event of the occupation of sites **s** and **t** by the same particle,

C: The event of the occupation of sites **r** and **s** by the same dumbbell.

Then $P(A)$ is the probability that site **s** is occupied, $P(B)$ and $P(C)$ are the probabilities that sites **s, t** and **r,s** are occupied by the same particle, respectively. We immediately see that events B and C are mutually exclusive ($B\cap C=\varnothing$) and exhaustive ($B\cup C=A$), therefore

$$\theta=P(A)=P(B)+P(C).$$

By symmetry

$$P(B)=P(C).$$

Therefore $P(B)=P(C)=\theta/2$.

$$\gamma_1=\theta/2.$$

We are going now to evaluate γ_2, γ_3, and γ_4. Let F be the event where sites **s** and **t** are never occupied by the same dumbbell. Then

$$P(F)=1-\gamma_1=1-\theta/2.$$

Within the universe defined by F the probability of an event $H\,(H\subset F)$ occuring is $\bar{P}(H)$;

$$\bar{P}(H)=P(H/F)=P(H)/P(F).$$

Assume now that H and K are two independent events within the F universe, then

$$\bar{P}(H\cap K)=\bar{P}(H)\cdot\bar{P}(K),$$

or

$$P(H\cap K)/P(F)=P(H)/P(F)\cdot P(K)/P(F).$$

Therefore the probability of events H and K occuring in the original universe will now be $P(H\cap K)$:

$$P(H\cap K)=P(H)\cdot P(K)/P(F).$$

We can now obtain γ_2, γ_3, and γ_4 by applying the previous equation with events H and K properly defined. In every case the conditions $H\subset F$ and $K\subset F$ must be fullfilled.

We proceed as follows to evaluate γ_2. Let H be the event site **s** is empty and K the event site **t** is empty. Then

$$\gamma_2=P(H\cap K)=P(H)\cdot P(K)/P(F)$$
$$=(1-\theta)(1-\theta)/(1-\theta/2)=(1-\theta)^2/(1-\theta/2).$$

To evaluate γ_3 we need to consider just one case and then we multiply by two. Now let H be the event of the occupation of sites **r** and **s** by the same dumbbell and K the event site **t** is empty. Then

$$\gamma_3=2\cdot P(H\cap K)=2\cdot P(H)\,P(K)/P(F),$$

$$\gamma_3=\frac{2\cdot\theta/2\cdot(1-\theta)}{1-\theta/2}=\frac{\theta(1-\theta)}{1-\theta/2}.$$

Finally, to obtain γ_4 we proceed as follows. H will now be the event of the occupation of sites **r** and **s** by the same dumbbell and similarly K will be the event of the occupation of sites **t** and **u** by another dumbbell. Then

$$\gamma_4=P(H\cap K)=P(H)\cdot P(K)/P(F),$$

$$\gamma_4=\frac{\theta/2\cdot\theta/2}{1-\theta/2}=\frac{\theta^2}{4(1-\theta/2)}.$$

2. Alternative procedure

Let us begin by considering a finite linear lattice consisting of N equivalent sites where k dumbbells are arranged, such as the one shown in Fig. 2(c). Let us now consider the occupational degeneracy $T(k,N)$ for dumbbells particles, that is, the number of ways **k** indistinguishable dumbbells can be arranged on a one-dimensional lattice space consisting of N equivalent sites.

$$T(k,N) = \left[\binom{N-k}{k}\right],$$

because there are k indistinguishable particles to be permuted and $k + N - 2k$ entities ($= N - k$).

The pair-site occupation probability for a finite linear lattice space will depend on the particular pair we are selecting. In other words, it depends on the position of the pair with respect to the ends of the lattice space.

We are going to assume that the "end effects" of the one-dimensional lattice will be negligible on the pair-site occupation probability as the number of sites (N) tends to infinity. This assumption is equivalent to the device of joining the ends of our linear lattice, Fig. 2(d). Now, any site is a typical site. If we choose at random two consecutive sites in a circular lattice space we can find the pair-site in:

(a) X different states where both sites are occupied by the same dumbbell (o − o)s:

$$X = \binom{N-k-1}{k-1},$$

because there are $k-1$ particles left to be permuted on $N-2$ equivalent sites, that is to say, $k-1$ dumbells and $k-1+N-2-2(k-1)$ entities ($= N-k-1$);

(b) or in Y different states where the chosen pair site is in states u − u, u − o, or (o − o)d:

$$Y = \binom{N-k}{k}.$$

Actually this is equivalent to a linear lattice space where k dumbbells are arranged on N equivalent sites.

Therefore the total number of ways we can order the circular lattice space is

$$X + Y = \binom{N-k-1}{k-1} + \binom{N-k}{k}.$$

The number of ways we choose a given pair site being occupied by a dumbbell is X, therefore

$$\gamma_1 = \frac{\binom{N-k-1}{k-1}}{\binom{N-k}{k} + \binom{N-k-1}{k-1}}.$$

Let now a given pair site be unoccupied. In such a case there are k dumbbells on $N-2$ sites, that is to say, k dumbbells and $k+N-2-2k$ entities ($= N-k-2$). Then the number of ways we can choose a given pair site being unoccupied is

$$\binom{N-k-2}{k}.$$

Then γ_2 is

$$\gamma_2 = \frac{\binom{N-k-2}{k}}{\binom{N-k}{k} + \binom{N-k-1}{k-1}}.$$

Similarly γ_3 and γ_4 are

$$\gamma_3 = 2\frac{\binom{N-k-2}{k-1}}{\binom{N-k}{k} + \binom{N-k-1}{k-1}},$$

$$\gamma_4 = \frac{\binom{N-k-2}{k-2}}{\binom{N-k}{k} + \binom{N-k-1}{k-1}}.$$

Let θ be the one-site occupation probability, then, when we allow $N \to \infty$ $2k/N \to \theta$. Taking the limits ($N \to \infty$) of γ_1, γ_2, γ_3, and γ_4 we find

$$\lim_{N\to\infty} \gamma_1 = \theta/2, \quad \lim_{N\to\infty} \gamma_2 = (1-\theta)^2/(1-\theta/2),$$

$$\lim_{N\to\infty} \gamma_3 = \theta(1-\theta)/(1-\theta/2),$$

$$\lim_{N\to\infty} \gamma_4 = \theta^2/4(1-\theta/2).$$

IV. CONCLUSIONS

We have presented two procedures for finding the pair-site occupation probability of a one-dimensional lattice space. The results we find are the same with either the first or the second procedure. The former is a more general treatment; the only assumption we have made is that there is an equal *a priori* one-site occupation probability (θ). In the second procedure, a combinatorial one, we begin with a finite linear lattice and in order to simplify the treatment we neglect all the "end effects" on the pair-site occupation probability by using the device of joining the ends of the linear lattice space. As N tends to infinity in this circular lattice space we come to the same pair-site occupation probabilities for states u − u; u − o; (o − o)d; and (o − o)s.

ACKNOWLEDGMENT

E. E. M. is grateful to the Consejo Nacional de Investigaciones Científicas y Técnicas for financial support.

[1]C. G. Goymour and D. A. King, J. C. S. Faraday **4**, 749 (1973).
[2]D. L. Adams, Surf. Sci. **42**, 12 (1974).
[3]E. Ising, Z. Phys. **31**, 253 (1925).
[4]A. Lenard, J. Math. Phys. **2**, 682 (1961).
[5]M. Kac, Phys. Fluids **2**, 8 (1959).
[6]S. F. Edwards and A. Lenard, J. Math. Phys. **3**, 778 (1962).
[7]J. M. J. Van Leeuwen, J. Chem. Phys. **41**, 2959 (1964).
[8]R. B. McQuistan and D. Lichtman, J. Math. Phys. **9**, 1680 (1968).
[9]R. B. McQuistan and D. Lichtman, J. Math. Phys. **11**, 3095 (1970).

Distributional and regulator limits for circuit breaker regularized Feynman amplitudes [a)]

Richard R. Hampton
Department of Mathematics, Villanova University, Villanova, Pennsylvania 19085

(Received 9 October 1980; accepted for publication 12 December 1980)

This paper continues an investigation of circuit breaker (sometimes called point split) regularized Feynman amplitudes and Taylor (BPHZ) subtraction terms. A circuit breaker regularized amplitude is essentially a Fourier transform with respect to internal (loop) momenta of the Feynman integrand. It is shown that the regularization produces tempered distributions in the external momenta in the limit as the propagator epsilon is taken to zero. It is also shown that for renormalized amplitudes the regularization may be removed interchangeably before or after the propagator epsilon is taken to zero. The regularization should be useful for studying time ordered functions and may lead to a more direct proof of the unitarity property of BPHZ renormalized perturbative field theory. A final incidental result in this paper is a new proof that the Feynman amplitudes produced by the forest formula of BPHZ are tempered distributions.

PACS numbers: 11.10.Gh

1. INTRODUCTION

In Ref. 1 the continuing importance of regularization as an analytic tool in renormalized perturbation theory was briefly surveyed. Particular attention was called to Ref. 2 where a heuristic proof of the unitarity of the Zimmermann formulation (BPHZ)[2,3] of BPH[4,5] is based upon the use of Pauli–Villars regularization. The unitarity demonstration remained incomplete because it included the unproven assumption that the propagator epsilon and regulator limits could be freely taken in either order.

Although BPHZ is known to be a unitary theory (to every order in perturbation theory) because it has been proven to be equivalent to analytic renormalization,[6,7] aspects of the unitarity property remain a topic of current research interest. The unitarity property is a relationship between vacuum expectation values of mixed products ($\bar{T}T$) of anti-time-ordered and time-ordered operators. Such mixed products are also important to the understanding of hadron production by pair annihilation.[8]

Circuit breaker regularization permits the propagator epsilon to be taken to zero, in the sense of distribution theory, before removing the regularization. For this reason, circuit breaker regularization may prove a useful tool for studying unitarity relations. In addition, circuit breaker regularization is a generalization of point splitting[1] and therefore relates to the Wilson expansion which has proven useful in studying hadron production.[8]

Point split or circuit breaker regularization was introduced in Ref. 1 and investigated for arbitrary terms of the Zimmermann forest formular for the renormalized Feynman amplitude. We briefly elaborate the notation of the reference.

The typical term of the forest formula contains factors, arising from Feynman propagators, of the form

$$[(k_\ell + q_\ell)^2 - m^2 + i\epsilon((k_\ell + q_\ell)_E^2 + m^2)]^{-1}. \tag{1.1}$$

These factors are assigned to, or in other words labeled by, the lines, ℓ or a circuit based graph[9] (CBG), G. The k_ℓ are the internal line momenta which define the structure of this CBG. We choose a g-chord, i.e., a set of lines whose internal momenta constitute a maximal linearly independent set, and express all of the internal line momenta in terms of this set by

$$k_\ell = \sum_{j=1}^{M} d^j_\ell k_j, \tag{1.2}$$

using a relabeling if necessary.

The external line momenta, q_ℓ, are linear combinations of the external momenta of the amplitude. As usual, m is the mass associated with a propagator, here taken to be the same for all lines of the CBG. We will later briefly discuss removing this restriction.

The Minkowski metric is used implicitly on all four vectors except where the subscript E indicates a Euclidean metric. The careful reader may notice that the coefficient of epsilon in Eq. (1.1) differes from that taken by Zimmermann, as Zimmermann's differs from the usual straight epsilon, by the addition of a non-negative polynomial in the components of the total line momentum. This was done for convenience in performing the internal momenta integrations to obtain the expressions in Feynman parameter space. Zimmermann found the changes he introduced useful in establishing the convergence of BPHZ amplitudes.[3,10] The convergence proof remains valid with propagators of the form (1.1) since the expressions used by Zimmermann obviously majorize those generated by (1.1). One expects that any positive quadratic form in the line momenta could be used without altering the final results since different positive quadratic coeffients of epsilon result in the same propagator, considering the propagator as a generalized function in the epsilon to

[a)]The research reported in this paper was supported, in part, by a summer grant from the Villanova University Office of Research and Patent Affairs.

 © 1981 American Institute of Physics

zero limit.[11] In any case, we do not further pursue this question here.

We assign to each line of the relevant CBG, G, the usual Feynman parameter α_ℓ, with the values of the parameters restricted to a $(\#(G)-1)$-dimensional region R_α defined by

$$0 \leqslant \alpha_\ell \leqslant 1,$$

and

$$\sum_{\ell=1}^{\#(G)} \alpha_\ell = 1. \tag{1.3}$$

We also assign a breaking (or splitting) parameter to each line of G.

The parameter space results are conveniently expressed using

$$A_{ij} = d^i \alpha d^j = \sum_{\ell=1}^{\#(G)} d^i_\ell \alpha_\ell d^j_\ell, \tag{1.4}$$

$$Q_\ell = \sqrt{\alpha_\ell}\, q_\ell, \tag{1.5}$$

$$Q^2_\mu = q_\mu \alpha q_\mu = \sum_{\ell=1}^{\#(G)} (q_\ell)_\mu \alpha_\ell (q_\ell)_\mu, \tag{1.6}$$

$$\chi_j = d^j \xi, \tag{1.7}$$

$$V = \sum_{i,j} (Qd^i)(A^{-1})_{ij}\chi_j, \tag{1.8}$$

$$\begin{aligned} W = {} & Q^2 - m^2 + i\epsilon(Q^2_E + m^2) \\ & - \sum_{i,j} [(Qd^i)(A^{-1})_{ij}(Qd^j) \\ & + i\epsilon(Qd^i)(A^{-1})_{ij}(Qd^j)_E], \end{aligned} \tag{1.9}$$

and

$$\begin{aligned} Z = {} & \frac{1}{4(1+\epsilon^2)} \\ & \times \sum_{i,j} [-\chi_i (A^{-1})_{ij}\chi_j + i\epsilon\chi_i (A^{-1})_{ij}\chi_{jE}]. \end{aligned} \tag{1.10}$$

In this notation, it follows from Ref. 1 that the typical point split forest term (psft) is of the form

$$\begin{aligned} T^{(\epsilon)} = {} & \int_{R_\alpha} (d\alpha) \frac{e^{iV}}{\det A^{2}} \int_{\lambda=0}^{\infty} d\lambda\, \lambda^{N-1} \\ & \times \mathscr{P}\left(\frac{l}{i}\frac{\partial}{\partial \chi}, q\right) \exp\left(i\lambda W + \frac{iZ}{\lambda}\right), \end{aligned} \tag{1.11}$$

where $\mathscr{P}$ is a polynomial in its arguments, and N is an integer which is the $\#(G) - 2M(G)$, with $\#G =$ the number of lines of G, and $M(G) =$ the number of lines in a g-chord of G.

The integrals in Eq. (1.11) were shown to converge absolutely providing that the breaking parameters do not lie in any of a finite number of hyperplanes. This condition permitted a regularizing choice of breaking parameters in every neighborhood of the origin.

The previous investigation was limited to the case of one fixed positive mass common to all propagators. The generalization to allow various positive masses on the various propagators is simple. One need only replace m by

$$m = \sum_{\substack{\ell:\\ \ell \in G}} \alpha_\ell m_\ell \leqslant \Big(\sum_{\substack{\ell:\\ \ell\in G}} \alpha_\ell\Big) \inf_{\ell \in G} m_\ell = \inf_{\ell\in G} m_\ell \geqslant 0).$$

Then the essential arguments of the reference and the present paper go through with practically no change.

In the present paper, we shall prove that the epsilon to zero limit of Eq. (1.1) is a tempered distribution in the external momenta. We will also investigate the point split renormalized Feynman amplitude for an arbitrary Feynman diagram (the Γ-psFa). This is given by

$$J^{(\epsilon)}_\Gamma(\xi,q) = \int (dk)\, e^{ik\xi} R^{(\epsilon)}_\Gamma(k,q), \tag{1.12}$$

where (dk) is the volume element for the independent loop momenta and $R^{(\epsilon)}_\Gamma$ is given by the forest formula. We will show that after folding in a test function, ϵ and ξ may be taken to zero interchangeably.

A final rather incidental result will be a new proof that the forest formula of BPHZ produces, in the ϵ to zero limit, tempered distributions in the external momenta. This is an old result. The method of proof in some respects resembles Hepp's in its reliance on parameter space representations of the amplitude, but is new in its utilization of the theory of circuit based graphs.

2. PROPERTIES OF THE QUADRATIC FORMS OF FEYNMAN PARAMETER SPACE

It will be useful to briefly review and extend some of the results of the references (1,9) for this section.

Theorem 2.1: If $M = M(G)$ is the (positive) number of lines of any g-chord of G and χ is any M-dimensional vector (each component of which may obviously be a 4-vector if the Euclidean product is understood) not lying in any of a finite number of specific hyperplanes intersecting at the origin, then for the matrix A of Eq. (1.4) the quantity

$$\left[(\det A)\left(\sum_{\substack{i,j\\=1}}^{M} \chi_i (A^{-1})_{ij}\chi_j\right)_E^{M}\right]^{-1}$$

is bounded almost everywhere in R_α.

The proof of this theorem and a detailed specification of the relevant hyperplanes may be found in Ref. 1.

We recall that a set of g-circuits is called fundamental (an f-set) iff no one of the g-circuits is contained in the union of the others in the set, and that a g-circuit is determined by a g-chord as follows. The internal momenta of the g-chord span the space which contains the internal momenta of the CBG. A circuit is then defined as the set of lines of the CBG which contain non-vanishing components of the particular line of the g-chord. It happens that each g-chord determines an f-set of g-circuits, and each f-set of g-circuits determines one or more g-chords consisting of one of the lines from each g-circuit which is unique to that g-circuit.

Lemma 2.2: If T^* is any g-chord of $G/C = G - C$, and C is a g-circuit of G, and $r \in C$, then $T^* \cup \{r\}$ is a g-chord of G.

Proof: For convenience, we call lines independent if the internal momenta assigned to these lines constitute a linearly independent set. Since T^* is a g-chord, by definition it consists of independent lines. If $T^* \cup \{r\}$ does not consist of independent lines, then the internal momentum of r may be expressed in terms of those of T^*. Consider a g-chord, T^*_1, of G

which has C amongst the f-set of circuit which it generates. Since $T^* \subset G/C$, the internal momenta of T^* may be expressed in terms of those of those of $T_1^* - C$. This would then require that $r \notin C$, which contradicts the hypothesis of the lemma. Therefore $T^* \cup \{r\}$ consists of independent lines. Since the number of independent lines in G/C is just $\#(T_1^* - C) = M(G) - 1$, because each g-chord intersects each circuit in its f-set in just one line, it follows that $T^* \cup \{r\}$ consists of $M(G)$ independent lines and is therefore a g-chord in G, proving the lemma.

The next theorem is a generalization of a result obtained by Speer.[12]

Theorem 2.3: The quantity V of Eq. (1.8) is a rational, bounded, and continuous function of α almost everywhere in R_α.

Proof: The matrix elements of A^{-1} are given by[9]

$$(A^{-1})_{ij} = \frac{1}{\det A} \sum_{\substack{C: \\ C \in \mathscr{C}(G)}} u_C^2 d_i^C d_j^C \times \sum_{\substack{T^*: \\ T^* \in \mathscr{T}^*(G/C)}} \left(d_{T^*}^2 \prod_{\substack{l: \\ l \in T^*}} \alpha_l \right) \tag{2.1}$$

with

$$\det A = \sum_{\substack{T^* \\ T^* \in \mathscr{T}^*(G)}} d_{T^*}^2 \prod_{\substack{l: \\ l \in T^*}} \alpha_l . \tag{2.2}$$

In these equation, $\mathscr{C}(G)$ is the class of all g-circuits in G, and $\mathscr{T}^*(G)$ is the class of g-chords in G. Of course, G is the CBG appropriate to the forest term in question. The number $d_{T^*}^2$ and u_C^2 are positive constants of the structure of G. In particular, u_C may also depend on the choice of g-chord, containing C amongst its f-set of g-circuits, used to generate C. Using that g-chord, the number d_i^C is the internal momentum routing coefficient indicating for the ith line the component of internal momentum belonging to the unique line of C in that g-chord. Changing the g-chord used to generate C then produces no actual change in the result since the resulting alterations in u_C and d_i^C cancel. Finally, we note the lines of G are numbered so that $T^* = \{1, 2, ..., M(G)\}$ is the g-chord used to define A in Eqs. (1.2) and (1.4). The reader wishing more detail is encouraged to consult Ref. 9.

As a result of Eqs. (2.1) and (2.2), V is a linear combination of terms of the form

$$V_{CT^*} = (q\alpha d^C)(\xi d^C)\Big(\prod_{l \in T^*} \alpha_l\Big)/\det A, \tag{2.3}$$

where now $T^* \in \mathscr{T}^*(G/C)$. By definition,

$$d_l^C = 0, \text{ for } l \notin C. \tag{2.4}$$

Therefore the factor $q\alpha d^C$ leads only to vanishing terms except for those of the form

$$V_{CT^*r} = \alpha_r\Big(\prod_{l \in T^*} \alpha_l\Big)/\det A, \tag{2.5}$$

with $r \in C$. This is clearly a continuous function of α at every point of R_α except those which make $\det A$ vanish. According to Eq. (2.2), $\det A$ vanishes iff α vanishes on at least one line of every possible g-chord. This defines a set in R_α of measure zero. At every other point of R_α, either V_{CT^*r} vanishes or, since $T^* \in \mathscr{T}^*(G/C)$, Lemma 2.2, and Eqs. (2.2) and (2.5) require that the denominator of V_{CT^*r}, $\det A$, is a sum of positive definite terms one of which exactly equals the numerator. This completes the proof of theorem 2.2.

Lemma 2.4: The function W defined by Eq. (1.9) obeys the following inequalities:

(1) $\operatorname{Re} iW > 0$,

(2) $\operatorname{Re}(-(1+i\epsilon)W)|_{\operatorname{Im} W=0} \geqslant m^2(1+\epsilon^2) > 0$, and

(3) $q_\mu \alpha q_\mu \geqslant q_\mu \alpha q_\mu - \sum_{i,j} (q_\mu \alpha d^i)(A^{-1})_{ij}(q_\mu \alpha d^j) \geqslant 0$.

Proof: Property (3) is proven in Ref. 1 by using Bessel's inequality in conjunction with the diagonalized version of Hermitian form A^{-1}. Property (3) with Eq. (1.9) results in property (1) and, after a simple calculation using the condition $\operatorname{Im} W = 0$, in property (2).

3. CIRCUIT BREAKER REGULARIZATION IN THE EPSILON TO ZERO LIMIT

We will evenutally show that the psft is a tempered distribution. Before proceding, we recall that a tempered distribution is a continuous linear functional on $\mathscr{S}$, the countably normed space of infinitely differentiable functions on $\mathbb{R}^n$ with $f \in \mathscr{S}$ iff $(\forall r \geqslant 0)\forall s \geqslant 0)$

$$\|f\|_{rs} = \sup_{\substack{0 \leqslant k \leqslant r \\ 0 \leqslant j \leqslant s}} \sup_{x \in R^n} |x^k \partial^j f(x)| < \infty . \tag{3.1}$$

We use here the multi-index notation[13]

$$|k| = \sum_{i=1}^n k_i, \quad |j| = \sum_{i=1}^n j_i,$$

$$x^k = \prod_{i=1}^n x_i^{k_i}, \text{ and } \partial^j = \prod_{i=1}^n \frac{\partial^{j_i}}{\partial x_i}. \tag{3.2}$$

Lemma 3.1: If V is given by Eq. (1.8) and $\phi(q) \in \mathscr{S}$, then $\phi e^{iV} \in \mathscr{S}$, $(\forall r \geqslant 0)(\forall s \geqslant 0)$ $\|\phi e^{iV}\|_{rs}$ is bounded by a function which is bounded and continuous almost everywhere in R_α.

Proof: The notation of Eq. (3.1) is extended to the $4n$-dimensional external momentum space in the obvious way. As a result

$$\|\phi e^{iV}\|_{rs} \leqslant \|\phi\|_{rs} \sup_{\substack{j: \\ j \leqslant s}} |\partial^j e^{iV}| , \tag{3.3}$$

since, by Eqs. (1.5) and (1.8), $\partial V / \partial q_{i\mu}$ is independent of q. In fact, since V is linear in q, Lemma (3.1) then follows from Theorem 2.3.

We proceede from Eq. (1.11) to rewrite the psft by first carrying out the indicated differentiations. Then Lemma 2.4(1) enables us to recognize in the lambda integral a well known representation of the modified Bessel function.[14] We redefine T to be a typical term in the result:

$$T^{(\epsilon)} = \int_{R_\alpha} (d\alpha) \frac{F^{(\epsilon)}(\alpha, q, \chi)}{(\det A)^P} \left[\frac{Z}{W} \right]^{N/2} K_{|N|}(2\sqrt{-ZW})\, e^{iV}. \tag{3.4}$$

Equations (1.8), (1.10), and (2.1) show that $F^{(\epsilon)}$ is a polynomial in its arguments; the entire ϵ dependence of $F^{(\epsilon)}$ is found in a factor which is defined and continuous over $[0,1]$. P is a positive integer.

To simplify the discussion, we now introduce the condition that the breaking parameters, ξ, Eq. (1.7), have vanishing time components. This enables us to write

$$2\sqrt{-ZW} = \Big\{\sum_{i,j}[\chi_i(A^{-1})_{ij}\chi_j]_E/(1+\epsilon^2)\Big\}^{\frac{1}{2}} \times[-(1+i\epsilon)W]^{\frac{1}{2}}. \tag{3.5}$$

The first factor is positive almost everywhere because A is positive definite except on a set of measure zero. For the second factor, we must take the branch of the square root function for which

$$-\pi/2 < \arg[-(1+i)W]^{\frac{1}{2}} < \pi/2. \tag{3.6}$$

In fact, by Lemma 2.4(2), the quantity $-(1+i\epsilon)W$ does not cross over the branch cut along the negative real axis, and we can therefore be assured of Eq. (3.6) with a choice of square root which will be continuous almost everywhere in R_α. We are thereby in a position to use a different representation of the modified Bessel function[14] and write

$$T^{(\epsilon)} = \int_{R_\alpha}(d\alpha)\frac{f^{(\epsilon)}(\alpha,q,\chi)}{(\det A)^P}BHe^{iV}, \tag{3.7}$$

where

$$B = \left[\frac{W}{a^2}\right]^{(|N|-N)/2}, \tag{3.8}$$

and

$$H = \int_0^\infty dt\frac{\cos at}{[t^2-(1+i\epsilon)W]^{|N|+\frac{1}{2}}}, \tag{3.9}$$

with

$$a^2 = \frac{1}{1+\epsilon^2}\sum_{i,j}[\chi_i(A^{-1})_{ij}\chi_j]_E. \tag{3.10}$$

The functions $f^{(\epsilon)}$ and $F^{(\epsilon)}$ differ only by a factor which depends on ϵ alone and is defined and continuous over [0,1].

We now fold in any $\phi\in\mathscr{S}$:

$$T^{(\epsilon)}(\phi) = \int dq\,\phi(q)T^{(\epsilon)}. \tag{3.11}$$

Since $T^{(\epsilon)}$ is polynomially bounded in q and the α integration is absolutely convergent for $\epsilon>0$ [cf. Eqs. (4.19), (4.20), (4.22), (4.35), and Theorem 4.1 of Ref. 1], the iterated integral of Eq. (3.11) is absolutely convergent. We use Fubini's theorem to exchange the order of integration:

$$T^{(\epsilon)}(\phi) = \int_{R_\alpha} d\alpha\,(\det A)^{-P} \times\int dq\,f^{(\epsilon)}(\alpha,q,\chi)\phi(q)e^{iV}BH. \tag{3.12}$$

Lemma 3.2: For every non-negative integer n, after $2n$ integrations by part, H becomes a sum of terms each of which is, up to multiplicative constants, of the form

$$H_{2n} = a^{-2n}\int_0^\infty dt\frac{t^r\cos at}{[t^2-(1+i\epsilon)W]^{|N|+s}};$$

after $2n+1$ integrations by parts, the terms are of the form

$$H_{2n+1} = a^{-2n-1}\int_0^\infty dt\frac{t^r\sin at}{[t^2-(1+i\epsilon)W]^{|N|+s}}.$$

In both cases, r will be a non-negative integer, and, if at least one integration by parts has been performed, then $s\geqslant 3/2$, and

$$r-2s-2N\leqslant -2.$$

Proof: For $n=0$, the first equation of the lemma is merely a re-statement of Eq. (3.9). This is integrated by parts once to obtain $H=(2|N|+1)H_1, s=3/2$, and $r=1$. After a second integration by part, we obtain

$$H = \frac{2|N|+1}{a^2}\int_0^\infty dt\,(\cos at)\left[\frac{1}{[t^2-(1+i\epsilon)W]^{|N|+3/2}} + \frac{(|N|+3/2)t}{[t^2-(1+i\epsilon)W]^{|N|+5/2}}\right]. \tag{3.13}$$

The inequalities claimed by the lemma are trivially verified. This demonstrates the first formula for $n=0$ and $n=1$, and the second formula for $n=0$, and the required inequalities. In a similar way, integration of H_{2n} and of H_{2n+1}, each by parts twice, verifies the lemma by mathematical induction on the n of each formula.

This lemma will prove to be useful both because it guarantees the absolute convergence of the t integration and because it extracts factors of $1/a$ which help us to demonstrate cancellation of the singularities of $1/\det A$.

As a result of Lemma (3.2), $T^{(\epsilon)}(\phi)$ can be written as a sum of terms of the form

$$T_1^{(\epsilon)}(\phi) = \int_{R_\alpha}(d\alpha)\,dq\,g^{(\epsilon)}(\alpha,q,\chi)\phi(q)e^{iV}I_{2n}, \tag{3.14}$$

with

$$I_{2n} = a^{2n}H_{2n} = \int_0^\infty dt\frac{t^r\cos at}{[t^2-(1+i\epsilon)W]^{|N|+s}}. \tag{3.15}$$

The function $g^{(\epsilon)}$ of Eq. (3.14) and the function $f^{(\epsilon)}$ of Eq. (3.7) differ only by a factor which is a continuous function of ϵ alone, defined on [0,1], a power of B, and, finally a factor of $(\det A)^{-P}a^{-2n}$. The factors of a^{-2n} contain a factor of a power of $(1+\epsilon^2)$ [cf. Eq. (3.10)] which we can absorb into the first of the three factors. What is left of the third factor is bounded almost everywhere in R_α, according to theorem (2.1), providing we take n big enough so that $2n\geqslant P$. The factor involving B is a non-negative power of B. Therefore, Eqs. (3.8), (3.10), and (1.9) together with theorem 2.1 guarantee that $g^{(\epsilon)}$ is polynomial in q with coefficients which are bounded and continuous almost everywhere in R_α and having the ϵ dependence confined to a factor which is a function of ϵ alone and is defined and continuous on [0,1].

The next theorem will enable us to recognize the psft as related to a well known distribution.

Theorem 3.3: For the psft represented by Eq. (3.14), for $\phi\in\mathscr{S}$ the q and t integrations may be exchanged so that a typical term is

$$T^{(\epsilon)}(\phi) = \int_{R_\alpha}(d\alpha)\,g^{(\epsilon)}(\alpha,\chi)\int_0^\infty dt\,t^r\cos at \times\int dq\frac{e^{iV}\Phi(q)}{[t^2-(1+i\epsilon)W]^{|N|+s}}.$$

where $q^\eta g^{(\epsilon)}(\alpha,\chi)$ is a typical term in $g^{(\epsilon)}(\alpha,\chi,q)$, and the factors of q^η have been absorbed onto $\Phi=q^\eta\phi$, $\Phi\in\mathscr{S}$.

Proof: For convenience we set

$$\sum_{imj}(Q_\mu ad^i)(A^{-1})_{ij}(Q_\mu ad^j) = Q_\mu'^2. \tag{3.16}$$

Then, by Lemma 2.4(3),

$$Q_\mu^2 - Q_\mu'^2 \geqslant 0 \tag{3.17}$$

and is, by Eqs. (1.5) and (1.6), quadratic in q_μ. Once again, Theorem 2.3 and Eq. (1.8) with (Qad^j) replacing x_j and Q_μ replacing Q implies $Q_\mu^2 - Q_\mu'^2$ is continuous and bounded almost everywhere in R_α. We conclude that $Q_\mu^2 - Q_\mu'^2$ is an infinitely differentiable function of q_μ which, with all its derivatives, is continuous and bounded almost everywhere in R_α.

Now let $E(x)$ be infinitely differentiable function on the real numbers such that

$$\begin{aligned} E(x) &= 1, \text{ for } x \geqslant 2m^2/3,\\ &= 0, \text{ for } x \leqslant m^2/3, \text{ and}\\ 0 &\leqslant E(x) \leqslant 1. \end{aligned} \tag{3.18}$$

The for any $\phi(q) \in \mathscr{S}: \phi(q)E(Q_0^2 - Q_0'^2) \in \mathscr{S}$, and $\phi(q)(1 - E(Q_0^2 - Q_0'^2)) \in \mathscr{S}$. We substitute

$$\phi(q) = \phi(q)E(Q_0^2 - Q_0'^2) + \phi(q)(1 - E(Q_0^2 - Q_0'^2)) \tag{3.19}$$

into Eq. (3.14).

By Eq. (3.1), for any s and for any sufficiently large n,

$$\left| \int dq\, \phi(q)(1 - E) \right| \leqslant \int dq\, |\phi(q)| < \|\phi\|_{ns}. \tag{3.20}$$

Then, by an argument of dominated convergence, the second term arising from Eqs. (3.14) and (3.19) converges absolutely to a function which is continuous and bounded almost everywhere in R_α;

$$\begin{aligned} &\left| \int dq\, \phi(q) e^{iV}(1 - E(Q_0^2 - Q_0'^2)) \right| \\ &\quad \times \int_0^\infty dt\, t^r (\cos at)(t^2 - (1 + i\epsilon)W)^{-N'} \\ &\leqslant \int dq\, |\phi(q)|\, 1 - E(Q_0^2 - Q_0'^2) \\ &\quad \times \int_0^\infty dt\, t^r (t^2 + m^2/3)^{-N'}. \end{aligned} \tag{3.21}$$

The right-hand side of Eq. (3.21), obtained by using Eq. (1.9) and Lemma 2.4(3), is obviously absolutely convergent, by Lemma 3.2, uniformly in ϵ. Furthermore, Eq. (3.20) then shows that this term of the psft is continuous in ϕ for $\epsilon \in [0,1]$.

We procede to study the first term arising from Eqs. (3.14) and (3.19). It is convenient to set

$$P = \sum_{\mu=1}^{3} (Q_\mu^2 - Q_\mu'^2),$$

and

$$P_0 = Q_0^2 - Q_0'^2. \tag{3.22}$$

Then the relevant term of the psft has an integrand which is dominated by the integrands of

$$\begin{aligned} &\int dq |\phi(q)E(P_0)| \\ &\quad \times \int_0^\infty dt\, t^r |(t^2 + P - P_0 + m^2 - 2i\epsilon P_0)^{-N'}| \\ &= \int dq |\phi(q)E(P_0)| \\ &\quad \times \int_0^\infty dt\, t^r [(t^2 + P - P_0 + m^2)^2 + 4\epsilon^2 P_0^2]^{-N'/2}, \end{aligned}$$

which, by Eq. (3.18),

$$\begin{aligned} &\leqslant \int dq |\phi(q)E(P_0)| \\ &\quad \times \int_0^\infty dt\, t^r [(t^2 + P - P_0 + m^2)^2 + 16\epsilon^2 m^2/9]^{-N'/2}. \end{aligned}$$

We break the t integration into two parts. The part including the origin is controlled by the $16\epsilon^2 m^2/9$ term of the denominator. If the point of breaking is taken to be

$$2(P_0 - P - m^2)^{1/2} \theta(P_0 - P - m^2),$$

where $\theta(x) = 1$ for non-negative x and is zero for negative x, then the rest of the t integration is bounded by

$$\begin{aligned} &\int dq\, |\phi(q)E(P_0)| \\ &\quad \times \int_{2(P_0 - P - m^2)^{1/2}}^\infty dt\, t^r (t^4/2 + 16\epsilon^2 m^4/9)^{N'/2} \\ &\leqslant \int dq\, |\phi(q)| \int_0^\infty dt\, t^r (t^4/2 + 16\epsilon^2 m^4/9)^{N'/2}, \end{aligned}$$

which, this time for ϵ in (0,1], is absolutely convergent by Lemma 3.2. Both parts of the t integration, again by Eq. (3.20), are continuous in ϕ.

Finally, Theorem 3.3 is then true by Fubini's theorem.

The remaining problems of this section are to show that we can take the epsilon to zero limit inside of the α and t integrals and to establish that the result is continuous in ϕ. For test functions $(1 - E(P_0))\phi, \phi \in \mathscr{S}$, this is already clear from Eqs. (3.21) and (3.22) and Lemma 3.1. For test functions from $E(P_0)\mathscr{S}$, the support properties of the chosen test function effectively keep P_0 positive definite. In this case, a discussion by Speer[12] of a similar expression shows that $(t^2 - (1 + i\epsilon)W)^{-N}$ defines, for every N, a distribution which is a continuous function of ϵ defined on [0,1]. We modify Speer's approach slightly to suit out present purpose.

Theorem 3.4: If $0 < \epsilon \leqslant 1$, $0 \leqslant t \leqslant 1$, and the coefficients of the quadratic forms P and P_0 are restricted to any compact region for which P and P_0 are non-negative, then the distribution $(t^2 - (1 + i\epsilon)W)^{-N}$ is uniformly Cauchy in all of these quantities. That is, there exist positive numbers r, s, and B such that

$$\begin{aligned} &|(t_1^2 + (1 + \epsilon_1^2)(m^2 + P') - (1 - \epsilon_1^2)P_0' + 2i\epsilon_i P_0')^{-N}(\phi) \\ &- (t^2 + (1 + \epsilon^2)(m^2 + P) - (1 - \epsilon^2)P_0 + 2i\epsilon P_0)^{-N}(\phi)| \\ &\leqslant B \|\phi\|_{rs}(|\epsilon_1 - \epsilon| + |t_1 - t|) + \sum_{ij} |P_{ij}' - P_{ij}|, \end{aligned}$$

where the (ij) subscripts on the quadratic form indicate the coefficients.

Proof: For test functions from $(1 - E(P_0)\mathscr{S})$, this is already obvious from the absolute convergence of the integrals, as previously discussed. Differentiation under the integral with respect to the quantities in question establishes the theorem for this case.

Since P_0 is a quadratic form in the time components of the external momenta, by direct calculation

$$\sum_{i=1}^{n} q_i^0 \frac{\partial}{\partial q_i^0}(t^2 - (1+i\epsilon)W)^{\lambda+1} = 2(\lambda+1)(1-\epsilon^2+2i\epsilon)(t^2-(1+i\epsilon)W)^{\lambda}P_0. \quad (3.23)$$

If $q\in\epsilon\,\mathrm{supp}E(P_0)\phi(q)$, then $P_0 \geqslant m^2/3 > 0$. Therefore, by repeated application of Eq. (3.23), provided λ is not a negative integer, and using integration by parts, as usual, to shift the differentiations to the test function, we obtain

$$(t^2-(1+i\epsilon)W)^{\lambda}(\phi E(P_0)) = (-2(1-\epsilon^2+2i\epsilon))^{-k}\frac{(t^2-(1+i\epsilon)W)^{\lambda+k}}{(\lambda+1)(\lambda+2)...(\lambda+k)} \times\left(\left(\frac{1}{P_0}\sum_{i=1}^{n}\frac{\partial}{\partial q_i^0}q_i^0\right)^k(\phi E(P_0))\right). \quad (3.24)$$

Since $E(P_0)$ and all of its derivatives are bounded and $1/P_0 \leqslant 3/m^2$ in the region which is relevant, we also obtain

$$\left|\left|\left(\frac{1}{P_0}\sum_{i=1}^{n}\frac{\partial}{\partial q_1^0}q_i^0\right)^k\left|(\phi(q)E(P_0))\right|\right|\right|_{ns'} \leqslant B_1\|\phi\|_{n+k,s}, \quad (3.25)$$

for some positive B_1.

We choose k sufficiently large that $\mathrm{Re}(\lambda+k)>1$. Then the derivatives of $(t^2-(1+i\epsilon)W)$ with respect to t,ϵ, or any of the coefficients of the quadratic forms, using Eq. (3.24), are uniformly polynomially bounded in q. For any complex value of λ the theorem then follows from successive application of the mean value theorem of differential calculus followed by an application of the maximum value theorem for continuous functions in a compact region. In particular, the right-hand side of Eq. (3.24) is now defined for all epsilon in [0,1]. The continuity in ϕ, including the claimed values of s and of $r=n+k$, is obtained from Eqs. (3.20) and (3.25).

In particular, the theorem is true for λ at any point on the circle of radius 1/2 centered at any negative integer in the complex plane. Since the distribution is, for positive epsilon, entire in λ [12], the maximum modulus principle implies that the left hand side of the inequality of the lemma cannot be bigger when λ is a negative integer than it is on the circle. Therefore Theorem 3.4 is true for every N.

Corollary to theorem 3.4: The distribution $(t^2-(1+i\epsilon)W)^{-N}$ may be extended to a distribution which is defined and continuous on the closure of the region of the theorem.

Proof: Theorem 3.4 guarantees the existance of the limit as epsilon approaches zero and that the convergence to this limit is uniform in the other variables (12).

In order to complete the integration over t, we introduce the change of variables, $u=1/t$,

Theorem 3.5: If $0<\epsilon\leqslant 1$, $0\leqslant u\leqslant 1$, and the coefficients of the quadratic forms P and P_0 are restricted to any compact region for which P and P_0 are non-negative, then the distribution $(1-u^2(1+i\epsilon)W)^{-N}$ is uniformly Cauchy in all of these quantities and may be extended to a distribution defined and continuous on the closure of the region indicated.

Proof: The proof is almost the same as that of Theorem 3.4 and its corollary, except that the (q) region $P_0\geqslant m^2/3$ is further partitioned into two regions:

(1) $m^2/3 + 1/(3u^2)\geqslant P_0\geqslant m^2/3$, and

(2) $P_0\geqslant m^2/3+1/(3u^2)$.

In the first of these regions, the integral over q is easily bounded uniformly in u, epsilon, and the quadratic form coefficients.

To complete the proof, we define $E_2(x)$ in the same way as $E(x)$, Eq. (3.18), but with its support translated to region 2. We then obtain, in analogy to Eq. (3.24),

$$(1-u^2(1+i\epsilon)W)^{\lambda}(\phi E_2(P_0)) = \frac{(2(\epsilon^2-1-2i\epsilon))^{-k}u^{-2k}(1-u^2(1+i\epsilon)W)^{\lambda+k}}{(\lambda+1)(\lambda+2)...(\lambda+k)} \times\left(\left(\frac{1}{P_0}\sum_{i=1}^{n}\frac{\partial}{\partial q_1^0}q_i^0\right)^k(\phi E_2(P_0))\right). \quad (3.26)$$

The proof is then completed by the same arguments as were made before.

We are now ready for the principal theorem of this section.

Theorem 3.6: Every circuit breaker regularized forest term (psft) with positive epsilon defines, in the epsilon to zero limit, a tempered distribtuion in the external momenta.

Proof: We need only prove the epsilon to zero limit exists and is continuous in ϕ. Using the form of the psft given by Theorem 3.3, we interrupt the t integration at $t=1$, and apply Lemma 2.3 to show that the coefficients of the quadratic forms lie in a closed and bounded region. For the first term, $0\leqslant t<1$, by the corollary to Theorem 3.4, the epsilon to zero limit of the integrand is approached uniformly in the region of integration. Therefore the limit exists and, by Theorem 3.4, is continuous in ϕ. For the second term, $1\leqslant t<\infty$, we make the change of variables $u=1/t$, $0<u\leqslant 1$. Using Theorem 3.5, the contribution from integrating from $u=0+$ to a sufficiently small positive number can be bounded uniformly in epsilon. Theorem 3.5 then enables us to conclude that the second term also has the required properties. This completes the proof of Theorem 3.6.

4. THE EXCHANGE OF REGULATOR AND DISTRIBUTIONAL LIMITS

We combine the terms of the point split forest formula (BPHZ) into a point split Feynman amplitude for the graph Γ (Γ-psFa), $J_\Gamma^{(\epsilon)}(\xi,q)$, which is then given by Eq. (1.12). In that equation we have been able to perform the Feynman parameter integrations before performing the internal momenta integrations because the integrals are now absolutely convergent (ac) for positive epsilon[3] and are therefore subject to Fubini's theorem. In addition, since the integrals converge uniformly in ξ.

$$\lim_{\xi\to 0} J_\Gamma^{(\epsilon)}(\xi,q) = J_\Gamma^{(\epsilon)}(0,q) = J_\Gamma^{(\epsilon)}(q).$$

It is known, and also will be shown below, that $J^{(\epsilon)}(q)$ converges to a tempered distribution in the epsilon to zero limit. It is, of course, this distribution which is the renormalized Feynman amplitude. On the other hand, $J_\Gamma^{(\epsilon)}(\xi,q)$ converges in the epsilon to zero limit to a tempered distribution, because, by Theorem 3.6, each psft does. It is the principal object of this section to show that this distribution is continuous in the breaking parameters. In other words, we will show that we can introduce the regularization and take the epsilon to zero limit before removing it without affecting the result.

We procede by combining all terms of the forest formula[3] into a single fraction. The denominator of this fraction is of the form of a product of propagators. In this way, $R_\Gamma^{(\epsilon)}$ is treated as any forest term, and $J_\Gamma^{(\epsilon)}(\xi,q)$ is a psft and is given by Eq. (1.11) with $T_\Gamma^{(\epsilon)}$ replaced by $J_\Gamma^{(\epsilon)}(\xi,q)$.

This result was not dependent, in its form, on the choice of basis for the internal momenta, i.e., not dependent on the choice of g-chord of G. We will make use of this by choosing a g-chord appropriate to each Hepp sector.

A Hepp sector[5] of R_α is a region of R_α in which

$$\alpha_{\pi(1)} \geqslant \alpha_{\pi(2)} \geqslant \alpha_{\pi(3)} \geqslant \cdots \geqslant \alpha_{\pi(\#(G))}, \tag{4.1}$$

where π is some permutation of the lines of G. We use this permutation to label the sector. Different sectors intersect in a region of measure zero; we can consider R_α to be, in effect, partitioned by the collection of sectors, at least as far as the integrals defining the psft are concerned.

Using the Hepp sectors, we perform an analysis of the Feynman parameter formulas of circuit based graph theory. This analysis is similar to that performed by Lowenstein[15] in a modification of the work of Hepp[5] and Speer.[12]

Definition 4.1: If $B \subseteq G = \{1,2,3,\ldots,\#(G)\}$, then by seq B we mean the first line of B in the sequence $(\pi(1), \pi(2),\ldots, \pi(\#G))$, where π is determined from the relevant Hepp sector through Eq. (4.1).

In each Hepp sector we choose a g-chord,

$$T_\pi^* = \{\ell_1, \ell_2, \ell_3, \ldots \ell_{(G)}\} \tag{4.2}$$

through the recursive relationship

$$\ell_i = \operatorname{seq}\{\ell\colon d_{\ell_1}, d_{\ell_2},\ldots, d_{\ell_{i-1}}, d_\ell \text{ is linearly independent.}\} \tag{4.3}$$

Equations (4.1) and (4.2) with Definition 4.1 imply

$$\alpha_{\ell_i} = \sup\{\alpha_\ell\colon d_{\ell_1}, d_{\ell_2}, \ldots, d_{\ell_{i-1}}, d_\ell \text{ is linearly independent.}\} \tag{4.4}$$

In particular ℓ_1 is the first line of the sequence $(\pi(1), \pi(2),\ldots, \pi(\#(G)))$ for which d_{ℓ_1} is not identically zero, i.e., the first line of the sequence which carries internal momentum. The quantity d_j is sometimes called a momentum routing vector; one is assigned to each line; the components of d_j are the coefficients in Eq. (1.2) for the internal momentum of that line.

We rewrite the Γ-psFa, Eq. (1.11), as a sum of the integrals over the various Hepp sectors. For each term, $T^{(\epsilon)}$, corresponding to a particular sector, π_α, we calculate Eq. (1.11) by doing the internal momenta integrations in the basis provided by the g-chord T_π^* of Eqs. (4.2) and (4.3). As usual, we freely omit factors which are defined, continuous, and bounded functions of ϵ alone for $\epsilon \geqslant 0$. The result is then still represented by Eq. (1.11) if we replace R_α by π_α.

We will find the analysis clarified by a change of the variables. It is convenient to define these new variables in terms of a singularity family developed from Eqs. (4.2) and (4.3). For $r \geqslant 0$, we let

$$G_r = G - \{\ell \in G\colon \{\ell, \ell_1, \ell_2, \ldots, \ell_r\} \text{ is linearly dependent.}\}, \tag{4.5}$$

so that

$$G_M = G - G = \varnothing,$$

the empty set, since T_π^* is a g-chord. We also note that if all the lines of G carry internal momentum, then $G_0 = G$.

The singularity family is then defined as

$$\mathscr{E}_\pi = \bigcup_{j=0}^{M(G)-1} \{G_j\}) \cup \{\{\ell\}\colon \ell \in G - T_\pi^*\}. \tag{4.6}$$

Clearly

$$\#(\mathscr{E}_\pi) = M(G) + \#(G) - M(G) = \#(G),$$

where, on the left-hand side, $\#$ is used to indicate the number of sets in the family.

For each $H \in \mathscr{E}_\pi$, we assign a variable $t_H \in [0,1]$ with

$$\alpha_\ell = \prod_{\substack{H: \\ l \in H \in \mathscr{E}_\pi}} t_H. \tag{4.7}$$

For the lines $l_i \in T_\pi^*$ this results in

$$\alpha_{\ell_i} = \prod_{j=0}^{i-l} t_{G_j}, \tag{4.8}$$

which implies

$$\begin{aligned} \prod_{\ell \in T_\pi^*} \alpha_{\ell_i} &= \prod_{i=1}^{M(G)} \prod_{j=0}^{i-1} t_{G_j} \\ &= \prod_{j=0}^{M(G)-1} t_{G_j}^{M(G)-j} = \prod_{j=0}^{M(G)-1} t_{G_j}^{M(G_j)}, \end{aligned} \tag{4.9}$$

where $M(G_j)$ is the number of number of complete g-circuits of G which are left in G_j.

Theorem 4.2: If A, defined by Eq. (1.4) is expressed in terms of the (t) variables, Eq. (4.7), then

$$\det A = \left(\prod_{i=0}^{M(G)-1} t_{G_i}^{M(G_i)} \right)(1 + P)S,$$

and S is a positive number and P is a non-negative polynomial in the variables t_H.

Proof: We note that for $\ell_i \in T_\pi^*$, $\{\ell_i\} \notin \mathscr{E}_\pi$. This allows us to define $t_{\{\ell_i\}} = 1$. Then, according to Eqs. (4.6) and (4.7),

$$\prod_{\ell \in T^*} \alpha_\ell = \left(\prod_{\ell \in T^*} t_{\{l\}} \right) \left(\prod_{j=0}^{M(G)-1} t_{G_j}^{M_{jT^*}} \right), \tag{4.10}$$

with

$$M_{jT^*} = \#(T^* \cap G_j). \tag{4.11}$$

Since G_j is defined by deleting from G the lines which are dependent on the first j elements T_π^*,

$$M_{jT^*} \geqslant M_{jT_\pi^*} = M(G_j). \tag{4.12}$$

Otherwise, $M(G_j) > M_{jT^*}$ would imply that more lines of T^* than of T_π^* have been deleted from G to form G_j/ In that case, more than j lines of T^* would be dependent on the first j lines of T_π^*. This would contradict the linear independence of the lines of T^*. Therefore, we conclude Eq. (4.12).

Theorem 4.2 is now immediate from Eqs. (2.2), (4.10), and (4.12).

Theorem 4.3: The Jacobian of the transformation defined Eq. (4.7) is

$$\left| \frac{\partial(\alpha)}{\partial(t)} \right| = \prod_{i=0}^{M(G)-1} t_{G_i}^{\#(G_i)-1}.$$

Proof: For convenience, we relabel the lines of G so that in $\pi_\alpha, T^*_\pi = \{1,2,...,M(G)\}$, and if $M(G) \geqslant i \geqslant j \geqslant 1$, then $\alpha_i \leqslant \alpha_j$.

We observe that for each of the lines $l \in G - T^*$, there is a number $r(\ell)$ defined by

$$j \leqslant r(\ell) \Leftrightarrow \ell \in G_j. \tag{4.13}$$

That is, there is a first G_j from which l has been deleted. We relabel these $\#(G) - M(G)$ lines in order of increasing $r(\ell)$ and in order of decreasing α_ℓ for each $r(\ell)$. The new labels for these will then be the integers from $M(G)+1$ to $\#(G)$. Then, for $1 \leqslant i \leqslant M(G)$,

$$\frac{\partial \alpha_1}{\partial t_{G_j}} = \delta_{j0}, \quad \frac{\partial \alpha_i}{\partial t_{G_{i-1}}} = \prod_{j=0}^{i-2} t_{G_j}, \tag{4.14}$$

and

$$\frac{\partial \alpha_i}{\partial t_{G_j}} = \frac{\partial \alpha_i}{\partial t_{\{\ell\}}} = 0 \tag{4.15}$$

for $j \geqslant i$ and $\{\ell\} \in \mathscr{E}_\pi$.

For $\#(G) \geqslant i \geqslant M(G)+1$, if $\{\ell\} \in \mathscr{E}_\pi$, we obtain

$$\frac{\partial \alpha_i}{\partial t_{\{\ell\}}} = \delta_{\ell i} \prod_{k=0}^{r(i)} t_{G_k}. \tag{4.16}$$

With an appropriate ordering of the t variables, then, all elements to the right of the diagonal vanish and the Jacobian is therefore

$$\frac{\partial(\alpha)}{\partial(t)} = \left(\prod_{i=2}^{M(G)} \left(\prod_{j=0}^{i-2} t_{G_j}\right)\right)\left(\prod_{l=M(G)+1}^{\#(G)} \left(\prod_{k=0}^{r(l)} t_{G_k}\right)\right) \tag{4.17}$$

$$= \left(\prod_{j=0}^{M(G)-1} t_{G_j}^{M(G)-j-1}\right)\left(\prod_{k=0}^{M(G)-1} t_{G_k}^{\#(G_k)-M(G)+k}\right). \tag{4.18}$$

Equation (4.17) is obtained by noting that the second product of that equation contains a factor of t_{G_k} for every line $\ell \in T^*_\pi$ for which $r(\ell) \geqslant k$. By the definition of r, this second condition on ℓ is equivalent to $\ell \in G_k$. We therefore obtain one factor of t_{G_k} for each line $\ell \in G_k$ for which $\ell \geqslant M(G)+1$, i.e., by our special labeling of the lines, one factor for each $\ell \in G_k - T^*_\pi$. Of course

$$\#(G_k - T^*_\pi) = \#(G_k) - M(G) + k. \tag{4.19}$$

Theorem 4.3 then results from combining the factors.

Theorem 4.4: Under the transformation from (α) to (t), Eq. (4.7), the sector π_α, described by Eq. (4.1) is mapped into the region described by

$$(\forall H \in \mathscr{E}_\pi) \quad 0 \leqslant t_H \leqslant 1.$$

Proof: We take the same labeling as in the proof of Theorem 4.3. From Eq. (4.8), $\alpha_1 = t_{G_0}$ implies $0 \leqslant t_{G_0} \leqslant 1$. Next, $\alpha_2 = t_{G_0} t_{G_1} = \alpha_1 t_{G_1}$ and $\alpha_1 \geqslant \alpha_2$ implies $t_{G_1} = \alpha_2/\alpha_1 \leqslant 1$ implies $0 \leqslant t_{G_1} \leqslant 1$. We continue in the same way until we have exhausted the parameters assigned to the lines of T^*_π. According to Eqs. (4.6) and (4.7), the parameters assigned to the remaining lines are of the form

$$\alpha_\ell = t_{G_0} t_{\{\ell\}} = \alpha_1 t_{\{\ell\}}.$$

Theorem 4.4 now results from Eq. (4.4).

Theorem 4.5: For any g-graph, G, there is a number B such that in any Hepp sector, π, for every j

$$|A_{ij}^{-1}| \leqslant B/\alpha_i = B \Big/ \left(\prod_{k=0}^{i-1} t_{G_k}\right),$$

where the lines of G have been labeled so that $T^*_\pi = \{1,2,...,M(G)\}$.

Proof: The definition of a g-circuit states that $d_i^C \neq 0$ iff i is an element of the g-circuit C. In that case, by Lemma 2.2, $\{i\} \cup T^*$ is a g-chord of G/C. With this, Theorem 4.5 then follows from Eqs. (2.1) and (2.2) and the positivity of the coefficients $d^2_{T^*}$.

Remark: The symmetry of A_{ij}^{-1} implies that theorem 4.5 is also true with i and j exchanged.

We now turn our attention back to Eq. (1.11) for the Γ-psFa with[1]

$$N = \#(G) - 2M(G). \tag{4.20}$$

We decompose the polynomial $\mathscr{P}(k,q)$ into parts $P_\delta^\eta(k,q)$ which are homogeneous separately in k and q with multi-index degrees δ and η, respectively.

From Eqs. (1.8) to (1.10) with $F = i(V + \lambda W + Z/\lambda)$, we calculate

$$\frac{\partial F}{\partial \chi_{j_\mu}} = \sum_{k=1}^{M(G)} (q_\mu \alpha d^k)(A^{-1})_{kj} g_{\mu\mu} + \frac{i\epsilon - g_{\mu\mu}}{2\lambda(1+\epsilon^2)} \chi_{k_\mu}(A^{-1})_{kj}, \tag{4.21}$$

and

$$\frac{\partial^2 F}{\partial \chi_{j_\mu} \partial \chi_{k_\mu}} = -\frac{\epsilon \delta_{\mu\mu} + i g_{\mu\nu}}{2\lambda(1+\epsilon^2)} (A^{-1})_{kj}. \tag{4.22}$$

Therefore $P((1/i)(\partial/\partial\chi), q)e^F$ consists of terms of the form [cf. Eq. (3.2)]

$$e^F \left(\frac{\partial F}{\partial \chi}\right)^n \left(\frac{\partial^2 F}{\partial \chi \partial \chi}\right)^{\bar{n}}, \tag{4.23}$$

where

$$|n| + 2|\bar{n}| = \delta = \deg r_p P_\delta^\eta(p,q). \tag{4.24}$$

The Γ-psFa therefore consists of a sum of terms each of which is of the form given by Eqs. (3.7)—(3.10) with π_α substituted for R_α, $P = 2$,

$$f^{(\epsilon)}(\alpha,q,\chi) = g(\epsilon) q^\eta \prod_{\substack{\nu=0 \\ k=1}}^{\substack{\nu=3 \\ k=M(G)}} \left(\sum_{i=1}^{M(G)} q_\nu \alpha d^i (A^{-1})_{ik}\right)^{r_\nu^k} \times \left(\sum_{i=1}^{M(G)} \chi_{i_\nu}(A^{-1})_{ik}\right)^{n_\nu^k - r_\nu^k} \prod_{kj=1}^{M(G)} ((A^{-1})_{jk})^{\bar{n}_{kj}}, \tag{4.25}$$

and

$$N = \#(G) - 2M(G) - |n| + |r| - |\bar{n}|, \tag{4.26}$$

and with $0 \leqslant r_\nu^k \leqslant n_\nu^k$, $|n|$ and $|\bar{n}|$ subject to Eq. (4.24), and $g(\epsilon)$ continuous and defined over [0,1].

According to the convergence proof for the forest formula,[3] the superficial degree of divergence of $R^{(\epsilon)}$ is

$$4M(G) - 2\#(G) + |n| + 2|\bar{n}| - 1. \tag{4.27}$$

Combining this with Eq. (4.26), we obtain

$$N > \tfrac{1}{2} - \tfrac{1}{2}|n| + |r| \geqslant \tfrac{1}{2} - \tfrac{1}{2}(|n| - |r|). \tag{4.28}$$

$$N = \#(G) - 2M(G) - |n| + |r| - |\bar{n}|, \tag{4.26}$$

and with $0 \leqslant r_v^k \leqslant n_v^k$, $|n|$ and $|\bar{n}|$ subject to Eq. (4.24), and $g(\epsilon)$ continuous and defined over [0,1].

According to the convergence proof for the forest formula,[3] the superficial degree of divergence of $R^{(\epsilon)}$ is

$$4M(G) - 2\#(G) + |n| + 2|\bar{n}| - 1. \tag{4.27}$$

Combining this with Eq. (4.26), we obtain

$$N > \tfrac{1}{2} - \tfrac{1}{2}|n| + |r| \geqslant \tfrac{1}{2} - \tfrac{1}{2}(|n| - |r|). \tag{4.28}$$

If N is positive, that is, if $|n| - |r| = 0$, Eqs. (3.7) and (3.8) simplify. This case will be discussed later. Otherwise, $|n| - |r| > 0$, and we integrate by parts S times using Lemma 3.2. After folding in a test function, we use Theorem 3.3 to obtain

$$T^{(\epsilon)}(\phi) = \int_{\pi_\alpha} (d\alpha) \frac{g^{(\epsilon)}(\alpha)}{\det A^2} \times \frac{\left[\prod_{k,v}\left(\sum_{i=1}^{M(G)} \chi_i (A^{-1})_{ik}\right)^{n^k - r^k}\right]\left[\prod_{kj}(A^{-1})_{jk}^{n^k}\right] I_S(e^{iV}\phi)}{a^{|n| - |r| - 1}}, \tag{4.29}$$

where $g^{(\epsilon)}(\alpha)$ is defined, continuous, and bounded almost everywhere in π_α with ϵ dependence restricted to a separate factor which is a function of ϵ alone, defined and continuous for $\epsilon \in [0,1]$; the quantity a is given by the positive root from Eq. (3.10),

$$I_S = a^S H_S, \tag{4.30}$$

H_S is given by Lemma 3.2, and S is the number of integrations by parts, i.e.,

$$S = |n| - |r| - 1, \text{ and} \tag{4.31}$$

$$n^k = |n_v^k|, \text{ and } r^k = |r_v^k|. \tag{4.32}$$

Lemma 4.6: There is a positive number B independent of the breaking parameters and of α, such that $(\forall i)(\forall k)$ and for $1 \leqslant v \leqslant 3$, if $a \neq 0$, then

$$\left| \frac{\chi_{iv}(A^{-1})_{ik}}{\left[\sum_{ij} \chi_i (A^{-1})_{ij} \chi_{j_k}\right]^{1/2}} \right| \leqslant \frac{B}{\sqrt{\alpha_k}}.$$

Proof: By Theorem 4.5, $(\forall i)(\forall k)(\exists B_1 > 0)$

$$(A^{-1})_{ik}^2 \leqslant B_1^2/(\alpha_k \alpha_i)$$

implies

$$|\chi_{iv}(A^{-1})_{ik}| \leqslant |\chi_{iv}| B_1/\sqrt{\alpha_k \alpha_i}\,. \tag{4.33}$$

Furthermore, by Lemma 4.1 of Ref. 1, for some $B_2 > 0$

$$\sum_{ij} (\chi_i (A^{-1})_{ij} \chi_j)_{kk} \geqslant \sum_i \chi_{i_k}^2 B_2/\alpha_i. \tag{4.34}$$

Since B_1 and B_2 are independent of χ and α, the theorem follows with $B = B_1/B_2$, and B depends only on the structure of the circuit-based graph G.

We are now ready to obtain the principal results of this section.

Theorem 4.7: The terms of $T^{(\epsilon)}(\xi,q)$, here denoted by $T^{(\epsilon)}$, for which $|n| - |r| \neq 0$ vanish in the regulator limit, whether that is taken before or after the distributional limit:

$$\lim_{\xi \to 0} T^{(\epsilon)}(\phi) = \lim_{\xi \to 0} \lim_{\epsilon \to 0} T^{(\epsilon)}(\phi) = \lim_{\xi \to 0} T^{(0)}(\phi) = 0,$$

for any $\phi \in \mathscr{S}$.

Proof: In Eq. (4.29) we make the change of variables defined by Eqs. (4.6) and (4.7). By Theorems 4.2–4.5 and Lemma 4.6, the resulting integrand is bounded over the region of integration for some positive constant B by

$$Bg^{(\epsilon)}(\alpha) \prod_{i=0}^{M(G)-1} t_{G_i}^{\{\#(G) - 1 - 2M(G) + 2i\}}$$
$$\times \prod_{k=1}^{M(G)} (\alpha_k)^{-1/2(n^k - r^k) + \sum_j n^{kj}} \frac{\sup_i |\chi_i|}{M(G)} I_S(e^{iV}\phi)$$
$$= \sup_i |\chi_i| B \prod_{i=0}^{M(G)-1} t_{G_i}^{\{\#(G_i) - 2M(G_i) - \frac{1}{2}\sum_{k=i+1}(n^k - r^k + \sum_j 2n^{kj}) - 2\}}$$
$$\times I_S(e^{iV}\phi). \tag{4.35}$$

If one additional integration by parts is done on Eq. (4.29), the integrand is in the same way bounded over the region of integration for some positive constant B by

$$B \prod_{i=0}^{M(G)-1} t_{G_i}^{\{\#(G_i) - 2M(G_i) - \frac{1}{2}\sum_{k=i+1}(n^k - r^k + \sum_j 2n^{kj}) - 1\}} I_{S+1}(e^{iV}\phi) \tag{4.36}$$

We distinguish two cases. In the first of these, $|n| - |r|$ is odd, and we perform the additional integration by parts. In this case, the bound of Eq. (4.36) applies, and I_{S+1} is the odd case of Lemma 3.2 and Eq. (4.30). It was shown in Theorems 3.4, 3.5, and 3.6 that $I_{S+1}(e^{iV}\phi)$ is continuous and bounded almost everywheres in R_α in the epsilon to zero limit. The t integral which defines I_{S+1} also clearly converges uniformly in ξ since the ξ dependence is restricted to a factor in the integrand of $\sin(at)$ and the proof of convergence eliminates this factor in the dominating integrand (cf. the proof of theorem 3.6).

The (α) integrals, now re-expressed in terms of the t_H variables, also converge uniformly in π_α. In the bound, Eq. (4.36), the power on each t_{G_i} is at least $-\frac{1}{2}$. To see this, note that setting the internal momenta of the lines of $G - G_i$ [cf., Eq. (4.5)] constant defines a hyperplane in internal momentum space. Since $G - G_i$ consists of the first i lines of T_π^* together with all lines of G dependent on these first i lines, the number of independent variable lines on this hyperplane is

$$M(G_i) = M(G) - i. \tag{4.37}$$

According to the convergence theorem of BPHZ (3), the superficial degree of divergence has been reduced on every hyperplane by the subtractions of the forest formula so as to give convergence on every hyperplane. The power on each t_{G_i} is just one less than minus one-half the superficial degree of divergence. Therefore this power is at least $-\frac{1}{2}$. Therefore the bound, Eq. (4.36), guarantees that the integral over π_α converges uniformly in ξ. Since

$$\lim_{\xi \to 0} a = 0,$$

and the t integral defining I_{s+1} contains a factor of $\sin at$, taking the ξ limit inside of the α (i.e., t_H) and t integrals implies Theorem 4.7 for the case that $|n| - |r|$ is odd.

Otherwise $|n| - |r|$ is even and we partition $0 \leqslant t_{G_i} \leqslant 1$ into two regions:

(1) $0 \leqslant t_{G_i} < \delta$, and

(2) $\delta \leqslant t_{G_i} \leqslant 1$.

In this case, by Eq. (4.31), $S + 1$ is even, and the factor of $\sin at$ is not available to help us if we do the additional integration by parts. On the other hand, if we do not do the additional integration by parts, the possibly negative powers of t_{G_i} in the bound of Eq. (4.35) makes the bound useless in neighborhoods of the t_{G_i} origin. The partition of the region enables us to overcome these difficulties.

There will then be $M(G)$ regions in which at least one of the t_{G_i} obeys (1) and one more region in which all of the t_{G_i} obey (2), and π_α is contained in the union of these regions.

For the regions of the first kind, we perform the additional integration by parts and use the bound of Eq. (4.36) to show that each region contributes at most

$$2B\delta^{\frac{1}{2}} I_{S+1}(e^{iV}\phi)$$

to the values of the integrals. By theorems 3.4 and 3.5, $I_{s+1}(e^{iV}\phi)$ is bounded and continuous over R_α for all $\epsilon \in [0,1]$. Therefore, for any positive ϵ_1, we may choose δ so small that the total contribution from the regions of the first kind is less than $\epsilon_1/2$.

For the remaining region, all of the t_{G_i} are subject to (2). In this case, we do not perform the additional integration by parts. The bound of Eq. (4.35) produces, on integration over the region (2), possibly negative powers of δ, but this bound also contains a factor of $\sup_i |\chi_i|$. We simply choose the ξ_l [cf. Eq. (1.7)] so small that the total contribution from this region is also less than $\epsilon_1/2$.

In other words,

$$(\forall \epsilon_1 > 0)(\forall \epsilon \in [0,1])(\exists b(\epsilon)) \sup_l \|\xi_l\| < b(\epsilon)$$

implies $T^{(\epsilon)}(\phi) < \epsilon_1$. This completes the proof of Theorem 4.7.

Theorem 4.8: For any subtracted amplitude of BPHZ, $J_\Gamma^{(\epsilon)}(q)$ with Γ-psFa $J^{(\epsilon)}(\xi, q)$, the regulator $(\xi \to 0)$ and distributional $(\epsilon \to 0+)$ limits may be taken interchangeably. That is, for any $\phi \in \mathscr{S}$, the test-function space of Eq. (3.1) of functions on $\mathbb{R}^N$ with N at least four times as big as the number of independent external momenta,

$$\lim_{\xi \to 0} \lim_{\epsilon \to 0} \int (dq)\, J_\Gamma^{(\epsilon)}(\xi, q)\phi(q)$$
$$= \lim_{\epsilon \to 0} \lim_{\xi \to 0} \int (dq)\, J^{(\epsilon)}(\xi, q)\phi(q)$$
$$= \lim_{\epsilon \to 0} \int (dq)\, J^{(\epsilon)}(q)\phi(q).$$

Proof: By Theorem 4.7, in whatever order the limits are taken, the terms for which $|n| - |r| \neq 0$ contribute nothing. Otherwise $|n| - |r| = 0$, and the Γ-psFa terms are given by Eqs. (3.7) to (3.10) with Eqs. (4.25) and (4.26). By Eq. (4.28),

$$N + 1 > 3/2.$$

This assures the necessary convergence properties of the integral of Eq. (3.9), and Theorems (3.4) and (3.5) lead to a bounding of $I_0(e^{iV}\phi) = H(e^{iV}\phi)$ over π_α for $\epsilon \in [0,1]$. Equation (3.8) now contributes only a factor of unity to the Hepp sector term because N is here positive.

Upon making the change of variables of Eqs. (4.6) and (4.7), with no integration by parts but otherwise following the procedure by which we obtained Eq. (4.35) and inserting the bound on $I_0(e^{iV}\phi)$, the integrand of the Hepp sector term is now bounded by

$$B \prod_{i=1}^{M(G)-1} t_{G_i}^{\{\#(G_i) - 2M(G_i) - 1 - \sum_{k,l;\, k \neq i+1} n^{kl}\}}, \tag{4.38}$$

where B is some positive constant.

By Eqs. (4.26) and (4.28), the integral of this bound over the bounded region π_α is absolutely convergent. Therefore the integral which is the Γ-psFa converges uniformly in ξ and $\epsilon \in [0,1]$. Since the integrand of the Γ-psFa is continuous in ξ, Theorem 4.8 then follows by taking the limits inside of all the t_{ll} and t integrals.

Since these limits have also been shown to exist and to be continuous in ϕ (which follows from Theorem 3.6 if one notes that I_0 is bounded independently of ξ), we have also proven

Theorem 4.9: A Feynman amplitude renormalized according to the Zimmermann forest formula is, in the epsilon to zero limit, a tempered distribution in the external momenta.

ACKNOWLEDGMENT

The author wishes to acknowlege his special gratitude to John Lowenstein for teaching the author about Hepp sectors and many other things.

[1] R. R. Hampton, J. Math. Phys. **21**, 1833 (1980).

[2] W. Zimmermann, "Local operator products and renormalization in quantum field theory," in *Lectures on Elementary Particles and Quantum Field Theory*, 1970 Brandeis lectures (MIT Press, Cambridge, MA, 1970), pp. 399–589.

[3] W. Zimmermann, Commun. Math. Phys. **15**, 208 (1969).

[4] N. N. Bogoliubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields* (Wiley, New York, 1959).

[5] K. Hepp, Commun. Math. Phys. **2**, 301 (1966).

[6] K. Hepp. "Renormalization Theory," in *Statistical Mechanics and Quantum Field Theory*, Université de Grenoble Summer School of Theoretical Physics, Les Houches 1970 (Gordon and Breach, New York, 1971).

[7] W. Zimmermann, "Remark on equivalent formulations for Bogoliubov's method of renormalization," in *Renormalization Theory*, Proceedings of the NATO Advanced Study Institute of August, 1975 at 'Ettore Majorana' in Erice, Italy (Reidel, Dordrecht, Holland, 1976), pp. 161–170.

[8] A. H. Mueller, Phys. Rev. D **9**, 963 (1974); **18**, 3705 (1978); and S. Gupta and A. H. Mueller, *ibid.* **20**, 118 (1979).

[9] R. R. Hampton, Commun. Math. Phys. **67**, 187 (1979).

[10] W. Zimmermann, Commun. Math. Phys. **15**, 208 (1969); Y. Hahn and W. Zimmermann, *ibid.* **10**, 330 (1968).

[11] I. M. Gel'fand and G. E. Shilov, *Generalized Functions* (Academic, New York, 1968), Vols. 1 and 2.

[12] E. R. Speer, *Generalized Feynman Amplitudes* (Princeton U. P., Princeton, NJ, 1969).

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

[14] W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer, Berlin, 1966).

[15] J. H. Lowenstein, "An operator-product light-cone expansion in renormalized perturbation theory," New York University preprint NYU/TR1/77 (1977).

Radiation gauge covariance

Jorge Krause

Instituto de Física, Universidad Católica de Chile, Casilla 114-D, Santiago, Chile
(Received 7 May 1980; accepted for publication 3 October 1980)

A manifestly covariant relation is discussed between electromagnetic gauge transformations and Lorentz transformations, while analyzing the contributions to the free field 4-potential coming exclusively from lightlike momenta. Transverse 4-potentials are thus introduced relative to each inertial frame which belong in the Lorentz gauge, behave as completely invariant objects under general gauge transformations, and give rise to one and the same frame-free electromagnetic field tensor. The (proper orthochronous) Lorentz covariant transformation law of the gauge-invariant transverse 4-potential is established, from the standpoint of the active transformation from one inertial observer to another.

PACS numbers: 11.10.Np, 11.30.Cp, 03.70. + k

1. INTRODUCTION

In this paper we present a free field relationship between the group of gauge transformations in electromagnetic theory and the Lorentz group. A basis for a description of free photons is thus obtained, which seems much more consistent with all other particles, since we drop the unphysical variables while retaining the relativistic invariance of the whole formalism.

It is well known that for zero-rest-mass particles, obeying the zero-mass Proca equations (Maxwell's equations), one has a larger set of symmetry groups than those holding for massive Proca particles. Indeed, the behavior of the electromagnetic potentials under the group of Poincaré transformations uniquely characterizes these quantities as components of a 4-vector in flat space-time. Photons have zero rest-mass, however, and therefore Maxwell's equations are also gauge invariant.[1] Thus, to the 10-parmetric Poincaré group (which describes the external symmetries of the system) one has to add the invariance group corresponding to gauge transformations of the second kind. These electromagnetic gauge transformations consitute a realization of an additive function group which maps the manifold of all allowable potentials into itself. In contrast with the Poincaré invariance, the gauge mapping corresponds to an internal symmetry of the system.[2]

A dynamical formalism containing gauge-dependent degrees of freedom leads to well-known difficulties in field theory. In order to overcome these difficulties, one has to separate the field variables into two sets, one being gauge-dependent quantities (i.e., gauge artifacts), while the others are gauge-invariant.[3] By the choice of some suitable gauge conditions, one may then drop the unphysical variables, retaining only those gauge-independent quantities associated with the true degrees of freedom of the system; these are unconstrained physical variables able to describe the configuration of the field. However, this approach works well if (and only if) the variable classification corresponds to a gauge-independent scheme and, moreover, if (and only if) it entails a Lorentz invariant classification.

A similar feature is quite familiar in general relativity, where the coordinate symmetry group is also a function group which operates as a gauge symmetry for the gravitational potentials.[4] Nevertheless, the whole invariance group of the general theory mixes the external with the internal symmetries, in such a strong way, that it becomes impossible to separate this group (in a well established manner) into a Lie group and an additive group involving arbitrary gauge functions.[5]

Thus, it seems to be important to search for the relations which might exist between the electromagnetic gauge group and the Poincaré group of special relativity,[6] in order to be able to identify the true configuration variables of the radiation field without spoiling its space-time symmetry. In general relativity this would be a very complicated undertaking (perhaps raising an impossible problem). It is the main purpose of this paper to show how this problem can be solved for the free radiation field of electrodynamics.[7]

The very problem posed by the internal symmetry of the second kind is that, when handling the potentials with their full gauge freedom, we are dealing with redundant variables for the description of the true degrees of freedom of the electromagnetic field.[8] When special gauges are imposed as subsidiary constraints, gauge subgroups of the full gauge group become realized in electromagnetic theory. Among these special gauges, the Lorentz gauge constraint presents two well known and important features: 1) it preserves the Lorentz invariance of the theory,[9] and 2) it brings Maxwell's equations into the form of the zero-mass Proca equations. The Lorentz gauge introduces just one subsidiary condition for the components of the 4-potential and, therefore, leaves three independent variables. Within the Lorentz gauge, however, the 4-potential is not uniquely defined, since restricted gauge transformations are still allowed. Hence, the Lorentz gauge constraint, by itself, does not completely disclose the true degrees of freedom of the Maxwell field. Further reduction is still necessary.

We wish to remark here that, from a group-theoretic point of view, the existence of both Poincaré and gauge invariance of the electromagnetic field means that the 4-potential plays a double geometric role: 1) It provides a basis for a linear irreducible representation of the Lorentz group, and 2) it furnishes a "basis" for a realization of the continuous abelian gauge group of the second kind.[10] It becomes clear then that in order to isolate the true degrees of freedom of the radiation field, we have to search for a completely "irreducible" reali-

 0022-2488/81/081759-08$1.00 © 1981 American Institute of Physics

zation of the full gauge group. In other words, we have to identify the part of $A^\mu(x)$ which remains completely invariant under general gauge transformations, while, moreover, we have to arrive at this irreducible basis (say) in a manifestly Lorentz invariant fashion. As we shall see, this gauge-invariant Lorentz-covariant realization can be attained within any given inertial frame.

The program of this note follows. In Sec. 2 we start with a brief review of the free-field-potential formalism in momentum representation. In Sec. 3 we analyze further the contributions to the free potentials coming exclusively from lightlike photons (i.e., physical free photons); we identify the locus of the Lorentz gauge, and we end up with a transverse 4-potential relative to a v-frame as the completely gauge-invariant potential within the Lorentz gauge. The kernel of the radiation's gauge projection formula is next analyzed in Sec. 4, while in Sec. 5 we briefly show how the Coulomb gauge is regained within the Lorentz gauge in the present formalism. In Sec. 6 we discuss the covariant transformation law of the completely gauge-invariant potential from one inertial observer to another. Finally, the two degrees of freedom of the electromagnetic field are briefly presented in Sec. 7.

2. THE FREE FIELD POTENTIALS IN MOMENTUM REPRESENTATION REVISITED

For the sake of completeness we begin our work with a brief review of the familiar momentum-space formalism for the free electromagnetic field. Let us consider the following plane wave superposition representing the free vector potential:

$$A^\mu(x) = (2\pi)^{-2} \int d^4k\, \tilde{A}^\mu(k) \exp(ikx), \tag{2.1}$$

where kx stands for $k_\mu x^\mu = k^0 t - \mathbf{k}\cdot\mathbf{x}$. The reality condition for $A\mu(x)$ requires $\tilde{A}^*_\mu(k) = \tilde{A}_\mu(-k)$. If no gauge condition is assumed, the free field equations take the form

$$k^2 \tilde{A}_\mu(k) - k_\mu k^\nu \tilde{A}_\nu(k) = 0, \tag{2.2}$$

which is manifestly invariant under general gauge transformations; namely,

$$\tilde{A}'_\mu(k) = \tilde{A}_\mu(k) + ik_\mu g(k), \tag{2.3}$$

where $g(k)$ is an arbitrary scalar function defined in momentum-space, provided its Fourier transform exists and conforms to the reality condition.

Equation (2.2) behaves rather singularly on the light cone. We thus analyze it first off this locus, where it has a simple geometric meaning; i.e., since $k^2 \neq 0$ there, we have

$$[\delta^\mu_\nu - (k_\lambda k^\lambda)^{-1} k^\mu k_\nu]\tilde{A}^\nu(k) = 0. \tag{2.4}$$

Hence, the Fourier amplitude $\tilde{A}^\mu(k)$ is a purely longitudinal field everywhere off the light cone, and we write

$$\tilde{A}^\mu_{(\text{off})}(k) = ik^\mu \gamma(k) \quad (k^2 \neq 0), \tag{2.5}$$

where $\gamma(k)$ is an arbitrary scalar density. As a matter of fact, it is clear that these off-light-cone longitudinal contributions to the free radiation field are devoid of physical meaning, since they correspond to a pure gauge artifact. Therefore, no true degrees of freedom of the free electromagnetic field can be obtained from these contributions.

Next, let us consider Eq. (2.2) *on* the light cone. Since now k^μ is an arbitrary lightlike vector, Eq. (2.2) splits into two homogeneous equations:

$$k^2 \tilde{A}^\mu(k) = 0, \tag{2.6}$$

and

$$k_\mu \tilde{A}^\mu(k) = 0, \tag{2.7}$$

where $k^2 = 0$. The first equation is singular, and its general solution is quite familiar:

$$\tilde{A}^\mu_{(\text{on})}(k) = \delta(k^2) f^\mu(k), \tag{2.8}$$

where the amplitudes $f^\mu(k)$ are defined everywhere on the light cone and satisfy $f^*_\mu(k) = f_\mu(-k)$ for a real field. The second equation [Eq. (2.7)] tells us that these cone-supported Fourier amplitudes are transverse. Thus we have

$$k_\mu f^\mu(k) = 0, \tag{2.9}$$

everywhere on the light cone.

Of course, Eq. (2.9) means that the light cone contribution to the free field comes from a 4-potential which automatically belongs in the Lorentz gauge. In this sense, one could say that the Lorentz gauge affords the natural gauge of the theory. Furthermore, it is clear that the same gauge transformation which fixes the Lorentz gauge also eliminates the longitudinal off-the-light-cone contributions to the free potentials.

Moreover, Eq. (2.9) shows that the $f^\mu(k)$ field has the following general structure (since k^μ is lightlike):

$$f^\mu(k) = \alpha^\mu(k) + ik^\mu \beta(k), \tag{2.10}$$

where the vector field $\alpha^\mu(k)$ has to be spacelike transverse, i.e.,

$$\alpha_\mu(k)\alpha^\mu(k) < 0, \tag{2.11}$$

$$k_\mu \alpha^\mu(k) = 0, \tag{2.12}$$

while the scalar $\beta(k)$ is an arbitrary field defined on the cone.

Hence, the most general solutions to the homogeneous Maxwell's equations in momentum representation are of the form:

$$\tilde{A}^\mu(k) = \delta(k^2)[\alpha^\mu(k) + ik^\mu\beta(k)] + ik^\mu\gamma(k); \tag{2.13}$$

i.e., correspondingly, the following general decomposition of the free 4-potential holds:

$$A^\mu(x) = A^\mu{}_{(\alpha)}(x) + A^\mu{}_{(\beta)}(x) + A^\mu{}_{(\gamma)}(x). \tag{2.14}$$

The $A^\mu{}_{(\alpha)}(x)$ and $A^\mu{}_{(\beta)}(x)$ components have their Fourier support on the light cone in momentum space. Each of them necessarily belongs in the Lorentz gauge and, therefore, $A^\mu{}_{(\alpha)}(x)$ and $A^\mu{}_{(\beta)}(x)$ separately satisfy the zero-mass Proca equation. The third contribution, $A^\mu{}_{(\gamma)}(x)$, comes completely from the off-cone region in momentum space, where it satisfies Maxwell's homogeneous equations in their general gauge-invariant form. Furthermore, it is clear that the two components $A^\mu{}_{(\beta)}(x)$ and $A^\mu{}_{(\gamma)}(x)$ are just gauge artifacts, since both are gradients of scalar fields.

3. THE COMPLETELY GAUGE-INVARIANT POTENTIALS

In this section we analyze further the contribution to the potentials coming exclusively from the light cone in momentum space. Clearly, it is only this part of the potentials which genuinely corresponds to zero-mass free photons, and therefore it must be enough for a complete description of the free field.

Recalling Eqs. (2.8) and (2.10), we have

$$\tilde{A}^{\mu}{}_{(\text{on})}(k) = \delta(k^2)[\alpha^{\mu}(k) + ik^{\mu}\beta(k)]. \tag{3.1}$$

This vector belongs in the Lorentz gauge. According to the previous discussion, one observes that the vector $f^{\mu}(k)$ [cf. Eq. (2.10)] lies in the hyperplane orthogonal to the lightlike vector k^{μ}. This hyperplane is a 3-flat tangent to the light cone along k^{μ}, and represents the locus of the Lorentz gauge in momentum space. Figure 1 is a sketchy representation of the Lorentz gauge hyperplane (i.e., plane $OKLT$ in that figure), with the vectors $f^{\mu}(k)$, $\alpha^{\mu}(k)$, k^{μ}, and $k^{\mu}\beta(k)$, all belonging in this locus, as shown.

In this sense, once the vector $f^{\mu}(k)$ is given, it becomes clear that every displacement along the lightlike line TL (of Fig. 1) corresponds to an allowable gauge transformation within the Lorentz gauge. Therefore, all vectors drawn from O up to the line TL are physically equivalent to $f^{\mu}(k)$, in the sense that they all belong to the same electromagnetic field. Moreover, it is easy to show (as we shall do presently) that these spacelike OF-vectors (with F moving along TL) all have the same Minkowski norm; i.e., Lorentz gauge motion (along TL) preserves the norm of the $f^{\mu}(k)$ vectors. Here an intimate relation reveals itself between Lorentz transformations and electromagnetic gauge transformations taking place within the Lorentz gauge. We will come back to this issue in Sec. 6.

Of course, we cannot claim for $A^{\mu}{}_{(\alpha)}(x)$ the property of being a completely gauge-invariant potential. In effect, it may happen that the spacelike amplitude $\alpha^{\mu}(k)$

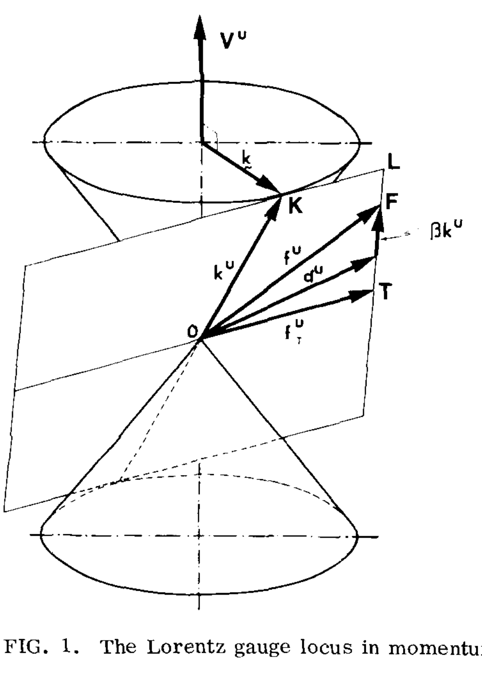


FIG. 1. The Lorentz gauge locus in momentum 4-space.

has a gauge component along k^{μ} relative to a given Cartesian frame (as a glance at Fig. 1 neatly shows). In other words, for the sake of having a completely gauge-irreducible 4-potential one has to introduce a fixed inertial frame, specified by a given 4-velocity v^{μ}. Indeed, the vector $f^{\mu}_T(k) = (OT)^{\mu}$ (shown in Fig. 1) is obviously the only vector, belonging in the same gauge as f^{μ}, which has no component along k^{μ} *from the point of view of the v-frame*. (As a matter of fact, this is the standpoint adopted in Fig. 1.) We observe that f^{μ}_T is a spacelike vector simultaneously orthogonal to k^{μ} and v^{μ}.

Let us discuss these features in a manifestly Lorentz covariant fashion. Once v^{μ} is given, we define a new decomposition of $f^{\mu}(k)$, instead of Eq. (2.10), according to the following scheme:

$$f^{\mu}(k) = f^{\mu}_T(k;v) + ik^{\mu} f_L(k;v), \tag{3.2}$$

such that

$$v_{\mu} f^{\mu}_T(k;v) = 0, \tag{3.3}$$

$$k_{\mu} f^{\mu}_T(k;v) = 0. \tag{3.4}$$

Covariant expressions for the new Fourier amplitudes are easily obtained if we use the identity

$$[\delta^{\mu}_{\nu} - (k_{\lambda}v^{\lambda})^{-1}k^{\mu}v_{\nu}]k^{\nu} = 0, \tag{3.5}$$

which holds everywhere in k-space. We thus have

$$f^{\mu}_T(k;v) = [\delta^{\mu}_{\nu} - (k_{\lambda}v^{\lambda})^{-1}k^{\mu}v_{\nu}]f^{\nu}(k), \tag{3.6}$$

$$f_L(k;v) = -i(k_{\lambda}v^{\lambda})^{-1}v_{\nu}f^{\nu}(k), \tag{3.7}$$

quite directly, that is, without previous recourse to decomposition (2.10). (These transformations obviously preserve the reality conditions.)

The rank-two tensor

$$\tilde{R}^{\mu}{}_{\nu}(k;v) = \delta^{\mu}_{\nu} - (k_{\lambda}v^{\lambda})^{-1}k^{\mu}v_{\nu} \tag{3.8}$$

behaves as the transverse left-projector within the Lorentz gauge hyperplane, since (from the left) it brings every transverse vector $f^{\mu}(k)$ into its gauge-irreducible form $f^{\mu}_T(k;v)$ relative to the v-frame. From the right $\tilde{R}^{\mu}_{\nu}(k;v)$ behaves as the identity on $f^{\mu}(k)$. One shows:

$$\tilde{R}^{\lambda}_{\mu}\tilde{R}_{\lambda\nu} = \tilde{R}_{\mu\nu} + \tilde{R}_{\nu\mu} - \eta_{\mu\nu} \tag{3.9}$$

(where $\eta_{\mu\nu}$ stands for the Minkowski metric), and therefore

$$\eta_{\mu\nu} f^{\mu}_T(k;v) f^{\nu}_T(k;v) = \eta_{\mu\nu} f^{\mu}(k) f^{\nu}(k), \tag{3.10}$$

as we have already remarked.

Representing the new decomposition (3.2) we have, in coordinate space,

$$A^{\mu}(x) = A^{\mu}_T(x;v) + A^{\mu}_L(x;v) \tag{3.11}$$

(relative to the v-frame), instead of decomposition (2.14), for the same electromagnetic potentials. Clearly, the field

$$A^{\mu}_T(x;v) = (2\pi)^{-2}\int d^4k\,\delta(k^2)[\delta^{\mu}_{\nu} - (k_{\lambda}v^{\lambda})^{-1}k^{\mu}v_{\nu}]f^{\nu}(k)\exp(ikx) \tag{3.12}$$

corresponds to the *transverse 4-potential relative to the v-frame*, and therefore it represents the only completely gauge-irreducible basis for the realization of the full electromagnetic gauge group we were searching for. (We further analyze this matter presently.)

After one has been able to identify the momentum geometry construct yielding the irreducible potentials, one may generalize the result stated in Eq. (3.12), for the sake of handiness. In effect, since the left-projection formula (3.5) holds quite generally for all kinds of k^μ wave vectors, one has

$$\delta(k^2) f^\mu_T(k; v) = [\delta^\mu_\nu - (k_\lambda v^\lambda)^{-1} k^\mu v_\nu] \tilde{A}^\nu(k) \tag{3.13}$$

everywhere in momentum 4-space, where $\tilde{A}^\mu(k)$ corresponds to a free Maxwell field without assuming any gauge [cf. Eq. (2.13)]. Hence, the following expression definitely obtains for the transverse potentials relative to a v-frame:

$$A^\mu_T(x; v) = (2\pi)^{-2} \int d^4k [\delta^\mu_\nu - (k_\lambda v^\lambda)^{-1} k^\mu v_\nu] \tilde{A}^\nu(k) \exp(ikx), \tag{3.14}$$

where the Fourier amplitudes $\tilde{A}^\nu(k)$ satisfy Eq. (2.2). Starting from definition (3.14), one immediately shows that $A^\mu_T(x; v)$ is endowed with the following properties:

$$v_\mu A^\mu_T(x; v) = 0, \tag{3.15}$$

$$\nabla_\mu A^\mu_T(x; v) = 0, \tag{3.16}$$

$$\Box A^\mu_T(x; v) = 0. \tag{3.17}$$

Moreover, $A^\mu_T(x; v)$ is a completely gauge-invariant 4-potential, in the sense that a gauge transformation of $A^\mu(x)$, according to Eqs. (2.3) and (3.5), induces no change on $A^\mu_T(x; v)$. On the other hand, if one gauge-transforms $A^\mu_T(x; v)$ directly, i.e., $A'^\mu_T(x; v) = A^\mu_T(x; v) + \nabla^\mu G(x; v)$, and requires that A'^μ_T is a new transverse 4-potential [cf. Eqs. (3.15), (3.16), and (3.17)], then it follows that $G(x; v) = 0$.

Equation (3.14) shows explicitly how to reduce a given 4-potential to its completely gauge-invariant transverse form, relative to any given inertial working frame. Equation (3.14) is a gauge transformation associated with the originally given 4-potential and inertial scaffold themselves. In Sec. 5 we shall discuss further the (already evident) fact that these $A^\mu_T(x; v)$ potentials belong in the Coulomb gauge of the v-observer. Henceforth we shall refer to Eq. (3.14) as the radiation gauge projection formula.

4. KERNEL OF THE RADIATION GAUGE PROJECTION

The fundamental result (3.14) invites us to introduce the following kernels:

$$R^\mu{}_\nu(x; v) = (2\pi)^{-4} \int d^4k [\delta^\mu_\nu - (k_\lambda v^\lambda)^{-1} k^\mu v_\nu] \exp(ikx), \tag{4.1}$$

$$G^\mu{}_\nu(x; v) = (2\pi)^{-4} \int d^4k (k_\lambda v^\lambda)^{-1} k^\mu v_\nu \exp(ikx), \tag{4.2}$$

corresponding to the decomposition of the identity:

$$R^\mu{}_\nu(x; v) + G^\mu{}_\nu(x; v) = \delta^\mu_\nu \delta^{(4)}(x). \tag{4.3}$$

In this manner we get the radiation gauge projection formulas

$$A^\mu_T(x; v) = \int d^4y R^\mu{}_\nu(x - y; v) A^\nu(y), \tag{4.4}$$

$$A^\mu_L(x; v) = \int d^4y G^\mu{}_\nu(x - y; v) A^\nu(y), \tag{4.5}$$

out of any free 4-potential $A^\mu(x)$ whatsoever.

Some immediate properties of these kernels follow:

$$G^\mu{}_\mu(x; v) = \delta^{(4)}(x), \tag{4.6}$$

$$v_\mu G^\mu{}_\nu(x; v) = v_\nu \delta^{(4)}(x), \tag{4.7}$$

$$\nabla_\nu G_\mu{}^\nu(x; v) = \nabla_\mu \delta^{(4)}(x). \tag{4.8}$$

Defining the scalar

$$G(x; v) = -i(2\pi)^{-4} \int d^4k (k_\lambda v^\lambda)^{-1} \exp(ikx), \tag{4.9}$$

one easily shows

$$G^\mu{}_\nu(x; v) = (v^\lambda \nabla_\lambda)^{-1} (v_\nu \nabla^\mu) \delta^{(4)}(x), \tag{4.10}$$

since clearly

$$G(x; v) = (v^\lambda \nabla_\lambda)^{-1} \delta^{(4)}(x). \tag{4.11}$$

The kernel $G(x; v)$ is an *invariant D-function*. It seems that the form presented in Eq. (4.9) cannot be reduced to any of the well-known invariant functions which occur in quantum field theory.[11] [The evaluation of $G(x; v)$, however, is quite simple *once we introduce its proper frame;* i.e., once we set $v^\mu = (1, \mathbf{0})$].

Of course, the problem already tackled in Sec. 3 can also be solved more easily in a symbolic manner. We have followed a more lengthy path of argumentation in momentum space, in order to explicitly show the underlying geometry of a free electromagnetic field. We think that approach more conducive to the physical understanding of the issue. However, for the sake of having a more handy notation, the statement of the problem and its symbolic operator solution briefly follows.

One has to solve Maxwell's free field equations for the potentials, when no gauge is assumed:

$$\Box A^\mu(x) - \nabla^\mu \nabla_\nu A^\nu(x) = 0, \tag{4.12}$$

while searching for a new 4-potential A^μ_T, out of A^μ exclusively, such that it satisfies condition (3.15). [We do not need to impose the Lorentz constraint (3.16), for it comes out automatically]. In order to solve this problem formally, it is enough to assume the existence of a linear gauge transformation able to produce the desired answer, i.e., $A^\mu_T(x; v) = A^\mu(x) + \nabla^\mu K_\nu(x; v; \nabla) A^\nu$, say, where A^μ is any solution to Eq. (4.12). Then, after some simple manipulations, one gets the answer:

$$A^\mu_T(x; v) = [\delta^\mu_\nu - (v_\lambda \nabla^\lambda)^{-1} \nabla^\mu v_\nu] A_\nu. \tag{4.13}$$

One shows next that, because of the homogeneous Maxwell's equations (4.12) satisfied by A^μ, the transverse field A^μ_T belongs in the Lorentz gauge. Equation (4.13) corresponds to a gauge transformation and, moreover, if we arbitrarily gauge-transform the potentials, the new $A^\mu_T(x; v)$ potentials remain quite the same. Thus, $A^\mu_T(x; v)$ is a completely gauge-invariant 4-potential for the free electromagnetic field. Of course, the inverse differentiation operators are defined, as usual, by means of the corresponding Fourier transforms. In this manner, we arrive back to the explicit solution (3.14), previously found in Sec. 3.

5. THE COULOMB GAUGE REGAINED AS A SUBGAUGE OF THE LORENTZ GAUGE

Before we go any further, it is important to examine the electromagnetic field tensor under the scope of the present approach. Once $A^\mu(x)$ is already decomposed as in Eq. (3.11), where $A_L^\mu(x;v)$ is just a gauge artifact, we have

$$F_{\mu\nu}(x)=\nabla_\mu A_\nu^T(x;v)-\nabla_\nu A_\mu^T(x;v). \tag{5.1}$$

Clearly, the electromagnetic field tensor must be an absolutely v-independent object. Indeed, from Eq. (4.13), we get

$$\partial A_T^\mu(x;v)/\partial v^\nu=-(v_\lambda\nabla^\lambda)^{-1}\nabla^\mu A_\nu(x)+(v_\lambda\nabla^\lambda)^{-2}\nabla^\mu\nabla_\rho A^\rho(x), \tag{5.2}$$

which holds formally for all $v_\mu v^\mu>0$. Therefore, the following identity obtains

$$\partial F_{\mu\nu}(x)/\partial v^\lambda=0 \tag{5.3}$$

(as it should), even when $v_\mu v^\mu=1$.

By the same token, if u^μ is another 4-velocity ($u^\mu\neq v^\mu$) one has, for the same field:

$$F_{\mu\nu}(x)=\nabla_\mu A_\nu^T(x;u)-\nabla_\nu A_\mu^T(x;u), \tag{5.4}$$

where now

$$A_T^\mu(x;u)=[\delta_\nu^\mu-(\nabla_\lambda u^\lambda)^{-1}\nabla^\mu u_\nu]A^\nu(x); \tag{5.5}$$

$$u_\mu A_T^\mu(x;u)=0, \tag{5.6}$$

$$\nabla_\mu A_T^\mu(x;u)=0, \tag{5.7}$$

$$\Box A_T^\mu(x;u)=0. \tag{5.8}$$

It is important to observe that these relations hold with respect to any inertial frame (not necessarily the u-frame). A gauge transformation relates these two transverse irreducible 4-potentials; namely, we have

$$A_T^\mu(x;u)=A_T^\mu(x;v)+A_L^\mu(x;v)-A_L^\mu(x;u),\ \text{quite simply.}$$

Nevertheless, a more enlightening form for this gauge transformation will be shown presently, in terms of the active Lorentz tranformation which relates v^μ with u^μ (cf. Sec. 6).

The physical meaning of the Fourier integral which corresponds to Eq. (5.1) is clear: no spurious fictitious free "photons" are explicitly exhibited when handling the Fourier integral with respect to the v-frame; that is, when one takes $v^\mu=(1,\mathbf{0})$ in order to evaluate the integral transform of Eq. (5.1). The $F_{\mu\nu}(x)$ tensor components, however, do not depend on this particular choice (and this is the main point in this approach). For this reason, we refer to the v-frame as the *proper frame* of the $A_T^\mu(x;v)$ potentials.

Therefore, for the description of a given free Maxwell field, every inertial observer introduces his own transverse potentials. This must be so, since no fictitious quanta of the free field may appear relative to any inertial observer whatsoever. Furthermore, it does not matter what particular frame we choose to work with, since all these equivalent potentials will give us the same electromagnetic field. Clearly, this property is related with the fact that photons have no rest frames.

As the Lorentz condition itself, the transverse subsidiary gauge condition [cf. Eq. (3.15)] represents a covariant constraint. The Lorentz condition, however, is an absolute constraint; while v-transversality is a *relative constraint,* for it anchors somehow the irreducible 4-potential to a given inertial frame. Therefore, a genuine transverse 4-potential associated with free photons satisfies two independent covariant gauge conditions, and we are left with only two degrees of freedom for the complete description of the free radiation field, as we must be.

In particular, if we Lorentz-transform this scheme to the proper frame of the $A_T^\mu(x;v)$ potentials [for that purpose, we set $v^\mu=(1,\mathbf{0})$ in Eq. (3.14)], we get

$$\mathbf{A}_T(x;1,\mathbf{0})=\mathbf{A}(x)-(2\pi)^{-2}\int d^4k\,(k^0)^{-1}\,\mathbf{k}\,\tilde{A}^0(k)\exp(ikx), \tag{5.9}$$

and also

$$A_T^0(x;1,\mathbf{0})=0, \tag{5.10}$$

which corresponds to the transverse gauge condition (3.15). So we get

$$\nabla\cdot\mathbf{A}_T(x;1,\mathbf{0})=0, \tag{5.11}$$

as corresponding to the Lorentz gauge condition (3.16). Hence, the irreducible 4-potential of the free electromagnetic field belongs in the Coulomb gauge with respect to its proper frame. In this formalism gauge invariance of the irreducible 4-potential appears as a projective property. This property enhances a covariant generalization of the radiation gauge, which thus behaves as a subgauge of the Lorentz gauge.

We wish here to remark on the fact that neither in Eq. (3.14), nor in the equivalent symbolic form (4.16), can we factor out the 4-velocity parameters v^μ which figure in the momentum integrals, for they also occur nonlinearly in the denominator of the projection operator. Formal as it is, this fact tells us that it is impossible to arrive at the present formalism by means of algebraic and/or differential manipulations performed exclusively in coordinate space-time. Recourse to momentum representation is an essential feature of the present approach.

This last formal comment is important because a different way of handling the Coulomb gauge, in a manifestly covariant fashion, can be found in the literature,[12] working exclusively on the local geometry of the coordinate representation. We repeat here some features of that formalism for the sake of comparison with our treatment. Given a free 4-potential $A^\mu(x)$ and a fixed 4-velocity v^μ, one defines two relative potentials:

$$A_S^\mu(x;v)=(\delta_\nu^\mu-v^\mu v_\nu)A^\nu(x), \tag{5.12}$$

and

$$A_I^\mu(x;v)=v^\mu v_\nu A^\nu(x). \tag{5.13}$$

These are projective transformations performed in coordinate space-time. Hence, $A_S^\mu(x;v)$ belongs in the

v-transverse gauge: $v_\mu A^\mu_S(x;v)=0$. Moreover, if $A^\mu(x)$ belongs in the Coulomb gauge with respect to the v-frame, then $A^\mu_S(x;v)$ also belongs in the Lorentz gauge, and *vice-versa*. On the v-frame (and only on that frame) we have: $A^0_S(x;1,\mathbf{0})=0$ and, also, $\mathbf{A}_S(x;1,\mathbf{0})=\mathbf{A}(x)$. Hence, transformations (5.12) and (5.13) afford a covariant generalization of the Coulomb gauge.

However, in order for the radiation gauge to appear as a subgauge of the Lorentz gauge, let us assume the constrain $\nabla_\mu A^\mu(x)=0$ from the beginning. Then we get

$$\nabla_\mu A^\mu(x;v)=-v^\mu v_\nu \nabla_\mu A^\nu(x)\,, \tag{5.14}$$

so that, in general, $A^\mu_S(x;v)$ does not belong in the Lorentz gauge [unless we have $v_\nu A^\nu(x)=0$; but then everything becomes trivial]. So we see that this approach does not present the Coulomb gauge as a subgauge of the Lorentz gauge, and, therefore, it fails to afford a realization of the electromagnetic radiation gauge group as a subgroup of the Lorentz gauge group. The disadvantages of this fact for quantum electrodynamics are immediate.

Moreover, one can see that [in contrast with Eq. (4.13)] Eq. (5.12) does not correspond to a guage transformation. Plainly, this means that $A^\mu_S(x;v)$ is not physically equivalent to $A^\mu(x)$, since both electromagnetic 4-potentials belong to different Maxwell fields. This fact makes a strong contrast with our result.

6. COVARIANT TRANSFORMATION LAW OF THE TRANSVERSE POTENTIALS

We now set ourselves the task of finding that transformation law which brings $A^\mu_T(x;u)$ into $A'^\mu_T(x';v)$, associated with the active Lorentz transformation $L(v;u)$ which brings u^μ into v^μ:

$$L^\mu{}_\nu(v;u)u^\nu=v^\nu. \tag{6.1}$$

Of course, the transformation law of these potentials must be such that it preserves both covariant gauge constraints.

The tensor $L(v;u)$ has the form[13]

$$L^\mu{}_\nu(v,u)=\delta^\mu{}_\nu-(1+v_\lambda u^\lambda)^{-1}(v^\mu+u^\mu)(v_\nu+u_\nu)+2v^\mu u_\nu\,, \tag{6.2}$$

and corresponds to an active proper orthochronous Lorentz transformation

$$L^\mu{}_\nu(v,u)x^\nu=x'^\nu. \tag{6.3}$$

The transformed event x' has precisely the same spacetime coordinates with respect to the v-frame as the object event x would have, once transformed to the new u-frame. It should be understood that all transformations considered in this paper are active.

Given any Galilean working frame and two timelike unit vectors, v^μ and u^μ, we have the potentials $A^\mu_T(x;v)$ and $A^\mu_T(x;u)$, obtained from a free 4-potential $A^\mu(x)$, as given in Eqs. (4.13) and (5.5), respectively. But then, since

$$[\delta^\mu_\nu-(\nabla_\rho u^\rho)^{-1}\nabla^\mu u_\nu][\delta^\nu_\lambda-(\nabla_\sigma v^\sigma)^{-1}\nabla^\nu v_\lambda]$$
$$\equiv\delta^\mu_\lambda-(\nabla_\rho u^\rho)^{-1}\nabla^\mu u_\lambda\,, \tag{6.4}$$

we have

$$A^\mu_T(x;u)=[\delta^\mu_\nu-(\nabla_\lambda v^\lambda)^{-1}\nabla^\mu u_\nu]A^\nu_T(x;v)\,. \tag{6.5}$$

Hence, under the active Lorentz transformation (6.3), one must set

$$A'^\mu_T(x';v)=L^\mu_\nu(v,u)A^\nu_T(x;u)\,, \tag{6.6}$$

i.e.,

$$A'^\mu_T(x';v)=L^\mu{}_\nu(v,u)[\delta^\nu_\lambda-(\nabla_\rho u^\rho)^{-1}\nabla^\nu u_\lambda]A^\lambda_T(x;v)\,, \tag{6.7}$$

where we have written $A'^\mu_T(x';v)$ for $A'^\mu_T(x';u')$, according to Eq. (6.1) (namely: $u'^\mu=v^\mu$, by definition). Equation (6.7) tells us that, as observer v uses the potentials $A^\mu_T(x;v)$ for the description of real photons in the v-frame of reference, so observer u uses the potentials $A'^\mu_T(x';v)$ for the description of the same photons in the u-frame of reference. (It is not necessary at all to set $v^\mu=(1,\mathbf{0})$ in order for this interpretation to hold!). Therefore, Eq. (6.7) states the covariant transformation law of the transverse potentials. It is easy to show that Eq. (6.7) may be also written in the form

$$A'^\mu_T(x';v)=[\delta^\mu_\nu-(\nabla'_\rho v^\rho)^{-1}\nabla'^\mu v_\nu]\,L^\nu{}_\lambda(v,u)A^\lambda_T(x;v)\,, \tag{6.8}$$

which makes it evident that $v_\mu A'^\mu_T(x';v)=0$ and $\nabla'_\mu A'^\mu_T(x';v)=0$, as required.

7. THE TWO DEGREES OF FREEDOM OF THE ELECTROMAGNETIC FIELD

Let us finally concentrate on the properties of the transverse potentials relative to the v-frame, as revealed in the momentum geometry of these fields. Since the Fourier amplitudes $f^\mu_T(k;v)$ have to conform simultaneously to the Lorentz gauge as well as to the v-transverse gauge, these are spacelike vectors pertaining in a 2-flat which is a subspace of the Lorentz gauge 3-flat. So let us introduce (as usual) two linearly independent *polarization* vectors: $\epsilon^\mu_A(k,v)$, $A=1,2$, which belong in this 2-flat gauge locus relative to v; i.e.,

$$v_\mu\epsilon^\mu_A(k,v)=0\,, \tag{7.1}$$

$$k_\mu\epsilon^\mu_A(k,v)=0\,, \tag{7.2}$$

and such that

$$\eta_{\mu\nu}\epsilon^\mu_A(k,v)\epsilon^\nu_B(k,v)=-\delta_{AB}\,, \tag{7.3}$$

corresponding to two mutually orthogonal spacelike unit vectors. Hence, the set $\{v^\mu, k^\mu, \epsilon^\mu_A(k,v),\ A=1,2\}$ is a tetrad whose completeness relation gives us

$$\delta_{AB}\epsilon^\mu_A(k,v)\epsilon^\nu_B(k,v)$$
$$=-\eta^{\mu\nu}-k^\mu k^\nu/(v^\lambda k_\lambda)^2+(v^\mu k^\nu+k^\mu v^\nu)/v^\lambda k_\lambda\,. \tag{7.4}$$

Next we define two quantities, which are (apparently) v-dependent

$$\rho_A(k,v)=-\eta_{\mu\nu}\epsilon^\mu_A(k,v)f^\nu_T(k,v)\,, \tag{7.5}$$

for $A=1,2$, so that

$$f^\mu_T(k,v)=\delta_{AB}\epsilon^\mu_A(k,v)\rho_B(k,v)\,. \tag{7.6}$$

Once the $\epsilon^\mu_A(k,v)$, $A=1,2$, have been suitable chosen, we also define the new local polarization base given by

$$\epsilon^\mu_A(k,u)=\{\delta^\mu_\nu-(k^\mu u_\nu)/(k^\lambda u_\lambda)\}\epsilon^\nu_A(k,v)\,, \tag{7.7}$$

which vectors have the properties

$$u_\mu \epsilon^\mu_A(k,u) = 0\,, \tag{7.8}$$

$$k_\mu \epsilon^\mu(k,u) = 0\,, \tag{7.9}$$

and also

$$\eta_{\mu\nu}\epsilon^\mu_A(k,u)\epsilon^\nu_B(k,u) = -\delta_{AB}\,, \tag{7.10}$$

and thus afford us a new complete tetrad $\{u^\mu, k^\mu, \epsilon^\mu_A(k,u)\}$, $A=1,2$. Therefore, we define two new u-dependent quantities, as in Eq. (7.5),

$$\rho_A(k,u) = -\eta_{\mu\nu}\epsilon^\mu_A(k,u) f^\nu_T(k,u)\,, \tag{7.11}$$

so as to have

$$f^\mu_T(k,u) = \delta_{AB}\epsilon^\mu_A(k,u)\rho_B(k,u)\,. \tag{7.12}$$

But then one has the following important result:

$$\begin{aligned}
&\rho_A(k,u)\\
&= -\eta_{\mu\nu}\{\delta^\mu_\lambda - k^\mu u_\lambda/k^\rho u_\rho\}\{\delta^\nu_\tau - k^\nu u_\tau/k^\rho u_\rho\}\epsilon^\lambda_A(k,v) f^\tau_T(k,v)\\
&= -\{\eta_{\lambda\tau} - (k_\lambda u_\tau + u_\lambda k_\tau)/k^\rho u_\rho\}\epsilon^\lambda_A(k,v) f^\tau_T(k,v)\\
&= -\eta_{\lambda\tau}\epsilon^\lambda_A(k,v) f^\tau_T(k,v) = \rho_A(k,v)\,,
\end{aligned} \tag{7.13}$$

since, clearly [cf. Eq. (6.5)],

$$f^\mu_T(k,u) = \{\delta^\mu_\nu - k^\mu u_\nu/k^\lambda u_\lambda\} f^\nu_T(k,v)\,. \tag{7.14}$$

Hence, we may define, quite generally,

$$\rho_A(k,v) = \rho_A(k,u) =: \rho_A(k)\,, \tag{7.15}$$

since these quantities are the same for all inertial observers. So we write

$$\begin{aligned}
f^\mu_T(k,v) &= \delta_{AB}\epsilon^\mu_A(k,v)\rho_B(k)\,,\\
f^\mu_T(k,u) &= \delta_{AB}\epsilon^\mu_A(k,u)\rho_B(k)\,,
\end{aligned} \tag{7.16}$$

relative to all 4-velocities, v^μ, u^μ, etc. We thus identify these $\rho_A(k)$ quantities with the true degress of freedom of the Maxwell field.[14] These are v-independent objects, since in the description of one and the same free field it is only the polarization vector that changes from one inertial observer to another.

The true degrees of freedom, as presented in Eq. (7.5) or (7.11), are manifestly scalar fields defined on the light cone in momentum space. Nevertheless, it is interesting to strengthen this fact, by explicitly showing how they behave under active Lorentz transformations. Of course, as in Eq. (6.6), we must have

$$\epsilon'^\mu_A(k',v) = L^\mu{}_\nu(v,u)\epsilon^\nu_A(k,u)\,, \tag{7.17}$$

and, therefore, we get [directly from Eq. (6.13)]

$$\begin{aligned}
&f'^\mu_T(k',v) = \delta_{AB}\epsilon'^\mu_A(k',v)\rho'_B(k')\\
&= L^\mu_\nu(v,u)\{\delta^\nu_\lambda - k^\nu u_\lambda/k^\rho u_\rho\}\delta_{AB}\epsilon^\lambda_A(k,v)\rho_B(k)\\
&= L^\mu{}_\nu(v,u)\delta_{AB}\epsilon^\nu_A(k,u)\rho_B(k)\\
&= \delta_{AB}\epsilon'^\mu_A(k',v)\rho_B(k)\,,
\end{aligned} \tag{7.18}$$

i.e.,

$$\rho'_A(k') = \rho_A(k)\,, \tag{7.19}$$

whenever $k'^\mu = L^\mu{}_\nu(v,u)k^\nu$, as it must be. Clearly, these $\rho_A(k)$, $A=1,2$, are the only true dynamical variables one should quantize.[14]

[1]Electrodynamics, in fact, could not be maintained gauge-invariant were the photon to have a finite mass. There are some interesting discussions on this point in the literature; cf. J. Schwinger, Phys. Rev. **125**, 397 (1962); V. I. Ogievetskij and I. V. Polubarinov, Ann. Int. Conf. High-Energy Phys. 11, 666 (1962) [CERN, Geneva (1962)]; P. Bandyopadhyay, Nuovo Cimento A **55**, 367 (1968). Also, on the zero mass of the photon, see L. Bass and E. Schrödinger, Proc. Roy. Soc. A 232, 1 (1955); for a relatively recent review of this most important matter, see A. G. Goldhaber and M. N. Nieto, Rev. Mod. Phys. **43**, 277 (1971).

[2]It is also well known that, besides these two symmetry groups, the electromagnetic theory is endowed with the invariance of the Maxwell field equations under transformations of the fifteen-parameter conformal group. It is still not clear, at this time, exactly what symmetries and conserved quantities are introduced into the physical picture by the group of conformal transformations; cf. F. Rohrlich, T. Fulton, and L. Witten, Rev. Mod. Phys. **34**, 442 (1962). We shall not touch on the intriguing issue of conformal symmetry in the present work.

[3]A different approach to overcome these difficulties can be found in the literature; cf. S. Mandelstam, Ann. Phys. (N.Y.) **19**, 1 (1962), where a Lorentz invariant theory is presented for quantum electrodynamics, written entirely in terms of gauge invariant quantities. [The same scheme is proposed for the quantization of the gravitational field in S. Mandelstam, Ann. Phys. (N.Y.) **19**, 25 (1962)]. However, Mandelstam shows that there is no Lorentz-invariant variational principle for this formalism, and so it fails to produce conservation laws associated with the symmetries of a Lagrangian density.

[4]See, for instance, P. G. Bergmann and A. Komar, Int. J. Theor. Phys. **5**, 15 (1972).

[5]An attempt to obtain such a separation can be found in M. Halpern and S. Malin, J. Math. Phys. **12**, 213 (1971), where the novel mathematical structure of "quasigroup" of coordinate transformations in curved space-time is introduced.

[6]The unification of the Poincaré group and the electromagnetic gauge group in a single invariant group has been tried by G. Rideau, "Gauge Group and Extensions of the Poincaré Group," in *Colloquium on Group Theoretical Methods in Physics* (CNRS, Université d'Aix, Marseille, 1972), p. 11–7. ff., where the continuous extensions of the Poincaré group by the gauge group are studied by means of the "extension of topological groups" tool. [See H. Nagao, Osaka Math. J. **1**, 36 (1949)]; The conclusions of Rideau's work, as remarked by himself, however, seem to be useless for the needs of quantum electrodynamics. Another relationship between gauge transformations and the relativistic invariance can be also found in C. G. Oliveira and A. Vidal, Progr. Theor. Phys. **43**, 510 (1970). The methods and conclusions of these authors are quite different from ours.

[7]The same problem for the classical field-source coupled system, as well as for quantum electrodynamics, will be discussed elsewhere according to the space-time geometric approach introduced in the present paper. [In this sense it is also interesting to recall the work of S. Osaki, Prog. Theor. Phys. **14**, 511 (1955)].

[8]This dynamical redundancy has been the origin of many difficulties in quantum electrodynamics, notwithstanding the fact that a new canonical formalism for dealing with redundant variables has been created by Dirac. [See P. A. M. Dirac, *Lectures on Quantum Mechanics* (Yeshiva U., New York, 1964)]. It is obvious, on physical grounds, that only the true dynamical variables (degrees of freedom) of a system should be quantized. The true degrees of freedom of the electromagnetic potentials have been covariantly identified by A. Valdes in a thesis "Formulación Covariante del Gauge de Radiación" submitted to the Faculty of Science, Universidad de Chile, Santiago 1976, in partial fulfillment to obtain the degree of Licenciado. The same issue has recently been dis-

cussed (for any adopted gauge) by S. Hojman, Ann. Phys. (N.Y.), **103**, 74 (1977); cf. also, R. Gambini and S. Hojman, Ann. Phys. (N.Y.) **105**, 407 (1977). Some important features of these analyses come very close to ours; there are substantial novelties, however, in the approach adopted in the present work.

[9]Other relativistic covariant gauges have been known for a long time in quantum electrodynamics (i.e., the Landau gauge, the Feynman gauge, the Yennie and Fried gauge). See, for instance, B. Zumino, J. Math. Phys. **1**, 1 (1960).

[10]It is well known that, according to gauge field theory, the group structure of gauge transformations as realized in electromagnetic theory corresponds to the Yang–Mills group extension of U(1); the 4-potential thus appears as the compensating field, while the electromagnetic tensor appears as the Riemann tensor, of the locally extended symmetry.

[11]The author has been unable to find any reference concerning this invariant $G(x;v)$ function in the current literature; as, for instance, in J. M. Jauch and F. Rohrlich, *The Theory of Photons and Electrons* (Addision–Wesley, Cambridge, Mass., 1955), Appendix A1, p. 419.

[12]F. Rohrlich, *Classical Charged Particles* (Addision–Wesley, Reading, Mass., 1965).

[13]This tensor, performing active (proper orthochronous) Lorentz transformations, is discussed by J. Krause, J. Math. Phys. **18**, 889 (1977); **19**, 370 (1978).

[14]Cf. S. Hojman, and also R. Gambini and S. Hojman (references given in Ref. 8).

Geometrical structure of Faddeev–Popov fields and invariance properties of gauge theories

M. Quirós and F. J. de Urries
Instituto de Estructura de la Materia Consejo Superior de Investigaciones Científicas, Serrano 119. Madrid 6, Spain and Departmento de Física, Universidad de Alcalá de Henares, Madrid, Spain

J. Hoyos, M. L. Mazón, and E. Rodriguez
Departmento de Matemáticas, Universidad de Alcalá de Henares, Madrid, Spain

(Received 16 July 1980; accepted for publication 20 February 1981)

Let $\mathscr{P}_n$ be the trivial principal bundle with structural group G and base space $\mathscr{P}_{n-1}$, $\mathscr{P}_1$ being the usual fiber bundle of gauge theories. In order to give a geometrical interpretation to the Faddeev-Popov fields, as well as to the Becchi, Rouet, and Stora transformations, we need to use the fiber bundle $\mathscr{P}_3$. The gauge fields and the Faddeev–Popov ghost and antighost fields appear as part of certain one-forms defined on the base space $\mathscr{P}_2$. The anticommuting character of the ghost and antighost fields is essentially due to their identification with one-forms. The Becchi, Rouet, and Stora transformations are identified with generalized infinitesimal gauge transformations on $\mathscr{P}_3$ of parameters related to the ghost fields. We obtain a further invariance of the action given by a similar generalized infinitesimal gauge transformation on $\mathscr{P}_3$ related to the antighost fields.

PACS numbers: 11.10.Np, 11.80.Jy

I. INTRODUCTION

Immediately after the introduction of nonabelian gauge field theories[1] by Yang and Mills,[2] the relationship between gauge fields and the internal geometry of the space of group parameters was becoming more and more evident. Gauge fields were indeed linked to the connection while the kinetic tensor played, with respect to the internal symmetry, the same role as the Riemann tensor did regarding the geometry of the space–time manifold.

The use of differential geometry techniques to handle gauge theories was revived by a paper by Wu and Yang[3] where the gauge group was interpreted as the structural group of a principal fiber bundle. Gauge fields were determined by the connection form on the fiber bundle while the kinetic tensor was, essentially, the curvature form of this connection.[4]

The quantization of gauge theories leads to the requirement of introducing a gauge-fixing term in the Lagrangian which allows the free propagators of gauge fields to be written. This term implies the presence of nonphysical degrees of freedom which must be cancelled by means of the Faddeev–Popov (FP) Lagrangian of ghost fields.[5] However the FP fields were not identified as geometrical objects —(see, for instance, the Wu and Yang dictionary[3]) so that the problem is to get a satisfactory geometrical interpretation for the FP fields, in the same sense as that given to the gauge fields. As we shall see below the FP ghost and antighost fields do not have a place in the usual principal fiber bundle of gauge theories, so that the geometrical structure must be enlarged in the way discussed in this paper.

A first attempt along this direction was made by Thierry-Mieg[6] who interpreted the FP ghost fields as one-forms belonging to the vertical part of the connection associated to the gauge fields in a principal fiber bundle. Nevertheless, Thierry-Mieg's method is not, in our opinion, entirely satisfactory, mainly for the following reasons.

(i) The FP ghost fields are given a space–time dependence by means of nongauge transformations

(ii) The FP and gauge fields are explicitly dependent on the parameters of the gauge group. This is essentially due to the fact that the defined transformations assign different elements of the group to points of the fiber bundle lying on the same fiber. In this way any relation with the physical meaning provided by the gauge transformation is lost.

(iii) The FP antighost fields have no geometrical interpretation.

(iv) The Becchi, Rouet, and Stora (BRS) transformations[7] are not related to gauge transformations in this scheme.

The aim of this work is to present a geometrical interpretation of FP fields, inspired by Thierry-Mieg, but solving, to our mind, the four difficulties mentioned above.

We introduce a principal fiber bundle of structural group G and whose base space is another principal fiber bundle. The gauge and FP ghost and antighost fields appear as parts of certain one-forms defined on the base space and associated to connections in the principal fiber bundle.

Point (iii) above is solved by the choice of the base fiber bundle while the gauge transformations on the overall principal fiber bundle dissipate the difficulties about points (i) and (ii). The BRS transformations may be interpreted as infinitesimal generalized gauge transformations whose parameters are two-forms related to the FP ghost fields.

As a consequence of our theoretical construction, the Lagrangian appears as a BRS invariant two-form. Furthermore, it is worth stressing that we obtain a new invariance property for the action under generalized gauge transformations whose parameters are related to the FP antighost fields.

In Sec. II we introduce the mathematical tools necessary to study gauge transformations. The construction of our particular principal fiber bundle and some mathematical results concerning gauge transformations on it are described

 0022-2488/81/081767-08$01.00 © 1981 American Institute of Physics

in Sec. III. Section IV is devoted to giving the definition of FP fields, to studying their change under gauge transformations, and to briefly reviewing Thierry-Mieg's approach. In Sec. V the structure equations are analyzed and we obtain expressions for the variations of the fields under exterior differentials. Last, but not least, in Sec. VII a new variance of the action (unknown in the previous literature on gauge theories) is proved.

II. GAUGE TRANSFORMATIONS: SOME MATHEMATICAL TOOLS

We begin by introducing two operations that will be of great relevance when dealing with gauge transformations.

Let G be a Lie group and $\mathscr{G}$ its Lie algebra. Let N be a smooth manifold and $C^\infty(N,G)$ the space of smooth maps from N into G. $C^\infty(N,G)$ inherits from G a group structure,

$$\gamma\cdot\delta(p)\equiv\gamma(p)\delta(p),\quad p\in N;\gamma,\delta\in C^\infty(N,G).$$

If $e\in G$ is the identity, then the constant map $p\to e$ is the identity of $C^\infty(N,G)$, and the inverse of $\gamma\in C^\infty(N,G)$, γ^{-1}, is defined by $\gamma^{-1}(p)=\gamma(p)^{-1}$, $p\in N$.

Let $\Lambda^k_{\mathscr{G}}(N)$ be the space of $\mathscr{G}$-valued k-forms on N. For each $\gamma\in C^\infty(N,G)$ we define a linear map from $\Lambda^k_{\mathscr{G}}(N)$ into itself, as follows.

$$(\mathrm{ad}(\gamma^{-1})\omega)_p(X_{1p},...,X_{kp})=\mathrm{ad}(\gamma(p)^{-1})\{\omega_p(X_{1p},...,X_{kp})\}.$$

For all $p\in N$, $X_{1p},...,X_{kp}\in T_p(N)$ and $\omega\in\Lambda^k_{\mathscr{G}}(N)$. The map $\gamma\to\mathrm{ad}(\gamma^{-1})$ has the following properties:

(i) $\forall\gamma,\delta\in C^\infty(N,G)$, $\mathrm{ad}((\gamma\cdot\delta)^{-1})=\mathrm{ad}(\delta^{-1})\,\mathrm{ad}(\gamma^{-1})$.

(ii) If N' is a second smooth manifold and $g\in C^\infty(N,N')$, then

$$\forall\gamma\in C^\infty(N,G),\quad g^*\mathrm{ad}(\gamma^{-1})=\mathrm{ad}((\gamma g)^{-1})g^*. \tag{2.1}$$

We have composition of functions everywhere (except when the dot appears explicitly) which is the group product of $C^\infty(N,G)$.

Let θ be the left invariant $\mathscr{G}$-valued canonical one-form on G[8] ($\theta_a=L_{a^{-1}*a}$, $a\in G$, where $L_{a^{-1}}$ is the left translation on G and $L_{a^{-1}*a}$ is its differential at the point a). Then $\gamma^*\theta\in\Lambda^1_{\mathscr{G}}(N)$ if $\gamma\in C^\infty(N,G)$.

Now we define an affine map from $\Lambda^1_{\mathscr{G}}(N)$ into itself as follows: $\gamma^\dagger\omega=\mathrm{ad}(\gamma^{-1})\omega+\gamma^*\theta$, $\omega\in\Lambda^1_{\mathscr{G}}(N)$.

The mapping has the following properties

(i) $\forall\gamma,\delta\in C^\infty(N,G)$, $(\gamma\cdot\delta)^\dagger=\delta^\dagger\gamma^\dagger$,

(ii) If N' is a second smooth manifold and $g\in C^\infty(N',N)$, then

$$\forall\gamma\in C^\infty(N,G),\ g^*\gamma^\dagger=(\gamma g)^\dagger g^*. \tag{2.2}$$

The next step will be to introduce the connections and the structure equations on fiber bundles.

Let M be a smooth manifold and let $\mathscr{P}=\mathscr{P}(M,G,\Pi)$ be a principal fiber bundle with base space M, group G, and projection Π. Let $\{U_i\}_{i\in I}$ be an open cover of M such that for each $i\in I$ there is a trivialization, $\psi_i:\Pi^{-1}(U_i)\to U_i\times G$, of $\mathscr{P}$. We have $\psi_i=(\Pi,\phi_i)$ with $\phi_i\in C^\infty(U_i,G)$ and $\phi_iR_a=R_a\phi_i$, $\forall a\in G$, R_a being the right translation on G. For each $i\in I$, let $\sigma_i:U_i\to\Pi^{-1}(U_i)$ be the preferred local section $\sigma_i(p)=\psi_i^{-1}(p,e)$, $p\in U_i$. It is possible to prove that ψ_i is fixed once ϕ_i or σ_i are given. The transition maps are denoted by $\psi_{ij}:U_i\cap U_j\to G$.

Given a connection form $\omega\in\Lambda^1_{\mathscr{G}}(\mathscr{P})$ on the principal fiber bundle $\mathscr{P}$, the curvature form Ω is related to ω through the structure equation

$$\Omega=d\omega+\tfrac{1}{2}[\omega,\omega], \tag{2.3}$$

where $([\omega,\omega])(X,Y)=[\omega(X),\omega(Y)]$ is the bracket product on $\mathscr{G}$.

The connection form ω is expressed by a family $\alpha=(\alpha_i)_{i\in I}$ of one-forms $\alpha_i\in\Lambda^1_{\mathscr{G}}(U_i)$ on each U_i, $\alpha_i=\sigma_i^*\omega$. These one-forms are related by $\alpha_j=\psi_{ij}^\dagger\alpha_i$, $i,j\in I$. Conversely, given any such family $\alpha=(\alpha_i)_{i\in I}$, $\alpha_i\in\Lambda^1_{\mathscr{G}}(U_i)$ with the property $\alpha_j=\psi_{ij}^\dagger\alpha_i$, there exists a unique connection form ω on $\mathscr{P}$ such that $\alpha_i=\sigma_i^*\omega$, with $\omega=\phi_i^\dagger\Pi^*\alpha_i$ on $\Pi^{-1}(U_i)$.

We get a similar expression for the curvature form Ω, which can be expressed by a family $R=(R_i)_{i\in I}$ of two-forms, $R_i\in\Lambda^2_{\mathscr{G}}(U_i)$, on each U_i, $R_i=\sigma_i^*\Omega$. These two-forms are related by $R_j=\mathrm{ad}(\psi_{ij}^{-1})R_i$ $i,j\in I$. We can regain Ω from the family R by means of $\Omega=\mathrm{ad}(\phi_i^{-1})\Pi^*R_i$ on $\Pi^{-1}(U_i)$. The structure equation for R reads

$$R_i=d\alpha_i+\tfrac{1}{2}[\alpha_i,\alpha_i],\quad i\in I, \tag{2.4}$$

so we get R_i directly from α_i.

Now we pass on to describing the gauge transformations on $\mathscr{P}$ and their action over connection and curvature forms.

A gauge transformation[4] on $\mathscr{P}$ is a smooth map $f:\mathscr{P}\to\mathscr{P}$ from $\mathscr{P}$ into itself such that $\Pi f=\Pi$ and $fR_a=R_af$, $\Lambda a\in G$, i. e., the fibers are preserved and the mapping f is equivariant. The gauge transformations are just the equivalences of $\mathscr{P}$.[9]

As before, a gauge transformation f can be completely characterized by a family $\gamma=(\gamma_i)_{i\in I}$ of maps $\gamma_i\in C^\infty(U_i,G)$ for each U_i, $\gamma_i=\phi_if\sigma_i$. Then $\gamma_j=\mathrm{ad}(\psi_{ij}^{-1})\gamma_i$; $i,j\in I$. Conversely, such a family $\gamma=(\gamma_i)_{i\in I}$ of maps $\gamma_i\in C^\infty(U_i,G)$ with the property $\gamma_j=\mathrm{ad}(\psi_{ij}^{-1})\gamma_i$ determines a unique gauge transformation f given by $f(u)=(\sigma_i\Pi(u))[(\gamma_i\Pi(u))\phi_i(u)]$, $u\in\Pi^{-1}(U_i)$, i. e., $f=(\sigma_i\Pi)[(\gamma_i\Pi)\cdot\phi_i]$ on $\Pi^{-1}(U_i)$.

It is worth noting that the composition of two gauge transformations f_1 and f_2 is a new gauge transformation f_1f_2 having as local expression $\gamma_1\cdot\gamma_2=(\gamma_{1i}\cdot\gamma_{2i})_{i\in I}$ if $\gamma_1=(\gamma_{1i})_{i\in I}$ and $\gamma_2=(\gamma_{2i})_{i\in I}$ are the local expressions of f_1 and f_2, respectively.

For our purposes it is very important to point out that the gauge transformation f acting on a connection form ω is a new connection $f^*\omega$. If ω is expressed locally by the family $\alpha=(\alpha_i)_{i\in I}$ and f by $\gamma=(\gamma_i)_{i\in I}$, $f^*\omega$ is given locally by $\gamma^\dagger\alpha\equiv(\gamma_i^\dagger\alpha_i)_{i\in I}$ as can be easily seen using (2.2). Also, if $\Omega=\Omega(\omega)$, then $\Omega(f^*\omega)=f^*\Omega$ and its local expression is given by $\mathrm{ad}(\gamma^{-1})R\equiv(\mathrm{ad}(\gamma_i^{-1})R_i)_{i\in I}$.

So we can handle the connections and curvature form on P, and their gauge transformations, directly on the base space M, without any explicit reference to P. This will simplify the expression of the gauge transformation in Secs. III and IV, and in the latter section will also permit us to give a space–time dependence to the FP ghost fields without introducing a new disturbing field.

III. A PARTICULAR CONSTRUCTION

In order to give a geometrical meaning to the FP fields we will need the construction of the particular principal fiber bundle developed below.

Let $\mathscr{P}_1 = \mathscr{P}(M,G,\Pi_1)$ be a general principal fiber bundle of base M, structural group G, and projection Π_1. Let $U = \{U_i\}_{i\in I}$ be an open cover of M such that there is a trivialization (U_i,ψ_i) for each $i\in I$. Let ψ_{ij} be the corresponding transition functions.

Let $\alpha = (\alpha_i)_{i\in I}$ be a family of one-forms $\alpha_i \in \Lambda^1_{\mathscr{G}}(U_i)$ such that $\alpha_j = \psi^\dagger_{ij}\alpha_i; i,j\in I$. We know from Sec. II that α determines a connection form $\omega_1 \in \Lambda^1_{\mathscr{G}}(\mathscr{P}_1)$.

Let $\mathscr{P}_2 = \mathscr{P}_1 \times G$ be the trivial principal fiber bundle of base $\mathscr{P}_1$ and structural group G. Let Π_2 be the fiber bundle projection from $\mathscr{P}_2$ onto $\mathscr{P}_1$ and $\Pi_G \in C^\infty(\mathscr{P}_2,G)$ be the direct product projection from $\mathscr{P}_2$ onto G.

In $\mathscr{P}_1$ we consider the open cover that has $\mathscr{P}_1$ as unique member and trivialization of $\mathscr{P}_2$ consisting of $(\mathscr{P}_1,I)$ where $I:\mathscr{P}_2 \to \mathscr{P}_1 \times G$ is the identity map. Then $\omega_1 \in \Lambda^1_{\mathscr{G}}(\mathscr{P}_1)$ determines a connection form $\omega_2 \in \Lambda^1_{\mathscr{G}}(\mathscr{P}_2)$ given by $\omega_2 = \Pi^\dagger_G \Pi^*_2 \omega_1$ on $\Pi_2^{-1}(\mathscr{P}_1) = \mathscr{P}_2$.

Since ω_1 is a connection form on $\mathscr{P}_1$ we have a corresponding curvature form $\Omega_1 = \Omega(\omega_1) \in \Lambda^2_{\mathscr{G}}(\mathscr{P}_1)$. The same is true for ω_2 and we have a curvature form $\Omega_2 = \Omega(\omega_2) \in \Lambda^2_{\mathscr{G}}(\mathscr{P}_2)$.

We now express ω_2 and Ω_2 in terms of the families α and R, where $R = (R_i)_{i\in I}$ is the family of curvature forms on U_i, $i\in I$, given in (2.4).

$$\omega_2 = [(\phi_i \Pi_2)\cdot\Pi_G]^\dagger(\Pi_1\Pi_2)^*\alpha_i, \quad \text{on } \Pi_1^{-1}(U_i)\times G, \tag{3.1}$$

$$\Omega_2 = \mathrm{ad}[((\phi_i\Pi_2)\cdot\Pi_G)^{-1}](\Pi_1\Pi_2)^* R_i, \quad \text{on } \Pi^{-1}(U_i)\times G.$$

We take another step and consider $\mathscr{P}_3 = \mathscr{P}_2 \times G$, the trivial principal fiber bundle of base $\mathscr{P}_2$ and structural group G. As before, we consider the family of local trivializations of $\mathscr{P}_3$ with the unique member $\{(\mathscr{P}_2,I)\}$, where $I:\mathscr{P}_3 \to \mathscr{P}_2 \times G$ is the identity map.

A gauge transformation on $\mathscr{P}_3$ is given by a map $\gamma \in C^\infty(\mathscr{P}_2,G)$ that transforms the one-form $\omega_2 \in \Lambda^1_{\mathscr{G}}(\mathscr{P}_2)$ and the two-form $\Omega_2 \in \Lambda^2_{\mathscr{G}}(\mathscr{P}_2)$ into $\omega'_2 = \gamma^\dagger\omega_2 \in \Lambda^1_{\mathscr{G}}(\mathscr{P}_2)$ and $\Omega'_2 = \mathrm{ad}(\gamma^{-1})\Omega_2 \in \Lambda^2_{\mathscr{G}}(\mathscr{P}_2)$. Note that ω_2 is a connection form of P_2 and Ω_2 its curvature form, but ω'_2 is not in general a connection form on $\mathscr{P}_2$. However, Ω'_2 satisfies the structure equation with respect to ω'_2.

$$\Omega'_2 = d\omega'_2 + \tfrac{1}{2}[\omega'_2,\omega'_2], \tag{3.2}$$

as follows from (2.4). We can express the new forms ω'_2 and Ω'_2 it terms of α and R as

$$\omega'_2 = \gamma^\dagger((\phi_i\Pi_2)\cdot\Pi_G)^\dagger(\Pi_1\Pi_2)^*\alpha_i, \quad \text{on } \Pi_1^{-1}(U_i)\times G, \tag{3.3}$$

$$\Omega'_2 = \mathrm{ad}(\gamma^{-1})\mathrm{ad}[((\phi_i\Pi_2)\cdot\Pi_G)^{-1}](\Pi_1\Pi_2)^*R_i, \quad \text{on } \Pi_1^{-1}(U_i)\times G.$$

Let us now consider the expressions in coordinates of the forms studied above. Let $U\subset\mathscr{R}^n$ be an open set and $a:U\to M$ be a chart such that $a(U)\subset U_i$ for some $i\in I$. We express α_i and R_i in this chart

$$\alpha_i(x) = \sum_{\mu=1}^{n} A_\mu(x)dx^\mu \qquad x\in U, \tag{3.4}$$

$$R_i(x) = \sum_{1\leqslant\mu<\nu\leqslant n} F_{\mu\nu}(x)dx^\mu\wedge dx^\nu \quad x\in U,$$

where $A_\mu, F_{\mu\nu}:U\to\mathscr{G}; \mu,\nu = 1,\dots,n$. Using Eq. (2.4) we obtain

$$F_{\mu\nu}(x) = \partial_\mu A_\nu(x) - \partial_\nu A_\mu(x) + [A_\mu(x),A_\nu(x)], \quad x\in U. \tag{3.5}$$

For the group G we consider $V,W\subset\mathscr{R}^m$ open sets such that $0\in V\cap W$, and $g:V\to G$, $h:W\to G$ are charts with $g(0) = h(0) = e$. Then $b:U\times V\times W\to P_2$ given by $b(x,y,z) = (\psi_i^{-1}(a(x),g(y)),h(z))$ is a chart for $\mathscr{P}_2$.

Hereafter we will always denote as $x = (x_\mu)$ the elements of U, as $y = (y_\alpha)$ the elements of V, and as $z = (z_\beta)$ the elements of W, and a summation over repeated indices will be understood, unless explicit mention is made otherwise.

We suppose that G is given in a matrix representation. In that case, if $\lambda = \gamma b:U\to G$ we have

$$(g^*\theta)_y = g(y)^{-1}\partial_\alpha g(y)dy^\alpha, \quad y\in V,$$

$$(h^*\theta)_z = h(z)^{-1}\partial_\beta h(z)dz^\beta, \quad z\in W,$$

$$(\lambda^*\theta)_{(x,y,z)} = \lambda(x,y,z)^{-1}\partial_\mu\lambda(x,y,z)dx^\mu + \lambda(x,y,z)^{-1}\times\partial_\alpha\lambda(x,y,z)dy^\alpha + \lambda(x,y,z)^{-1}\partial_\beta\lambda(x,y,z)dz^\beta,$$

$(x,y,z)\in U\times V\times W$.

Using the equalities (2.1) and (2.2) we obtain the expression for $\omega_2,\omega'_2,\Omega_2$, and Ω'_2 in these particular coordinates.

$$\begin{aligned}\omega_2(x,y,z) &= \mathrm{ad}(h(z)^{-1})\mathrm{ad}(g(y)^{-1})A_\mu(x)dx^\mu \\ &\quad + \mathrm{ad}(h(z)^{-1})g(y)^{-1}\partial_\alpha g(y)dy^\alpha + h(z)^{-1}\partial_\beta h(z)dz^\beta,\\ \omega'_2(x,,y,z) &= \mathrm{ad}(\lambda(x,y,z)^{-1})\omega_2(x,y,z) + \lambda(x,y,z)^{-1}\\ &\quad\times\partial_\mu\lambda(x,y,z)dx^\mu + \lambda(x,y,z)^{-1}\partial_\alpha\lambda(x,y,z)dy^\alpha\\ &\quad + \lambda(x,y,z)^{-1}\partial_\beta\lambda(x,y,z)dz^\beta,\end{aligned} \tag{3.6}$$

$$\Omega_2(x,y,z) = ad(h(z)^{-1})\mathrm{ad}(g(y)^{-1})F_{\mu\nu}(x)dx^\mu\wedge dx^\nu,$$

$$\Omega'_2(x,y,z) = \mathrm{ad}(\lambda(x,y,z)^{-1})\Omega_2(x,y,z).$$

Note that the curvature forms Ω_2 and Ω'_2 have all components null except, perhaps, those corresponding to the (μ,ν) coordinates.

IV. THE FADDEEV–POPOV GHOST FIELDS

In this section we give a geometrical interpretation of the FP ghost fields. In order to do that we first dicuss the work of Thierry-Mieg[6] pointing out some features that, in our opinion, are not satisfactory. We can modify these features using construction described in Sec. III.

We begin by giving another interpretation of gauge transformations. To do it we came back to $\mathscr{P}_1$ and the connection form ω_1. We consider the chart $r:U\times V\to\mathscr{P}_1$ given by $r(x,y) = \psi_i^{-1}(a(x),g(y))$. Then

$$\omega_1(x,y) = \mathrm{ad}(g(y)^{-1})A_\mu(x)dx^\mu + g(y)^{-1}\partial_\alpha g(y)dy^\alpha. \tag{4.1}$$

Let f be a gauge transformation on $\mathscr{P}_1$ given by the family $\gamma = (\gamma_i)_{i\in I}$, where $\gamma_i \in C^\infty(U_i,G)$ $i\in I$, $\omega'_1 = f^*\omega_1$ is the gauge transformed. Then if $\lambda = \gamma_i a$,

$$\omega_1'(x,y) = \mathrm{ad}(g(y)^{-1})[\mathrm{ad}(\lambda(x)^{-1})A_\mu(x) + \lambda(x)^{-1} \times \partial_\mu \lambda(x)]dx^\mu + g(y)^{-1}\partial_\alpha g(y) dy^\alpha, \quad (4.2)$$

expressed in the chart r.

Next we consider another chart $s: U \times V \to P_1$ given by $s(x,y) = \psi_i^{-1}(a(x), \lambda(x) g(y))$ and we express ω_1 in this new chart,

$$\omega_1(x,y) = \mathrm{ad}(g(y)^{-1})[\mathrm{ad}(\lambda(x)^{-1})A_\mu(x) + \lambda(x)^{-1} \times \partial_\mu \lambda(x)]dx^\mu + g(y)^{-1}\partial_\alpha g(y) dy^\alpha, \quad (4.3)$$

that coincides with the expression of ω_1' in r.

Thierry-Mieg[6] uses the expression $g(y) = \lambda(x) g'(y')$ to indicate the change of coordinates from r to s given above.

Comparison of (4.2) with (4.3) tells us that ω_1 and ω_1' are equal at different points of the fiber. Also we have that $sr^{-1} = f$, so that the gauge transformation can be considered as a change of coordinates.

In what follows the base space M will be the space–time manifold $\mathscr{R}^4$ (or $\mathscr{S}^4$) and G any of the usual gauge groups of order m, e.g., SU(M), $m = M^2 - 1$.

We take from the connection form ω_1 the vertical part, that in r-coordinates reads as

$$c(y) = c_\alpha(y) dy^\alpha, \quad (4.4)$$

where

$$c_\alpha(y) = g(y)^{-1}\partial_\alpha g(y), \quad 1 \leqslant \alpha \leqslant m. \quad (4.5)$$

Since $c_\alpha(y)$ belongs to the Lie algebra $\mathscr{G}$, we consider a basis $\{-i\lambda_a\}$ of $\mathscr{G}$, and then

$$c_\alpha(y) = c_\alpha^a(y)(-i\lambda_a). \quad (4.6)$$

We identify the FP ghost fields with the real one-forms

$$c^a(y) = c_\alpha^a(y) dy^\alpha, \quad (4.7)$$

so that

$$c(y) = c^a(y)(-i\lambda_a). \quad (4.8)$$

With this notation

$$\omega_1(x,y) = \mathrm{ad}(g(y)^{-1})A_\mu(x) dx^\mu + c(y), \quad (4.9)$$

and the gauge fields $A_\mu(x) = A_\mu^a(x)(-i\lambda_a)$ are related, as usual, to the horizontal part of the connection form.

We express the transformed connection form ω_1' in a similar way to that of ω_1 in (4.9).

$$\omega_1'(x,y) = \mathrm{ad}(g(y)^{-1})A_\mu'(x) dx^\mu + c'(y), \quad (4.10)$$

where

$$A_\mu'(x) = \mathrm{ad}(\lambda(x)^{-1})A_\mu x) + \lambda(x)^{-1}\partial_\mu \lambda(x), \quad (4.11)$$

and

$$c'(y) = g(y)^{-1}\partial_\alpha g(y) dy^\alpha = c(y). \quad (4.12)$$

From (4.12) we see that the FP ghost fields do not get any space–time dependence through gauge transformations. This fact gives us the clue to think that the fiber bundle $\mathscr{P}_1$ and the connection form ω_1 are not the correct mathematical formalism for introducing the FP ghost fields, because we know, from the physical picture provided by the path-integral quantization, that they must possess a nontrivial space–time dependence. In order to resolve this difficulty Thierry-Mieg did a coordinate change in $\mathscr{P}_1$, illustrated by $g(y) = \lambda(x,y') g'(y')$, that is not equivalent to a gauge transformation on $\mathscr{P}_1$. Under this coordinate change, the FP ghost fields are transformed according to

$$c_\alpha'(x,y') = \mathrm{ad}(g'(y')^{-1})\lambda(x,y')^{-1}\partial_\alpha \lambda(x,y') + c_\alpha(y'), \quad (4.13)$$

while the gauge fields transform as in (4.11), but $\lambda(x)$ is replaced by $\lambda(x,y')$. This procedure is not satisfactory, however, mainly due to the following reasons:

(a) The transformation used is not a gauge transformation and hence has nothing to do with the invariance properties of the known physical systems.

(b) The ghost and anitghost of FP play a very dissimilar role. Indeed the antighost has not any geometrical interpretation in $\mathscr{P}_1$.

Next, we see we can to resolve (a) by doing a gauge transformation of the connection form ω_2 on P_2, and we get a space–time dependence for the ghost field equal to that given by Eq. (4.13). We write ω_2 with respect to the b-coordinates in the form

$$\omega_2(x,y,z) = \mathrm{ad}[(g(y)h(z))^{-1}]A_\mu(x) dx^\mu + \mathrm{ad}(h(z)^{-1})c(y) + \bar{c}(z), \quad (4.14)$$

where $c(y)$ is given by (4.8) and

$$\bar{c}(z) = \bar{c}_\beta(z) dz^\beta, \quad (4.15a)$$

with

$$\bar{c}_\beta(z) = h(z)^{-1}\partial_\beta h(z). \quad (4.15b)$$

As before we take the components of the one-form $\bar{c}(z)$ with respect to the basis $\{-i\lambda_a\}$ of the Lie abgebra $\mathscr{G}$, obtaining m real one-forms $\bar{c}^a(z)$ which will be identified with the FP antighosts.

A gauge transformation on $\mathscr{P}_2$ is given by a map $\gamma \in C^\infty(P_1, G)$, that in b-coordinates can be expressed by the map $\lambda \in C^\infty(U \times V, G)$. If ω_2' is the connection form of P_2 obtained from ω_2 through this gauge transformation, we have

$$\omega_2'(x,y,z) = \mathrm{ad}[(g(y)h(z))^{-1}]A_\mu'(x,y) dx^\mu + \mathrm{ad}(h(z)^{-1})c'(x,y) + \bar{c}'(z),$$

where

$$\begin{aligned} A_\mu'(x,y) &= \mathrm{ad}(\delta(x,y)^{-1})A_\mu(x) + \delta(x,y)^{-1}\partial_\mu \delta(x,y), \\ c_\alpha'(x,y) &= \mathrm{ad}(g(y)^{-1})\delta(x,y)^{-1}\partial_\alpha \delta(x,y) + c_\alpha(y), \quad (4.16) \\ \bar{c}_\beta' &= \bar{c}_\beta(z), \end{aligned}$$

with $\delta(x,y) = \mathrm{ad}(g(y))\lambda(x,y)$.

Let us remark that we can do the gauge transformation on $\mathscr{P}_2$ directly on its base $\mathscr{P}_1$ (as seen in Sec. II) obtaining the first two equations of (4.16). Then the antighosts do not appear at all, but the ghosts do get space–time dependence. Obviously, Eq. (4.16) are equivalent to Eq. (4.13), but now they are a consequence of a gauge transformation. So we have resolved point (a).

Looking at (4.14) and (4.16) we see that

(i) There is a new object, $\bar{c}(z)$, which may serve to give a geometrical interpretation of the FP antighost fields, as seen before, that have no place in $\mathscr{P}_1$. This solves in part point (b) above.

(ii) The antighost field would be, however, invariant under gauge transformations on $\mathscr{P}_2$ and hence it does not acquire any space–time dependence.

Now it is evident, from the above remarks, that if we wish to get a geometrical interpretation for the FP ghost and antighost fields we need, at least, to consider ω_2, on $\mathscr{P}_2$, and if we want to give them a space–time dependence, by means of gauge transformations, we need to consider $\mathscr{P}_2$ as the base of the principal fiber bundle $\mathscr{P}_3$, cited in Sec. III, and perform the gauge transformation on $\mathscr{P}_3$. As noted in Sec. III, and in view of the remark after (4.16), we can do the gauge transformation directly on $\mathscr{P}_2$ with the additional advantage that a new field, undesirable from the physical point of view, does not appear.

We proceed just as indicated. Let $\gamma \in C^\infty(\mathscr{P}_2, G)$ be the local form of a gauge transformation on $\mathscr{P}_3$ and $\lambda = \gamma b$ its expression in b-coordinates. Let $\omega_2' = \gamma^\dagger \omega_2$ be the gauge transform of ω_2. We write ω_2' in the form

$$\omega_2'(x,y,z) = \mathrm{ad}[(g(y)h(z))^{-1}]A_\mu'(x,y,z)dx^\mu + \mathrm{ad}(h(z)^{-1})c'(x,y,z) + \bar{c}'(x,y,z),$$

where

$$\begin{aligned} A_\mu'(x,y,z) &= \mathrm{ad}(\delta(x,y,z)^{-1})A_\mu(x) + \delta(x,y,z,)^{-1}\partial_\mu\delta(x,y,z), \\ c_\alpha'(x,y,z) &= \mathrm{ad}(g(y)^{-1})\delta(x,y,z)^{-1}\partial_\alpha\delta(x,y,z) + c_\alpha(y), \\ \bar{c}_\beta'(x,y,z) &= \mathrm{ad}[(g(y)h(z))^{-1}]\delta(x,y,z)^{-1}\partial_\beta\delta(x,y,z) + \bar{c}_\beta(z), \end{aligned} \tag{4.17}$$

with

$$\delta(x,y,z) = \mathrm{ad}(g(y)h(z))\lambda(x,y,z).$$

Here we see that both the ghost and the antighost acquire a space–time dependence. The laws of transformation (4.17) are deduced from the initial fields $(A_\mu(x), c_\alpha(y), \bar{c}_\beta(z))$. Now we will try to reproduce them after a new gauge transformation, i.e., considering as initial the fields $(A_\mu'(x,y,z), c_\alpha'(x,y,z), \bar{c}_\beta'(x,y,z))$ basing all of them on the three paramenters. Let $\gamma' \in C^\infty(\mathscr{P}_2, G)$ be a new gauge transformation on $\mathscr{P}_3$ with λ' its expression in b-coordinates. Let $\omega_2'' = \gamma'^\dagger\omega_2'$ be the gauge transformed of ω_2'. Then

$$\omega_2''(x,y,z) = \mathrm{ad}[(g(y)h(z))^{-1}]A_\mu''(x,y,z)dx^\mu + \mathrm{ad}[h(z)^{-1}]c''(x,y,z) + \bar{c}''(x,y,z),$$

with

$$\begin{aligned} A_\mu''(x,y,z) &= \mathrm{ad}(\delta'(x,y,z)^{-1})A_\mu'(x,y,z) + \delta'(x,y,z)^{-1}\partial_\mu\delta'(x,y,z), \\ c_\alpha''(x,y,z) &= \mathrm{ad}\{[\mathrm{ad}(g(y)^{-1})\delta'(x,y,z)]^{-1}\} \\ &\quad \times (c_\alpha'(x,y,z) - c_\alpha(y)) + \mathrm{ad}(g(y)^{-1})\delta'(x,y,z)^{-1} \\ &\quad \times \partial_\alpha\delta'(x,y,z) + c_\alpha(y), \end{aligned} \tag{4.18}$$

$$\bar{c}_\beta''(x,y,z) = ad\{[\mathrm{ad}[(g(y)h(z))^{-1}]\delta'(x,y,z)]^{-1}\}(\bar{c}_\beta'(x,y,z) - \bar{c}_\beta(z)) + \mathrm{ad}[(g(y)h(z))^{-1}]\delta'(x,y,z)^{-1}\partial_\beta\delta'(x,y,z) + \bar{c}_\beta(z),$$

where

$$\delta'(x,y,z) = \mathrm{ad}(g(y)h(z))\lambda'(x,y,z).$$

We must point out that (4.18) is different from (4.17); however, we see that if $c_\alpha' = c_\alpha$ and $\bar{c}_\beta' = \bar{c}_\beta$ they coincide. On the other hand (4.17) and (4.18) are very complicated. To avoid all this we make a final choice for the physical fields. The gauge field will be the coefficients of dx^μ and the FP ghost (antighost) fields will be the components of $dy^\alpha(dz^\beta)$ in the one-form ω_2. Explicitly,

$$\begin{aligned} \tilde{A}_\mu(x,y,z) &= \mathrm{ad}\{[g(y)h(z)]^{-1}\}A_\mu(x), \\ \tilde{c}(x,y,z) &= \mathrm{ad}[h(z)^{-1}]c(y), \\ \tilde{\bar{c}}(x,y,z) &= \bar{c}(z), \end{aligned} \tag{4.19}$$

so that the one-form ω_2 is written, in the initial configuration, as $\omega_2(x,y,z) = \tilde{A}_\mu(x,y,z)dx^\mu + \tilde{c}(x,y,z) + \tilde{\bar{c}}(x,y,z)$. For this choice for the physical fields, the gauge transformed ω_2' is given by $\omega_2'(x,y,z) = \tilde{A}_\mu'(x,y,z)dx^\mu + \tilde{c}'(x,y,z)\tilde{\bar{c}}'(x,y,z)$, where

$$\begin{aligned} \tilde{A}_\mu'(x,y,z) &= \mathrm{ad}[\lambda(x,y,z)^{-1}]\tilde{A}_\mu(x,y,z) + \lambda(x,y,z)^{-1}\partial_\mu\lambda(x,y,z), \\ \tilde{c}'(x,y,z) &= \mathrm{ad}[\lambda(x,y,z)^{-1}]\tilde{c}(x,y,z) + \lambda(x,y,z)^{-1}d_y\lambda(x,y,z), \\ \tilde{\bar{c}}'(x,y,z) &= \mathrm{ad}[\lambda(x,y,z)^{-1}]\tilde{\bar{c}}(x,y,z) + \lambda(x,y,z)^{-1}d_z\lambda(x,y,z). \end{aligned} \tag{4.20}$$

As it should be, (4.20) are nothing else than another form of (3.6). The curvature forms were extensively discussed in Sec. III and we shall not dwel upon them any longer.

Let us finally remark that our choice of the FP fields is such that they are one-forms and hence anticommuting quantities. This is the crucial properlty from the physical point of view. Thierry-Mieg was the first to identify the FP fields with one-forms.[6]

V. THE STRUCTURE EQUATIONS

From here on we will work with a fixed one-form $\rho \in A_g^1(\mathscr{P}_2)$ coming from the gauge transformed of a connection form on $\mathscr{P}_2$, obtained in the form described in Sec. III.

To simplify the notation we will write

$$\rho(x,y,z) = A_\mu(x,y,z)dx^\mu + c_\alpha(x,y,z)dy^\alpha + \bar{c}_\beta(x,y,z)dz^\beta, \tag{5.1}$$

and we do not want to distinguish between the coefficients of ρ that will be referred to as ρ_i, where the index i can take the values μ, α, and β.

We remind the reader that A_μ, c_α, *and* $\bar{c}_\beta$ are elements of the Lie algebra $\mathscr{G}$. Hence if $\{-i\lambda_a\}$ is a basis of $\mathscr{G}$, then $A_\mu = A_\mu^a(-i\lambda_a)$, $c_\alpha = c_\alpha^a(-i\lambda_a)$, and $\bar{c}_\beta = \bar{c}_\beta^a(-i\lambda_a)$. In this way the gauge fields are $A_\mu^a(x,y,z)$, the FP ghost fields are $c^a = c_\alpha^a dy^\alpha$, and the FP anti-ghost fields are $\bar{c}^b = \bar{c}_\beta^b dz^\beta$, $a = 1,\ldots,m$.

The commutation relations between the generators of $\mathscr{G}$ are given by $[-i\lambda_b, -i\lambda_c] = f_{bc}^a(-i\lambda_a)$, f_{bc}^a being the structure constants. We designate R by the curvature form corresponding to (5.1) and R_{ij} $(i,j = \mu,\alpha,\beta)$ its components. Then the structure equations (2.3) and (2.4) can be written as

$$R_{ij} = \partial_i\rho_j - \partial_j\rho_i + [\rho_i, \rho_j], \tag{5.2}$$

and we know from (3.6) that $R_{ij} = 0$ except, perhaps, when $(i,j) = (\mu,\nu)$. In this case

$$R_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu] = F_{\mu\nu}, \tag{5.3}$$

is the well-known strength tensor field $F_{\mu\nu}$ of gauge theories.

As far as the vanishing components of the curvature are concerned, they lead to the following equations.

$$d_y c^a = -\tfrac{1}{2} f^a_{bc} c^b \wedge c^c, \tag{5.4a}$$

$$d_y A^a_\mu = (D_\mu c)^a, \tag{5.4b}$$

$$d_z \bar{c}^a = -\tfrac{1}{2} f^a_{bc} \bar{c}^b \wedge \bar{c}^c, \tag{5.4c}$$

$$d_z A^a_\mu = (D_\mu \bar{c})^a, \tag{5.4d}$$

$$d_y \bar{c}^a + d_z c^a + f^a_{bc} c^b \wedge \bar{c}^c = 0, \tag{5.4e}$$

where $(D_\mu c)^a \equiv \partial_\mu c^a + f^a_{bc} A^b_\mu c^c$ is the covariant derivative of the FP ghost field, and similarly for $(D_\mu \bar{c})^a$. These expressions are obtained by multiplying the equations R_{ij} by adequate differentials. Thus, for instance, (5.4a) is obtained by taking the product of $R_{\alpha\alpha'} = 0$ by $dy^\alpha \wedge dy^{\alpha'}$ and summing over all α and α'. With respect to the differentials d_y and d_z, they are the exterior derivatives $d_y = \partial_\alpha dy^\alpha$ and $d_z = \partial_\beta dz^\beta$.

We also refer to Eq. (5.4) as structure equations. They contain all the relevant geometrical information of the theory and we will make use of them later.

VI. THE GEOMETRICAL MEANING OF THE BECCHI, ROUET, AND STORA TRANSFORMATIONS.

As is well known, the BRS transformations[7] leave invariant the Lagrangian of nonabelian gauge field theories, including the gauge-fixing term and the Faddeev–Popov Lagrangian. The presence of these two terms, gauge-fixing and FP, in the Lagrangian is implemented by the covariant quantization of the theory, Indeed, the gauge invariance of total Lagrangian no longer holds, but is replaced by the BRS invariance, which plays a fundamental role in the proof of the renormalizability of gauge theories.[10,11]

In this section we shall give a geometrical meaning to the BRS transformations by means of the structure equations (5.4). On the other hand, the BRS transformation will be interpreted as a "gauge" transformation in a wide sense to be explained later.

We write for the Lagrangian of pure gauge theories, in the absence of any matter field, the two-form

$$\mathscr{L} = \{-\tfrac{1}{4}(F^a_{\mu\nu})^2 - \tfrac{1}{2}(\partial^\mu A^a_\mu)^2\}\mathscr{F} - \bar{c}^a\partial^\mu(D_\mu c)^a, \tag{6.1}$$

where the gauge-fixing and FP terms are given in the Lorentz gauge. In what follows we limit ourselves, for simplicity, to this gauge but it is clear that our results must be true for a more general class of gauges.

The Lagrangian (6.1) differs from the usual one only in the (nonzero) constant two-form

$$\mathscr{F} = c_{\beta\alpha} dz^\beta \wedge dy^\alpha. \tag{6.2}$$

Nevertheless, the presence of $\mathscr{F}$ does not affect the physical results of the theory since it is a constant (with constant coefficients) and commuting object.

The Lagrangian density (6.1) is a two-form depending on x^μ, y^α, and z^β, where x^μ are space–time coordinates while y^α and z^β are group coordinates without any physical meaning. They have been introduced to give an anticommuting character to the FP ghost and antighost fields. To get this character we do not need the whole group G, but only a neighborhood of a fixed point of G, that can be taken to be the identity $e \in G$, corresponding to $y^\alpha = z^\beta = 0$.

In order to completely avoid the undesired dependence of the fields on y^α and z^β we take their values at $y^\alpha = 0$ and $z^\beta = 0$. Thus the forms become fields with values in the exterior algebra of the tangent space of $G \times G$ at the point (e,e), $T_{(e,e)}(G \times G)$, the corresponding commuting or anticommuting character being preserved.

However, to conserve the meaning of the operators d_y and d_z we define the action in the above exterior algebra as the action on the Grassman algebra of $\mathscr{P}_2$ followed by the particularization to the point (e,e). The same for any other operator which we shall do act on the Lagrangian density, as for instance Δ and $\bar{\Delta}$ below.

In view, of this, the Lagrangian density is only dependent on the space–time, so that we define the action as usual

$$\mathscr{S} = \int_{R^4} \mathscr{L} d^4x. \tag{6.3}$$

We now prove that the Lagrangian (6.1) is a closed form with respect to the y-variables. For this we must apply d_y to $\mathscr{L}$.

The FP Lagrangian may be written as

$$\mathscr{L}_{\rm FP} = -\bar{c}^a d_y \partial^\mu A^a_\mu, \tag{6.4}$$

thanks to (5.4b), and using $d_y^2 = 0$. we arrive at

$$d_y \mathscr{L} = -(\partial^\mu A^a_\mu) d_y(\partial^\mu A^a_\mu) \wedge \mathscr{F} - (d_y \bar{c}^a) \wedge d_y(\partial^\mu A^a_\mu). \tag{6.5}$$

The derivative of the gauge-invariant Lagrangian density $-\frac{1}{4}(F^a_{\mu\nu})^2$ is null in the standard way since (5.4b) is formally an infinitesimal gauge transformation for A^a_μ with parameters c^a.

We observe that the structure equations (5.4) give us the freedom to fix $d_y \bar{c}^a$ arbitrarily, provided that $d_z c^a$ would be constrained by Eq. (5.4e). Thus we choose

$$d_y \bar{c}^a = -\partial^\mu A^a_\mu \mathscr{F}, \tag{6.6}$$

and we get immediately from (6.5) that $d_y \mathscr{L} = 0$.

We obtain the usual BRS transformation[7] as

$$\delta = \xi d_y, \tag{6.7}$$

where ξ is an arbitrary nonzero constant real one-form, introduced only to give to the exterior derivative d_y of the fields (5.4a), (5.4b), and (6.6) the same commuting/anticommuting character as the fields themselves.

We end this section by showing the relation between BRS and gauge transformations.

We take as gauge function

$$\lambda(x,y,z) = e^{\xi c(x,y,z)} = 1 + \xi c(x,y,z). \tag{6.8}$$

The expression (6.8) has a sense in the exterior algebra of the forms and can be considered as a gauge transformation, in a wide sense, since ξc is a commuting object (but not a function) with values in the Lie algebra of G.

Using Eq. (4.20) for the gauge-transformed fields, and with the present notation, the infinitesimal transformations generated by (6.8) are given by

$$\Delta A^a_\mu = \partial_\mu \xi c^a + f^a_{bc} A^b_\mu \xi c^c = (D_\mu \xi c)^a, \tag{6.9a}$$

$$\Delta c^a = d_y \xi c^a + [c, \xi c]^a, \tag{6.9b}$$

$$\Delta \bar{c}^a = d_z \xi c^a + [\bar{c}, \xi c]^a. \tag{6.9c}$$

From (5.4b) and (6.9a) we get

$$\Delta A^a_\mu = \xi d_y A^a_\mu = \delta A^a_\mu, \tag{6.10}$$

which illustrates to us the well-known fact the BRS transform of A_μ^a is nothing but the gauge transform with parameter ξc^a.

As we shall prove, the same holds for the fields c^a and $\bar{c}^a$. Equation (5.4a) can be written, after multiplication by ξ, as

$$\xi d_y c^a = \tfrac{1}{2}[c,\xi c]^a.$$

Hence,

$$\Delta c^a = d_y \xi c^a + 2\xi d_y c^a = \delta c^a = \xi d_y c^a. \tag{6.11}$$

Similarly, from (5.4e) we have

$$[\bar{c},\xi c]^a = \xi d_y \bar{c}^a + \xi d_z c^a,$$

so that

$$\Delta \bar{c}^a = \xi d_y \bar{c}^a = \delta \bar{c}^a. \tag{6.12}$$

Equations (6.10)–(6.12) show that we can identify the BRS transformations with infinitesimal gauge transformations, $\delta = \Delta$. This is not a trivial fact but a direct consequence of the structure equations in the fiber bundle $\mathscr{P}_3$, in sharp contrast with the usual treatment where the BRS transformations for the fields c and $\bar{c}$ are imposed in order to get the invariance of the Lagrangian, but without any reference to gauge transformations.

The infinitesimal "gauge" transformation of the Lagrangian

$$\Delta \mathscr{L} = \mathscr{L}((A + \Delta A, c + \Delta c, \bar{c} + \Delta \bar{c}) - \mathscr{L}(A,c,\bar{c})$$

can be obtained (at first order) by the action of ξd_y over $\mathscr{L}$

$$\Delta \mathscr{L} = \xi d_y \mathscr{L}. \tag{6.13}$$

This is true because ξd_y behaves as a derivation of degree two. In fact, d_y is a skew-derivation of degree one, so that when it acts over a term like $\bar{c}\wedge Mc$ we get

$$d_y(\bar{c}\wedge Mc) = (d_y\bar{c})\wedge Mc - \bar{c}\wedge d_y(Mc);$$

but the one-form ξ restores the sign

$$\xi d_y(\bar{c}\wedge Mc) = (\xi d_y\bar{c})\wedge Mc + \bar{c}\wedge \xi d_y(Mc),$$

and we get that ξd_y is a true derivation of degree two.

From the above remarks we obviously obtain the invariance of the Lagrangian under Δ, as

$$\Delta \mathscr{L} = \xi d_y \mathscr{L} = 0. \tag{6.14}$$

VII. AN ADDITIONAL INVARIANCE OF THE ACTION

We saw in Sec. VI that the well-known invariance of the Lagrangian, and hence of the action, under BRS transformations is closely related to the behavior of the Lagrangian under the exterior differential d_y. So we can ask what is the role played by d_z. Since d_y is associated with the FP ghost fields, as given by (6.8), it seems reasonable to associate d_z with the FP antighost fields.

In this section we shall study the current relation between d_z and the FP antighost fields, through an infinitesimal gauge transformation, and prove the invariance of the action under d_z.

Let us first introduce the infinitesimal "gauge" transformation

$$\bar{\lambda}(x,y,z) = e^{\xi\bar{c}(x,y,z)} = 1 + \xi\bar{c}(x,y,z), \tag{7.1}$$

whose parameters are associated with FP antighost fields.

Making use again of (4.20), we obtain from (7.1) the infinitesimal transformations

$$\bar{\Delta} A^a = D_\mu \xi \bar{c}^a, \tag{7.2a}$$

$$\bar{\Delta} c^a = d_y \xi \bar{c}^a + [c,\xi\bar{c}]^a, \tag{7.2b}$$

$$\bar{\Delta} \bar{c}^a = d_z \xi \bar{c}^a + [\bar{c},\xi\bar{c}]^a. \tag{7.2c}$$

With the aid of Eqs. (5.4c), (5.4d), and (5.4e) and proceeding in a similar way as in Sec. VI, we get from (7.2)

$$\bar{\Delta} A_\mu^a = \xi d_z A_\mu^a, \tag{7.3a}$$

$$\bar{\Delta} c^a = \xi d_z c^a, \tag{7.3b}$$

$$\bar{\Delta} \bar{c}^a = \xi d_z \bar{c}^a. \tag{7.3c}$$

Next, we study the transformation properties of the Lagrangian under d_z and we apply them to the invariance properties of the action.

We write the Lagrangian in the form

$$\mathscr{L} = \bar{\mathscr{L}} - \partial^\mu(\bar{c}^a \wedge d_y A_\mu^a),$$

where

$$\bar{\mathscr{L}} = \{ -\tfrac{1}{4}(F_{\mu\nu}^a)^2 - \tfrac{1}{2}(\partial^\mu A_\mu^a)^2\}\mathscr{T} + (\partial^\mu \bar{c}^a)\wedge d_y A_\mu^a, \tag{7.4}$$

since Lagrangians differing only by a four-divergence are associated to the same action.

Because of (5.4d), $(F_{\mu\nu}^a)^2$ is invariant under d_z. Therefore we can write

$$d_z\bar{\mathscr{L}} = -(\partial^\mu A_\mu^a) d_z(\partial^\nu A_\nu^a)\mathscr{T} + (\partial^\mu d_z \bar{c}^a)\wedge d_y A_\mu^a - (\partial^\mu\bar{c}^a)\wedge d_z d_y A_\mu^a. \tag{7.5}$$

By means of Eqs. (5.4) and (6.6), fixed to obtain the usual BRS invariance, we obtain the decomposition

$$d_z\bar{\mathscr{L}} = \mathrm{A}\wedge\mathscr{T} + B, \tag{7.6}$$

where A is the one-form

$$A = -(\partial^\mu A_\mu^a)\Box\bar{c}^a - f_{bc}^a(\partial^\mu A_\mu^a) A_\nu^b \partial^\nu \bar{c}^c - (\partial^\nu\bar{c}^a)\partial_\nu(\partial^\mu A_\mu^a) - f_{de}^a A_\nu^d(\partial^\mu A_\mu^e)\partial^\nu\bar{c}^a,$$

and B the three-form

$$\begin{aligned} B = {} & -\tfrac{1}{2} f_{de}^b \partial^\mu(\bar{c}^d\wedge\bar{c}^e)\wedge(\partial_\mu c^b + f_{gh}^b A_\mu^g c^h) \\ & + (\partial^\mu\bar{c}^b)\wedge\{f_{de}^b\partial_\mu(c^d\wedge\bar{c}^e) \\ & - f_{de}^b(\partial_\mu\bar{c}^d)\wedge c^e - f_{de}^b f_{gh}^d A_\mu^g \bar{c}^h\wedge c^e \\ & + f_{de}^b f_{gh}^e A_\mu^d c^g\wedge\bar{c}^h\}. \end{aligned}$$

After an extensive use of the Jacobi identities and the anticommuting character of the fields c^a and $\bar{c}^a$, we find

$$A = -\partial^\mu\{(\partial_\mu\bar{c}^a)(\partial^\nu A_\nu^a)\},$$

$$B = 0. \tag{7.7}$$

That, together with (6.6), gives us

$$d_z\bar{\mathscr{L}} = \partial^\mu(\partial_\mu\bar{c}^a\wedge d_y\bar{c}^a), \tag{7.8}$$

while for the Lagrangian $\mathscr{L}$ we get

$$d_z\mathscr{L} = \partial^\mu \mathrm{T}_\mu, \tag{7.9}$$

where

$$T_\mu = \partial_\mu\bar{c}^a\wedge d_y\bar{c}^a - d_z(\bar{c}^a\wedge d_y A_\mu^a).$$

As in Sec. VI we have the ξd_z is a derivation of degree two, so that

$$\bar{\Delta}\mathscr{L} = \xi d_z \mathscr{L} = \partial^\mu(\xi \wedge T_\mu). \tag{7.10}$$

From there we get for the action the invariance property

$$\bar{\Delta} S = 0. \tag{7.11}$$

Let us remark that the BRS invariance was the only known invariance for the action in nonabelian gauge theories. We have shown the existence of a new invariance, $\bar{\Delta}$, for the action, which is associated with gauge transformations of parameter $\xi\bar{c}$. The physical consequences of the invariance (7.11) with respect to generalized Ward–Takahashi identities, and properties derived from them, will be considered in a forthcoming paper.

VIII. CONCLUSION

In this work we have reached a geometrical interpretation for the FP ghost and antighost fields. The anticommuting character of these fields was essentially due to their identification with one-forms. The gauge transformations on the principal fiber bundle $\mathscr{P}_3$ provide us with the space–time dependence for the FP fields. The generalized gauge transformations of parameters related to the FP fields provide us with the BRS trnasformations and the corresponding invariance for the Lagrangian. Furthermore, we obtain a new generalized gauge transformation, with parameter related to FP antighost fields, leading to a new invariance property for the action.

As a consequence of the mathematical framework which has been developed in this paper, there arise several open questions which might be of interest from the physical point of view and are worthy of further investigation.

First of all, the new invariance property for the action, which was discovered in this paper, might have physical consequences providing identities between Green's functions as, for instance, some kind of generalized Ward-Takahashi identities.

On the other hand, a possible physical interpretation of the group parameters y and z can be drawn as internal degrees of freedom while they play, in the present study, a subsidiary role, since they have been fixed so that the physical fields remain only as functions of the space–time coordinates. Concerning these additional internal degrees of freedom we introduce, let us remember that a similar situation also occurs in supersymmetries[12] with the introduction of superspace.

Finally it might be of interest to study the global structure of the overall principal fiber bundle $\mathscr{P}_3$ whose base, $\mathscr{P}_2$, is essentially different from the usual space–time manifold. This study may be relevant in relation with nonperturbative objects (as for instance the topological solitons) where the global structure of the base space plays a fundamental role.

Note added in proof: After completion of this work we found that a particular $\bar{\Delta}$ transformation satisfying our Eq. (5.4) was introduced by Curci and Ferrari [G. Curci and R. Ferrari, Phys. Lett. B **63**, 91 (1976)] and by Ojima [I. Ojima, Prog. Theor. Phys. **64**, 625 (1980)] without any refernce to the geometrical structure of gauge theories.

[1]For a general review on gauge theories, see E. S. Abers and B. W. Lee, Phys. Rep. C **9**, 1 (1973).

[2]R. L. Mills and C. N. Yang, Phys. Rev. **96**, 191 (1954); see also Ref. 1.

[3]T. T. Wu and C. N. Yang, Phys. Rev. D **12**, 3854 (1975).

[4]M. Daniel and C. M. Viallet, Rev. Mod. Phys. **52**, 175 (1980).

[5]L. D. Faddeev and V. N. Popov, Phys. Lett. B **25**, 29 (1967); see also Ref. 1.

[6]J. Thierry-Mieg, Nuovo Cimento A **56**, 396 (1980); J. Math. Phys. **21**, 2834 (1980); J. Thierry-Mieg and Y. Ne'eman, Ann. Phys. (NY) **123**, 247 (1979).

[7]C. Becchi, A. Rouet, and R. Stora, Comm. Math. Phys. **42**, 127 (1975); see also Refs. 10 and 11.

[8]S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963), Vol. I.

[9]N. Steenrod, *The Topology of Fiber Bundles* (Princenton U. P., Princeton, NJ, 1951).

[10]B. Lee, in *Methods in Field Theory* (North-Holland, Amsterdam, 1976).

[11]J. C. Taylor, *Gauge Theories of Weak Interactions* (Cambridge U. P., Cambridge, 1976).

[12]J. Wess and B. Zumino, Nucl. Phys. B **70**, 39, (1974); J. Wess, in *Topics in Quantum Field Theory and Gauge Theories*, Lecture Notes in Physics **77**, (Springer-Verlag, 1978).

Field equations invariant under indecomposable representations of the Lorentz group

L. Hlavatý and J. Niederle
Institute of Physics, Czechoslovak Academy of Sciences, Prague, Czechoslovakia

(Received 28 May 1980; accepted for publication 24 June 1980)

Relativistic equations in which the fields cotransform under the direct sum of indecomposable representations of the Lorentz group are derived and discussed.

PACS numbers: 11.10.Qr

I. INTRODUCTION

The theory of linear relativistically invariant field equations has had a long and tangled history. It began with Schrödinger's proposal of the relativistic wave equation in 1926, Dirac's wave equation for a massive spin 1/2 particle in 1928 and Majorana's pioneering study in 1932 and has always been a subject of interest of many scientists. Consequently, several systematic approaches to relativistic field equations are now available[1], the most popular of which are perhaps those worked out by Gel'fand and Yaglom,[2] by Bruhat[3] and those based on the Gårding's technique.[4]

In this work, relativistic equations of the form

$$(\Gamma_\mu \partial^\mu + iM)\psi(x) = 0$$

are constructed in which field $\psi(x)$, under space-time transformations from the proper Lorentz group SL(2,ℂ), cotransforms as a direct sum of indecomposable representations of SL(2,ℂ). The case in which the field $\psi(x)$ cotransforms according to a direct sum of irreducible representations of SL(2,ℂ) was worked out by Gel'fand and Yaglom.[2] (A generalization to a direct integral of irreducible representations of SL(2,ℂ) was done by Cantoni.[5]) Thus the present work can be viewed as an extension of the Gel'fand–Yaglom method to indecomposable so-called Harish–Chandra representations of SL(2,ℂ). These representations were studied and classified by Gel'fand and Ponomarev[6] in 1968. They will be briefly reviewed in Sec. 2. Section 3 contains the construction of field equations invariant under a direct sum of Harish–Chandra indecomposable representations. Finally, Sec. 4 is devoted to conclusions and to a brief comparison with other relativistic equations based on indecomposable representations of the Poincaré group.

2. INDECOMPOSABLE REPRESENTATIONS OF SL(2,ℂ)

Let $\mathscr{L} = \{h_+, h_-, h_3, f_+, f_-, f_3\}$ denote the complexification of the Lie algebra of the proper Lorentz group or SL(2,ℂ) and $\mathscr{L}_c = \{h_+, h_-, h_3\}$ the complexification of SU(2)—the maximal compact subalgebra of SL(2,ℂ). Generators of $\mathscr{L}$ satisfy the usual commutation relations:

$$[h_+, h_-] = 2h_3, \quad [h_\pm, h_3] = \mp h_\pm, \tag{2.1a}$$

$$[f_+, h_+] = [f_-, h_-] = [f_3, h_3] = 0,$$
$$[h_\pm, f_3] = \mp f_\pm, \quad [f_\pm, h_3] = \mp f_\pm, \tag{2.1b}$$
$$[h_+, f_-] = [f_+, h_-] = 2f_3,$$

$$[f_+, f_-] = -2h_3, \quad [f_\pm, f_3] = \pm h_\pm. \tag{2.1c}$$

Let T be a representation of $\mathscr{L}$ which when restricted to $\mathscr{L}_c$ decomposes into the direct sum of irreducible representations T_i of $\mathscr{L}_c$,

$$T = \bigoplus_i T_i, \tag{2.2}$$

with finite multiplicities (i.e. for every T_{i_0} there is only a finite number of T_i in T equivalent to T_{i_0}). The representation T of $\mathscr{L}$ is said to be a Harish–Chandra representation. They were studied by Gel'fand and Ponomarev.[6] The carrier space R of Harish–Chandra's representation T can be written as

$$R = \bigoplus_l \bigoplus_{m=-l}^{l} R_{l,m}, \tag{2.3}$$

where $R_{l,m}$ are subspaces formed by eigenvectors of the operators H_3 and H^2 (which correspond to h_3 and $h^2 = \frac{1}{2}(h_+h_- + h_-h_+) + h_3^2$ in representation T) with eigenvalues $m = l, l-1, \cdots, -l$ and $l(l+1), l = l_0, l_0+1, \cdots$, respectively. Representations of $h_\pm$, $f_\pm$, f_3 on R can be determined by means of the auxiliary operators $E_\pm$, $D_\pm$, D_0 mapping one subspace $R_{l,m}$ into another in the following way:

$$\begin{aligned} E_+ &: R_{l,m} \to R_{l,m+1}, \quad m = -l, -l+1, \cdots, l-1, \\ &: R_{l,l} \to 0, \\ E_- &: R_{l,m} \to R_{l,m-1}, \quad m = -l+1, \cdots, l, \\ &: R_{l,-l} \to 0, \end{aligned} \tag{2.4a}$$

$$\begin{aligned} D_+ &: R_{l,m} \to R_{l+1,m}, \quad m = -l, -l+1, \cdots, l, \\ D_0 &: R_{l,m} \to R_{l,m}, \quad m = -l, -l+1, \cdots, l, \\ D_- &: R_{l,m} \to R_{l-1,m}, \quad m = -l+1, \cdots, l-1, \\ &: R_{l,l} \to 0, \\ &: R_{l,-l} \to 0. \end{aligned} \tag{2.4b}$$

Operators E_+ and E_- are isomorphisms of the spaces $R_{l,m}$ with $R_{l,m+1}$ and $R_{l,m-1}$ respectively and commute with the operators D_+, D_- and D_0. The explicit form of the operators $H_\pm$, $F_\pm$, F_3 on $R_{l,m}$ in terms of $E_\pm$, $D_\pm$, and D_0 are given by the following expressions:

$$H_\pm \xi = [(l \pm m + 1)(l \mp m)]^{1/2} E_\pm \xi, \quad \xi \in R_{l,m}, \quad m \neq \pm l, \tag{2.5a}$$

$$H_\pm \xi = 0, \quad \xi \in R_{l,\pm l},$$

 0022-2488/81/081775-06$01.00 © 1981 American Institute of Physics

$$F_\pm \xi = \pm[(l\mp m)(l\mp m-1)]^{1/2}D_-E_\pm\xi - [(l\mp m)(l\pm m+1)]^{1/2}D_0E_\pm\xi \pm[(l\pm m+1)(l\pm m+2)]^{1/2}D_+E_\pm\xi,\ \xi\in R_{l,m}, \tag{2.5b}$$

$$F_3\xi = [l^2-m^2]^{1/2}D_-\xi - mD_0\xi - [(l+1)^2-m^2]^{1/2}D_+\xi,\quad \xi\in R_{l,m}. \tag{2.5c}$$

The explicit form of $D_\pm, D_0$ is determined from (2.1c) and will be specified in the next subsections. It was shown in Ref. 6 that each of the Casimir operators

$$\Delta_1 = \tfrac{1}{2}(H_-F_+ + H_+F_-) + H_3F_3,$$
$$\Delta_2 = H_-H_+ + F_-F_+ + H_3^2 - F_3^2 + 2H_3 \tag{2.6}$$

has only one eigenvalue λ_i, $i=1,2$, for every indecomposable representation T of $\mathscr{L}$. It is given by

$$\lambda_1 = il_0l_1,\quad \lambda_2 = l_0^2 + l_1^2 - 1, \tag{2.7}$$

where $2l_0$ is an integer and l_1 a complex number. According to the values of l_0 and l_1 we shall distinguish so-called ordinary and special indecomposable representations which will be specified in the next subsections.

Notice that the forms of the eigenvalues of λ_1 and λ_2 in (2.7) are the same as for irreducible representations of $\mathscr{L}$. For indecomposable representations there are, however, also other invariant numbers besides λ_1 and λ_2 which fully specify the considered representations (the Casimir operators contain nilpotent operators).

A. The ordinary indecomposable representations

These are the indecomposable representations T of $\mathscr{L}$ for which $l_1 - l_0$ is not an integer. The carrier space R of the ordinary indecomposable representation is of the form

$$R = \bigoplus_{l\geqslant|l_0|}\ \bigoplus_{m=-l}^{l} R_{l,m}, \tag{2.8}$$

where all $R_{l,m}$ are mutually isomorphic. We may choose the basis in $R_{l,m}$, $m\neq\pm l$, in which $E_\pm$, $D_\pm$ and D_0 are given by

$$[E_\pm]_{l,m} = [D_+]_{l,m} = [I_n],\quad m\neq\pm l,$$
$$[D_0]_{l,m} = \frac{il_0l_1}{l(l+1)}([I_n]+[a_0]), \tag{2.9}$$
$$[D_-]_{l,m} = \frac{l_0^2-l^2}{l^2(4l^2-1)}\{l^2[I_n] - l_1^2([I_n]+[a_0])^2\},\quad m\neq\pm l$$

where $[I_n]$ is an $n\times n$ unit matrix ($n=\dim R_{l,m}$) and $[a_0]$ is the $n\times n$ nilpotent matrix of the form

$$[a_0] = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0\\ 0 & 0 & 1 & \cdots & 0\\ - & - & - & - & -\\ 0 & 0 & 0 & \cdots & 1\\ 0 & 0 & 0 & \cdots & 0\end{pmatrix}. \tag{2.10}$$

Let us recall that subspaces $R_{l,\pm l}$ are mapped by $E_\pm$, $D_\pm$ to zero.

Thus we can see that any ordinary indecomposable representation is characterized by three numbers namely by l_0, l_1 and n which is equal to dimension of $R_{l,m}$. If $n=1$ the corresponding ordinary indecomposable representation reduces to an irreducible representation.

B. Special indecomposable representations

Indecomposable representations for which $l_1 - l_0$ is an integer are said to be special. Their carrier spaces R are of the form

$$R = \bigoplus_{l\geqslant|l_0|}\ \bigoplus_{m=-l}^{l} R_{l,m}. \tag{2.11}$$

All $R_{l,m}$ for $|l_0|\leqslant l\leqslant|l_1|-1$ (resp. $l\geqslant|l_1|$) are isomorphic and of dimension n_0 (respectively n_1). It is possible to choose the basis in R in such a way that the matrix representations of $e_\pm$, $d_\pm$, d_0 on R are given by

$$[E_\pm]_{l,m} = \begin{cases}[I_{n_0}] & \text{for } |l_0|\geqslant l\geqslant|l_1|-1,\\ [I_{n_1}] & \text{for } |l\geqslant|l_1|,\end{cases} \tag{2.12a}$$

$$[D_+]_{l,m} = \begin{cases}[I_{n_0}] & \text{for } |l_0|\leqslant l<|l_1|-1,\\ [d_+] & \text{for } l=|l_1|-1,\\ [I_{n_1}] & \text{for } l\geqslant|l_1|,\end{cases} \tag{2.12b}$$

$$[D_-]_{l,m} = \begin{cases}\dfrac{(l^2-l_0^2)(l_1^2-l^2)}{(4l^2-1)l^2}\left([I_{n_0}] + \dfrac{l_1^2}{l_1^2-l^2}[a_0]\right) & \text{for } |l_0|\leqslant l\leqslant|l_1|-1,\\ \dfrac{(l^2-l_0^2)(l_1^2-l^2)}{(4l^2-1)l^2}\left([I_{n_1}] + \dfrac{l_1^2}{l_1^2-l^2}[a_1] - \dfrac{l_0^2}{l_1^2-l^2}[\delta]\right) & \text{for } l>|l_1|,\\ [d_-] & \text{for } l=|l_1|,\end{cases} \tag{2.12c}$$

$$[D_0]_{l,m} = \begin{cases}\dfrac{il_0l_1}{l(l+1)}([a_0]+[I_{n_0}])^{1/2} & \text{for } l_0\leqslant l\leqslant|l_1|-1,\\ \dfrac{il_0l_1}{l(l+1)}([a_1]+[I_{n_1}]+[\delta])^{1/2} & \text{for } l\geqslant|l_1|.\end{cases} \tag{2.12d}$$

Here, $[I_{n_i}]$, $i=0,1$, are $n_i \times n_i$ unit matrices, $[\delta]$ and $[a_i]$, $i=0,1$, are nilpotent matrices, $[\delta]$, $[d_\pm]$ are specified by a finite chain of integers and by one complex number. $[a_i]$ are expressed by means of the operators

$$\begin{aligned} d_+ &: R_{|l_1|-1,m} \to R_{|l_1|,m}\,, \\ d_- &: R_{|l_1|m} \to R_{|l_1|-1,m}\,. \end{aligned} \tag{2.13}$$

These numbers together with l_0 and l_1 characterize a concrete special indecomposable representation (for details see Ref. 6).

Later we shall use two simplest special indecomposable (reducible) representations $(l_0,l_1,+)$ for which

$$d_+ = \mathbb{1}, \quad n_0 = n_1 = 1, \quad d_- = a_0 = a_1 = \delta = \mathbb{0}\,, \tag{2.14a}$$

and $(l_0,l_1,-)$ for which

$$d_- = \mathbb{1}, \quad n_0 = n_1 = 1, \quad d_+ = a_0 = a_1 = \delta = \mathbb{0}\,. \tag{2.14b}$$

Notice that the cases in which $d_+ = d_- = a_1 = a_0 = \delta = \mathbb{0}$ correspond either to the finite-dimensional ($|l_0| < |l_1|, n_0 = 1, n_1 = 0$) or infinite-dimensional ($|l_0| \geqslant |l_1|, n_0 = 0, n_1 = 1$) irreducible representations of $\mathscr{L}$.

3. RELATIVISTICALLY INVARIANT EQUATIONS

We shall construct now relativistic equations for the wave function which transforms according to the Harish–Chandra representation T of $\mathscr{L}$, i.e., according to a direct sum of indecomposable representations τ_j of $\mathscr{L}$,

$$T = \bigoplus_j \tau_j\,. \tag{3.1}$$

Since any linear higher order partial differential equation is equivalent to a system of linear first-order partial differential equations we shall restrict ourselves to the equation of the form

$$(\Gamma_\mu \partial^\mu + iM)\psi(x) = 0\,, \tag{3.2}$$

where $\psi(x)$ transforms according to (3.1) and, consequently, is an infinite component function on the Minkowski space and Γ_μ, $\mu = 0,1,2,3$ and M are infinite matrices acting in the component space of ψ.

The Eq. (3.2) is relativistically invariant if there exists the representation T of $\mathscr{L}$ such that the equations

$$\begin{aligned} T(A)^{-1}\Gamma_\mu T(A) &= \Lambda(A)_\mu^\nu \Gamma_\nu\,, \\ T(A)^{-1} M T(A) &= M \end{aligned} \tag{3.3}$$

hold whenever

$x_\mu \mapsto x'_\mu = \Lambda(A)_\mu^\nu x_\nu + a_\nu$, $\Lambda(A), a \in$ the Poincaré group, and field $\psi(x)$ cotransform according to the rule

$$\psi(x) \mapsto \psi'(x') = T(A)\psi(x)\,.$$

The Eq. (3.3) lead to the relations (see e.g. Ref. 7)

$$\Gamma_i = [\Gamma_0, B_i], \quad i = 1,2,3\,, \tag{3.4a}$$

which express the operators Γ_i, $i = 1,2,3$ in terms of Γ_0 and to the relations:

$$[\Gamma_0, H_+] = [\Gamma_0, H_-] = [\Gamma_0, H_3] = 0 \tag{3.4b}$$

and

$$\Gamma_0 = [[F_3, \Gamma_0], F_3]\,, \tag{3.4c}$$

which specify the transformation properties of Γ_0 under $\mathscr{L}_c$ and the boost F_3. Notice that the operators $H_\pm, H_3, F_\pm$ and F_3 denote the operators $h_\pm, h_3, f_\pm$ and f_3 in the representations of $\mathscr{L}$ described in the previous sections and

$$B_1 = (1/2i)(F_+ + F_-), \quad B_2 = \tfrac{1}{2}(F_- - F_+)$$

and

$$B_3 = -iF_3\,. \tag{3.4d}$$

A. Equations for the matrix elements of Γ_0

First let us specify the basis in the carrier space $R = \bigoplus_j R^{\tau_j}$ of the Harish–Chandra representation $T = \bigoplus_j \tau_j$ in (3.1). By virtue of (2.2) the carrier spaces R^{τ_j} of the indecomposable representations τ_j may be written as

$$R^{\tau_j} = \bigoplus_{l,m} R^{\tau_j}_{l,m}\,.$$

By taking the basis $\{\,|\tau_j;l,m,i\rangle$, $i = 1,2,\ldots,\dim R^{\tau_j}_{l,m}$, $l = l_0, l_0+1,\ldots, m = -l, -l+1,\ldots,l\,\}$ in every $R^{\tau_j}_{l,m}$, the basis in R can be chosen as a union of the bases of all $R^{\tau_j}_{l,m}$ occurring in R.

In this basis the relations (3.4b) imply that the matrix elements of matrix Γ_0 are of the form

$$\langle \tau;l,m,i|\Gamma_0|\tau';l',m'i'\rangle = \delta_{ll'}\delta_{mm'}[X_l^{\tau\tau'}]_{ii'}\,, \tag{3.5}$$

i.e. Γ_0 is diagonal with respect to the SU(2) indices m and l. From (3.4c) [with the use of (2.5c)] we obtain the following, rather complicated, system of algebraic homogeneous equations for matrices $[X_l^{\tau\tau'}]$:

$$[X_{l+1}^{\tau\tau'}][P_{l+1}^{\tau'}][P_l^{\tau'}] - 2[P_{l+1}^{\tau}][X_l^{\tau\tau'}][P_l^{\tau'}] + [P_{l+1}^{\tau}][P_l^{\tau}][X_{l-1}^{\tau\tau'}] = 0\,, \tag{3.6a}$$

$$[M_l^{\tau}][M_{l+1}^{\tau}][X_{l+1}^{\tau\tau'}] - 2[M_l^{\tau}][X_l^{\tau\tau'}][M_{l+1}^{\tau'}] + [X_{l-1}^{\tau\tau'}][M_l^{\tau'}][M_{l+1}^{\tau'}] = 0\,, \tag{3.6b}$$

$$\begin{aligned} &[X_l^{\tau\tau'}]([P_l^{\tau'}][Z_{l-1}^{\tau'}] + [Z_l^{\tau'}][P_l^{\tau'}]) - 2[Z_l^{\tau}][X_l^{\tau\tau'}][P_l^{\tau'}] \\ &\quad = 2[P_l^{\tau}][X_{l-1}^{\tau\tau'}][Z_{l-1}^{\tau'}] - ([P_l^{\tau}][Z_{l-1}^{\tau}] + [Z_l^{\tau}][P_l^{\tau}])[X_{l-1}^{\tau\tau'}]\,, \end{aligned} \tag{3.6c}$$

$$\begin{aligned} &[X_{l-1}^{\tau\tau'}]([M_l^{\tau'}][Z_l^{\tau'}] + [Z_{l-1}^{\tau'}][M_l^{\tau'}]) - 2[Z_{l-1}^{\tau}][X_{l-1}^{\tau\tau'}][M_l^{\tau'}] \\ &\quad = 2[M_l^{\tau}][X_l^{\tau\tau'}][Z_l^{\tau'}] - ([M_l^{\tau}][Z_l^{\tau}] + [Z_{l-1}^{\tau}][M_l^{\tau}])[X_l^{\tau\tau'}]\,, \end{aligned} \tag{3.6d}$$

$$[X_l^{\tau\tau'}]([P_l^{\tau'}][M_l^{\tau'}] + [M_{l+1}^{\tau'}][P_{l+1}^{\tau'}] + [Z_l^{\tau'}][Z_l^{\tau'}])$$
$$= 2([P_l^{\tau}][X_{l-1}^{\tau\tau'}][M_l^{\tau'}] + [M_{l+1}^{\tau}][X_{l+1}^{\tau\tau'}][P_{l+1}^{\tau'}] + [Z_l^{\tau}][X_l^{\tau\tau'}][Z_l^{\tau'}])$$
$$- ([P_l^{\tau}][M_l^{\tau}] + [M_{l+1}^{\tau}][P_{l+1}^{\tau}] + [Z_l^{\tau}][Z_l^{\tau}])[X_l^{\tau\tau'}], \tag{3.6e}$$

$$(l+1)^2([X_l^{\tau\tau'}][M_{l+1}^{\tau'}][P_{l+1}^{\tau'}] - 2[M_{l+1}^{\tau}][X_{l+1}^{\tau\tau'}][P_{l+1}^{\tau'}] + [M_{l+1}^{\tau}][P_{l+1}^{\tau}][X_l^{\tau\tau'}])$$
$$+ l^2([X_l^{\tau\tau'}][P_l^{\tau'}][M_l^{\tau'}] - 2[P_l^{\tau}][X_{l-1}^{\tau\tau'}][M_l^{\tau'}] + [P_l^{\tau}][M_l^{\tau}][X_l^{\tau\tau'}])) = [X_l^{\tau\tau'}], \tag{3.6f}$$

where $[Z_l^{\tau}]$, $[P_l^{\tau}]$, $[M_l^{\tau}]$ are matrices matrix elements of which are defined by

$$[Z_l^{\tau}]_{ii'} = \langle \tau;l,m,i|D_0|\tau;l,m,i'\rangle, \tag{3.7a}$$

$$[P_l^{\tau}]_{ii'} = \langle \tau;l+1,m,i|D_+|\tau;l,m,i'\rangle, \tag{3.7b}$$

and

$$[M_l^{\tau}]_{ii'} = \langle \tau;l-1,m,i|D_-|\tau;l,m,i'\rangle, \tag{3.7c}$$

If τ and τ' are irreducible representations of $\mathscr{L}$ the system (3.6a–f) reduces to that solved by Gel'fand and Yaglom.[2] Here, we shall discuss the cases in which τ and τ' are indecomposable representations of $\mathscr{L}$.

B. Matrix elements of Γ_0 for ordinary indecomposable representations

This case was briefly discussed in Ref. 7. By virtue of (3.7a)–(3.7c) and (2.9) and after some algebraic operations the solutions of Eqs. (3.6a)–(3.6c) can be written as

$$[X_l^{\tau\tau'}] = l(l+1)\{l[Z_l^{\tau}][\mathscr{H}^{\tau\tau'}] - (l+1)[\mathscr{H}^{\tau\tau'}][Z_l^{\tau'}]\}, \tag{3.8}$$

where $[\mathscr{H}^{\tau\tau'}]$ is an arbitrary $n'\times n$ matrix. The matrix $[X_l^{\tau\tau'}]$ in (3.8) represents the solution of the whole system (3.6a)–(3.6f) if and only if

i. $\tau = (l_0,l_1,n)$ and $\tau' = (l_0',l_1',n')$ are interlocked, i.e. either $l_0' = l_0$ and $l_1' = l_1 \pm 1$ or $l_0' = l_0 \pm 1$ and $l_1' = l_1$ and
ii. $[\mathscr{H}^{\tau\tau'}][a_0^{\tau'}] = [a_0^{\tau}][\mathscr{H}^{\tau\tau'}]$. By virtue of (2.10) the last condition leads

$$[\mathscr{H}^{\tau\tau'}] = \begin{pmatrix} 0 & 0 & \dots & 0 & \mathscr{H}_1 & \mathscr{H}_2 & \dots & \mathscr{H}_n \\ 0 & 0 & \dots & 0 & 0 & \mathscr{H}_1 & \dots & \mathscr{H}_{n-1} \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & \mathscr{H}_1 \end{pmatrix} \quad \text{for } n' \geqslant n \tag{3.9a}$$

or

$$[\mathscr{H}^{\tau\tau'}] = \begin{pmatrix} \mathscr{H}_1 & \mathscr{H}_2 & \dots & \mathscr{H}_{n'} \\ 0 & \mathscr{H}_1 & \dots & \mathscr{H}_{n'-1} \\ 0 & 0 & \dots & \mathscr{H}_1 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \end{pmatrix} \quad \text{for } n' \leqslant n. \tag{3.9b}$$

Thus we have proved the following result:

Theorem: The Eq. (3.2) in which $\psi(x)$ cotransforms as $\psi(x) \mapsto \psi(x') = T(A)\psi(x)$ under the Poincaré transformation (A,a) of x is relativistically invariant if the matrix elements of Γ_0 is of the form

$$\langle \tau;l,m,i|\Gamma_0|\tau',l',m',i'\rangle = \delta_{ll'}\delta_{mm'}\{l^2(l+1)[Z_l^{\tau}]_{ij}[\mathscr{H}^{\tau\tau'}]_{ji'} - l(l+1)^2[\mathscr{H}^{\tau\tau'}]_{ij}[Z_l^{\tau'}]_{ji'}\}, \tag{3.10}$$

where matrices $[Z_l^{\tau}]$ and $[\mathscr{H}^{\tau\tau'}]$ are defined in (3.7a), (2.9), and (3.9), respectively. The matrices Γ_i are given by (3.4a).

The peculiar form of $\mathscr{H}^{\tau\tau'}$ yields further restrictions on the choice of τ and τ' as stated in the following lemma.

Lemma 1: Let the wave function ψ be a nontrivial solution of the relativistic equation (3.2) with $M \neq 0$ which transforms according to a direct sum of ordinary indecomposable representations. Then to any indecomposable component $\tau = (l_0,l_1,n)$ in the direct sum there is an indecomposable representation $\tau' = (l_0',l_1',n')$ which is interlocked with τ and for which $n' \geqslant n$.

Proof: Let there be a representation $\tau = (l_0,l_1,n)$ such that for every $\tau' = (l_0',l_1',n')$ interlocked with it n' is smaller than n. Then, by virtue of (3.10), we have

$$\langle \tau;l,m,i|\Gamma_0|\tau';l',m',i'\rangle = 0 \quad \text{for } i > n'.$$

Consider Eq. (3.2) expressed in p-representation. Then, in the rest frame, we obtain

$$0 = \langle \tau;l,m,i|\,p_0\Gamma_0|\psi\rangle = -M\langle \tau;l,m,i|\psi\rangle$$

for any l,m and $i > m'$. This means that ψ actually transforms according to a direct sum of representations $\tilde{\tau}$ smaller than τ. In other words, ψ transforms according to $\tilde{\tau} = (l_0,l_1,\tilde{n})$ interlocked with τ' and for which $\tilde{n} \leqslant n'$.

Corollary 1: There must be at least two indecomposable interlocked representations with maximal n in representation T.

Corollary 2: If $T = \tau \oplus \tau'$ then

i. $\tau = (l_0,l_1,n)$ must be interlocked with $\tau' = (l_0',l_1',n')$ (i.e. either $l_0' = l_0 \pm 1$, $l_1' = l_1$ or $l_0' = l_0$, $l_1' = l_1 \pm 1$) and
ii. $n = n'$.

Example: There are two well-known field equations in which the fields transform according to infinitely dimensional irreducible representations of the Lorentz group—the Majorana equations. The fields in them transform according to one of the self-interlocked representations $(0,\frac{1}{2})$ or $(\frac{1}{2},0)$ (see Ref. 8). Let us consider their generalizations, i.e. the equations based on the indecomposable self-interlocked representations $(\frac{1}{2},0,n)$ or $(0,\frac{1}{2},n)$. We call them the extended Majorana equations. The corresponding matrix Γ_0 has a diagonal block form in l,m consisting of

$$[X_l^{\tau\tau}] = (l+\tfrac{1}{2})[\mathscr{H}^{\tau\tau}] \quad \text{or} \quad [X_l^{\tau\tau}] = \{(l+\tfrac{1}{2})[I_n] + [a_n]\} \times [\mathscr{H}^{\tau\tau}] \tag{3.11}$$

respectively, where $[\mathscr{H}^{\tau\tau}]$ is $n\times n$ matrix given by (3.9) and $[a_0]$ by (2.10). Thus, taking into account (3.5), (3.4a), the extended Majorana equation based, for instance, on $(\frac{1}{2},0,n)$ is of the form

$$\frac{\partial}{\partial x_0}(l+\tfrac{1}{2})[\mathscr{H}]\psi_{lm} + i\frac{\partial}{\partial x_3}\{(l^2-m^2)^{1/2}[\mathscr{H}]\psi_{l-1,m} + ((l+1)^2-m^2)^{1/2}[M_l^\tau][\mathscr{H}]\psi_{l+1,m}\}$$
$$+\frac{1}{2}\left(i\frac{\partial}{\partial x_1}-\frac{\partial}{\partial x_2}\right)\{((l-m-1)(l-m))^{1/2}[\mathscr{H}]\psi_{l-1,m+1} - ((l+m+2)(l+m+1))^{1/2}[M_l^\tau][\mathscr{H}]\psi_{l+1,m+1}\}$$
$$-\frac{1}{2}\left(i\frac{\partial}{\partial x_1}+\frac{\partial}{\partial x_2}\right)\{((l+m-1)(l+m))^{1/2}[\mathscr{H}]\psi_{l-1,m-1} - ((l-m+2)(l+m+1))^{1/2}[M_l^\tau][\mathscr{H}]\psi_{l+1,m-1}\}$$
$$+iM\psi_{l,m}=0. \quad (3.12)$$

Here, ψ_{lm} is a column vector consisting of n components; $[\mathscr{H}]$ and $[M_l^\tau]$ are $n\times n$ matrices given by

$$[\mathscr{H}] = \begin{pmatrix} \mathscr{H}_1 & \mathscr{H}_2 & \cdots & \mathscr{H}_n \\ 0 & \mathscr{H}_1 & \cdots & \mathscr{H}_{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathscr{H}_2 \\ 0 & 0 & \cdots 0 & \mathscr{H}_1 \end{pmatrix} \quad (3.13)$$

and

$$[M_l^\tau] = \frac{1}{4}\left\{\frac{[a_0]^2+2l_1[a_0]}{l^2-l_1^2} - [I_n]\right\}, \quad (3.14)$$

where $\mathscr{H}_i$ are unspecified constants and $[a_0]$, $[I_n]$ are defined in (2.10), (2.9) respectively.

Concerning the mass spectrum of Eq. (3.12) we can easily find that to any "spin" l there corresponds the mass $M/(l+\frac{1}{2})$. Thus the situation is the same as for the Majorana equations except that, now, the "mass operator" consists also of nilpotent operators.

C. Matrix elements of Γ_0 for special indecomposable representations

In contrast to the previous case we do not know the general explicit solution of system (3.6a)–(3.6f) for special indecompossable representations since the generators of $\mathscr{L}$ are represented by more complicated structures. We can solve, however, system (3.6) in some particular cases, e.g. when ψ cotransforms under the representation $(\frac{1}{2},l_1,+)\oplus(-\frac{1}{2},l_1,-)$. Let us denote $(\frac{1}{2},l_1,+)$ and $(-\frac{1}{2},l_1-)$ by τ and τ' respectively. Then the matrices (3.7) are numbers of the form

$$Z_l^\tau = -Z_l^{\tau'} = il_1/2(l+1)l, \quad (3.15a)$$

$$P_l^\tau = 1, \quad (3.15b)$$

$$P_l^{\tau'} = \begin{cases} 1 & \text{for } l\neq|l_1|, \\ 0 & \text{for } l=|l_1|, \end{cases} \quad (3.15c)$$

$$M_l^\tau = \begin{cases} (l^2-l_0^2)(l_1^2-l^2)/l^2(4l^2-1) & \text{for } l\neq|l_1|, \\ 0 & \text{for } l=|l_1|, \end{cases} \quad (3.15d)$$

$$M_l^{\tau'} = \begin{cases} (l^2-l_0^2)(l_1^2-l^2)/l^2(4l^2-1) & \text{for } l\neq|l_1|, \\ 1 & \text{for } l=|l_1|. \end{cases} \quad (3.15e)$$

The corresponding solutions of system (3.6) are given by

$$X_l^{\tau\tau'} = \begin{cases} l_1(l+\frac{3}{2})\mathscr{H}^{\tau\tau'}, & \text{for } l<|l_1|, \\ 0, & l\geqslant|l_1|, \end{cases} \quad (3.16a)$$

and

$$X_l^{\tau'\tau} = \begin{cases} 0, & l<|l_1|, \\ l_1(l+\frac{3}{2})\mathscr{H}^{\tau'\tau}, & l\geqslant|l_1|. \end{cases} \quad (3.16b)$$

Here, $\mathscr{H}^{\tau\tau'}$, $\mathscr{H}^{\tau'\tau}$ are constants and τ and τ' are interlocked. However, the interrelations of special indecomposable representations for a nontrivial relativistic equation might be more tricky than in the case of ordinary indecomposable representations as the following lemma indicates.

Lemma 2: If function ψ in Eq. (3.2) cotransforms under the Poincaré transformation according to representation $(\frac{1}{2},l_1,+)\oplus(-\frac{1}{2},l_1,-)$ then the corresponding relativistic Eq. (3.2) has only a trivial solution.

Proof: From Eq. (3.2) in the rest frame and the matrix elements (3.16) of Γ_0 we obtain

$$0=\frac{p_0}{M}\langle\tau;l,m|\Gamma_0|\psi\rangle = \langle\tau;l,m|\psi\rangle \quad \text{for } l\geqslant|l_1|, \quad (3.17a)$$

and

$$0=\frac{p_0}{M}\langle\tau';l,m|\Gamma_0|\psi\rangle = \langle\tau';l,m|\psi\rangle \quad \text{for } l<|l_1|. \quad (3.17b)$$

Since the matrix elements of Γ_0 can be nonvanishing only for τ and τ' interlocked then by virtue of (3.2), (3.16), and (3.17b) we get

$$\begin{aligned}\langle\tau;l,m|\psi\rangle &= \frac{p_0}{M}\langle\tau;l,m|\Gamma_0|\psi\rangle \\ &= \frac{p_0}{M}\langle\tau;l,m|\Gamma_0|\tau';l,m\rangle\langle\tau';l,m|\psi\rangle \\ &= \frac{p_0}{M}\mathscr{H}^{\tau\tau'}l_1(l+\tfrac{3}{2})\langle\tau';l,m|\psi\rangle = 0 \quad \text{for } l<|l_1|\end{aligned}$$

and analogously

$$\langle\tau';l,m|\psi\rangle = \frac{p_0}{M}\mathscr{H}^{\tau'\tau}l_1(l+\tfrac{3}{2})\langle\tau;l,m|\psi\rangle = 0 \quad \text{for } l\geqslant|l_1|.$$

Consequently $|\psi\rangle\equiv 0$.

4. DISCUSSION

The indecomposable representations of the Poincaré group appear naturally in the case of massless particles.[9] As emphasized in Ref. 10 the indecomposable representations provide also a suitable framework for description of unstable particles. Moreover, quantum field theory of fields transforming according to indecomposable representations reveals new interesting features.[11,12] Thus, it seems to be interesting to investigate the field equations associated with the indecomposable representations.

The indecomposable representations of the Poincaré group $T^4 \circledS \mathrm{SL}(2,\mathbb{C})$ arise in three possible ways:

i). by taking indecomposable representations of T^4 and usual decomposable representations of SL(2,$\mathbb{C}$),

ii). by taking representations of T^4 decomposable but representations of SL(2,$\mathbb{C}$) indecomposable, and

iii). by taking both representations of T^4 and of SL(2,$\mathbb{C}$) indecomposable.

The first possibility was investigated by Adamczyk and Raczka[13] by using the technique of induced representations. In particular the new finite-component field equations associated with indecomposable representations for spin $\frac{1}{2}$ and 1 were discussed in detail.

The second possibility was treated in the present paper (see also Ref. 7). As we have seen the corresponding new relativistic field equations are associated with the Gel'fand–Ponomarev indecomposable representations of SL(2,$\mathbb{C}$). They were classified by using Gel'fand–Yaglom's algebraic technique. All appear to be infinite-dimensional multispin equations. Although they carry new quantum numbers their physical relevance is not clear.

[1] Proceedings of the "Ettore Majorana" International School of Mathematical Physics, Erice 1977, edited by G. Velo and A. S. Wightman, *Lecture Notes in Physics*; Vol. 73 (Springer–Verlag, Berlin, 1978).

[2] I. M. Gel'fand and Y. Yaglom, J. Exper. Theor. Phys. **18**, 703, 1096 (1948).

[3] F. Bruhat, Bull. Soc. Math. France **84**, 97 (1956).

[4] L. Garding, Medd. Lunds Mat. Sem. **6**, 1 (1944).

[5] V. Cantoni, Symposia Math. **12**, 13 (1973).

[6] I. M. Gel'fand and V. A. Ponomarev, Uspechi Mat. Nauk **23**, 3 (1968), see also G. Pinczon and J. Simon, Lett. Math. Phys. **1**, 83 (1975).

[7] L. Hlavatý and J. Niederle, Czech. J. Phys. B **29**, 283 (1979).

[8] I. M. Gel'fand, R. A. Minlos, and Z. Y. Shapiro; *Representations of the rotation group and the Lorentz group* (Pergamon New York, 1963).

[9] A. O. Barut and R. Raczka, Ann. Inst. H. Poincaré A **17**, 111 (1972).

[10] R. Raczka, Ann. Inst. H. Poincaré A **19**, 341 (1973); A. O. Barut and R. Raczka: *Theory of Group Representations and Applications* (PWN, Warsaw, 1977).

[11] E. Iacopini, Nuovo Cimento A **18**, 205 (1973).

[12] J. Lukierski, Nuovo Cimento Ser. X A **54**, 583 (1968).

[13] J. Adamczyk and R. Raczka, *New Relativistic Wave Equations Associated with Indecomposable Representations of the Poincaré Group*, Preprint IC/78/100.

On the existence of neutrino "zero-modes" in vacuum spacetimes [a)]

Amitabha Sen [b)]

Center for Theoretical Physics, [c)] *Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742 and Enrico Fermi Institute, University of Chicago, Chicago, Illinois 60637*

(Received 3 December 1980; accepted for publication 23 January 1981)

Neutrino "zero modes" in curved spacetime, the analog of static solutions of the neutrino equation in flat space, are defined as the kernel of an elliptic operator obtained from a "3 + 1" decomposition of the neutrino equation relative to a spacelike hypersurface. In this paper, vacuum, globally hyperbolic spacetimes that admit "zero modes" are characterized.

PACS numbers: 11.10.Qr, 14.60.Gh

I. INTRODUCTION

In flat spacetime "zero modes" refer to static or zero-frequency normalizable solutions of a given field equation. Our purpose here is to introduce a notion of "zero modes" of the neutrino equation in a background curved spacetime and to address the following problem: Characterize vacuum spacetimes, i.e., solutions of vacuum Einstein's equations, that admit neutrino "zero modes."

There are several physical situations that motivate this investigation. We mention two. The first is directed towards understanding the structure of the "vacuum state" in quantum gravity. The idea follows that of Jackiw and Rebbi,[1] who consider (in flat spacetime) a theory of Dirac field (gauge invariantly) coupled to a Yang–Mills theory of spontaneously broken isospin symmetry with a triplet of spinless mesons. The latter, in the absence of the fermion, is known to possess static monopole (soliton) solutions with finite energy. In the background field of this soliton, the Dirac equation has (c-number) "zero mode" solutions. Jackiw and Rebbi[1] have shown how the existence of these "zero mode" solutions can be consistently interpreted at the quantum level as signalling a degenerate soliton state with fermion number $\pm\frac{1}{2}$. Their analysis follows a general strategy[2] for extracting information about solutions of a quantum field theory by studying fluctuations about a classical background field (such as the soliton). One might expect to gain insight about the "vacuum state" in quantum gravity in a similar fashion. Thus, for instance, consider a neutrino field minimally coupled to gravity. In the absence of the fermion, one solves the vacuum Einstein's equations (possibly with a cosmological constant) for a c-number solution which will serve as a background field (spacetime). Next look for "zero modes" of the neutrino equation in this background spacetime, and, if they exist, try to interpret the result in the manner described by Jackiw and Rebbi. Clearly, the key step in this program is to search for (vacuum) spacetimes that admit neutrino "zero modes."

Another situation where neutrino "zero modes" play a role is in a theory of a spin-3/2 field coupled to gravity (supergravity theory). In a semiclassical approximation, one is led to consider a quantum theory of a massless spin-3/2 (Rarita–Schwinger) field on a background Einstein spacetime, i.e., solutions of vacuum Einstein's equations with or without cosmological constant. The spin-3/2 field has a gauge freedom up to addition of a gradient of a neutrino field. It turns out that in the quantum theory of the spin-3/2 field the gauge can be unambiguously "fixed" if and only if the underlying spacetime does *not* admit neutrino "zero modes."[3] If the spacetime does admit the "zero modes," then one is forced to consider a quantum theory on an indefinite metric Hilbert space, the physical consequences of which are yet unclear.

The strategy for the analysis of the problem is simple. The key step is to obtain a "3 + 1" decomposition of the neutrino equation relative to a spacelike hypersurface Σ in the spacetime. Setting the "time" derivative of the field to zero, the notion of "zero mode" emerges as the kernel of an elliptic operator on Σ. The conditions imposed on the spacetime by requiring the kernel to be nonzero then characterize the spacetimes that admit neutrino "zero modes." Since our method crucially involves the use of the vacuum Einstein's field equations, we have restricted attention to only vacuum spacetimes.[4]

The central result of this note, contained in the theorem in Sec. III, is that vacuum, globally hyperbolic spacetimes with a complete Cauchy hypersurface that admit neutrino "zero modes" are algebraically special of Petrov type III (hence, also type N or flat). The proof of this result rests on an assumption, presented here in the form of a conjecture, that on a complete Cauchy hypersurface the boundary terms, resulting from certain intergration by parts, are negligilble. For closed or asymptotically flat hypersurfaces, this assumption is indeed valid.

We begin by summarizing the technique of "3 + 1" decomposition of spinor field equations in Sec. II, primarily to fix our notation and to assemble the results needed for the main theorem. Details of this technique is given in Ref. 3. The main theorem is proved in Sec. III, and in the final section we conclude with a discussion of the result.

II. PRELIMINARIES

We briefly review here the technique for obtaining a "3 + 1" decomposition of spinor field equations in curved spacetime and assemble the main results which we shall use in the next section. (For details see Appendix A of Ref. 3.) We shall throughout work with two-component or Weyl

a) Supported in part by the NSF, under Contract PHY 78-24275 with the University of Chicago.

b) Fellow of Center for Theoretical Physics, University of Maryland.

c) Present address.

 © 1981 American Institute of Physics

spinors (rather than Dirac 4-spinors) and our notation will be that of Pirani.[5]

We briefly review here the technique for obtaining a "3 + 1" decomposition of spinor field equations in curved spacetime and assemble the main results which we shall use in the next section. (For details see Appendix A of Ref. 3.) We shall throughout work with two-component or Weyl spinors (rather than Dirac 4-spinors) and our notation will be that of Pirani.[5]

A. Space spinors

Consider a globally hyperbolic, orientable spacetime[6] (M, g_{ab}) and let Σ denote a Cauchy hypersurface in M with an everywhere timelike future-directed unit normal vector field t^a. Given a spinor field on M, its restriction to Σ is defined to be a spinor field on Σ or a *space spinor* field. Let S_Σ denote the collection of all space spinor fields on Σ. One property of space spinors is that they can be described entirely by spinors of one kind, say unprimed spinors, on Σ. The reason for this is that the vector field t^a, regarded as a Hermitian (space) spinor $t^{A'A}$, provides a natural isomorphism between primed and unprimed spinors. This isomorphism is displayed in the following notation which we shall adopt. Given $\lambda_{A'} \in S_\Sigma$, the corresponding unprimed space spinor will be denoted by

$$\lambda^{A+} = \sqrt{2}\, t^{AA'}\lambda_{A'}. \tag{1}$$

Since t^a is unit,

$$t^{AA'}t_{BA'} = \tfrac{1}{2}\delta^A{}_B, \quad t^{AA'}t_{AB'} = \tfrac{1}{2}\delta^{A'}{}_{B'}, \tag{2}$$

where $\delta^A{}_B$ is the Kronecker delta symbol. From (1) and (2) note that

$$(\lambda^{A+})^+ = -\lambda^A. \tag{3}$$

The Hermitian spinor $t^{A'A}$ is positive definite,[7] and thus one obtains a natural Hermitian, positive-definite inner product (,) on the space V of unprimed space spinors defined by

$$(\lambda, \eta) = \lambda^{+A}\eta_A \quad , \quad \lambda_A \ , \eta_A \in V. \tag{4}$$

The group that preserves the structure of $(V, (\ ,\ ), \epsilon_{AB})$ (where ϵ_{AB} is the usual symplectic form on V) is SU(2). So space spinors are in fact SU(2) spinors on Σ. It is this feature which allows us to relate space spinors to the geometry of the three-dimensional hypersurface Σ with the induced (negative-definite) metric $h_{ab} := g_{ab} - t_a t_b$.[8] Henceforth, we shall denote elements of S_Σ by unprimed spinor fields on Σ and shall raise and lower spinor indices with ϵ_{AB} in the usual way.[5]

Spatial tensors can be expressed in terms of space spinors on Σ. [A tensor on M is said to be spatial relative to Σ if the contraction of any of its indices with t^a or $t_b (= t^a g_{ab})$ vanishes.] A spatial covector, for example, is represented by

$$S_{AB} := (\sqrt{2}) t_A{}^{A'} S_{BA'} = S_{(AB)}, \tag{5}$$

where $S_{AA'}$ is the usual spinor form of the covector S_a. The symmetry of the indices AB in S_{AB} follows from the fact that $t^{AA'}S_{AA'} = 0$, i.e., S_a is spatial. In general, to express any spatial tensor in terms of SU(2) spinors on Σ, the rule is to replace each tensor index by a pair of symmetrized (unprimed) spinor indices.

Two spatial tensors of particular interest are the spatial metric h_{ab} on Σ and the extrinsic curvature $\pi_{ab} (= \pi_{(ab)})$ of Σ defined by $\pi_{ab} = h_a{}^m h_b{}^n \nabla_m t_n$, where ∇_a is the derivative operator defined by g_{ab}. (Tensor indices are raised and lowered using g_{ab}.) In terms of SU(2) spinors,

$$h_{ab} \equiv h_{AC\,BD} = h_{(AC)(BD)} = -\epsilon_{B(A}\epsilon_{C)D}, \tag{6}$$

$$\pi_{ab} \equiv \pi_{AC\,BD} = \pi_{(AC)(BD)}. \tag{7}$$

Also, since $\pi_{ab} = \pi_{ba}$,

$$\pi_{AC\,BD}\epsilon^{AB} = \tfrac{1}{2}\epsilon_{CD}\pi, \tag{8}$$

where $\pi = \pi_{ab}h^{ab}$ is the trace of the extrinsic curvature of Σ. The last step in (6) is obtained from the definition

$$h_{AC\,BD} := 2 t_C{}^{A'} t_D{}^{B'} h_{A'A\,B'B}$$

and

$$h_{A'A\,B'B} := \epsilon_{A'B'}\epsilon_{AB} - t_{A'A}t_{B'B}.$$

B. Derivative operator

We introduce a "spatial" derivative operator on space spinors, which refers only to the intrinsic geometry of Σ. Let λ_C be a space spinor, and consider a derivative operator $D_{(AB)}$ whose action on λ_C is defined by

$$D_{AB}\lambda_C := (\sqrt{2}) t_{(A}{}^{A'}\nabla_{B)A'}\lambda_C + (1/\sqrt{2})\pi_{ABCD}\lambda^D, \tag{9}$$

where $\nabla_{A'A}$ is the spinor form of the torsion-free covariant derivative operator ∇_a on M defined by g_{ab}. The action of $D_{(AB)}$ on scalar fields ϕ on Σ is

$$D_{AB}\phi = (\sqrt{2}) t_{(A}{}^{A'}\nabla_{B)A'}\phi,$$

and its action on spinors of higher valence is extended by Leibniz's rule. It can be shown[3] that D_{AB} defined by (7) is indeed the unique derivative operator on Σ defined by the metric h_{ab}.

The following two important properties of D_{AB} are useful. (We state them without proof, referring the reader to Ref. 3 for details.)

$$\text{(i)}\ D_{M(A}D_{B)}{}^M\lambda_C = \tfrac{1}{4}(2R_{ACBD} - \epsilon_{AB}\epsilon_{CD}R/2)\lambda^D, \tag{10}$$

where R_{ACBD} is the spinor form of the Ricci tensor R_{ab} of Σ and $R = R_{ab}h^{ab}$ is the scalar curvature of Σ.

$$\text{(ii)}\ (D_{AB}\lambda_C)^+ = -D_{AB}\lambda_C^+, \tag{11}$$

where

$$(D_{AB}\lambda_C)^+ = (2)^{3/2} t_A{}^{A'} t_B{}^{B'} t_C{}^{C'} D_{A'B'}\bar{\lambda}_{C'},$$

$$\lambda_C^+ = (\sqrt{2}) t_C{}^{C'}\bar{\lambda}_{C'}.$$

C. "3 + 1" decomposition

To express a field equation in "3 + 1" form, we need to write the covariant derivative operator $\nabla_{A'A}$ in terms of a spatial derivative and a suitable time derivative. The first step is to "unprime" the primed index A' in $\nabla_{A'A}$:

$$(\sqrt{2}) t_B{}^{A'}\nabla_{A'A} = \tfrac{1}{2}(\sqrt{2})\epsilon_{AB}\, t\cdot\nabla + (\sqrt{2}) t_{(A}{}^{A'}\nabla_{B)A'}, \tag{12}$$

where $t\cdot\nabla = t^{M'M}\nabla_{M'M}$ is the time derivative. The second term in (12) can be expressed as a spatial derivative D_{AB} defined by (9). We give two examples.

1. Neutrino equation

$$\nabla_{A'A}\lambda^{A} = 0; \tag{13}$$

"unpriming" the primed index in the neutrino equation (13) and using (12) and (9), one obtains

$$[(t\cdot\nabla)/\sqrt{2}]\lambda_A + \{D_{AM}\lambda^{M} + [(\pi/2\sqrt{2})]\lambda_A\} = 0 \tag{14}$$

This is the "3 + 1" form of the neutrino equation. Setting $t\cdot\nabla\lambda_A = 0$, one obtains the *neutrino "zero modes"* as the (normalizable) solutions of the equation

$$(L\lambda)_A := [D_{AB}\lambda^{B} + (\pi/2\sqrt{2})\lambda_A] = 0. \tag{15}$$

2. Spin-3/2 equation

Consider a spinor field $\psi_{AA'B'}$ satisfying

$$\psi_{AA'B'} = \psi_{A(A'B')}, \tag{16a}$$

$$\nabla_{B}{}^{A'}\psi_{AA'B'} = 0. \tag{16b}$$

Equations (16) describe a massless spin-3/2 field (Rarita–Schwinger) field of a particular helicity. We are interested in this particular field because, as we shall show, some properties of (16b) play a role in our analysis of neutrino "zero modes."

To obtain a "3 + 1" form of (16b), it is convenient ot express $\psi_{AA'B'}$ on Σ as a space spinor. Define

$$\psi_{ABC} = 2t_{B}{}^{B'}t_{C}{}^{C'}\psi_{AB'C'} \tag{17}$$

which can be decomposed into irreducible pieces as

$$\psi_{ABC} = \psi_{(ABC)} - \tfrac{2}{3}\epsilon_{A(B}\eta_{C)}, \tag{18}$$

where $\eta_C := \epsilon^{AB}\psi_{ABC}$. The pair $(\Psi_{(ABC)},\eta_A)$ can be viewed as the initial data on the slice Σ for the field equation (16b). The "3 + 1" equations for (16b) can be expressed as equations on this pair $(\psi_{(ABC)},\eta_A)$. One begins by writing

$$T_{ABC} = (\sqrt{2})t_{C}{}^{C'}\nabla_{B}{}^{B'}\psi_{AB'C'} = 0, \tag{19}$$

and decomposing T_{ABC} into irreducible pieces as

$$T_{ABC} = T_{(ABC)} + \tfrac{2}{3}\epsilon_{C(A}\mu_{B)} + \tfrac{2}{3}\lambda_{(A}\epsilon_{C)B},$$

where $\mu_B := \epsilon^{CA}T_{ABC}$, $\lambda_A := \epsilon^{AB}T_{ABC}$. Since $T_{ABC} = 0$, each irreducible piece must vanish separately. Thus one obtains a set of three equations. Two of these equations involve $t\cdot\nabla\psi_{(ABC)}$ and $t\cdot\nabla\eta_C$; these are not interesting for our purposes here so we avoid writing them explicitly. The third equation is

$$D^{AB}\psi_{(CAB)} - (1/\sqrt{2})\pi_{ABCD}\psi^{(DAB)} - \tfrac{2}{3}[D_{CM}\eta^{M} + (\pi/2\sqrt{2})\eta_C = 0]. \tag{20}$$

Equation (20) involves no time derivatives and is, in fact, a constraint equation on the pair $(\psi_{(ABC)},\eta_A)$. The important point to observe in (20) is that a data of the form $(0, \eta_A)$ satisfies the equation

$$(L\eta)_A = 0. \tag{20'}$$

In other words η_A is a neutrino "zero mode" [see Eq. (15)].

Another property of the spin-3/2 field which one needs is the following. Given a solution $\psi_{AA'B'}$ of (16b) define an inner product $\gamma(\ ,\ )$ by

$$\gamma(\psi,\psi) := (-\sqrt{2})\int_{\Sigma}\bar{\psi}^{A'AB}\psi_{AA'}{}^{B'}d\sigma_{B'B}. \tag{21}$$

One may check that the right-hand side of (21) is independent of the choice of hypersurface Σ, by virtue of the field equation (16b). Since (21) is obtained by restricting $\psi_{AA'B'}$ to Σ, one can express (21) in terms of the pair $(\psi_{(ABC)}, \eta_A)$. The result is

$$\gamma(\psi, \psi) = \int_{\Sigma}(\psi^{+(ABC)}\psi_{ABC} - \tfrac{1}{3}\eta^{+A}\eta_A)d\Sigma\ , \tag{22}$$

where $d\Sigma$ is the volume element on Σ defined by $d\sigma_{A'A} = t_{A'A}d\Sigma$.

For data on Σ of the form $(0, \eta_A)$, (22) shows that the γ inner product is negative definite on Σ. Since the definition of γ is independent of the choice of Σ, γ must be negative on another slice $\hat{\Sigma}$. Hence, on $\hat{\Sigma}$, the spin-3/2 field must induce data of the form $(0, \hat{\eta}_A)$,[9] which must satisfy the constraint equation similar to (20.1) on $\hat{\Sigma}$. In otherwords, if there is a neutrino "zero mode" relative to one Cauchy hypersurface Σ, then there must exist a neutrino "zero mode" relative to all Cauchy hypersurfaces.

III. MAIN RESULT

In this section we investigate the restrictions imposed on vacuum spacetimes by the requirement that they admit neutrino "zero modes."

An important feature of our analysis will be to integrate by parts certain expressions on Σ and neglect the resulting boundary terms. Thus our results will apply to only those manifolds Σ (Cauchy surfaces) which guarantee that the boundary terms vanish. Such manifolds will be said to have negligible boundaries. Closed or asymptotically flat Cauchy hypersurfaces have negligible boundaries. In the latter case, one can check by observing that, in Minkowski space, the leading term in the multipole expansion of solutions of $(L\lambda)_A = 0$ (that vanish at infinity) is of $O(1/r^2)$.[10] The relevant boundary terms that one encounters are then zero. In general, it is difficult to characterize Cauchy hypersurfaces (three-dimensional Riemannian manifolds) with negligible boundaries. It seems plausible, however, that

Conjecture 1: All complete,[11] three-dimensional, Riemannian, smooth manifolds have negligible boundaries.

With this caveat about the boundary terms in mind, we proceed to address the problem of this section.

Consider a vacuum spacetime (M, g_{ab}), i.e., one whose Ricci curvature vanishes: $R_{ab} = 0$. Let Σ be a complete, Cauchy hypersurface with extrinsic curvature π_{ab} and metric h_{ab} [of signature $(-\ -\ -)$]. Let H be the space of smooth spinor fields λ_A of compact support on Σ, with an inner product $\langle\lambda_A,\eta_A\rangle := \int_{\Sigma}\lambda^{A+}\eta_A d\Sigma$. Denote by $\bar{H}$ the Cauchy completion of $(H, \langle\ ,\ \rangle)$. Next consider the linear operator $L := D_{AB}\epsilon^{BC} + (\pi/2\sqrt{2})\ \delta_A{}^{C}$ defined on the dense domain $D(L) = H$.

Definition: λ_A is a neutrino "zero mode" if $\lambda_A \in \ker L$. (Note: L is an elliptic differential operator and it can be shown that $\ker L$ consists of C^{∞} spinor fields.[3])

From the initial value formulation of general relativity,[12] we know that the extrinsic curvature π_{ab} is constrained by the following equations in a vacuum spacetime:

$$-R-\pi^{ab}\pi_{ab}+\pi^2=0, \tag{23a}$$

$$D_a(\pi^{ab}-\pi h^{ab})=0, \tag{23b}$$

where R is the scalar curvature of the 3-manifold Σ. Assuming Conjecture 1 to hold, Eqs. (23a) and (23b) give us the following:

Lemma 1: In a vacuum spacetime admitting a complete Cauchy surface Σ,

$(L\lambda)_A=0 \Leftrightarrow D_{AB}\lambda_C-(1/\sqrt{2})\pi_{ABCD}\lambda^D=0$ on Σ.

Proof:

$$\begin{aligned}\langle (L\lambda)_A,(L\lambda)_A\rangle &= \langle D_{AB}\lambda^B+(\pi/2\sqrt{2})\lambda_A, D_{AC}\lambda^C+(\pi/2\sqrt{2})\lambda_A\rangle\\ &= \langle D_A{}^B\lambda_B, D_A{}^C\lambda_C\rangle+\tfrac{1}{8}\langle\pi\lambda_A,\pi\lambda_A\rangle-(1/2\sqrt{2})\langle\pi\lambda_A, D_A{}^C\lambda_C\rangle-(1/2\sqrt{2})\langle D_A{}^B\lambda_B,\pi\lambda_A\rangle\\ &= -\langle\lambda_B, D_B{}^A D_A{}^C\lambda_C\rangle+\tfrac{1}{8}\langle\lambda_A,\pi^2\lambda_A\rangle+(1/2\sqrt{2})\langle\lambda_B,(D_B{}^A\pi)\lambda_A\rangle\\ &= \langle\lambda_B,(\tfrac{1}{2}D^{MN}D_{MN}-\tfrac{1}{8}R)\lambda_B\rangle+\tfrac{1}{8}\langle\lambda_A,\pi^2\lambda_A\rangle+(1/2\sqrt{2})\langle\lambda_B,(D_B{}^A\pi)\lambda_A\rangle.\end{aligned} \tag{24}$$

In the second step, the first term is obtained from the fact that D_{AB} is skew-symmetric in the norm $\langle\,,\rangle$ [using (11)], and the last term is a result of simple integration by parts. The first term in (24) is obtained by simplifying $D_B{}^A D_{AC}$ using Eq. (10). Since, from (23a), $-R+\pi^2=\pi^{ab}\pi_{ab}$,

$$\langle (L\lambda)_A,(L\lambda)_A\rangle=\tfrac{1}{2}\langle\lambda_B, D^{MN}D_{MN}\lambda_B\rangle+\tfrac{1}{8}\langle\lambda_B,\pi^{ab}\pi_{ab}\lambda_B\rangle+(1/2\sqrt{2})\langle\lambda_B,(D_B{}^A\pi)\lambda_A\rangle \tag{25}$$

Using the second constraint equation (23b) to replace $D_B{}^A\pi$ by $D_{CD}\pi^{CD}{}_B{}^A$ in the last term in (25) and integrating by parts, we have

$$\langle (L\lambda)_A,(L\lambda)_A\rangle=\tfrac{1}{2}\langle D_{MN}\lambda_B, D_{MN}\lambda_B\rangle+\tfrac{1}{8}\langle\pi_{ab}\lambda_B,\pi_{ab}\lambda_B\rangle+(1/2\sqrt{2})[\langle D_{MN}\lambda_B,\pi_{MNB}{}^A\lambda_A\rangle+\langle\pi_{MNB}{}^A\lambda_A, D_{MN}\lambda_B\rangle$$

Since $\pi^{ABC}{}_D\pi_{ABCE}=\tfrac{1}{2}\epsilon_{ED}\pi^{ab}\pi_{ab}$, it is easy to check $\langle\pi_{ab}\lambda_B,\pi_{ab}\lambda_B\rangle=2\langle\pi_{MNB}{}^A\lambda_A,\pi_{MNB}{}^A\lambda_A\rangle$.

Then

$$\begin{aligned}&\langle (L\lambda)_A,(L\lambda)_A\rangle\\ &=\tfrac{1}{2}\langle D_{MN}\lambda_B, D_{MN}\lambda_B\rangle+\tfrac{1}{4}\langle\pi_{MNB}{}^A\lambda_A,\pi_{MNB}{}^A\lambda_A\rangle+(1/2\sqrt{2})[\langle D_{MN}\lambda_B,\pi_{MNB}{}^A\lambda_A\rangle+\langle\pi_{MNB}{}^A\lambda_A, D_{MN}\lambda_B\rangle]\\ &=\tfrac{1}{2}\langle[D_{MN}\lambda_B-(1/\sqrt{2})\pi_{MNBA}\lambda^A],[D_{MN}\lambda_B-(1/\sqrt{2})\pi_{MNBA}\lambda^A]\rangle.\end{aligned} \tag{26}$$

From (26), $(L\lambda)_A=0 \Rightarrow [D_{MN}\lambda_B-(1/\sqrt{2})\pi_{MNBA}\lambda^A]=0$.

The converse is trivial. This completes the proof of Lemma 1.

Lemma 2: A spinor field λ_A on Σ which vanishes at a point and satisfies $D_{AB}\lambda_C-(1/\sqrt{2})\pi_{ABCD}\lambda^D=0$ vanishes everywhere on Σ.

Proof[13]:

$$D_{AB}\lambda_C-(1/\sqrt{2})\pi_{ABCD}\lambda^D=0 \tag{27a}$$

$$D_{AB}\lambda^+_C+(1/\sqrt{2})\pi_{ABCD}\lambda^{D+}=0. \tag{27b}$$

Hence

$$\lambda^{C+}D_{AB}\lambda_C-(1/\sqrt{2})\pi_{ABCD}\lambda^{C+}\lambda^D=0, \tag{28a}$$

$$\lambda^C D_{AB}\lambda^+_C+(1/\sqrt{2})\pi_{ABCD}\lambda^{D+}\lambda^C=0. \tag{28b}$$

From (28a) and (28b) follows

$$D_{AB}(\lambda^{C+}\lambda_C)=(\sqrt{2})\pi_{ABCD}\lambda^{C+}\lambda^D. \tag{29}$$

Let $\phi=\lambda^{C+}\lambda_C$ and define a spatial vector $v^a\equiv\lambda^{(A}\lambda^{B)+}$. Then (29), in tensor form, is

$$D_a\phi=(\sqrt{2})\pi_{ab}v^b. \tag{30}$$

Suppose λ_A (hence ϕ) vanishes at some point $p\in\Sigma$. Choose a C^1 curve Γ passing through p and let n^a be the tangent to Γ. Then from (30)

$$n^aD_a\phi=(\sqrt{2})n^a\pi_{ab}v^b$$

so

$$|n^aD_a\phi|=(\sqrt{2})|n^a\pi_{ab}v^b|\leqslant\sqrt{2}\|n\|\,\|\hat{v}\|\leqslant C|\phi| \tag{31}$$

for some real nonnegative constant C, where $\hat{v}_a=\pi_{ab}v^b$ and $\|\hat{v}\|=(-\hat{v}^a\hat{v}_a)^{1/2}$. Note that $-v^av_a=\phi^2$. Now, for a real function $f(s)$ of a real variable s, (*i*) $|df/ds|\leqslant c|f|$, $c>0$, and (ii) $f>0$ at $s=s_0$ implies $f(s)$ is positive for all s. [If $f(s_0)=f_0>0$, then, on the interval containing s_0 over which $f(s)$ is positive, one has $f_0-cs\leqslant\ln f\leqslant f_0+cs$, so $f(s)$ can vanish only as $s\to\infty$.] We apply this result to (31). Since $\phi=\lambda^{C+}\lambda_C$ is positive, ϕ is positive on Γ except (at least) at p where it is zero. But this contradicts (31). Hence ϕ must vanish everywhere on Γ. It is now easy to see that ϕ must vanish everywhere on Σ since Γ is arbitrary. Hence if ϕ vanishes at a point p on Σ, ϕ (hence λ_A) vanishes everywhere on Σ. This completes the proof of Lemma 2.

Theorem: Vacuum globally hyperbolic spacetimes with a complete Cauchy hypersurface which admit neutrino "zero modes," i.e., $\ker L\neq0$, are algebraically special of Petrov type III (hence of type N or flat.)

Proof: From Lemma 1, a neutrino "zero mode" λ_A must satisfy

$$D_{AB}\lambda_C-(1/\sqrt{2})\pi_{ABCD}\lambda^D=0. \tag{27a}$$

Regard the spinor index pair AB as a spatial tensor index a, and let $T_{aA}:=D_a\lambda_A-(1/\sqrt{2})\pi_{aAB}\lambda^B$. The integrability condition for Eq. (27.1) is obtained from $D_{[b}T_{a]A}=0$ or equivalently

$$\epsilon_a{}^{mn}D_mT_{nA}=0, \tag{32}$$

where ϵ_{abc} is the three-dimensional alternating tensor defined by the metric on Σ. Using Eq. (10) and the spinor form of ϵ^{mea}, one gets

$$+ [\epsilon_{MN}{}^{pq} D_p \pi_{q\,CD}]\lambda^D$$
$$+ i[R_{MNCD} + \pi_{aMN}\pi^a{}_{CD} - \pi\pi_{MNCD}]\lambda^D = 0. \qquad (33)$$

(*Note*: We are using mixed tensor and spinor indices.) To understand the integrability condition (33), one needs the following known results. Recall that the electric and magnetic parts of the Weyl tensor C_{abcd} are definded respectively by

$$E_{ab} = t^m t^n C_{ambn}$$
$$B_{ab} = t^m t^n {}^*C_{ambn},\ {}^*C_{ambn} = \tfrac{1}{2}\epsilon_{am}{}^{pq} C_{pq\,bn}$$

and can be expressed as [14] (in vacuum spacetimes)

$$E_{ab} = -R_{ab} - \pi_a{}^m \pi_{mb} + \pi\pi_{ab},$$
$$B_{ab} = \epsilon_{(a}{}^{pq} D_p \pi_{q\,b)}. \qquad (34)$$

Further, since

$$C_{abcd} \equiv \Psi_{ABCD}\epsilon_{A'B'}\epsilon_{C'D'} + \bar{\Psi}_{A'B'C'D'}\epsilon_{AB}\epsilon_{CD}$$

and

$$C_{abcd} + i\,{}^*C_{abcd} \equiv \epsilon_{A'B'}\epsilon_{C'D'}\Psi_{ABCD},$$

one has

$$E_{MNCD} + i\,B_{MNCD} = \Psi_{MNCD}, \qquad (35)$$

where

$$E_{MNCD} = 2t_M{}^{N'} t_C{}^{D'} E_{N'ND'D},$$
$$B_{MNCD} = 2t_M{}^{N'} t_C{}^{D'} B_{N'ND'D}$$

are just the SU(2) spinor form of E_{ab} and B_{ab}, respectively. Using (34) and (35), one can express (33) as

$$\Psi_{MNCD}\lambda^D + \tfrac{1}{2}i[\epsilon_{MN}{}^{pq} D_p \pi_{q\,CD} - \epsilon_{CD}{}^{pq} D_p \pi_{q\,MN}]\lambda^D = 0. \qquad (36)$$

Further simplication is achieved by defining

$$\Pi_{(AB)} := \epsilon_{N(A}{}^{pq} D_p \pi_{q\,B)}{}^N.$$

Then (36) takes the form

$$\Psi_{MNCD}\lambda^D + \tfrac{1}{2}i(\epsilon_{MC}\Pi_{(ND)} + \epsilon_{ND}\Pi_{(MC)})\lambda^D = 0. \qquad (37)$$

Contract indices MC in (37) to obtain

$$\Pi_{(ND)}\lambda^D = 0 \Rightarrow \Pi_{(ND)} = \alpha\lambda_N\lambda_D,\ \alpha\in\mathbb{C}. \qquad (38)$$

Hence (37) reduces to

$$\Psi_{MNCD}\lambda^D + \tfrac{1}{2}i(\epsilon_{ND}\alpha\lambda_M\lambda_C)\lambda^D = 0$$

or

$$\Psi_{MNCD}\lambda^D = \beta\lambda_M\lambda_N\lambda_C,\ \beta\in\mathbb{C}. \qquad (39)$$

From Lemma 2, if $\lambda_A \neq 0$, then $\lambda_A \neq 0$ everywhere on Σ. Hence, from (39), Ψ_{MNCD} is of Petrov type III (see, for example, Ref. 5) on the entire surface Σ. To complete the proof, we show that Ψ_{MNCD} must be of type III on *every* slice Σ. Recall, from Sec. IIC, that if there is a neutrino "zero mode" on one slice Σ, then there is a "zero mode" relative to every slice. Thus we can repeat the above analysis on every slice and hence establish that Ψ_{MNCD} must be type III everywhere on the spacetime. The proof is now complete.

In the above proof, we have regarded type III to include integration by parts. Although this step anticipates the validity of Conjecture 1, the result holds at least for a large class of physically interesting spacetimes, namely closed or asymptotically flat spaces. The second fact was that the spacetime satisfied Einstein's equations so that the constraint equations (23a) and (23b) could be used. However, the requirement that the spacetime be vacuum is *not* crucial at this stage as the following argument shows. Consider, as before, a globally hyperbolic spacetime (M, g_{ab}) satisfying Einstein's equations with matter. Then, the constraint equations of general relativity on a spacelike (Cauchy) hypersurface are

$$-R - \pi^{ab}\pi_{ab} + \pi^2 = 2\mu, \qquad (23a')$$
$$D_a(\pi^{ab} - \pi h^{ab}) = J^b, \qquad (23b')$$

where μ and J^a are the energy density and momentum density, respectively, of the matter, as measured by an observer whose 4-velocity is normal to Σ. One further requires that

$$\mu \geqslant |J^a J_a|^{1/2}, \qquad (40)$$

which says that the apparent energy–momentum of the matter is timelike. Thus (40) is a physical requirement. If one uses (23a') and (23b') instead of in (23a) and (23b) in Eq. (24), one obtains the following equation instead of (26):

$$\langle (L\lambda)_A, (L\lambda)_A\rangle = \tfrac{1}{2}\langle [D_{MN}\lambda_B - (1/\sqrt{2})\pi_{MNBA}\lambda^A],$$
$$[D_{MN}\lambda_B - (1/\sqrt{2})\pi_{MNBA}\lambda^A]\rangle$$
$$+ \tfrac{1}{4}\langle\lambda_A, (\mu\delta_A{}^B - (\sqrt{2})J_A{}^B)\lambda_B\rangle, \qquad (41)$$

where J^{AB} is the spinor form of J^a. Now the condition (40) ensures that the last term in (41) is nonnegative. Hence $(L\lambda)_A = 0$ implies each term on the right in (41) must vanish separately, whence Lemma 1 is seen to be valid even for spacetimes with matter satisfying (40).

The integrability of (27a) is the final step in the proof. Here the requirement that the spacetime be vacuum enters at two stages. First, in the interpretation of Eq. (33) leading to Eq. (39). In the presence of matter, there are additional terms in the expression (34) for E_{ab}. Consequently, not much insight about the underlying spacetime can be obtained. Second, to establish that the Weyl tensor Ψ_{ABCD} must be of type III everywhere, one uses some property of the spin-3/2 field discussed in Sec. IIC. Now, the spin-3/2 equation can be shown to be free from inconsistencies (Buchdahl conditions)[3] if the underlying spacetime is Einstein, i.e., $R_{ab} = \Lambda g_{ab}$. Thus, on vacuum spacetimes, one can legitimately use the required properties on th spin-3/2 field. Incidentally, for $\Lambda > 0$, one can show, by a method similar to the one given in this paper, that there are *no* neutrino "zero modes." For $\Lambda < 0$, nothing can be said about the existence of "zero modes" by our methods.

We end this note by suggesting a stronger version of our theorem: Vacuum, globally hyperbolic spacetimes admitting neutrino "zero modes" are *flat*. The reason is that type III and type N solutions represent gravitational waves in general relativity. Now, the simplest type N solution, the plane wave,[5] does not admit any Cauchy surface.[15] One might find a piece of the plane wave spacetime with a Cauchy surface, but it will not be complete. One expects this situation to persist for the general type III or N spacetimes. Thus the only candidates must be flat.

type N or type 0 (flat) as special cases, corresponding respectively to $\beta = 0$ or $\Psi_{MNCD} = 0$ in Eq. (39).

IV. DISCUSSION

We summarize the important steps involved in arguments given above. Neutrino "zero modes" relative to a Cauchy surface Σ in the spacetime are normalizable solutions of the (elliptic) equation $(L\lambda)_A = 0$ on Σ. Lemma 1 showed that the "zero modes" must satisfy Eq. (27a) $D_{AB}\lambda_C - (1/\sqrt{2})\pi_{ABCD}\lambda^D = 0$ as well. Two facts were crucial in establishing this result. The first was the validity of

ACKNOWLEDGMENTS

I thank Bob Geroch and Rafael Sorkin for helpful discussions.

[1] R. Jackiw and C. Rebbi, Phys. Rev. D **13**, 3398 (1976).

[2] R. Jackiw, Rev. Mod. Phys. **49**, 681 (1977).

[3] A. Sen, "A Quantum Theory of Spin 3/2 Field in Einstein Spaces,"to appear in Int. J. Theo. Phys. (1980).

[4] The method is, in fact, also applicable with a cosmological constant. However, in a spacetime satisfying $R_{ab} = \Lambda g_{ab}$ with $\Lambda > 0$ there are no neutrino "zero modes" while for $\Lambda < 0$ it is possible to have "zero modes."

[5] F. Pirani, *Brandeis Lectures in General Relativity* (Prentice-Hall, Englewood Cliffs, N.J., 1964).

[6] Our conventions are the following: g_{ab} has signature $(+ - - -)$ and the curvature tensors are defined by $[\nabla_a\nabla_b - \nabla_b\nabla_a]v_c = R_{abc}{}^d v_d, R_{ab} = R_{amb}{}^m, R = R_{ab}g^{ab}$.

[7] To see this, consider a future-directed null vector $l^a \equiv \lambda^{A'}\lambda^A$. Then $\lambda^{A'}t_{A'A}\lambda^A = l^a t_a \geqslant 0, = 0$ iff $l^a = 0$.

[8] See, for example, Ref. 3 or J. Friedman and R. D. Sorkin, Comm. Math. Phys. **2**, 161 (1980).

[9] It can be shown that on any Cauchy surface Σ, the space w of $\mathscr{L}^2$ data (on Σ) of the spin 3/2 field [with nonzero $\gamma(\,,\,)$ norm] can be written as a direct sum: $w = w_1 \oplus w_2$, where w_1 and w_2, respectively, contain data of the form $(\psi_{(ABC')}, 0)$ and $(0, \eta_A)$.

[10] In Minkowski space, if F_+ and F are the components of the neutrino field in a basis, then the solutions of $D_{AB}\lambda^B = 0$ (choosing a $\pi_{ab} = 0$ slice) are given by $F_{+} = (1/r)R_{+}(r)S_{+}(\theta,\phi)$, where $S(\theta,\phi)$ is an angular function and $R = A_l r^{l+1/2} + B r^{(l+1/2)}$, $R_{+} = 1/r\,R$, where $l > 1/2$ and A_l, B_l are constants.

[11] A Riemannian manifold is complete if it is complete as a metric space, i.e., every Cauchy sequence must converge.

[12] See, For example, R. Geroch J. Math. Phys. **7**, 956 (1972).

[13] I thank Bob Geroch for suggesting the proof.

[14] See, for example, R. Geroch, *Asymptotic Structure of Spacetime*, edited by F. P. Esposito and L. Witten (Plenum, New York, 1977).

[15] R. Penrose, Rev. Mod. Phys. **37**, 215 (1965).

A new integral equation for summing Feynman graph series (φ^3 ladder graph case)

C. Gilain and D. Levy
CEN-Saclay, Division de la Physique, Boite Postale No. 2, 91190 Gif-sur-Yvette, France

(Received 13 June 1980; accepted for publication 30 July 1980)

The Schwinger parameter formalism is used to derive a new integral equation verified by the sum of the "open amplitudes" of the ladder graph series with a φ^3 interaction. We prove the existence of a solution to this equation and of the corresponding Green's function. This solution, for any finite value of the coupling constant, is a finite sum of solutions to Fredholm integral equations plus the sum of a convergent series. Its set of singularities is a set of poles from which the spectrum of Regge poles is obtained.

PACS numbers: 11.20.Dj, 11.10.Ef

INTRODUCTION

The question of studying particle interactions in terms of Lagrangian field theory is one of the great up-to-date problems. Lagrangian field theory gives us the amplitude as a perturbative series in terms of the coupling constant g, each term of this series being defined as a sum of Feynman integrals

$$A(P,g) = \sum_{n=0}^{\infty} g^n a_n(P)\,, \tag{1}$$

where P denotes collectively the external momenta [or the Lorentz invariants which can be built from them, e.g., in the 2 particles $\to$2 particles case the Mandelstam variables s, t, u (see Fig. 1) and the square of the external leg momenta].

From the four point amplitude, we are able to build the physical observables. In some situations, the amplitude $A(P, g)$ can be approximated by the low order terms of the series (1). This is the case if the particles interact weakly ($g \ll 1$) and the series is convergent or asymptotic (in fact, in general this series is known to diverge, the asymptotic case being the best we can hope for). Sometimes this is also the case when we are dealing with a strong interaction (via the property of asymptotic freedom). However, many physical problems force us to face the whole series: The correct treatment of most of the strong interaction phenomena (Regge behavior of cross section, Bjorken limit, etc.) requires the summation of the entire series. We are then faced with an infinite summation problem.

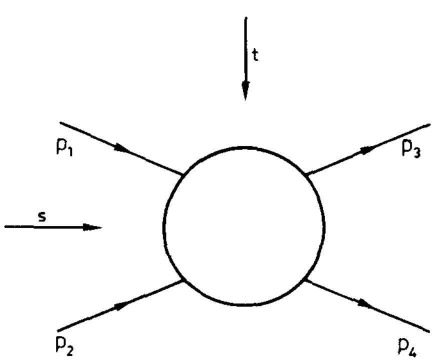


FIG. 1. The two particles $\to$ two particles reaction. The Mandelstam variables are defined as $s = (p_1 + p_2)^2$, $t = (p_1 - p_3)^2$, $u = (p_1 - p_4)^2$, the p_i's being the quadrimomenta attached to the external legs.

The question we want to contribute to in this work is the property of the four point strong interaction amplitude. In fact we will work not with $A(s,t, g)$ but with its s-Mellin transform $\bar{A}(x,t, g)$

$$\bar{A}(x,t,g) = \int_0^{\infty} s^{-x-1} A(s,t,g)\, ds\,. \tag{2}$$

The reasons for working with $\bar{A}(x,t, g)$ are of two different types: (1) First, there are technical reasons which are linked to the Wick rotation problem and to the Landau singularities. These points and also the mathematical aspects (singularities of the integrand, convergence of the integral) of the exact definition of $\bar{A}(x,t, g)$ will be discussed in Sec. II. (2) On the other hand it is well known that the Mellin space is very well adapted for the discussion of the amplitude at high energy, where the Regge model is relevant. The Regge model, which has a strong theoretical foundation and which reproduces well the principal features of the experimental data, gives a phenomenological form of the amplitude. In the high energy domain, the hadronic amplitude behaves as

$$A(s,t,g) = \sum_i \gamma_i(t) s^{\alpha_i(t)}\,, \tag{3}$$

where the index i runs over a finite number of values and where the Regge trajectories $\alpha_i(t)$ are the positions of the moving poles of the amplitude in the complex space of the angular momentum of the crossed channel. A Regge trajectory $\alpha_i(t)$ is associated, in general, with a physical particle of mass M_i and spin J_i and satisfies the relation

$$\alpha_i(M_i^2) = J_i\,. \tag{4}$$

The set of singularities of $\bar{A}(x,t, g)$ in the Mellin space embodies the same physical information as the set of singularities in the angular momentum space.

The positions of the singularities of $\bar{A}(x,t, g)$ in the variable x are functions of the other two variables t and g: $\bar{A}(x,t, g)$ is singular when

$$x = x_i(t, g) \quad (i = 1,2,\cdots)\,. \tag{5}$$

 0022-2488/81/081787-23$01.00 © 1981 American Institute of Physics

In fact, it is strictly equivalent to look for the singularities of $\bar{A}(x,t,g)$ in the form

$$g = g_i(x,t), \qquad (6)$$

where $g_i(x,t)$ is the inverse of the function $x_i(t,g)$. The strength of the method presented in this paper is to give us entirely this set of functions $g_i(x,t)$.

Series (1) has been studied for a long time, essentially in momentum (or energy s) space. Two general methods can be applied for that purpose: The first one relies on the properties of the $a_n(P)$ and consists of attempting to actually perform the summation. The alternative method is nonperturbative; it exploits global properties of the sum $A(P,q)$ such as the Bethe–Salpeter structure of the four point Green's function, to provide us with an integral equation.

(1) The perturbative method, applied for the Regge limit problem, requires as a first step knowledge of the dominant part $a_n^{as}(P)$ for large incoming energy s and fixed momentum transfer t, for all n. This first step is very difficult. This asymptotic problem consists in giving explicitly the $r = \Omega$ term in

$$a_n(P) = a_n(s,t) = \sum_{r=-\infty}^{\Omega} s^r \left(\sum_{q=1}^{Q(r)} (\log s)^{-q} \tilde{a}_{r,q}(t) \right). \qquad (7)$$

This problem, in the scalar Lagrangian case, together with the subsequent summation, was first solved for an infinite subseries of (1), the ladder graph series. In that case the convergence radius is nonzero; approximating the asymptotic terms by retaining only the leading logarithm part, or even all the logarithms of the leading power terms of (7), the Regge behavior of the amplitude was proved a long time ago.[1] The first attempts to get rid of this approximation and to go further than the ladder case was done by Zav'Yalov, who derived prescriptions for (7) for general graphs.[2] For scalar φ^3 and φ^4 Lagrangians and planar graphs, the complete summation was finally achieved a few years ago, exhibiting Regge behavior.[3]

This approach is confronted by several intrinsic limitations.

(i) First, it cannot be checked that we actually obtain the asymptotic part of the amplitude through the sum of the asymptotic part of each graph: Behavior of the infinite sum of subdominant terms cannot be controlled directly.

(ii) Of course only information inside the convergence radius (if any) can be obtained: Nothing can be said after the first singularity.

(iii) Finally, the more complete calculation[3] proves the reggeization of the four point amplitude but gives trajectories and residues in terms of a perturbative series.

(2) The Bethe–Salpeter structure of the four point Green's function $\Gamma(P,g)$, provides us with an integral equation

$$\Gamma(P,g) = \Gamma_1(P,g) + g \int K(P,P')\, \Gamma(P',g)\, dP', \qquad (8)$$

and the nonperturbative approach consists of solving (8). To be more precise we should speak of a family of integral equations. The first term and the kernel are indeed dependent on the Lagrangian, and of course on the eventual approximations. Under conditions of sufficient regularity of the first term and of the kernel, the analytic structure of the solution of an integral equation is controlled. The best example is the Fredholm case: Provided the kernel and the inhomogeneous term are of finite $\mathscr{L}_2$ norm, the solution $\Gamma(P,g)$ is a ratio of two holomorphic functions of g. More precisely, under the above-mentioned conditions, if $\Gamma(P,g)$ satisfies Eq. (8), then

$$\Gamma(P,g) = N(P,g)/D(g), \qquad (9)$$

where $D(g)$ depends only on the kernel $K(P,P')$. General results of this kind make the integral equation method for our problem very powerful. Equation (8) has been extensively studied in the ladder graph approximation. Complete solutions can be given only for rather special mass cases (Wick–Cutkovsky model).[4,5] (For a review of the work on the Bethe–Salpeter equation, see Nakanishi.[6])

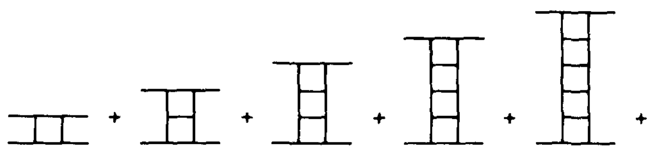


FIG. 2. The ladder graph series.

In this work also, and as first step in the course of our study, we will restrict ourself to the ladder graph case. In that case, only the terms even in n contribute in (1), and each term corresponds to a single graph. It is convenient to use the following notation for the amplitude $\bar{M}$ corresponding to the infinite sum of the ladder graphs (see Fig. 2)

$$\bar{M}(x,t) = \sum_{n=0}^{\infty} \bar{M}_n(x,t,\lambda),$$

where $\lambda = g^2$, and n denotes now the number of loops of the ladder graph (of course, the λ dependence of $\bar{M}_n$ is through a multiplicative factor λ^{n+1}).

Now, in order to exhibit our integral equation, we make extensive use of tools employed for the perturbative approach, in particular the Schwinger integral representation of Feynman amplitudes [we recall that we are working in the Mellin space, see (2)]

$$\bar{M}_n(x,t,\lambda) = \int_0^\infty \left(\prod_i d\alpha_i \right) I_n(\{\alpha_i\}, x,t,\lambda), \qquad (10)$$

where α_i are scalar variables attached to each internal line of the graph. Recurrence relations are exhibited for the integrand $I_n(\{\alpha_i\},x,t,\lambda)$ of the Feynman amplitudes [see Eq. (I.6)]. These recurrence relations are the fundamental ingredient upon which this work relies.

Let us call external the quantities attached to external legs of the graphs (e.g., incoming momenta) and internal ones those attached to internal lines. The Feynman amplitude is obtained through integration upon all internal variables. Let us define the "open" Feynman amplitude of a given graph as being the function obained when some of these integrations are left over. The Feynman amplitude is then unambigously obtained by performing these remaining integrations. In our case for the graph of order n in λ,

$$\bar{M}_n(x,t,\lambda) = \int_0^\infty d\alpha_J \, \bar{F}_n(\alpha_J, x,t,\lambda) , \quad (11)$$

where $\bar{F}_n$ is the open Feynman amplitude and J denotes a subset of the set $\{i\}$ of internal lines. Of course, $\bar{F}_n$ is defined only for n sufficiently large,

$$n \geqslant n_J ,$$

where n_J depends on the choice of J. In our problem, n_J is equal to 1 [see Eq. (I.4)].

Let us finally define the "open" four point amplitude $\bar{F}(\alpha_J, x,t,\lambda)$ as the following infinite sum

$$\bar{F}(\alpha_J, x,t,\lambda) = \sum_{n=1}^{\infty} \bar{F}_n(\alpha_J, x,t,\lambda) . \quad (12)$$

The recurrence relations satisfied by the integrand $I_n(\{\alpha_i\}, x,t,\lambda)$ induce for the open amplitude an integral equation

$$\bar{F}(\alpha_J, x,t,\lambda) = \bar{F}_1(\alpha_J, x,t,\lambda) + \lambda \int_0^\infty K(\alpha_J, \alpha'_J, x,t) \times \bar{F}(\alpha'_J,x,t,\lambda) \, d\alpha'_J , \quad (13)$$

where, as we are in the ladder approximation, the inhomogeneous term $\bar{F}_1$ and the kernel K are explicit functions.

This integral equation is not equivalent to the Bethe–Salpeter one [see (8)], which would imply for an "open" Green's function defined analogously, an equation of the type

$$\widetilde{\Gamma}(P,\alpha_J) = \widetilde{\Gamma}_1(P,\alpha_J) + \lambda^2 \int K(P,P',\alpha_J) \, \widetilde{\Gamma}(P',\alpha_J) \, dP' , \quad (14)$$

which integral equation does not involve the same category of variables as Eq. (13).

After deriving integral equation (13), we present in this paper a method for solving it. We find that the kernel K is too singular to fall within the scope of the Fredholm theory. However, we prove that, as would be the case if the integral kernel was sufficiently regular, the solution $\bar{F}(\alpha_J, x,t,\lambda)$ has only poles and can be written in the form

$$\bar{F}(\alpha_J, x,t,\lambda) = N(\alpha_J, x,t,\lambda)/D(x,t,\lambda) . \quad (15)$$

Moreover, the function $N(\alpha_J, x,t,\lambda)$ is such that the integration upon α_J is possible, and such that the equation

$$D(x,t,\lambda) = 0 \quad (16)$$

gives the set of poles of the amplitude $\bar{M}(x,t,\lambda)$.

This paper is the first one of a series of publications and is devoted to the exposition of the method. After the statement of the basic formalism, we obtain in Sec. I a recurrence relation verified by the topological polynomials in the ladder case [Eq. (I.3)], and we end this section by establishing the integral equation verified by the open amplitude (I.11). Some comments and fundamental properties of this integral equation are grouped in Sec. II, where particular cases are also shown. Section III is devoted to explaining and proving a general theorem on the solution of a wide class of integral equations. Then the open amplitude is shown in Sec. IV to lie within the scope of this theorem. Finally, the physical amplitude is built, and its singularity structure given (Sec. V). Let us stress that the strength of the method presented here, as compared to the Bethe–Salpeter one, is that the same analysis provides us with both the spectrum of Regge singularities and the expression of the amplitude itself. Indeed, from the method initiated by Lee and Sawyer, the Regge singularity analysis is obtained from an analytic continuation of the partial waves, the problem of the summation of the partial wave expansion, which gives the amplitude, being left over. If one is interested in the amplitude, other methods must be used (such as the perturbation-theoretical integral representation[6] for instance), and so the complete study of the properties of the amplitude through the Bethe–Salpeter equation is difficult and lengthy.

A following paper will then be devoted to quantitative results on the dominant Regge trajectory, and next the set of daughter trajectories will be studied. The last step will be to extend our theorem for general graphs of φ^3.

Let us finally end this introduction by stressing the contribution of the above-described method. One of its strengths is that it allows quantitative work. In particular for the dominant trajectory, we are able to check various existing approximations. The most natural one, the trace aproximation, whose use in the Bethe–Salpeter case is questionable, is here very simple and turns out to be an accurate approximation.

Moreover, the solution of Eq. (16)

$$D(x,t,\lambda) = 0 ,$$

provides us with the entire set of subdominant trajectories (daughters), and in that sense this method allows us to go further than other types of approaches.

Another interesting point is that it gives us an external check of the perturbative approach. We recall that the most complete calculation done in that framework provides us with a dominant Regge trajectory given in terms of a series. If we work with that series in the same way that we treat the perturbation series, we find an integral equation which is in some sense a first order approximation of a particular development of the two sides of Eq. (13). Comparing the singularity structure of the two solutions, we observe that this "first order" provides us with part of the complete set of trajectories (and among them, as expected, the dominant one). This proves first that the asymptotic approximation is correct, and also that one obtains correctly part of the set of the subdominant trajectories.

I. DERIVATION OF THE INTEGRAL EQUATION

We recall that in this work, among the whole series of graphs generated from the interacting Lagrangian $g\varphi^3$, we keep only the ladder graphs. The Schwinger-integral representation for the Feynman amplitude M_n of a ladder graph with $(n+1)$ rungs (see Fig. 3) is

$$M_n(s,t) = (\lambda i)^{n+1} \times \int_0^\infty \prod_{i=1}^{n} \left(d\alpha_i \, d\alpha'_i \, e^{-i(\alpha_i + \alpha'_i)m^2} \right) \times \prod_{i=0}^{n} \left(d\beta_i \, e^{-i\beta_i m^2} \right) \frac{1}{P_n^2} \times \exp\left(i \frac{sA_n^s + tA_n^t + \Sigma_{i=1}^4 p_i^2 A_n^i}{P_n} \right), \quad (I.1)$$

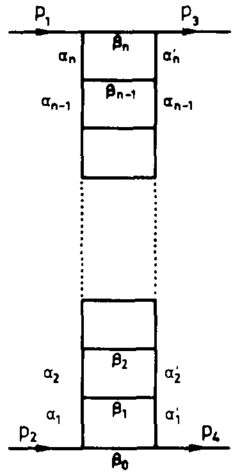


FIG. 3. The ladder graph with $(n+1)$ rungs (n loops).

where the scalar variables α_i, α'_i, and β_i are attached to each internal line as shown in Fig. 3. In the expression (I.1), the invariants s, t, and p_i^2 must be thought of as having a small imaginary part when needed for the convergence of the integral, i.e., above the Landau singularities.

P_n, A_n^s, A_n^t, A_n^i are polynomials, homogeneous in all variables α_i, α'_i, β_i, of degree n for P_n and $(n+1)$ for the others. Their complete definitions are given in Appendix A.

We now arrive at an important point: The existence of the integral equation is the consequence of the recurrence relations verified by the polynomials.

Let f be a function of $(3n+1)$ scalar variables. Then we will denote

$$f = f(\beta_n, \alpha_n, \alpha'_n, \beta_{n-1}, \alpha_{n-1}, \dots, \beta_0),$$

and

$$f^* = f(\beta_n^*, \alpha_n^*, \alpha_n'^*, \beta_{n-1}, \alpha_{n-1}, \dots, \beta_0),$$

where

$$\begin{aligned} \beta_n^* &= (\beta_n / c_{n+1}) \beta_{n+1}, \\ \alpha_n^* &= \alpha_n + (\beta_n / c_{n+1}) \alpha_{n+1}, \\ \alpha_n'^* &= \alpha'_n + (\beta_n / c_{n+1}) \alpha'_{n+1}, \end{aligned} \tag{I.2}$$

with

$$c_{n+1} = \beta_{n+1} + \alpha_{n+1} + \alpha'_{n+1} + \beta_n .$$

Then we can write the recurrence relations

$$P_{n+1} = c_{n+1} P_n^*, \tag{I.3a}$$

$$\frac{A_{n+1}^s}{P_{n+1}} = \left(\frac{A_n^s}{P_n}\right)^*, \tag{I.3b}$$

$$\frac{A_{n+1}^t}{P_{n+1}} = \frac{\alpha_{n+1} \alpha'_{n+1}}{c_{n+1}} + \left(\frac{A_n^t}{P_n}\right)^*, \tag{I.3c}$$

$$\frac{A_{n+1}^i}{P_{n+1}} = \left(\frac{A_n^i}{P_n}\right)^* \quad \text{if } i = 2,4, \tag{I.3d}$$

$$\frac{A_{n+1}^1}{P_{n+1}} = \frac{\alpha_{n+1} \beta_{n+1}}{c_{n+1}} + \left(\frac{A_n^1}{P_n}\right)^*, \tag{I.3e}$$

$$\frac{A_{n+1}^3}{P_{n+1}} = \frac{\alpha'_{n+1} \beta_{n+1}}{c_{n+1}} + \left(\frac{A_n^3}{P_n}\right)^*. \tag{I.3f}$$

These relations, which reflect the topological properties of the graphs, are proved in Appendix A.

The recurrence relations (I.3) do not allow us to write a recurrence involving the amplitude $M_n(s,t)$ itself. Now let us define the "open" Feynman amplitude [see Introduction, equality (11)] by choosing α_j as being the scalar variables attached to the lines constituting the upper loop [i.e., $\{\alpha_n, \alpha'_n, \beta_n\}$ for $M_n(s,t)$]. We will denote them collectively as the closing variables, but we maintain the subscripts n wherever an ambiguity is possible. We then have

$$\begin{aligned} &F_n(\alpha_n, \alpha'_n, \beta_n; s,t) \\ &\quad = (\lambda i)^{n+1} \int_0^\infty \prod_{i=1}^{n-1} \left[d\alpha_i \, d\alpha'_i \, e^{-i(\alpha_i + \alpha'_i)m^2}\right] \\ &\qquad \times \prod_{i=0}^{n-1} (d\beta_i \, e^{-i\beta_i m^2}) \frac{1}{P_n^2} \\ &\qquad \times \exp\left(i \frac{sA_n^s + tA_n^t + \Sigma_{i=1}^4 p_i^2 A_n^i}{P_n}\right). \end{aligned} \tag{I.4}$$

Then the amplitude will be obtained by performing the last three integrations left over

$$\begin{aligned} M_n(s,t) = \int_0^\infty & d\alpha_n \, d\alpha'_n \, d\beta_n \, e^{-i(\alpha_n + \alpha'_n + \beta_n)m^2} \\ & \times F_n(\alpha_n, \alpha'_n, \beta_n; s,t). \end{aligned} \tag{I.5}$$

Now the structure of the recurrence relation (I.3) clearly leads to distinguishing between two sets of internal variables, the closing variables on one hand, the remaining internal variables on the other: The * operation leaves invariant this lost subset, while on the closing variables it induces the transformation (I.2). This fact provides us with a recurrence for the functions F_n

$$\begin{aligned} &F_{n+1}(\alpha_{n+1}, \alpha'_{n+1}, \beta_{n+1}; s,t) \\ &= (\lambda i) \int_0^\infty d\alpha_n \, d\alpha'_n \, d\beta_n \, F_n(\alpha_n^*, \alpha_n'^*, \beta_n^*; s,t) \left(\frac{1}{c_{n+1}}\right)^2 \\ &\quad \times \exp[-im^2(\alpha_n + \alpha'_n + \beta_n)] \\ &\times \exp\left(i \frac{t\alpha_{n+1}\alpha'_{n+1} + p_1^2 \alpha_{n+1} \beta_{n+1} + p_3^2 \alpha'_{n+1} \beta_{n+1}}{c_{n+1}}\right). \end{aligned} \tag{I.6}$$

Equation (I.6) is a straightforward consequence of the relations (I.3). It is convenient to rewrite it under a form that will make obvious the existence of an integral equation verified by the sum of the open ladder [see Introduction, Eq. (12)]. For that let us re-express the integrals in (I.6) in terms of the *-transformed variables. From now on, the only internal scalar variables that will appear will be (i) the closing variables of the ladder with $(n+1)$ rungs, which we will refer to hereafter as $\{\alpha, \alpha', \beta\}$; (ii) the *-transform of the closing variables of the ladder with n rungs, whose notation will be $\{\alpha^*, \alpha'^*, \beta^*\}$.

The domain of variation of the variables α^*, α'^*, β^* is

$$\left\{ \begin{array}{c} 0 \leqslant \alpha^* < \infty, \\ 0 \leqslant \alpha'^* < \infty, \\ 0 < \beta^* < \beta \inf\left(1, \dfrac{\alpha^*}{\alpha}, \dfrac{\alpha'^*}{\alpha'}\right) = \beta U . \end{array} \right\} \tag{I.7}$$

Taking into account the Jacobian of this transformation, we rewrite (I.7) under the form

$$F_{n+1}(\alpha,\alpha',\beta;s,t) = \lambda \int_0^\infty d\alpha^* \, d\alpha'^* \, d\beta^* \, J(\alpha,\alpha',\beta;\alpha^*,\alpha'^*,\beta^*) \times F_n(\alpha^*,\alpha'^*,\beta^*;s,t), \tag{I.8}$$

where

$$J(\alpha,\alpha',\beta;\alpha^*,\alpha'^*,\beta^*) = i\frac{1}{\beta}\frac{\theta(U-\beta^*/\beta)}{\alpha+\alpha'+\beta} \times \exp\left[\mathrm{L} - im^2(\alpha^*+\alpha'^*+\beta^*) - im^2(\beta+\alpha+\alpha') \times \frac{(\beta^*/\beta)^2}{1-\beta^*/\beta}\right] \exp\left[i\frac{t\alpha\alpha' + p_1^2\alpha\beta + p_3^2\alpha'\beta}{\alpha+\alpha'+\beta}\left(1-\frac{\beta^*}{\beta}\right)\right]. \tag{I.9}$$

If

$$F(\alpha,\alpha',\beta;s,t) = \sum_{n=1}^{\infty} F_n(\alpha,\alpha',\beta;s,t), \tag{I.10}$$

then (I.8) implies

$$F(\alpha,\alpha',\beta;s,t) = F_1(\alpha,\alpha',\beta;s,t) + \lambda\int_0^\infty d\alpha^*\,d\alpha'^*\,d\beta^*\,J(\alpha,\alpha',\beta;\alpha^*,\alpha'^*,\beta^*) \times F(\alpha^*,\alpha'^*,\beta^*;s,t), \tag{I.11}$$

where J is given in (I.9), and where $F_1(\alpha,\alpha',\beta)$ is the term $n=1$ in (I.10) and corresponds to the square box graph

$$F_1(\alpha,\alpha',\beta;s,t) = (\lambda i)^2\int_0^\infty d\beta_0\, e^{-i\beta_0 m^2}\frac{1}{(\alpha+\alpha'+\beta_0+\beta)^2} \times \exp\left(i\frac{s\beta_0\beta + t\alpha\alpha' + p_1^2\alpha'\beta + p_2^2\alpha\beta_0 + p_3^2\alpha'\beta + p_4^2\alpha'\beta_0}{(\alpha+\alpha'+\beta_0+\beta)}\right). \tag{I.12}$$

Equation (I.11) is the fundamental equation upon which all this work relies. Before ending this first part, we are going to give two generalizations of it.

1. The integral equation in dimension d

The expression of the Feynman amplitude M_n in dimension d is

$$M_n^d(s,t) = (\lambda i)^{n+1}\int_0^\infty \prod_{i=0}^{n}\{d\alpha_i\,d\alpha_i'\exp[-i(\alpha_i+\alpha_i')m^2]\} \times \prod_{i=0}^{n}[d\beta_i\exp(-i\beta_i m^2)] \times (1/P_n)^{d/2}\exp\left(i\frac{sA_n^s + tA_n^t + \Sigma_{i=1}^4 p_i^2 A_n^i}{P_n}\right). \tag{I.13}$$

If we apply exactly the same method as previously we find that the open amplitude $F^d(\alpha,\alpha',\beta)$ in dimension d verifies the integral equation

$$F^d(\alpha,\alpha',\beta;s,t) = F_1^d(\alpha,\alpha',\beta;s,t) + \lambda\int_0^\infty d\alpha^*\,d\alpha'^*\,d\beta^*\,J^d(\alpha,\alpha',\beta;\alpha^*,\alpha'^*,\beta^*) \times F^d(\alpha^*,\alpha'^*,\beta^*;s,t), \tag{I.14}$$

with

$$J^d(\alpha,\alpha',\beta;\alpha^*,\alpha'^*,\beta^*) = \left(\frac{\beta-\beta^*}{\beta(\alpha+\alpha'+\beta)}\right)^{(d-4)/2} J(\alpha,\alpha',\beta;\alpha^*,\alpha'^*,\beta^*), \tag{I.15}$$

and where F_1^d is obtained from (I.12) through the only change

$$\left(\frac{1}{\alpha+\alpha'+\beta_0+\beta}\right)^2 \to \left(\frac{1}{\alpha+\alpha'+\beta_0+\beta}\right)^{d/2}. \tag{I.16}$$

2. The integral equation with different mass on horizontal lines and vertical lines

Let us call m the mass of the particle in a vertical internal line and μ that in a horizontal internal line. The Feynman amplitude $M_n^{(m,\mu)}$ of the $(n+1)$ rungs ladder is

$$M_n^{(m,\mu)}(s,t) = \lambda^{n+1}\int_0^\infty \prod_{i=1}^{n}\{d\alpha_i\,d\alpha_i'\exp[-i(\alpha_i+\alpha_i')m^2]\} \times \prod_{i=0}^{n}[d\beta_i\exp(-i\beta_i\mu^2)] \times \left(\frac{1}{P_n}\right)^2\exp\left(i\frac{sA_n^s + tA_n^t + \Sigma_{i=1}^4 p_i^2 A_n^i}{P_n}\right). \tag{I.17}$$

The integral equation verified by the corresponding open amplitude is

$$F^{(m,\mu)}(\alpha,\alpha',\beta;s,t) = F_1^{(m,\mu)}(\alpha,\alpha',\beta;s,t) + \lambda\int J^{(m,\mu)}(\alpha,\alpha',\beta;\alpha^*,\alpha'^*,\beta^*) \times F^{(m,\mu)}(\alpha^*,\alpha'^*,\beta^*;s,t)\,d\alpha^*\,d\alpha'^*\,d\beta^*. \tag{I.18}$$

The only difference between $J^{(m,\mu)}$ and $J = J^{(m,m)}$ is the replacement of the factor

$$\left(-m^2\frac{(\beta^*/\beta)^2}{1-\beta^*/\beta}\right)$$

by

$$\left(-\mu^2\frac{\beta^*/\beta}{1-\beta^*/\beta} + m^2\frac{\beta^*}{\beta}\right)$$

in the definition (I.9). In the same way $F_1^{(m,\mu)}$ is obtained from $F_1 = F_1^{(m,m)}$ by the replacement of the factor $(\beta_0 m^2)$ by $(\beta_0\mu^2)$ in Eq. (I.12).

II. COMMENTS, FUNDAMENTAL PROPERTIES, AND PARTICULAR CASES

We obtained in the first part an integral equation verified by the infinite sum of the open amplitudes F_n. The aim of this part is to obtain a problem equivalent to (I.11) but more easily soluble.

Let us make some comments on the singularities of Eq. (I.11).

(1) The function $F(\alpha,\alpha',\beta;s,t)$ has two kinds of singularities. In addition to the singularities already discussed in the introduction and coming from the zeros of the function $D(x,t,\lambda)$ [see Eq. (16)], $F(\alpha,\alpha',\beta;s,t)$ also has all the Landau singularities of each $F_n(\alpha,\alpha',\beta;s,t)$. In Sec. II.1 some results about these singularities and about the Wick rotation will be recalled. In Sec. II.2, in order to get rid of these singularities,

we will consider the Mellin transform of the amplitude, and finally an equation in which all the quantities are real is obtained.

(2) The kernel $J(\alpha,\alpha',\beta;\alpha^*,\alpha'^*,\beta^*)$ also presents singularities when the variables α, α', and β go to zero [see Eq. (I.9)]. (i) The $1/\beta$ pole is a true singularity which prevents the Fredholm theory of $\mathscr{L}^2$ kernel to be directly used. It will be the object of Sec. III of this paper to give a general theorem which generalizes the Fredholm results to our case. (ii) The $1/(\alpha+\alpha'+\beta)$ term is not an actual difficulty, and it will be suppressed in Sec. II.3 by making the simple change of variables

$$(\alpha,\alpha',\beta)\to(\sigma=\alpha+\alpha',\ \delta=\alpha/\sigma,\ \gamma=\beta/\sigma).$$

In fact the real necessity of this change of variable will be completely clear only in Sec. IV where we will apply the general theorem of Sec. III to our integral equation.

At last, in Sec. II.4 we write the integral equation in two particular cases: First, for the value $t=0$ of the transfer and $p_i^2=m^2$ of the squared external momentum, and second for the case $\gamma=0$. In these two cases it happens that the integral equation becomes much simpler because the number of integration variables is reduced from three to two.

Due to the large number of formal manipulations, the notations needed in this section are quite laborious. Once our problem is correctly stated we will return to simple notations.

1. Landau singularities and Wick rotation

It is well known that each amplitude $M_n(s,t)$ is singular for some value of the invariants s, t, and $p_i^2 (i=1\to 4)$. If the transfer t is less than the elastic threshold of the t channel and if the external masses p_i^2 are not too different from m^2, the only Landau singularity (LS) of $M_n(s,t)$ is the threshold singularity in the direct channel for $s=[(n+1)m]^2$. The LS comes from the limit $(\alpha_i,\alpha_i',\beta_i)\to\infty$ in the integral of Eq. (I.1).

Let us make the change of variables

$$(\alpha_i,\alpha_i',\beta_i)\to(\bar\alpha_i=i\alpha_i,\bar\alpha_i'=i\alpha_i',\bar\beta_i=i\beta_i),\qquad\text{(II.1)}$$

in the integral (I.1) [we recall that P_n (resp. A_n^s, A_n^t, A_n^i) is an homogeneous polynomial of degree n (resp. $(n+1)$)]:

$$M_n(s,t)=\lambda^{n+1}\times\int_0^{i\infty}\left(\prod_{i=1}^{n}d\bar\alpha_i\,d\bar\alpha_i'\,e^{-(\bar\alpha_i+\bar\alpha_i')m^2}\right)\times\left(\prod_{i=0}^{n}d\bar\beta_i\,e^{-\bar\beta_i m^2}\right)\frac{1}{P_n^2}\times\exp\left(\frac{sA_n^s+tA_n^t+\Sigma_{i=1}^4 p_i^2A_n^i}{P_n}\right).\qquad\text{(II.2)}$$

The integrand becomes a real function of the new variables. If the integrand goes to zero sufficiently rapidly when the variables go to infinity, and as we do not encounter singularities in the first quadrant, it is possible to make the Wick rotation (WR), that is to say, to replace the integration on the imaginary axis by an integration on the real axis.[4]

Thus, when the WR is allowed, the amplitude $M_n(s,t)$ is real and has no singularity. It is a well-known property of the WR[6] that it can be carried out only when the values of s are less than the value of the first LS: $s<4m^2$.

The open amplitudes $F_n(\alpha,\alpha',\beta;s,t)$ possess singularities similar to the LS. When the energy s is less than the two particles threshold, it is possible to make the WR in the integral (I.4) and, if we define new functions $\hat F_n$ by the relation

$$\hat F_n(\bar\alpha,\bar\alpha',\bar\beta;s,t)=iF_n\left(\frac{\bar\alpha}{i},\frac{\bar\alpha'}{i},\frac{\bar\beta}{i};s,t\right),\qquad\text{(II.3)}$$

we find that $\hat F_n$ is a real function of $\bar\alpha,\bar\alpha',\bar\beta$

$$\hat F_n(\bar\alpha,\bar\alpha',\bar\beta;s,t)=\lambda^{n+1}\int_0^{\infty}\left(\prod_{i=1}^{n-1}d\bar\alpha_i\,d\bar\alpha_i'\,e^{-(\bar\alpha_i+\bar\alpha_i')m^2}\right)\times\left(\prod_{i=0}^{n-1}d\bar\beta_i\,e^{-\bar\beta_i m^2}\right)\frac{1}{P_n^2}\times\exp\left(\frac{sA_n^s+tA_n^t+\Sigma_{i=1}^4 p_i^2A_n^i}{P_n}\right),\qquad\text{(II.4)}$$

and is related to M_n by

$$M_n(s,t)=\int_0^\infty d\bar\alpha\,d\bar\alpha'\,d\bar\beta\exp[-(\bar\alpha+\bar\alpha'+\bar\beta)m^2]\times\hat F_n(\bar\alpha,\bar\alpha',\bar\beta;s,t).\qquad\text{(II.5)}$$

The integral recurrence relation (I.8) becomes a relation between real functions, and the integral equation (I.11) verified by

$$\hat F=\sum_{n=1}^{\infty}\hat F_n$$

is

$$\hat F(\bar\alpha,\bar\alpha',\bar\beta;s,t)=\hat F_1(\bar\alpha,\bar\alpha',\bar\beta;s,t)+\lambda\int_0^\infty d\bar\alpha^*\,d\bar\alpha'^*\,d\bar\beta^*\,\hat J(\bar\alpha,\bar\alpha',\bar\beta;\bar\alpha^*,\bar\alpha'^*,\bar\beta^*)\times\hat F(\bar\alpha^*,\bar\alpha'^*,\bar\beta^*;s,t),\qquad\text{(II.6)}$$

where $\hat J(\alpha,\alpha',\beta;\alpha^*,\alpha'^*,\beta^*)$ is deduced from $J(\alpha,\alpha',\beta;\alpha^*,\alpha'^*,\beta^*)$ of Eq. (I.9) through the change $i\to 1$, and similarly for $\hat F_1(\alpha,\alpha',\beta;s,t)$ [see Eq. (I.12)].

We see again that $\hat J$ and $\hat F_1$ are real functions.

2. Mellin transform

The Mellin transform $\bar f(x)$ of a function $f(s)$ which is integrable and regular when s goes to zero is defined by the relation

$$\bar f(x)=\int_0^\infty ds\,s^{-x-1}f(s).\qquad\text{(II.7)}$$

where $-1<\mathrm{Re}x<0$.

For the other values of x, $\bar f(x)$ can be defined either by analytic continuation or by an explicit relation different from the previous one and which takes into account the singularities of the integral.

In this paper we need the Mellin transform $\bar M_n$ of the amplitude M_n. If we use the Schwinger representation (I.1) of M_n, it is possible to explicitly perform the integration (II.7), and we obtain

$$\bar M_n(x,t)=(\lambda i)^{n+1}e^{-i\pi x}\Gamma(-x)\times\int_0^\infty\left(\prod_{i=1}^{n}d\alpha_i\,d\alpha_i'\,e^{-i(\alpha_i+\alpha_i')m^2}\right)$$

$$\times\Big(\prod_{i=0}^{n} d\beta_i\, e^{-i\beta_i m^2}\Big)$$
$$\times\frac{1}{P_n^2}\Big(\frac{iA_n^s}{P_n}\Big)^x \exp\Big(i\,\frac{tA_n^t+\Sigma_{i=1}^4 p_i^2 A_n^i}{P_n}\Big). \qquad \text{(II.8)}$$

Part of the singularities of $M_n(x,t)$ are set explicitly into the Γ function. The integral has no singularity in x for $x>-1$ and defines an analytic continuation of $\bar{M}_n(x,t)$ for these values. In the following, all the calculations are done inside the domain

$$-1<\mathrm{Re}x<\infty.$$

The results for $\mathrm{Re}x<-1$ will be obtained by analytic continuation. As the LS come from the limit $(\alpha_i,\alpha_i',\beta_i)\to\infty$, the factor $(iA_s/P_n)^x$ which grows less quickly than an exponential introduces no new singularities in the integral: The exponential behavior of the integrand when the variables go to infinity is the same as the one of $M_n(s,t)$ when $s=0$ and thus the WR can always be done when $\mathrm{Re}(x)$ is larger than (-1). The amplitude $\bar{M}_n(x,t)$ is the product of a factor $e^{-i\pi x}\Gamma(-x)$ which is independent of n by a real amplitude $\hat{M}$ which we call the regular part of the Mellin transform $\bar{M}$

$$\bar{M}_n(x,t)=e^{-i\pi x}\Gamma(-x)\,\hat{M}_n(x,t), \qquad \text{(II.9)}$$

with

$$\hat{M}(x,t)=\lambda^{n+1}\times\int_0^\infty\Big(\prod_{i=1}^{n} d\bar{\alpha}_i\, d\bar{\alpha}_i'\, e^{-(\bar{\alpha}_i+\bar{\alpha}_i')m^2}\Big)\times\Big(\prod_{i=0}^{n} d\bar{\beta}_i\, e^{-\bar{\beta}_i m^2}\Big)\frac{(A_n^s)^x}{P_n^{x+2}}\times\exp\Big(\frac{tA_n^t+\Sigma_{i=1}^4 p_i^2 A_n^i}{P_n}\Big), \qquad \text{(II.10)}$$

where the WR has been performed.

The Mellin transform of the open amplitude presents no new difficulties. It is possible to commute the integration on s which defines the Mellin transform [Eq. (II.7)] and the one on (α,α',β) which links F_n and M_n [Eq. (I.5)], and to carry out the WR. We find that, if $\hat{F}$ is the regular part of the Mellin transform of (iF_n), then

$$\hat{\bar{F}}_n=\lambda^{n+1}\times\int_0^\infty\Big(\prod_{i=1}^{n-1} d\bar{\alpha}_i\, d\bar{\alpha}_i'\, e^{-(\bar{\alpha}_i+\bar{\alpha}_i')m^2}\Big)\times\Big(\prod_{i=0}^{n-1} d\bar{\beta}_i\, e^{-\bar{\beta}_i m^2}\Big)\frac{(A_n^s)^x}{P_n^{x+2}}\times\exp\Big(\frac{tA_n^t+\Sigma_{i=1}^4 p_i^2 A_n^i}{P_n}\Big), \qquad \text{(II.11)}$$

and $\hat{M}_n$ and $\hat{F}_n$ are related by

$$\hat{M}_n(x,t)=\int_0^\infty d\bar{\alpha}\, d\bar{\alpha}'\, d\bar{\beta}\, \hat{\bar{F}}_n(\bar{\alpha},\bar{\alpha}',\bar{\beta};x,t)\times\exp[-(\bar{\alpha}+\bar{\alpha}'+\bar{\beta})m^2]. \qquad \text{(II.12)}$$

Let us now come to the important fact that the function $\hat{F}=\Sigma_{n=1}^\infty \hat{F}_n$ verifies the same integral equation as $\hat{F}$. As has been seen in Sec. I the kernel J of the recurrence relation (I.8) comes, on the one hand, from the Jacobian of the change variables

$$(\alpha_n,\alpha_n',\beta_n)\to(\alpha^*,\alpha'^*,\beta^*),$$

and, on the other hand, through the part of the recurrence laws for the ratios $A_n^{s,t,i}/P_n$ which deviates from the *-operation [see (I.3)]. The only difference between $\hat{\bar{F}}_n$ and $\hat{F}_n$ comes from the replacement of $\exp[s(A_n^s/P_n)]$ by $(A_n^s/P_n)^x$, which does not change anything in the kernel because A_n^s/P_n obeys exactly the *-law [Eq. (I.3b)].

The structure of the kernel $\hat{J}$ [Eq. (II.6)] is such that it leaves invariant some subspaces of functions. For example, due to the term $\theta(U-\beta^*/\beta)$ in $\hat{J}$, the subspace of functions with a β^x singularity when β goes to zero is stable. $\hat{\bar{F}}_1$ has such a singularity [see (II.11) and (A12)], thus $\hat{\bar{F}}_2$, $\hat{\bar{F}}_3$,$\cdots$ and their sum $\hat{\bar{F}}$ have the same singularity. It is convenient to define new functions $\bar{\bar{F}}_n$ which are regular when β goes to zero (we recall that we work in the domain $\mathrm{Re}x>-1$)

$$\hat{\bar{F}}_n(\alpha,\alpha',\beta;x,t)=\beta^x\,\bar{\bar{F}}_n(\alpha,\alpha',\beta;x,t). \qquad \text{(II.13)}$$

The function

$$\bar{\bar{F}}=\sum_n \bar{\bar{F}}_n$$

verifies an integral equation

$$\bar{\bar{F}}(\alpha,\alpha',\beta;x,t)=\bar{\bar{F}}_1(\alpha,\alpha',\beta;x,t)+\lambda\int_0^\infty d\alpha^*\,d\alpha'^*\,d\beta^*\,\bar{\bar{J}}(\alpha,\alpha',\beta;\alpha^*,\alpha'^*,\beta^*)\times\bar{\bar{F}}(\alpha^*,\alpha'^*,\beta^*;x,t), \qquad \text{(II.14)}$$

whose kernel $\bar{\bar{J}}$ is

$$\bar{\bar{J}}(\alpha,\alpha',\beta;\alpha^*,\alpha'^*,\beta^*)=(\beta^*/\beta)^x\,\hat{J}(\alpha,\alpha',\beta;\alpha^*,\alpha'^*,\beta^*). \qquad \text{(II.15)}$$

3. Final form of the integral equation

Finally, we are going to perform a change of variables. Let us define a new set of variables $\{\sigma,\delta,\gamma\}$ by the relations

$$\left\{\begin{array}{lll}\sigma=\alpha+\alpha' & & 0\leqslant\sigma<\infty,\\ \delta=\alpha/\sigma & \text{with} & 0\leqslant\delta\leqslant 1,\\ \gamma=\beta/\sigma & & 0\leqslant\gamma<\infty.\end{array}\right\} \qquad \text{(II.16)}$$

We also define a new function $\bar{\bar{\bar{F}}}$ by:

$$\bar{\bar{\bar{F}}}(\sigma,\delta,\gamma;x,t)=(\sigma/\lambda^2)\,\bar{\bar{F}}(\sigma\delta,\sigma(1-\delta),\sigma\gamma;x,t). \qquad \text{(II.17)}$$

In order to simplify the writing of the following parts, the function $\bar{\bar{\bar{F}}}(\sigma,\delta,\gamma;x,t)$ will be written as $F(\sigma,\delta,\gamma)$.

We have now reached the definitive shape of the integral equation. $F(\sigma,\delta,\gamma)$ verifies

$$F(\sigma,\delta,\gamma)=F_1(\sigma,\delta,\gamma)+\lambda\int_0^\infty d\sigma^*\int_0^1 d\delta^*\times\int_0^\infty d\gamma^*\,J(\sigma,\delta,\gamma;\sigma^*,\delta^*,\gamma^*)F(\sigma^*,\delta^*,\gamma^*), \qquad \text{(II.18)}$$

with

$$J(\sigma,\delta,\gamma;\sigma^*,\delta^*,\gamma^*)$$

$$= \frac{\sigma^*}{\sigma}\frac{1}{\gamma} L\left(\sigma,\delta,\gamma;\sigma^*,\delta^*,\frac{\sigma^*}{\sigma}\frac{\gamma^*}{\gamma}\right) \times\theta\left(U(\sigma,\delta;\sigma^*,\delta^*) - \frac{\sigma^*}{\sigma}\frac{\gamma^*}{\gamma}\right), \quad (II.19)$$

where

$$L(\sigma,\delta,\gamma;\sigma^*,\delta^*,u) = [u^x/(1+\gamma)]\exp[-m^2\sigma^* + \sigma A(\delta,\gamma,u)], \quad (II.20)$$

with

$$A(\delta,\gamma,u) = -m^2\gamma\frac{u}{1-u} - m^2\frac{u^2}{1-u} + \frac{t\delta(1-\delta) + p_1^2\delta\gamma + p_3^2(1-\delta)\gamma}{1+\gamma}(1-u), \quad (II.21)$$

and

$$U(\sigma,\delta;\sigma^*,\delta^*) = \inf\left(1, \frac{\sigma^*}{\sigma}\frac{\delta^*}{\delta}, \frac{\sigma^*}{\sigma}\frac{1-\delta^*}{1-\delta}\right). \quad (II.22)$$

In (II.18), F_1 is

$$F_1 = \int_0^\infty d\gamma_0 \frac{\gamma_0^x}{(1+\gamma+\gamma_0)^{x+2}}\exp[\sigma B(\delta,\gamma,\gamma_0)], \quad (II.23)$$

with

$$B(\delta,\gamma,\gamma_0) = -m^2\gamma_0 + \frac{t\delta(1-\delta) + p_1^2\delta\gamma + p_2^2\delta\gamma_0 + p_3^2(1-\delta)\gamma + p_4^2(1-\delta)\gamma_0}{1+\gamma+\gamma_0}. \quad (II.24)$$

Before ending this part we must give the way to come back from $F(\sigma,\delta,\gamma)$ to the physical amplitude $M(s,t)$. First $\widehat{M}(x,t)$ is obtained from $F(\sigma,\delta,\gamma)$ by

$$\widehat{M}(x,t) = \lambda^2\int_0^\infty d\sigma\int_0^1 d\delta\int_0^\infty d\gamma\,\gamma^x\sigma^{x+1} \times\exp[-\sigma(1+\gamma)m^2]\,F(\sigma,\delta,\gamma). \quad (II.25)$$

Then we use the inverse Mellin transform to build $M(s,t)$,

$$M(s,t) = \frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty} dx\, s^x\,\Gamma(-x)\,e^{-i\pi x}\,\widehat{M}(x,t), \quad (II.26)$$

with $-1<\sigma<0$.

In fact, this inversion formula is valid only when the real part of the poles $x_i(t,g)$ of $\widehat{M}$ are less than σ. If this is not the case, Eq. (II.26) must be replaced by

$$M(s,t) = \frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty} dx\, s^x\,\Gamma(-x)\,e^{-i\pi x}\,\widehat{M}(x,t) + \sum_{\substack{\text{poles with}\\ \mathrm{Re}[x_i(t,g)]>\sigma}}\left\{\Gamma[-x_i(t,g)]\,e^{-i\pi x_i(t,g)}\operatorname*{Residue}_{x\to x_i(t,g)}[\widehat{M}(x,t)]\right\}. \quad (II.27)$$

4. Particular cases

In this subsection the expression of the integral equation is given in two particular cases; it is shown that the number of integration variables is reduced from three to two.

a. The integral equation with $t=0$ and $p_i^2=m^2$

This first particular case is of deep physical interest as the poles of the amplitude F taken at the value zero of the transfer are related with the intercept of the Regge trajectories.

If in the expression of A [Eq. (II.21)] and B [Eq. (II.24)] we put $t=0$ and $p_i^2=m^2$ for $i=1$ to 4, we find that the functions L and F_1 do not depend on the variable δ. This is not the case for the kernel $J(\sigma,\delta,\gamma;\sigma^*,\delta^*,\gamma^*)$ which depends on δ, but only through the limits of the integration region (the θ function). These limits in the plane (δ^*,γ^*) at fixed σ^* are shown in Fig. 4.

The integration region has an important property: The length of the segment DD', that is to say the value of the integral $\int_{\delta_+^*}^{\delta_-^*} d\delta^*$ when σ^* and γ^* are kept fixed, is equal to $(1-\gamma^*/\gamma)$ and is independent of δ. Thus if we integrate in this region a function $f(\sigma,\gamma)$ which is independent of δ, the result is also independent of δ

$$\int d\sigma^*\,d\delta^*\,d\gamma^*\,f(\sigma^*,\gamma^*)\,\theta\left(U(\sigma,\delta;\sigma^*,\delta^*) - \frac{\sigma^*}{\sigma}\frac{\gamma^*}{\gamma}\right) = \int_0^\infty d\sigma^*\,d\gamma^*\,f(\sigma^*,\gamma^*)\left(1-\frac{\gamma^*}{\gamma}\right) \times\theta\left(V(\sigma,\sigma^*) - \frac{\sigma^*}{\sigma}\frac{\gamma^*}{\gamma}\right), \quad (II.28)$$

with

$$V(\sigma,\sigma^*) = \inf(1,\sigma^*/\sigma). \quad (II.29)$$

So, since $F_1(\sigma,\delta,\gamma)$ does not depend on δ, $F_2,F_3,\ldots,F_n$ and thus their sum $F(\sigma,\delta,\gamma)$ does not depend on δ. Let us write $F(\sigma,\gamma)$, this last function, which verifies the integral equation

$$F(\sigma,\gamma) = F_1(\sigma,\gamma) + \lambda\int_0^\infty d\sigma^*\,d\gamma^*\,j(\sigma,\gamma;\sigma^*,\gamma^*)\,F(\sigma^*,\gamma^*), \quad (II.30)$$

with

$$j(\sigma,\gamma;\sigma^*,\gamma^*) = \frac{\sigma^*}{\sigma}\frac{1}{\gamma}\left(1-\frac{\gamma^*}{\gamma}\right) l\left(\sigma,\gamma;\sigma^*,\frac{\sigma^*}{\sigma}\frac{\gamma^*}{\gamma}\right) \times\theta\left(V(\sigma,\sigma^*) - \frac{\sigma^*}{\sigma}\frac{\gamma^*}{\gamma}\right), \quad (II.31)$$

where

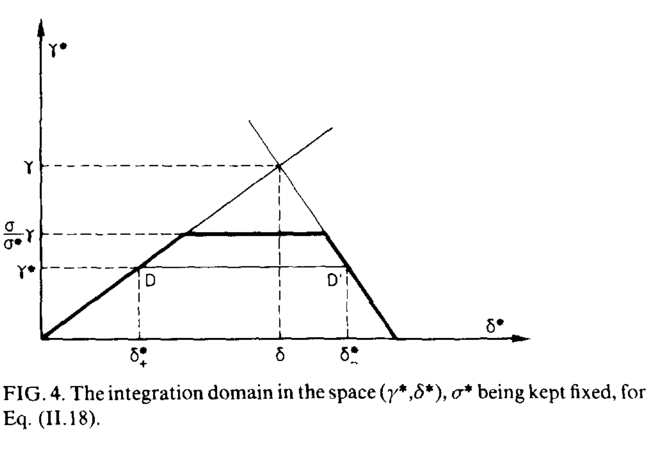


FIG. 4. The integration domain in the space (γ^*,δ^*), σ^* being kept fixed, for Eq. (II.18).

$$l(\sigma,\gamma,\sigma^*,u) = \frac{u^x}{1+\gamma}\exp[-m^2\sigma + a(\sigma,\gamma,u)\,\sigma],$$

$$a(\sigma,\gamma,u) = -m^2\gamma\frac{u}{1-u} - m^2\,\frac{u^2}{1-u} + m^2\frac{\gamma}{1+\gamma}(1-u), \tag{II.32}$$

$$V(\sigma,\sigma^*) = \inf\left(1, \frac{\sigma^*}{\sigma}\right),$$

and

$$F_1(\sigma,\gamma) = \int_0^\infty d\gamma_0\,\frac{\gamma_0^x}{(1+\gamma+\gamma_0)^{x+2}} \times\exp\left(-\sigma m^2\gamma_0 + \sigma m^2\frac{\gamma+\gamma_0}{1+\gamma+\gamma_0}\right). \tag{II.33}$$

The reduction of the number of integration variables from three to two is a consequence of the well-known result[7] that in the equal mass case and at $t = 0$, the φ^3 ladder graphs have a supplementary symmetry.

b. The integral equation with $\gamma = 0$

Usually when an integral equation is written for a particular value of a variable the number of integration variables does not diminish. Here if we put $\gamma = 0$, the interval of integration for the variable γ^* disappears. More precisely, if in Eq. (II.18) we put $\gamma = 0$, we find that the function

$$F(\sigma,\delta) = F(\sigma,\delta,\gamma=0)$$

verifies an integral equation

$$F(\sigma,\delta) = F_1(\sigma,\delta) + \lambda\int_0^\infty d\sigma^*\int_0^1 d\delta^*\,j(\sigma,\delta,\sigma^*,\delta^*)\,F(\sigma^*,\delta^*), \tag{II.34}$$

with

$$j(\sigma,\delta;\,\sigma^*,\delta^*) = \int_0^U L(\sigma,\delta,\gamma=0;\,\sigma^*,\,\delta^*,\,u)\,du, \tag{II.35}$$

$$L(\sigma,\delta,\gamma=0;\,\sigma^*,\,\delta^*,u) = u^x\exp\left(-m^2\sigma^* - m^2\frac{u^2}{1-u}\sigma + t\delta(1-\delta)(1-u)\,\sigma\right), \tag{II.36}$$

and

$$F_1(\sigma,\delta) = \int_0^\infty d\gamma_0\,\frac{\gamma_0^x}{(1+\gamma_0)^{x+2}} \times\exp\left(-\sigma m^2\gamma_0 + \sigma\,\frac{t\delta(1-\delta) + p_2^2\,\delta\gamma_0 + p_4^2\,(1-\delta)\,\gamma_0}{1+\gamma_0}\right). \tag{II.37}$$

This result will be very useful in the Secs. III and IV.

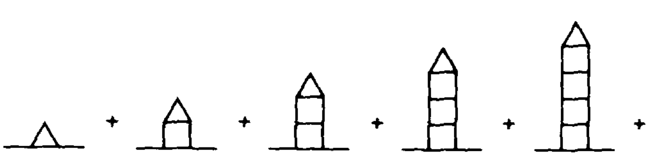


FIG. 5. The series deduced from the ladder series with γ put equal to zero.

Let us remark that the integral equation does not depend on the variables p_1^2 and p_3^2. $F(\sigma,\delta)$ represents the open amplitude of the sum of graphs shown in Fig. 5.

c. The integral equation with $t = 0$, $p_i^2 = m^2$ and $\gamma = 0$

In this case we obtain a one-variable integral equation for the function

$$F(\sigma) = F(\sigma, \gamma=0) \quad \text{at } t = 0, p_i^2 = m^2,$$

$$F(\sigma) = F_1(\sigma) + \lambda\int_0^\infty j(\sigma,\sigma^*)\,F(\sigma^*)\,d\sigma^*, \tag{II.38}$$

with

$$j(\sigma,\sigma^*) = \int_0^V du\left(1 - \frac{\sigma}{\sigma^*}u\right)u^x \times\exp\left(-m^2\sigma^* - m^2\frac{u^2}{1-u}\sigma\right), \tag{II.39}$$

and

$$F_1(\sigma) = \int_0^\infty d\gamma_0\,\frac{\gamma_0^x}{(1+\gamma_0)^{x+2}}\exp\left(-\sigma m^2\frac{\gamma_0^2}{1+\gamma_0}\right). \tag{II.40}$$

III. SOLUTION OF A FAMILY OF SINGULAR INTEGRAL EQUATIONS

In this section we want to study the family of integral equations which can be written

$$f(x,y) = g(x,y) + \lambda\int_0^X dx^*\int_0^y dy^*\,K(x,y;\,x^*,y^*)f(x^*,y^*), \tag{III.1}$$

where the kernel K has a $1/y$ singularity when the variable y goes to zero. The variable x and the limit of integration can stand for a set of variables $x = \{x_1,\dots,x_n\}$ and $X = \{X_1,\dots,X_n\}$. The X_i can be either finite or infinite. The integral operator corresponding to the kernel K will be denoted $\mathscr{K}$. These integral equations cannot be solved in general by applying directly the classical methods: Due to the $1/y$ singularity the kernel K is neither bounded nor square integrable, and the classical Fredholm theory cannot be applied. It can be verified that the theory of compact operator[8] which generalizes the Fredholm results to a larger number of situations is also unusable.

Here we prove the existence of solutions of the family of singular integral equations and we give their analytic structure. In a first step we define the function spaces and the operator spaces in which we are going to work and we give some of their properties. Then our main theorem can be correctly presented. The proof of the theorem needs the demonstration of several intermediary lemmas.

1. The function spaces C_y and D. The operator spaces C_y^* and D^*

Let us now present the function space C_y and D in which we are looking for solutions.

Definition: C_y is the set of function $f(x, y)$ such that

$$\|f\|_Y = \left[\int_0^X dx \left(\operatorname*{Max}_{0\leqslant y\leqslant Y} |f(x,y)|\right)^2\right]^{1/2} \tag{III.2}$$

is finite. The space C_∞ is defined as the intersection of all the C_y space for Y finite

$$C_\infty = \bigcap_{Y\text{ finite}} C_y.$$

It is easy to see that C_y is a vectorial space and $\|\ \ \|_Y$ is a norm on the space.

Definition: D is the set of functions $f(x, y)$ which are indefinitely differentiable with respect to the second variable y near $y = 0$ and such that the expression

$$|||f|||_n = \left[\int_0^X dx \left(\frac{1}{n!}\frac{\partial^n f(x,y=0)}{\partial y^n}\right)^2\right]^{1/2} \tag{III.3}$$

is finite for $n = 0,1,2,\cdots$.

D is a vectorial space and $|||\ \ |||_n$ is a seminorm on D. It is not a norm because $|||f|||_n$ can be equal to zero with f not being the zero function.

It is also necessary to define the spaces of integral operators. First, let us write the kernel $K(x, y; x^*, y^*)$ in terms of the reduced kernel $M(x, y; x^*, v)$

$$K(x,y,x^*,y^*) = \frac{1}{y} M\left(x,y;x^*,\frac{y^*}{y}\right). \tag{III.4}$$

If we make the change of variable

$$y^* \to v = \frac{y^*}{y}, \tag{III.5}$$

the action of the operator $\mathcal{K}$ on a function f becomes

$$f(x,y) \to \mathcal{K} f(x,y) = \int_0^X dx^* \int_0^1 dv\, M(x,y,x^*,v) f(x^*, yv). \tag{III.6}$$

The spaces of operators C_y^* and D^* can now be defined.

Definition: An operator $\mathcal{K}$ belongs to the C_y^* space if its norm $\|\mathcal{K} f\|_Y^2$, defined by

$$\|\mathcal{K}\|_Y = \left[\int_0^X dx \int_0^X dx^* \left(\int_0^1 dv \operatorname*{Max}_{0\leqslant y\leqslant Y} |M(x,y;x^*,v)|\right)^2\right]^{1/2}, \tag{III.7}$$

is finite.

We define C_∞^* by

$$C_\infty^* \bigcap_{Y\text{ finite}} C_y.$$

We shall also say that the kernel K or even the reduced kernel M belongs to C_y^*.

It must be noted that the norm $\|\ \ \|_Y$ is not the usual norm $\mathcal{N}(\mathcal{K})$ of a linear operator on a normed vectorial space which is defined by

$$\mathcal{N}(\mathcal{K}) = \operatorname{Max}\|\mathcal{K} f\|_Y,$$

where the Max must be taken on the unit ball of C_y: $\|f\|_Y = 1$.

However, it can be shown that

Lemma 1: The norm $\|\ \ \|_Y$ is always an upper bound for the usual norm $\mathcal{N}$

$$\mathcal{N}(\mathcal{K}) \leqslant \|\mathcal{K}\|_Y. \tag{III.8}$$

If $\mathcal{K}$ belongs to C_y^*, $\mathcal{K}$ is a bounded operator in the usual sense. If, moreover, the function f belongs to C_y, so also does the function $\mathcal{K} f$.

The proof of this proposition is straightforward. By definition one has

$$\|\mathcal{K} f\|_Y^2 = \int_0^X dx \left(\operatorname*{Max}_{0\leqslant y\leqslant Y} \left|\int_0^X dx^* \int_0^1 dv\, M(x,y;x^*,v) f(x^*,yv)\right|\right)^2.$$

The maximum of an integral is less than the integral of the maximum of the integrand, and using $v \leqslant 1$, one obtains

$$\|\mathcal{K} f\|_Y^2 \leqslant \int_0^X dx \left[\int_0^X dx^* \left(\int_0^1 dv \operatorname*{Max}_{0\leqslant y\leqslant Y} |M(x,y,x^*,v)|\right) \times \left(\operatorname*{Max}_{0\leqslant y\leqslant Y} |f(x,y)|\right)\right]^2,$$

which can be bounded using the Schwarz inequality

$$\left(\int dx\, f(x)\, g(x)\right)^2 \leqslant \left(\int dx\, f^2(x)\right)\left(\int dx\, g^2(x)\right);$$

and finally

$$\|\mathcal{K} f\|_Y < \|\mathcal{K}\|_Y \|f\|_Y,$$

which means

$$\mathcal{N}(\mathcal{K}) = \operatorname{Max}\left(\frac{\|\mathcal{K} f\|_Y}{\|f\|_Y}\right) < \|\mathcal{K}\|_Y.$$

It is also useful to define the D^* space.

Definition: A linear operator $\mathcal{K}$, with a kernel

$$K(x,y;x^*,y^*) = (1/y)\, M(x,y;x^*,y^*/y)$$

belongs to the space D^* if for any $n \geqslant 0$, the expression

$$|||\mathcal{K}|||_n = \left[\int_0^X dx \int_0^X dx^* \left(\int_0^1 dv \left|\frac{1}{n!}\frac{\partial^n M(x,y=0,x^*,v)}{\partial y^n}\right|\right)^2\right]^{1/2} \tag{III.9}$$

is finite.

The generalization of Lemma 1 to the new space is

Lemma 2: If f belongs to D and $\mathcal{K}$ to D^* then $\mathcal{K} f$ belongs to D and, for any $n \geqslant 0$

$$|||\mathcal{K} f|||_n < \sum_{l=0}^{n} |||\mathcal{K}|||_{n-l}\, |||f|||_l. \tag{III.10}$$

The proof of this lemma is similar to that of the first lemma.

It is now possible to state the main theorem of this section.

2. The theorem

Theorem 1: If an operator $\mathcal{K}$ belongs to C_y^* and D^* and if the function g belongs to C_y and D, then the equation

$$f(x,y) = g(x,y) + \lambda \int_0^X dx^* \int_0^y dy^*\, K(x,y;x^*,y^*) f(x^*,y^*), \tag{III.11}$$

$$f(x,y) = g(x,y)$$

$$+\lambda\int_0^X dx^*\int_0^1 dv\, M(x,y;x^*,v)f(x^*,yv)\,, \tag{III.12}$$

has a unique solution which belongs to C_y and D for any value of λ. The only λ dependent singularities of $f(x,y)$ considered as a function of x, y and λ are an infinite set of fixed poles in λ: $1/(\lambda-\Lambda_{n,i})$ with $n=0,1,2,\cdots$ and $i=1,2,\cdots$. The position of the poles is independent of x and y, independent of the function g, and only depends on the operator $\mathscr{K}$. More precisely, $\Lambda_{n,i}$ for $i=1,2,\cdots$ is the set of eigenvalues of the operator defined by the kernel

$$j^{(n,n)}(x,x^*)=\int_0^1 dv\, v^n\, M(x,0,v^*,v)\,.$$

This kernel is an $\mathscr{L}^2$ kernel and thus has only a discrete spectrum. The function $f(x,y)$ can also have singularities in the variable x, independent of λ, if ever g or K has such singularities.

The method we use to solve the equation is to decompose $f(x,y)$ into a sum of $N+1$ terms

$$f(x,y)=\sum_{n=0}^{N-1} f_n(x)\frac{y^n}{n!}+\bar f(x,y)\frac{y^N}{N!} \tag{III.13}$$

and to write integral equations verified by the functions $f_n(x)$ $(n=0,1,\ldots,N-1)$ and $\bar f(x,y)$. Then we prove that the integral equations verified by $f_n(x)$ are of the Fredholm type and that the function $\bar f(x,y)$ can be obtained as the sum of a convergent perturbative series of λ. Before beginning the demonstration we must define $\bar g$ and $\bar M$, the Taylor remainder of the functions f and M, and we give their properties.

3. Taylor remainders $\bar g$ and $\bar M$

Definition: The Taylor remainder $\bar g(x,y)$ of g is defined by the relation

$$g(x,y)=\sum_{n=0}^{N-1} g_n(x)\frac{y^n}{n!}+\bar g(x,y)\frac{y^N}{N!}\,, \tag{III.14}$$

where

$$g_n(x)=\frac{\partial^n g(x,y=0)}{\partial y^n}\,,$$

and the Taylor remainder $\bar M(x,y;x^*,v)$ of M is defined by

$$M(x,y;x^*,v)=\sum_{n=0}^{N-1} M_n(x,x^*,v)\frac{y^n}{n!}+\bar M(x,y;x^*,v)\frac{y^N}{N!}\,, \tag{III.14'}$$

where

$$M_n(x,x^*,v)=\frac{\partial^n M(x,y=0,x^*,v)}{\partial y^n}\,.$$

The functions $\bar g$ and $\bar M$ verify the following properties.

Lemma 3: If the function g belongs to D so does its Taylor remainder $\bar g$. If the reduced kernel M belongs to D* so does its Taylor remainder $\bar M$.

Proof: Let $t(y)$ be a function which is indefinitely derivable and $\bar t(y)$ its Taylor remainder defined by

$$t(y)=\sum_{n=0}^{N-1}\frac{\partial^n t(y=0)}{\partial y^n}\frac{y^n}{n!}+\bar t(y)\frac{y^N}{N!}\,.$$

The function $\bar t$ admits an integral representation

$$\bar t(y)=N\int_0^1\frac{\partial^N t(uy)}{\partial y^N}(1-u)^{N-1}du\,, \tag{III.15}$$

and since t is indefinitely differentiable

$$\frac{\partial^k\bar t(y)}{\partial y^k}=N\int_0^1\frac{\partial^{N+k}t(uy)}{\partial y^{N+k}}u^k(1-u)^{N-1}du\,. \tag{III.16}$$

If we put $y=0$ in the previous equation we have

$$\frac{\partial^k\bar t(y=0)}{\partial y^k}=\frac{\Gamma(k+1)\Gamma(N+1)}{\Gamma(N+k+1)}\frac{\partial^{N+k}t(y=0)}{\partial y^{N+k}}\,. \tag{III.17}$$

The $|||\ \ |||_k$ norm of $\bar g(x,y)$ [see Eq. (III.3)] can be obtained using this last equality

$$|||\bar g|||_k=\frac{\Gamma(k+1)\Gamma(N+1)}{\Gamma(N+k+1)}|||g|||_{N+k}\,,$$

which proves that $\bar g$ belongs to D if g belongs to D. In the same manner we find

$$|||\bar M|||_k=\frac{\Gamma(k+1)\Gamma(N+1)}{\Gamma(N+k+1)}|||M|||_{N+k}\,,$$

and thus $\bar M$ belongs to D* if M belongs to D*.

4. Integral equations verified by $f_n(x)$ $(n=0,1,\ldots,N-1)$ and $\bar f(x)$

Let us put the decomposition (III.13) of $f(x,y)$ and the expression (III.14) and (III.14') of $g(x,y)$ and $M(x,y;x^*,v)$ into Eq. (III.12)

$$\begin{aligned}\sum_{n=0}^{N-1} f_n(x)\frac{y^n}{n!}+\bar f(x,y)\frac{y^N}{N!}&=\sum_{n=0}^{N-1} g_n(x)\frac{y^n}{n!}+\bar g(x,y)\frac{y^N}{N!}\\&+\lambda\int_0^X dx^*\int_0^1 dv\left(\sum_{n=0}^{N-1}M_n(x,x^*,v)\frac{y^n}{n!}+\bar M(x,y,x^*,v)\frac{y^N}{N!}\right)\\&\times\left(\sum_{n=0}^{N-1}f_n(x^*)\frac{(yv)^n}{n!}+f(x^*,yv)\frac{(yv)^N}{N!}\right).\end{aligned} \tag{III.18}$$

The identification of the coefficients of y^n $(n=0,1,\ldots,N-1)$ of the two members of the equation gives the set of integral equations for $f_n(x)$

$$f_n(x)=h_n(x)+\lambda\int_0^X dx^*\, j^{(n,n)}(x,x^*)f_n(x^*)\,, \tag{III.19}$$

with

$$h_n(x)=g_n(x)+\lambda\sum_{l=0}^{n-1}C_n^l\int_0^X dx^*\, j^{(n,l)}(x,x^*)f_l(x^*)\,, \tag{III.20}$$

and

$$j^{(n,l)}(x,x^*)=\int_0^1 dv\, v^l M_{n-l}(x,x^*,v)\,. \tag{III.21}$$

The identification of the remainder in the two sides of (III.18) gives an integral equation for $\bar f(x,y)$

$$\bar f(x,y)=\bar h(x,y)$$

$$+\lambda \int_0^X dx^* \int_0^y dy^* \left(\frac{y^*}{y}\right)^N K(x,y;x^*,y^*)\bar{f}(x^*,y^*),$$

with

$$\begin{aligned}\bar{h}(x,y) &= \bar{g}(x,y)\\ &+\lambda \int_0^X dx^* \int_0^1 dv\, \bar{M}(x,y;x^*,v)\left(\sum_{n=0}^{N-1} f_n(x^*)\frac{(yv)^n}{n!}\right)\\ &+\lambda N! \sum_{n=N}^{2(N-1)} \frac{y^{n-N}}{n!}\\ &\times\left(\sum_{l=n-N+1}^{N-1} C_n^l \int_0^X dx^* j^{(n,l)}(x,x^*) f^{(l)}(x^*)\right). \quad \text{(III.23)}\end{aligned}$$

All the identifications are possible because the functions $\bar{g}(x,y)$ and $\bar{M}(x,y,x^*,v)$ have no singularity when y goes to zero (see Lemma 3 and the definition of the D-space).

The first interest of this set of coupled integral equations is the possibility of obtaining the solution step by step: The inhomogeneous term $h_0(x)$ of the integral equation verified by $f_0(x)$ is a known function and thus the equation can be solved independently of the functions $f_n(x)$ $(n=1,2,...,N-1)$ or $\bar{f}$; once $f_0(x)$ is known, the inhomogeneous term $h_1(x)$ of the integral equation verified by $f_1(x)$ is also known and it is possible to solve the equation and to obtain $f_1(x)$, and so on for all the functions $f_n(x)$ and $\bar{f}(x,y)$.

5. Solution of Eq. (III.19) verified by $f_n(x)$, $n = 0,1,...,N-1$

We prove in this subsection the following results:

Lemma 4: The kernel $j^{(n,n)}(x,x^*)$ and the inhomogeneous term $h_n(x)$ are square integrable functions and thus the integral equation (III.19) can be solved by the usual Fredholm methods. The functions $f_n(x)$ are square integrable and their only singularities, which depend on λ, are fixed poles $1/(\lambda - \Lambda_{l,i})$ $(l=0,1,...,n$ and $i=1,2,\cdots)$, where $\Lambda_{l,i}$ is the ith eigenvalue of the operator $j^{(l,l)}(x,x^*)$.

Proof: We do not recall here the classical Fredholm theorems which can be found in Ref. 8 for the case of $\mathscr{L}^2$ functions. In order to apply these theorems it is necessary to show that the kernels $j^{(n,n)}(x,x^*)$, and the inhomogeneous term $h_n(x)$ are square integrable. The kernels $j^{(n,n)}(x,x^*)$, and, more generally, all the kernels $j^{(n,l)}(x,x^*)$ are bounded by

$$|j^{(n,l)}(x,x^*)| < \int_0^1 dv\, |M_{n-l}(x,x^*,v)|\,, \qquad \text{(III.24)}$$

which are square integrable because the operator $\mathscr{K}$ belongs to D*, and thus are square integrable.

We now prove simultaneously by recurrence that $f_n(x)$ and $h_n(x)$ are square integrable. The functions $g_n(x)$ are square integrable because g belongs to D. Since $h_0(x) = g_0(x)$, $h_0(x)$ is square integrable and it is possible to apply the Fredholm theorems to the first integral equation and then to prove that $f_0(x)$ is also square integrable. Let us suppose now that for $l=0,1,...,n-1$, $f_l(x)$ is square integrable; as $j^{(n,l)}(x,x^*)$ is also square integrable, so (due to the Schwarz inequality) is the term

$$\int_0^X dx^* j^{(n,l)}(x,x^*) f_l(x^*)\,,$$

and $h_n(x)$ being a finite sum of square integrable terms, is also square integrable. Finally, using the Fredholm theorem, $f_n(x)$ being the solution of the integral equation is also square integrable.

The functions $f_n(x)$ have a set of singularities in λ which consists of two parts.

(i) The set of singularities of the first term $h_n(x)$, which is the set of all the singularities of all the $f_l(x)$ for $l=0,1,...,n-1$.

(ii) the new singularities coming from the Fredholm determinant. These singularities are fixed poles in λ, the positions of which are the eigenvalues of $j^{(n,n)}$.

In other words the set of all singularities of f_n is a set of fixed poles $1/(\lambda - \Lambda_{l,i})$, $l=0,1,...,n$, where $\Lambda_{l,i}$ is the ith eigenvalue of $j^{(l,l)}(x,x^*)$.

6. Solution of Eq. (III.22) verified by $\bar{f}(x,y)$

Before giving the main results of this subsection a preliminary lemma must be proved.

Lemma 5: Let be $\mathscr{K}$ an operator of C_y^* and Λ a fixed finite number. Then there exists an integer number N which depends only on Λ and on $\mathscr{K}$ and such that for any $\lambda < \Lambda$ the $\mathscr{N}$-norm of the operator $\lambda\mathscr{K}_N \equiv \lambda (y^*/y)^N \mathscr{K}$ is less than one

$$\mathscr{N}(\lambda\mathscr{K}_N) < 1. \qquad \text{(III.25)}$$

Proof: Due to the inequality (III.8) proved in the Lemma 1, it is enough to show that $\lambda^2 I_N < 1$ where

$$\begin{aligned}I_n &= \|\mathscr{K}_n\|_Y^2\\ &= \int_0^X dx \int_0^X dx^* \left(\int_0^1 dv\, v^n \operatorname*{Max}_{0\leqslant y\leqslant Y} |M(x,y;x^*,v)|\right)^2. \qquad \text{(III.26)}\end{aligned}$$

I_n can be written

$$I_n = \int_0^1 dv_1\, v_1^n \int_0^1 dv_2\, v_2^n\, H(v_1,v_2)\,,$$

with

$$\begin{aligned}H(v_1,v_2) = \int_0^X dx \int_0^X dx^* &\left(\operatorname*{Max}_{0\leqslant y\leqslant Y} |M(x,y,x^*,v_1)|\right)\\ &\times\left(\operatorname*{Max}_{0\leqslant y\leqslant Y} |M(x,y,x^*,v_2)|\right). \qquad \text{(III.27)}\end{aligned}$$

H is a positive integrable function, and thus for any $\epsilon > 0$, there exists $\phi > 0$ such that for any $\varphi < \phi$, we have

$$I_0 - \frac{\epsilon}{2} < \int_0^{1-\varphi} dv_1 \int_0^{1-\varphi} dv_2\, H(v_1,v_2) < I_0\,.$$

The variables v_1 and v_2 being less than one, we have for any φ less than the same ϕ as above

$$I_n - \frac{\epsilon}{2} < \int_0^{1-\varphi} dv_1\, v_1^n \int_0^{1-\varphi} dv_2\, v_2^n\, H(v_1,v_2)\,.$$

In this last integral, we can bound v_1^n, and v_2^n by $(1-\varphi)^n$, and finally we find

$$I_n < (1-\varphi)^{2n} I_0 + \epsilon/2\,.$$

By taking N such that $(1-\varphi)^{2N}<\epsilon/2I_0$, we see that $I_n<\epsilon$ when $n\geqslant N$. Finally choosing $\epsilon=1/\Lambda^2$ we achieve the proof of the lemma.

Lemma 6: In the integral equation (III.22), the inhomogeneous term $\bar{h}(x,y)$ belongs to C_y and the operator $\mathscr{K}_N$ belongs to C_y^*. Thus the function $\bar{f}(x,y)$ is the sum of the convergent Neumann series

$$\bar{f}=\sum_{l=0}^{\infty}(\lambda\,\mathscr{K}_N)^l\,\bar{h}, \tag{III.28}$$

for all the values of λ less than Λ in modulus and belongs to C_y. Furthermore $\bar{h}$ belongs to D and K_N to D^* and thus $\bar{f}$ belongs to D. Finally the only singularities of $\bar{f}(x,y)$ for $|\lambda|$ less than Λ, are the poles of $f_n(x)$ $(n=0,1,\ldots,N-1)$.

Proof: The function $\bar{h}(x,y)$ is given by Eq. (III.23). The function $\bar{g}(x,y)$ can be written

$$\bar{g}(x,y)=\frac{N!}{y^N}\left(g(x,y)-\sum_{n=0}^{N-1}f_n(x)\frac{y^n}{n!}\right)\quad \text{if } y>0,$$

$$\bar{g}(x,0)=\frac{\partial^N g(x,y=0)}{\partial y^N}\quad \text{if } y=0.$$

Using the fact that g belongs to C_y and to D, the continuity of $\bar{g}(x,y)$ when y goes to zero, Lemma 4, and the relation

$$\underset{0\leqslant y\leqslant Y}{\mathrm{Max}}\,(y^n)=Y^n<\infty, \tag{III.29}$$

it is easy to prove that $\bar{g}$ belongs to C_y.

In the same manner we prove that $\bar{M}$ belongs to C_y^*. The proof that $\bar{h}(x,y)$ belongs to C_y is now straightforward by again using Lemma 4 and Eq. (III.29).

The proof that $\bar{h}(x,y)$ belongs to D, follows from Lemma 3 and Lemma 4 for $n=0$. The kernel $\mathscr{K}_N$ has a reduced kernel M_N which is equal to

$$M_N(x,y;x^*,u)=u^N M(x,y;x^*,u), \tag{III.30}$$

and thus belongs to the same spaces (C_y^* and D^*) as $\mathscr{K}$.

By taking the N defined in Lemma 5, one sees that the Neumann series converges for all the values of λ less than Λ in modulus. The function $\bar{f}$ has the same set of singularities as $\bar{h}$, that is to say all the singularities of the f_n for $n=0,1,\ldots,N-1$. Each term of the Neumann series belongs to C_y and D (see Lemmas 1 and 2) and since the convergence is uniform, this is also the case for its sum $\bar{f}(x,y)$.

Theorem 1 is a consequence of Lemmas 4 and 6 and of these two last points: (i) In Lemma 5, Λ has been chosen in an arbitrary way and thus can be taken arbitrarily large. (ii) The function $f(x,y)$ is a finite sum of terms which belongs to C_y and D [see Lemmas 4 and 6 and Eq. (III.29)], and thus belongs to C_y and D.

7. Generalization of Theorem 1

Theorem 1 cannot be directly applied to the integral equation (II.18) and it is necessary to slightly generalize it by imposing less restrictive conditions on the inhomogeneous term and on the kernel. If we define a new unknown function $\tilde{f}$ by the relation

$$\tilde{f}(x,y)=\rho(x)f(x,y)$$

where $\rho(x)$ is any function of x, $\tilde{f}$ is solution of a modified integral equation

$$\tilde{f}(x,y)=\tilde{g}(x,y)+\lambda\int_0^X dx^*\int_0^y dy^*\,\tilde{K}(x,y;x^*,y^*)\tilde{f}(x^*,y^*), \tag{III.31}$$

where the functions $\tilde{g}$ and $\tilde{K}$ are defined by

$$\tilde{g}(x,y)=\rho(x)\,g(x,y),$$

$$\tilde{K}(x,y;x^*,y^*)=\frac{\rho(x)}{\rho(x^*)}K(x,y;x^*,y^*). \tag{III.32}$$

In the same way, the function

$$\tilde{\tilde{f}}(x,y)=\rho'(x,y)f(x,y)$$

verifies an integral equation with an inhomogeneous term $\tilde{\tilde{g}}$ and a kernel $\tilde{\tilde{K}}$ equal to

$$\tilde{\tilde{g}}(x,y)=\rho'(x,y)\,g(x,y), \tag{III.33}$$

$$\tilde{\tilde{K}}(x,y;x^*,y^*)=\frac{\rho'(x,y)}{\rho'(x^*,y^*)}K(x,y;x^*,y^*). \tag{III.34}$$

The generalization of Theorem 1 is

Theorem 1′: Let us consider the integral equation (III.1). If there exist two functions $\rho(x)$ and $\rho'(x,y)$ such that

$$\rho'(x,y)\leqslant\rho(x)\leqslant 1, \tag{III.35}$$

and such that $\tilde{g}$ belongs to D, $\tilde{K}$ belongs to D^*, $\tilde{\tilde{g}}$ belongs to C_y, and $\tilde{\tilde{K}}$ belongs to C_y^*, then Eq. (III.1) has an unique solution $f(x,y)$ such that $\tilde{f}$ belongs to D and $\tilde{\tilde{f}}$ belongs to C_y.

The only λ dependent singularities of $f(x,y)$ considered as a function of x, y, and λ are an infinite set of fixed poles in λ: $1/(\lambda-\Lambda_{n,i})$ with $n=0,1,2,\cdots$ and $i=1,2,\cdots$. The position of the poles is independent of x and y, independent of the function g, and only depends on the operator $\mathscr{K}$. More precisely, $\Lambda_{n,i}$ $i=1,2,\cdots$, is the set of eigenvalues of the operators defined by the kernel $\tilde{j}^{(n,n)}(x,x^*)$ $=[\rho(x)/\rho(x^*)]\int_0^1 dv\,v^n M(x,0,x^*,v)$.

These kernels April are $\mathscr{L}^2$ kernels and thus have only a discret spectrum. The function $f(x,y)$ can also have singularities in the variable x, independent of λ if g or K ever have such singularities.

Proof: The demonstration of the theorem is a direct generalization of the one of Theorem 1. We are not going to redo all of it but only present the main changes.

(1) The demonstration of Lemma 4 must be done on the set of equations

$$\tilde{f}_n(x)=\tilde{h}_n(x)+\lambda\int_0^X dx^*\,\tilde{j}^{(n,n)}(x,x^*)\,\tilde{f}_n(x^*), \tag{III.36}$$

obtained by multiplying Eq. (III.19) by $\rho(x)$.

(2) In the same way, the equation

$$\tilde{\tilde{\bar{f}}}(x,y)=\tilde{\tilde{\bar{h}}}(x,y)+\lambda\int_0^X dx^*\int_0^y dy^*\left(\frac{y^*}{y}\right)^N\tilde{\tilde{K}}(x,y;x^*,y^*)\times\tilde{\tilde{\bar{f}}}(x^*,y^*), \tag{III.37}$$

with

$$\tilde{\tilde{\bar{f}}}(x,y)=\rho'(x,y)\bar{f}(x,y),$$

$$\bar{\bar{\bar{h}}}(x,y) = \rho'(x,y)\,\bar{h}(x,y),$$

and which has been obtained by multiplying Eq. (III.22) by $\rho'(x,y)$ must be used in Lemmas 5 and 6. The only delicate point comes from the proof that $\bar{\bar{\bar{h}}}(x,y)$ belongs to C_Y. For example, let us consider the first term $\bar{\bar{\bar{g}}}$ of the finite sum which defines $\bar{\bar{\bar{h}}}$ [see (III.23)]. We have

$$\bar{\bar{\bar{g}}}(x,y) = \left(\bar{\bar{g}}(x,y) - \sum_{n=0}^{N-1} \frac{\partial^n g(x,0)}{\partial y^n}\frac{y^n}{n!}\rho'(x,y)\right)\frac{N!}{y^N} \quad \text{if } y>0,$$

$$\bar{\bar{\bar{g}}}(x,0) = \rho'(x,0)\frac{\partial^N g(x,0)}{\partial y^N} \quad \text{if } y=0.$$

Using the hypotheses of Theorem 1′, the continuity of $\bar{\bar{\bar{g}}}(x,y)$ when y goes to zero, Lemma 4, and the inequalities (III.29) and (III.35), it can be proved that $\bar{\bar{\bar{g}}}$ belongs to C_Y.

It must be noted that we have proved that the Neumann series

$$\bar{\bar{\bar{f}}} = \sum_{n=0}^{\infty} (\lambda \bar{\bar{\bar{\mathscr{K}}}}_N)^n \bar{\bar{\bar{h}}}, \tag{III.38}$$

of the integral equation (III.37), is convergent. But, since we have

$$(\bar{\bar{\bar{\mathscr{K}}}}_N)^n \bar{\bar{\bar{h}}}(x,y) = \rho'(x,y)\,(\mathscr{K}_N)^n \bar{h}(x,y), \tag{III.39}$$

the series (III.38) is nothing other than the Neumann series (III.28) of Eq. (III.22) itself which thus is also convergent.

Before ending this part let us make a last remark. All the results proved here are also valid when the functions belong to the spaces C_∞ and C^*_∞ instead of C_Y and C^*_Y for Y finite (see the definitions of C_∞ and C^*_∞ in Sec. III.1).

IV. SOLUTION OF THE INTEGRAL EQUATION VERIFIED BY THE OPEN AMPLITUDE

In this section it is shown that Theorem 1′ can really be applied to the integral equation verified by the open amplitude $F(\sigma,\delta,\gamma)$

$$F(\sigma,\delta,\gamma) = F_1(\sigma,\delta,\gamma) + \lambda\int_0^\infty d\sigma^*\int_0^1 d\delta^*\int_0^\infty d\gamma^* \times J(\sigma,\delta,\gamma;\sigma^*,\delta^*,\gamma^*)\,F(\sigma^*,\delta^*,\gamma^*), \tag{IV.1}$$

with $F_1(\sigma,\delta,\gamma)$ and $J(\sigma,\delta,\gamma;\sigma^*,\delta^*,\gamma^*)$ given by Eqs. (II.23) and (II.19). In order to apply Theorem 1′, it is necessary to begin with verifying that Eq. (IV.1) is of the type (III.1). x stands for the set of variables $\{\sigma,\delta\}$ and X for the set $\{\infty,1\}$; to y corresponds the variable γ, and to $v = y^*/y$ corresponds $\gamma^*/\gamma = (\sigma/\sigma^*)\,u$. The limit of integration is determined by the θ function in Eq. (II.19). Using the inequality

$$U(\sigma,\delta,\sigma^*,\delta^*) \leqslant V(\sigma,\sigma^*) = \inf(1,\sigma^*/\sigma), \tag{IV.2}$$

we find that, when the change of variable (III.5) is performed, the upper limit of integration on $v = (\sigma/\sigma^*)\,u$ is bounded by

$$\frac{\sigma}{\sigma^*}U(\sigma,\delta,\sigma^*,\delta^*) \leqslant \frac{\sigma}{\sigma^*}V(\sigma,\sigma^*) \leqslant 1. \tag{IV.3}$$

Finally we find that to $M(x,y;x^*,v)$ corresponds $(\sigma^*/\sigma)\,L(\sigma,\delta,\gamma;\sigma^*,\delta^*,(\sigma^*/\sigma)\,v)$.

We can now prove the following theorem:

Theorem 2: When the conditions

$$-1 < x, \tag{IV.4}$$

$$t < 4m^2, \tag{IV.5}$$

$$\tfrac14 t - 2m^2(1+\sqrt{1-t/4m^2}) < p_i^2 < 2m^2(1+\sqrt{1-t/4m^2}) \quad i=1,2,3,4, \quad \text{if } 0\leqslant t<4m^2, \tag{IV.6}$$

or

$$-4m^2 < p_i^2 < 4m^2 \quad \text{if } t\leqslant 0,$$

are verified, there exist two functions

$$r(\sigma) = \exp(-\omega m^2\sigma),$$

$$r'(\sigma,\gamma) = \exp(-\omega m^2\sigma - \omega' m^2\sigma\gamma),$$

where ω and ω' are positive numbers which depend on t, p_i^2, and x, and such that the hypotheses of Theorem 1′ are verified. Then $F(\sigma,\delta,\gamma)$ exists for any value of λ; the function

$$\tilde{F}(\sigma,\delta,\gamma) = r(\sigma)\,F(\sigma,\delta,\gamma), \tag{IV.7}$$

belongs to D and the function

$$\tilde{\tilde{F}}(\sigma,\delta,\gamma) = r(\sigma,\gamma)\,F(\sigma,\delta,\gamma), \tag{IV.8}$$

belongs to C_Y. The singularities of $F(\sigma,\delta,\gamma)$ are a set of fixed poles $1/(\lambda - \Lambda_{n,i})$ with $n = 0,1,2,\cdots$ and $i = 1,2,\cdots$. The poles $\Lambda_{n,i}$ $i=1,2,\cdots$ are the eigenvalues of the $\mathscr{L}^2$ kernel

$$\begin{aligned}\tilde{j}^{(n,n)}(\sigma,\delta,\sigma^*,\delta^*) &= \exp[-m^2(1-\omega)\,\sigma^* - m^2\omega\sigma] \\ &\times \int_0^{U(\sigma,\delta,\sigma^*,\delta^*)} du\left(\frac{\sigma}{\sigma^*}\right)^n u^{n+x} \\ &\times \exp\left(-\sigma m^2\frac{u^2}{1-u} + t\delta(1-\delta)(1-u)\right).\end{aligned} \tag{IV.9}$$

Proof:

Definition and properties of the constant P^2: Let us define P^2 by

$$P^2 = \mathrm{Max}[p_i^2,\ (i=1,2,3,4);\ \tfrac14 t - p_i^2,\ (i=1,2,3,4);\ m^2+\tfrac14 t] \quad \text{if } 0\leqslant t<4m^2 \tag{IV.10}$$

or

$$P^2 = \mathrm{Max}[p_i^2,\ (i=1,2,3,4);\ -p_i^2,\ (i=1,2,3,4);\ m^2] \quad \text{if } t\leqslant 0. \tag{IV.10′}$$

Using Eqs. (IV.6), (IV.10), and (IV.10′), we show that P^2 verifies

$$\tfrac14 t - P^2 < p_i^2 < P^2,$$

$$m^2 + \tfrac14 t \leqslant P^2 < 2m^2(1+\sqrt{1-t/4m^2}) \quad \text{if } 0\leqslant t<4m^2,$$

$$m^2 \leqslant P^2 < 4m^2 \quad \text{if } t\leqslant 0. \tag{IV.11}$$

(ii) *Proof that* $\tilde{J}$ *belongs to* D*: To show $\tilde{J}$ belongs to D*, that is to say that for any value of n $|||\tilde{J}|||_n$ is finite, we must prove that for any value of n

$$I = \int_0^{(\sigma/\sigma^*)U} dv\left|\frac{1}{n!}\frac{\partial^n}{\partial\gamma^n}\left(\frac{\sigma^*}{\sigma}\tilde{L}(\sigma,\delta,\gamma=0,\sigma^*,\delta^*,\frac{\sigma^*}{\sigma}v)\right)\right|$$

belongs to $\mathscr{L}^2(\sigma,\delta,\sigma^*,\delta^*)$. Using the bounds (C3) and (C4) verified by the nth derivatives of the reduced kernel $\tilde{L}$, and the inequality (IV.3), we find

$$I<\frac{1}{x+1}\exp[-m^2(1-\omega)\sigma^*-(m^2\omega-\tfrac14 t)\sigma]$$
$$\times\sum_{k=0}^{n}\frac{(\sigma P^2)^k}{k!}C_{n+1}^{k+1}\quad\text{if } 0\leqslant t<4m^2,$$
$$I<\frac{1}{x+1}\exp[-m^2(1-\omega)\sigma^*-m^2\omega\sigma)]$$
$$\times\sum_{k=0}^{n}\frac{[\sigma(P^2-\tfrac14 t)]^k}{k!}C_{n+1}^{k+1}\quad\text{if } t\leqslant 0.$$

These bounds are integrable when σ and σ^* go to infinity if

$$t/4m^2<\omega<1\quad\text{if } 0\leqslant t<4m^2,\qquad\qquad (\text{IV}.12)$$
$$0<\omega<1\quad\text{if } t\leqslant 0.$$

If the condition (IV.12) is verified, we can use inequality (B2) and obtain

$$|||\tilde{J}|||_n<\frac{1}{x+1}\left(\frac{1}{2m^2(1-\omega)}\frac{1}{2(m^2\omega-\tfrac14 t)}\right)^{1/2}\times\left(1+\frac{P^2}{m^2\omega-\tfrac14 t}\right)^{n+1}\quad\text{if } 0\leqslant t<4m^2,$$
$$|||\tilde{J}|||_n<\frac{1}{x+1}\left(\frac{1}{2m^2(1-\omega)}\frac{1}{2m^2\omega}\right)^{1/2}\times\left(1+\frac{P^2-\tfrac14 t}{m^2\omega}\right)^{n+1}\quad\text{if } t\leqslant 0,\qquad (\text{IV}.13)$$

which are finite.

(iii) *Proof that* $\tilde{\tilde{J}}$ *belongs to* C_∞^*: In order to prove that the kernel $\tilde{\tilde{J}}$ belongs to C_∞^* it is enough to show that

$$I'=\int_0^{(\sigma/\sigma^*)U}dv\,\underset{0<\gamma<\infty}{\mathrm{Max}}\left|\tilde{\tilde{L}}\left(\sigma,\delta,\gamma,\sigma^*,\delta^*,\frac{\sigma^*}{\sigma}v\right)\right|$$

belongs to $\mathscr{L}^2(\sigma,\delta,\sigma^*,\delta^*)$.

Using Eqs. (C10), (C4), and (IV.3), we find

$$I'<\frac{1}{x+1}D_0,$$

which is square integrable if the conditions (IV.12) are verified.

(iv) *Proof that* $\widetilde{F}_1$ *belongs to* D: The function $\widetilde{F}_1$ belongs to D if for any value of n, $|(1/n!)\partial^n\widetilde{F}_1(\sigma,\delta,\gamma=0)/\partial\gamma^n|$, is square integrable. The nth derivative of $\widetilde{F}_1$ verifies the bound (D16).

$$\left|\frac{1}{n!}\frac{\partial^n\widetilde{F}_1(\sigma,\delta,\gamma=0)}{\partial\gamma^n}\right|<\overline{C}_n\exp(-\sigma\overline{A}),\qquad(\text{IV}.14)$$

where $\overline{A}$ and $\overline{C}_n$ are given by Eqs. (D9), (D9′), and (D17). A sufficient condition for $|(1/n!)\,\partial^n\widetilde{F}_1(\sigma,\delta,\gamma=0)/\partial\gamma^n|$ to be square integrable is that the coefficient $\overline{A}$ of σ in the exponential of (IV.14) is strictly positive. When $0\leqslant t<4m^2$, the function $2[m^2(P^2-\tfrac14 t)-P^2]^{1/2}$ being a strictly positive function of P^2 for $m^2+(t/4)<P^2<2m^2[1+(1-t/4m^2)^{1/2}]$, it is always possible to choose ω verifying the condition (IV.12) and such that

$$\overline{A}>0.\qquad(\text{IV}.15)$$

For example one can take

$$\omega=\mathrm{Max}\left(\frac{t}{4m^2},\ 1-\frac12\frac{\sqrt{m^2(P^2-\tfrac14 t)}-P^2}{m^2}\right)\quad\text{if } 0\leqslant t<4m^2.\qquad(\text{IV}.16)$$

The case $t\leqslant 0$ is identical to the case $t=0$.

Finally, if the conditions (IV.4) are verified and if we take suitable ω, we find that

$$|||\widetilde{F}_1|||_n<\frac{\overline{C}_n}{\sqrt{2\overline{A}}}.\qquad(\text{IV}.17)$$

(v) *Proof that* $\tilde{\tilde{F}}_1$ *belongs to* C_∞: In order to prove that $\tilde{\tilde{F}}_1$ belongs to C_∞, it is enough to show that

$$\underset{0<\gamma<\infty}{\mathrm{Max}}\ |\tilde{\tilde{F}}_1(\sigma,\delta,\gamma)|$$

is square integrable. It can be shown [see Eq. (D20)] that the previous quantity is bounded by

$$[1/(x+1)]\exp(-\sigma\overline{A}),$$

where $\overline{A}$ is given by Eq. (D9) or (D9′). If ω is taken as before, $\overline{A}$ is positive [Eq. (IV.15)] and

$$\|\tilde{\tilde{F}}_1\|_\infty<\frac{1}{x+1}\frac{1}{\sqrt{2\overline{A}}}<\infty.$$

V. RECONSTRUCTION OF THE PHYSICAL AMPLITUDE

The regular part $\hat{\hat{M}}(x,t)$ of the Mellin transform of the physical amplitude M can be expressed as an integral of the function $F(\sigma,\delta,\gamma)$ [see Eqs. (II.25)]

$$\hat{\hat{M}}(x,t)=\lambda^2\int_0^\infty d\sigma\int_0^1 d\delta\int_0^\infty d\gamma\,\gamma^x\sigma^{x+1}\times\exp[-\sigma(1+\gamma)m^2]\,F(\sigma,\delta,\gamma).\qquad(\text{V}.1)$$

It must be verified that this integral is convergent for any value of λ, including the one for which the perturbative series diverges.

In this section we will use the following notation: $F_m^n(\sigma,\delta)$ is the n^{nth} derivative with respect to γ, taken at $\gamma=0$, of the open amplitude $F_m(\sigma,\delta,\gamma)$ associated to the graph with $(m+1)$ rungs

$$F_m^n(\sigma,\delta)=\frac{\partial^n F_m(\sigma,\delta,\gamma=0)}{\partial\gamma^n}$$

(let us remark that the notation of the derivative differs from the one of Sec. III).

The decomposition (III.13) can be used order by order in the number of rungs. (The constant N depends on the value of the arbitrary constant Λ—see Lemma 5.) We get the following list:

$$F_1(\sigma,\delta,\gamma)=F_1^0(\sigma,\delta)+F_1^1(\sigma,\delta)\gamma+\cdots+F_1^{N-1}(\sigma,\delta)\frac{\gamma^{N-1}}{(N-1)!}+\overline{F}_1(\sigma,\delta,\gamma)\frac{\gamma^N}{N!},$$
$$F_2(\sigma,\delta,\gamma)=F_2^0(\sigma,\delta)+F_2^1(\sigma,\delta)\gamma+\cdots+F_2^{N-1}(\sigma,\delta)\frac{\gamma^{N-1}}{(N-1)!}+\overline{F}_2(\sigma,\delta,\gamma)\frac{\gamma^N}{N!},$$
$$\vdots$$
$$F(\sigma,\delta,\gamma)=F^0(\sigma,\delta)+F^1(\sigma,\delta)\gamma+\cdots+F^{N-1}(\sigma,\delta)\frac{\gamma^{N-1}}{(N-1)!}+\overline{F}(\sigma,\delta,\gamma)\frac{\gamma^N}{N!}.$$

Let us recall the results of the previous sections.

(1) The integral

$$\lambda^2 \int_0^\infty d\sigma \int_0^1 d\delta \int_0^\infty d\gamma\, \sigma^{x+1} \gamma^x \exp[-\sigma(1+\gamma) m^2] \times F_m(\sigma,\delta,\gamma)$$

is finite, as it represents nothing but the regular part of the Mellin transform of the amplitude of the ladder graph with $(m+1)$ rungs, which has no singularity (see Sec. II).

(2) Each function $\widetilde{F}^n(\sigma,\delta) = r(\sigma) F^n(\sigma,\delta)$ is the solution of a Fredholm integral equation (see Theorem 2, Theorem 1', and Lemma 4).

(3) The series $\Sigma_{m=1}^\infty F_m(\sigma,\delta,\gamma)$ converges in the domain

$\lambda < \Lambda_{\min}$

where $\Lambda_{\min}$ is the smallest eigenvalue of the set $\Lambda_{n,i}$, $n = 0,1,2,\cdots$ and $i = 1,2,\cdots$ (see Theorem 2). $\Lambda_{\min}$ is positive as all the F_m's are positive.

In the same domain, the series

$$F^n(\sigma,\delta) = \sum_{m=1}^\infty F_m^n(\sigma,\delta) \quad (n = 0,1,\ldots,N-1),$$

and

$$\overline{F}(\sigma,\delta,\gamma) = \sum_{m=1}^\infty \overline{F}_m(\sigma,\delta,\gamma), \tag{V.3}$$

are also convergent.

(4) The function $\overline{F}(\sigma,\delta,\gamma)$ is solution of an integral equation for which the Neumann series converges in the domain [cf. Lemmas 5 and 6 and Eqs. (III.38) and (III.39)]

$\lambda < \Lambda$,

and consequently $\overline{F}$ is regular in λ in this domain.

Thus the series (V.3) is convergent not only when $\lambda < \Lambda_{\min}$ but in the larger domain $\lambda < \Lambda$.

We will now build $\hat{\hat{M}}(x,t)$ [cf. (V.1)] in two steps. First we define the partially integrated amplitude $f(\sigma)$

$$f(\sigma) = \int_0^1 d\delta \int_0^\infty d\gamma\, \gamma^x \sigma^{x+1} \exp[-\sigma(1+\gamma) m^2] \times F(\sigma,\delta,\gamma), \tag{V.4}$$

and we prove the existence of $f(\sigma)$ in the domain $[0,\infty]$. Then we show that the last integration

$$\hat{\hat{M}}(x,t) = \lambda^2 \int_0^\infty d\sigma f(\sigma), \tag{V.5}$$

is convergent.

One of the difficulties of the demonstration comes from the fact that, due to the γ^n factor, each term of the sum $\Sigma_{n=0}^m F^n(\sigma,\delta)(\gamma^n/n!) + \overline{F}(\sigma,\delta,\gamma)(\gamma^N/N!)$, is not integrable.

In Sec. V.1, bounds are exhibited for the functions $\widetilde{F}^n$ and $\widetilde{F}_m^n$. These bounds are used in Sec. V.2 to prove the existence of $f(\sigma)$ in the domain $]0,\infty]$. Then the existence of $f(0)$ is directly established. Finally integral (V.5) is performed and the existence of $\hat{\hat{M}}$ is proved.

1. Bounds on $\widetilde{F}^n(\sigma,\delta)$ and $\widetilde{F}_m^n(\sigma,\delta)$

Bounds on $\widetilde{F}^n(\sigma,\delta)$

The function $\widetilde{F}^n$ verifies the integral equation

$$\widetilde{F}^n(\sigma,\delta) = \widetilde{H}^n(\sigma,\delta) + \lambda \int_0^\infty d\sigma^* \int_0^1 d\delta^* \tilde{j}^{(n,n)}(\sigma,\delta;\sigma^*,\delta^*) \widetilde{F}^n(\sigma^*,\delta^*), \tag{V.6}$$

with $\tilde{j}^{(n,n)}$ given by Eq. (IV.9) and $\widetilde{H}^n$ by Eq. (III.20) with the needed changes of notations.

To obtain a bound on $\widetilde{F}^n$ we first prove the following theorem:

Theorem 3: Let

$$f(\sigma,\delta) = h(\sigma,\delta) + \lambda \int_0^\infty d\sigma^* \int_0^1 d\delta^*\, k(\sigma,\delta;\sigma^*,\delta^*) f(\sigma^*,\delta^*), \tag{V.7}$$

be a Fredholm integral equation.

If the kernel k and the first term h verify the bounds

$$|k(\sigma,\delta;\sigma^*,\delta^*)| < c \exp(-a\sigma - b\sigma^*), \tag{V.8}$$

$$|h(\sigma,\delta)| < P_l(\sigma) \exp(-a\sigma), \tag{V.8'}$$

where $P_l(\sigma)$ is a polynomial of degree l in σ, and where a, b, and c are three positive constants, then for any value of λ, the solution $f(\sigma,\delta)$ verifies the inequality

$$|f(\sigma,\delta)| < \left(P_l(\sigma) + \frac{\mathscr{N}(\lambda)}{\mathscr{D}(\lambda)}\right) \exp(-a\sigma), \tag{V.9}$$

where $\mathscr{N}(\lambda)$ and $\mathscr{D}(\lambda)$ are two entire functions of λ, $\mathscr{D}(\lambda)$ being the Fredholm determinant.

Proof: In the following, $P_n(\sigma)$ means any polynomial of degree n in σ. The solution of a Fredholm integral equation is given by the formula

$$f(\sigma,\delta) = h(\sigma,\delta) - \lambda \int_0^\infty d\sigma^* \int_0^1 d\delta^*\, H(\sigma,\delta;\sigma^*,\delta^*,\lambda)\, h(\sigma^*,\delta^*). \tag{V.10}$$

The resolvent kernel H is the ratio of two entire functions of λ

$$H(\sigma,\delta;\sigma^*,\delta^*,\lambda) = \frac{\mathscr{D}(\sigma,\delta;\sigma^*,\delta^*,\lambda)}{\mathscr{D}(\lambda)} = \frac{\Sigma_\infty^{m=0} d_m(\sigma,\delta;\sigma^*,\delta^*,\lambda)}{\mathscr{D}(\lambda)}, \tag{V.11}$$

where $\mathscr{D}(\lambda)$ is the Fredholm determinant and

$$d_m(\sigma,\delta;\sigma^*,\delta^*,\lambda) = -\frac{(-\lambda)^m}{m!} \int\int\cdots\int K\begin{pmatrix} x, & \xi_1,\ldots,\xi_m \\ x^*, & \xi_1,\ldots,\xi_m \end{pmatrix} d\xi_1\cdots d\xi_m,$$

with

$$K\begin{pmatrix} x_1,\ldots,x_n \\ x_1^*,\ldots,x_n^* \end{pmatrix} = \begin{vmatrix} k(x_1,x_1^*), k(x_2,x_1^*),\ldots, k(x_n,x_1^*) \\ \cdot \qquad\qquad\qquad\qquad \cdot \\ \cdot \qquad\qquad\qquad\qquad \cdot \\ \cdot \qquad\qquad\qquad\qquad \cdot \\ \cdot \qquad\qquad\qquad\qquad \cdot \\ \cdot \qquad\qquad\qquad\qquad \cdot \\ k(x_1,x_n^*) \quad ,\ldots, \quad kk(x_n,x_n^*) \end{vmatrix}.$$

In the two last equations, x means the set of variables $\{\sigma,\delta\}$. The kernel k can be written

$$k(\sigma,\delta;\sigma^*,\delta^*)=\exp(-a\sigma-b\sigma^*)\,k_c(\sigma,\delta;\sigma^*,\delta^*), \tag{V.12}$$

where k_c is a bounded kernel

$$k_c(\sigma,\delta;\sigma^*,\delta^*)<c\,. \tag{V.13}$$

Using the expression (V.12) of k, we have

$$K\begin{pmatrix}x_1,\dots,x_n\\ x_1^*,\dots,x_n^*\end{pmatrix}=\exp\Big(-a\sum_{i=1}^n\sigma_i-b\sum_{i=1}^n\sigma_i^*\Big)\times\begin{vmatrix}k_c(x_1,x_1^*),\dots,k_c(x_n,x_1^*)\\ k_c(x_1,x_n^*),\dots,k_c(x_n,x_n^*)\end{vmatrix},$$

and using the bound (V.13) on k_c and the Hadamard's theorem,[8] we obtain

$$K\begin{pmatrix}x_1,\dots,x_n\\ x_1^*,\dots,x_n^*\end{pmatrix}<\exp\Big(-a\sum_{i=1}^n\sigma_i-b\sum_{i=1}^n\sigma_i^*\Big)c^n\,n^{n/2}\,.$$

Finally as $a+b$ is positive we see that d_m is bounded by

$$|d_m(\sigma,\delta;\sigma^*,\delta^*,\lambda)|<\exp(-a\sigma-b\sigma^*)\,c\times\frac{(\lambda c/(a+b))^m\,(m+1)^{(m+1)/2}}{m!}, \tag{V.14}$$

and thus

$$|\mathscr{D}(\sigma,\delta;\sigma^*,\delta^*,\lambda)|<\exp(-a\sigma-b\sigma^*)\mathscr{N}'(\lambda), \tag{V.15}$$

where $\mathscr{N}'(\lambda)$ is an entire function of λ.

To obtain the bound on $f(\sigma,\delta)$ it remains to put the inequalities (V.8′) and (V.15) in Eqs. (V.11) and (V.10) and we find that

$$\mathscr{N}(\lambda)=c'\mathscr{N}'(\lambda), \tag{V.16}$$

where c' is a constant.

In order to apply this theorem to the functions $\widetilde{F}^n(\sigma,\delta)$, we are going to prove that the kernels $\widetilde{j}^{(n,n)}$ and the first term $\widetilde{H}^n(\sigma,\delta)$ admit the bounds (V.8) and (V.8′). First, using the inequalitites (C7) and (D10) or (D10′) we find

$$\left|\frac{1}{(n-l)!}j^{(n,l)}(\sigma,\delta;\sigma^*,\delta^*)\right|<P_{n-l}(\sigma)\exp[-m^2(1-\omega)\sigma^*-\sigma\bar{A}]\,. \tag{V.17}$$

Then using the expression (III.20) for $\widetilde{H}^n$ and (D16), and applying successively the previous theorem for $n=0,1,\dots,N-1$, we prove by recurrence that

$$|\widetilde{H}^n(\sigma,\delta)|<P_n(\sigma)\exp(-\sigma\bar{A}), \tag{V.18}$$

$$|\widetilde{F}^n(\sigma,\delta)|<P_n(\sigma)\exp(-\sigma\bar{A})\,. \tag{V.19}$$

Bounds on $F_m^n(\sigma,\delta)$

We are going to prove by recurrence on m that

$$|\widetilde{F}_m^n(\sigma,\delta)|<P_n(\sigma)\exp(-\sigma\bar{A})\,. \tag{V.20}$$

It is true for $m=1$ [see the bounds (D16)].

If we decompose the relation

$$\widetilde{F}_{m+1}^n(\sigma,\delta,\gamma)=\lambda\int_0^\infty d\sigma^*\int_0^1 d\delta^*\times\int_0^U du\,L(\sigma,\delta,\gamma;\sigma^*\,\delta^*,u)\,\widetilde{F}_m^n\Big(\sigma^*,\delta^*,\frac{\sigma}{\sigma^*}\gamma u\Big),$$

in powers of γ and we identify in the two members the term with the same power, we obtain

$$\widetilde{F}_{m+1}^n(\sigma,\delta)=\lambda\sum_{l=0}^n C_n^l\int_0^\infty d\sigma^*\times\int_0^1 d\delta^*\,j^{(n,l)}(\sigma,\delta;\sigma^*,\delta^*)\,\widetilde{F}_m^l(\sigma^*,\delta^*)\,. \tag{V.21}$$

Using the bound (C7), and assuming that the functions $\widetilde{F}_m^l$ verify the inequalities (V.20), we find that $\widetilde{F}_{m+1}^n$ also verify the same inequalities.

2. Existence of $f(\sigma)$ and of $\bar{M}$

It is now quite easy to prove the existence of $f(\sigma)$ in the domain $]0,\infty]$. Let us consider the last line of Eq. (V.2). Using bound (V.19) and Eq. (V.4), we find that the contribution to $f(\sigma)$ of

$$F^n(\sigma,\delta)\frac{\gamma^n}{n!}=\frac{1}{r(\sigma)}\widetilde{F}^n(\sigma,\delta)\frac{\gamma^n}{n!}$$

is bounded by

$$\frac{1}{n!}\frac{\Gamma(x+n+1)}{m^{2(x+n+1)}}\frac{1}{\sigma^n}\exp[-\sigma(m^2(1-\omega)+\bar{A})]\,P_n(\sigma), \tag{V.22}$$

which is finite for $\sigma>0$.

Using bound (V.20), the partial integral (V.4) on δ and γ can be performed for the function $F_m^n(\sigma,\delta)\,\gamma^n/n!$ and it is bounded by the same expression as (V.22).

The functions $\widetilde{F}_m(\sigma,\delta,\gamma)\,\gamma^N/N!$ being finite sums of partially integrable terms are themselves partially integrable in γ and δ.

The functions $\widetilde{F}(\sigma,\delta,\gamma)\,\gamma^N/N!$ being the sum of the absolutely convergent series (V.3) for $\lambda<\Lambda$, is also partially integrable in γ and δ in the same domain.

This achieves the proof of the existence of $f(\sigma)$ for $\sigma>0$.

When σ goes to zero, the limit of $f(\sigma)$ is

$$\lim_{\sigma\to0}[f(\sigma)]=\lim_{\sigma\to0}\Big(\int_0^1 d\delta\int_0^\infty d\gamma\,\sigma^{x+1}\gamma^x\exp(-\sigma\gamma m^2)\,F(\sigma=0,\delta,\gamma)\Big)\,. \tag{V.23}$$

The value of the function F for $\sigma=0$ can be obtained using the integral equation (IV.1). We have

$$F(\sigma=0,\delta,\gamma)=\frac{1}{1+\gamma}\frac{1}{x+1}\times\Big(1+\int_0^\infty d\sigma^*\int_0^1 d\delta^*\exp(-m^2\sigma^*)\,F(\sigma^*,\delta^*,0)\Big)\,.$$

The integral in σ^* and δ^* is convergent because we already know that the integral equation has a solution. We have thus obtained the bound

$$\widetilde{F}(\sigma=0,\delta,\gamma)<\text{cst}. \tag{V.24}$$

Finally, if we combine (V.23) and (V.24) we find that $\lim_{\sigma\to0}[f(\sigma)]$ exists and is finite.

Since $f(\sigma)$ is always bounded when $0\leqslant\sigma\leqslant\infty$, the only point we have to examine is the convergence of the integral (V.5) when σ goes to infinity. The bound (V.22) of the contribution of $F^n(\sigma,\delta)\,\gamma^n/n!$ to $f(\sigma)$ is integrable when σ goes to infinity. It is the same situation for the function $F^n_m(\sigma,\delta)\,\gamma^n/n!$, and since the integral (V.1) exists for $F_m(\sigma,\delta,\gamma)$, the contribution of $\overline{F}_m(\sigma,\delta,\gamma)\,\gamma^N/N$ to $f(\sigma)$ is integrable when σ goes to infinity. Thus the contribution of $\overline{F}(\sigma,\delta,\gamma)\,\gamma^N/N!$ is also integrable when σ goes to infinity. The function $f(\sigma)$ being a finite sum of integrable functions when σ goes to infinity, is itself integrable when σ goes to infinity and the integral (V.5) which defines $\hat{M}$ exists.

CONCLUSION

We summarize here our steps and their main results. We want to obtain information on the hadronic amplitude from Lagrangian field theory. The study is carried out here for the interacting Lagrangian $g\varphi^3$, and a drastic approximation is done by retaining only the ladder graph series. The method used to obtain the sum of the perturbative series is based on a nonperturbative property of this sum: the fact that it is the solution of an integral equation. Contrary to perturbative method, which could give information only inside the convergence circle, we can build the amplitude in the whole plane of the coupling constant and the position of the first singularity does not play any crucial part; in that sense, it is legitimate to hope that the ladder approximation, for which the number of graphs contributing at each order in the squared coupling constant stays equal to one and thus the convergence radius is finite, is technical and not basic.

The method proceeds through three stages:

(1) First, we show that the "regular part" of the Mellin transform of the "open amplitude" [see (Eq. 2.9)] is the solution of an integral equation. This amplitude being a real quantity turns out to be a very good and simple tool. Integral equations of the same type can also be written when the dimension of the space is not equal to 4 or when there are different masses on vertical and horizontal lines. It must be noted that the equation is written in terms of the Mandelstam invariants and not in terms of the momenta as for Bethe–Salpeter equation, and thus the solution is manifestly Lorentz invariant.

(2) Though the integral equation is singular, we prove the existence and unicity of its solution and show explicitly its singularity structure. Our fundamental result is that for each given value of the squared coupling constant λ, the solution can be written as a finite sum of solutions F_n of Fredholm equations plus a function F which is the sum of an convergent series in λ

$$F(\sigma,\delta,\gamma) = \sum_{n=0}^{N-1} F_n(\sigma,\delta)\,\frac{\gamma^n}{n!} + \overline{F}(\sigma,\delta,\gamma)\,\frac{\gamma^N}{N!},$$

where N depends on λ: $N = N(\lambda)$. The function $D(x,t,\lambda)$ [see Introduction, Eq. (16)] can be given explicitly

$$D(x,t,\lambda) = \prod_{n=0}^{N-1} D_n(x,t,\lambda),$$

where $D_n(x,t,\lambda)$ is the Fredholm determinant of the operator $\tilde{j}^{(n,n)}$ of Eq. (IV.9).

It can be easily shown that all the demonstrations can be performed in arbitrary dimension d, provided d is strictly less than 6. This reflects the well-known fact that at $d = 6$, the sum of the φ^3 ladder graphs possesses fixed cuts instead of Regge poles.[9]

(3) Finally we integrated the open amplitude to obtain the amplitude M itself. The singularities of M are the same as the ones of $F(\sigma,\delta,\gamma)$ and are the zeros of the equation $D(x,t,\lambda) = 0$, which provides us with a set of moving poles in the x plane

$$x = X_{n,i}(t,\lambda) \quad n = 0,1,2,\cdots; \quad i = 1,2,\cdots.$$

The expression of D gives us a natural classification: The first index refers to the Fredholm kernel $\tilde{j}^{(n,n)}$ and the second index is the number of the eigenvalues of this kernel.

In the present paper we have given only the first results of a wider work.[10] In a future publication the dominant trajectory will be extensively studied: limits $g\to 0$, $g\to\infty$, lower and upper bounds, trace approximation; then results will be established, such as multiplicity of daughter trajectories. A step of first importance will also be to get rid of the ladder restriction and to write an integral equation verified by the sum of all the planar graphs.

APPENDIX A: RECURRENCE RELATIONS VERIFIED BY P_n and A^q_n/P_n for $q = s, t, 1, 2, 3,$ or 4

We recall the following definitions:

(1) A tree is a connected graph with no closed loop.

(2) A cut is a set of lines of a connected graph such that this graph falls into two connected subgraphs when they are removed, and such that none of its subsets possess the same property.

(3) A s-cut is a cut such that vertices 1 and 2 are attached to one subgraph, vertices 3 and 4 to the other one (see Fig. 6). The corresponding definition for a t-cut is straightforward. Similarly, an i-cut ($i = 1,2,3,4$) is a cut for which the vertex i is isolated in one subgraph, the three other vertices being attached to the other subgraph.

Let us define

$$\begin{aligned} \beta^*_n &= \beta_n\,\beta_{n+1}/c_{n+1}, \\ \alpha^*_n &= \alpha_n + \alpha_{n+1}\,\beta_n/c_{n+1}, \\ \alpha'^*_n &= \alpha'_n + \alpha'_{n+1}\beta_n/c_{n+1}, \end{aligned} \tag{A1}$$

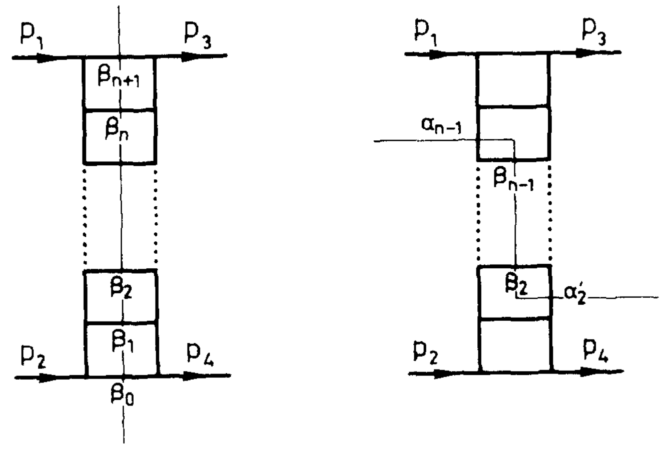


FIG. 6. (a) The one s-cut for a ladder graph. (b) An example of t-cut.

where $c_{n+1} = \beta_{n+1} + \alpha_{n+1} + \alpha'_{n+1} + \beta_n$.

Let f be a function of $(3n+1)$ scalar variables. Then we will denote

$$f = f(\beta_n, \alpha_n, \alpha'_n, \beta_{n-1}, \alpha_{n-1}, \ldots, \beta_0),$$
$$f^* = f(\beta_n^*, \alpha_n^*, \alpha_n'^*, \beta_{n-1}, \alpha_{n-1}, \ldots, \beta_0), \tag{A2}$$

and we define the recurrence operation r as

$$r(f) = c_{n+1} f^*. \tag{A3}$$

We will denote G_n the ladder graph with $(n+1)$ rungs (n loops).

1. The recurrence for P_n

The polynomial P_n is defined by

$$P_n = \sum_{\text{tree}} \Big(\prod \alpha_j\Big), \tag{A4}$$

where Σ_{tree} means the sum over all the connected subgraphs with no closed loop (the trees) (see Fig. 7). $(\Pi\alpha_j)$ is the product of the Schwinger parameters α_j of all the lines which do not belong to the tree.

P_n is a homogeneous polynomial of degree n, and of first degree in each variable.

P_n can be written

$$P_n = a\beta_n + b, \tag{A5}$$

where a and b do not depend on β_n. The coefficient of β_n is nothing else than P_{n-1}

$$a = P_{n-1}. \tag{A6}$$

In order to express P_{n+1} in functions of P_n, we must know how all the trees of G_{n+1} can be constructed from the trees of G_n. If in the tree of G_n, the upper loop $(\beta_{n+1}, \alpha_n, \alpha'_n, \beta_n)$ has been opened by cutting the β_n line, the upper loop of G_{n+1} can be opened only by cutting one of the lines α_{n+1}, α'_{n+1}, or β_{n+1}. If in the tree of G_n, the upper loop has been opened by cutting one of the lines α_n, α'_n, or β_{n-1}, the upper loop of G_{n+1} can be opened by cutting α_{n+1}, α'_{n+1}, β_{n+1}, or β_n. In other words, if P_n is given by Eq. (A5) then

$$P_{n+1} = a\beta_n(\alpha_{n+1} + \alpha'_{n+1} + \beta_{n+1}) + b(\alpha_{n+1} + \alpha'_{n+1} + \beta_{n+1} + \beta_n). \tag{A7}$$

In fact P_n depends on the variables $\{\beta_n, \alpha_n, \alpha'_n\}$ only through their sum

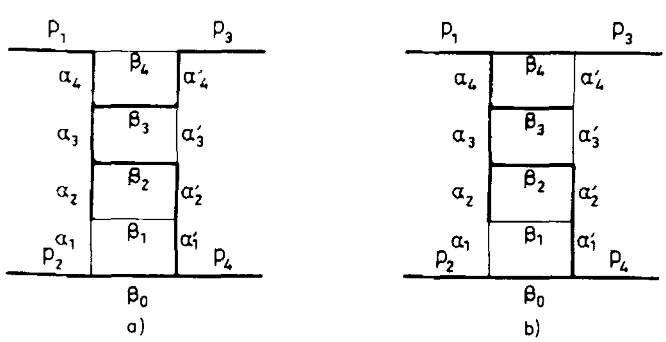


FIG. 7. Two examples of contribution to the polynomials P_n. The first one is equal to $\alpha_1\beta_1\alpha'_3\beta_4$ and contributes to the term $a\beta_n$ of Eq. (A5). The second one is equal to $\alpha_1\beta_1\alpha'_3\alpha'_4$ and contributes to the term b.

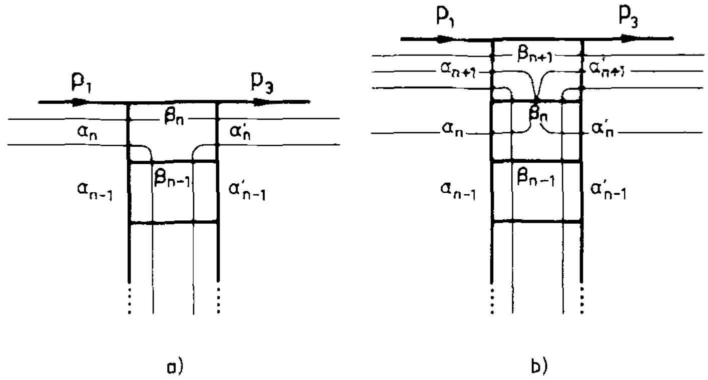


FIG. 8. Building A^t_{n+1}. (a) The three t-cut destroying the upper loop of the graph with n loops. (b) How to destroy the supplementary loop when building A^t_{n+1}.

$$P_n = a(\alpha_n + \alpha'_n + \beta_n) + b', \tag{A8}$$

where a and b' do not depend on α_n, α'_n, or β_n.

Then

$$\begin{aligned} P_{n+1} &= a\beta_n(\alpha_{n+1} + \alpha'_{n+1} + \beta_{n+1}) \\ &\quad + [a(\alpha_n + \alpha'_n) + b'](\alpha_{n+1} + \alpha'_{n+1} + \beta_{n+1} + \beta_n) \\ &= a(\alpha_n^* + \alpha_n'^* + \beta_n^*)c_{n+1} + b'c_{n+1}, \end{aligned} \tag{A9}$$

or $P_{n+1} = r(P_n)$

Using Eqs. (A7) and (A6) we find

$$P_{n+1} = c_{n+1}P_n - \beta_n^2 P_{n-1}. \tag{A10}$$

2. The recurrence relation for A^s_n/P_n

The polynomials A^s, A^t, and A^i $(i = 1,2,3,4)$ are defined by

$$A^q = \sum_{q\text{-cut}} P_1 P_2 \prod_{\{j\}} \alpha_j, \tag{A11}$$

where q stands for s, t, or i $(i = 1,2,3,4)$, $\{j\}$ labels the set of lines consituting the cut and where P_1 and P_2 are the polynomials P attached to each remaining subgraph.

For a ladder graph there is only one s-cut, which, following the notation of Fig. (3), is the product over all the β. The remaining subgraphs being trees, the polynomials P_1 and P_2 of Eq. (A11) are equal to 1. Then

$$A^s_n = \prod_{i=0}^{n} \beta_i, \tag{A12}$$

from which we deduce

$$A^s_{n+1} = r(A^s_n), \tag{A13}$$

and

$$\frac{A^s_{n+1}}{P_{n+1}} = \left(\frac{A^s_n}{P_n}\right)^*. \tag{A14}$$

3. The recurrence relation for A^t_n/P_n

Let we write A^t_n under the form (see Fig. 8)

$$A^t_n = \alpha_n\alpha'_n P_{n-1} + \sum_{i=1}^{n-1} \alpha_n \Big(\prod_{j=i}^{n-1} \beta_j\Big) \alpha'_i P_{i-1} + \sum_{i=1}^{n-1} \alpha'_n \Big(\prod_{j=i}^{n-1} \beta_j\Big) \alpha_i P_{i-1} + A''_n \tag{A15}$$

where $A_n^{\prime t}$ corresponds to all the terms of A_n^t for which the t-cut does not destroy the last loop. The dependence of $A_n^{\prime t}$ on the variable $(\alpha_n, \alpha_n', \beta_n)$ is through a polynomial $P_{\rm up}$ attached to the upper subgraph.

When we want to build A_{n+1}^t from A_n^t, we can group the different obtained terms in four classes [see Fig. 8(b)].

(i) The cut remains unchanged. Then the only change is in the polynomial corresponding to the upper subgraph: $P_{\rm up}$ becomes $r(P_{\rm up})$.

When we modify the cuts, we obtain four supplementary cuts:

(ii) The cut $\alpha_n \alpha_n'$ may be modified in two ways

$\alpha_n \alpha_n' \rightarrow (\alpha_{n+1} \alpha_n' + \alpha_n \alpha_{n+1}') \beta_n$.

(iii) The cuts $\alpha_n \beta_{n-1} \cdots$ and $\alpha_n' \beta_{n-1} \cdots$ become respecively $\alpha_{n+1} \beta_n \beta_{n-1} \cdots$ and $\alpha_{n+1}' \beta_n \beta_{n-1} \cdots$.

(iv) Finally we have to add a new term: $\alpha_{n+1} \alpha_{n+1}' P_n$.

We may try *a priori* the same recurrence as for A_n^s and P_n. Using (A9), we write easily

$$
\begin{aligned}
r(A_n^t)
&= c_{n+1} \left(\alpha_n + \alpha_{n+1} \frac{\beta_n}{c_{n+1}}\right)\left(\alpha_n' + \alpha_{n+1}' \frac{\beta_n}{c_{n+1}}\right) P_{n-1} \\
&+ \sum_{i=1}^{n-1} c_{n+1} \left(\alpha_n + \alpha_{n+1} \frac{\beta_n}{c_{n+1}}\right)\left(\prod_{j=i}^{n-1} \beta_j\right) \alpha_i' P_{i-1} \\
&+ \sum_{i=1}^{n-1} c_{n+1} \left(\alpha_n' + \alpha_{n+1}' \frac{\beta_n}{c_{n+1}}\right)\left(\prod_{j=i}^{n-1} \beta_j\right) \alpha_i P_{i-1} \\
&+ c_{n+1} A_n^{\prime t *}.
\end{aligned}
\tag{A16}
$$

Comparing (A16) with the four classes of terms of A_{n+1}^t, we see that $r(A_n^t)$ possesses an extra term $[\alpha_{n+1} \alpha_{n+1}' (\beta_n^2 / c_{n+1}) P_{n-1}]$, and on the other hand that the term $\alpha_{n+1} \alpha_{n+1}' P_n$ is missing.

We have then obtained

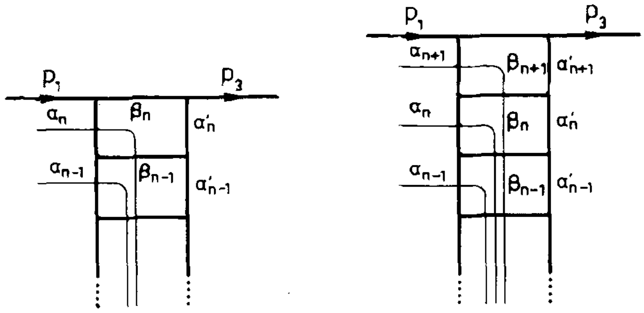
$$
A_{n+1}^t = r(A_n^t) + \alpha_{n+1} \alpha_{n+1}' [P_n - (\beta_n^2 / c_{n+1}) P_{n-1}],
$$

which we rewrite, using (A10),

$$
A_{n+1}^t = \frac{\alpha_{n+1} \alpha_{n+1}'}{c_{n+1}} P_{n+1} + r(A_n^t),
\tag{A17}
$$

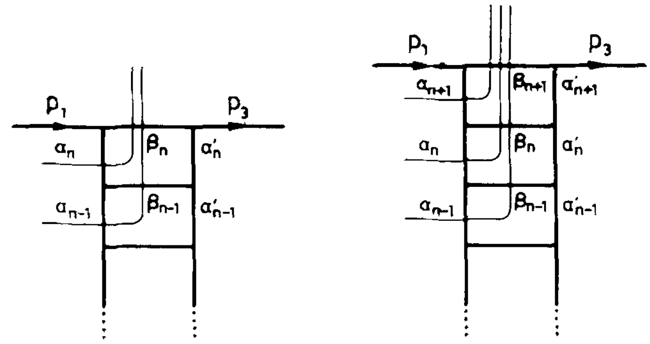
or

$$
\frac{A_{n+1}^t}{P_{n+1}} = \frac{\alpha_{n+1} \alpha_{n+1}'}{c_{n+1}} + \left(\frac{A_n^t}{P_n}\right)^*.
\tag{A18}
$$



a) b)

FIG. 9. Building A_{n+1}^2 from A_n^2. The thin lines are the materialization of the 2-cuts.



a) b)

FIG. 10. Building A_{n+1}^1 from A_n^1. The thin lines are the materialization of the 1-cuts.

4. The recurrence relation for A_n^2/P_n and A_n^4/P_n

Clearly the polynomials A_n^2 and A_n^4 will obey the same recurrence. The expression of A_n^2 is [see Fig. 9(a)]

$$
A_n^2 = \sum_{i=0}^{n-1} \alpha_{i+1} \left(\prod_{j=0}^{i} \beta_j\right) \times P_{n-i-1}(\beta_n, \alpha_n, \alpha_n', \ldots, \alpha_{i+1}, \alpha_{i+1}', \beta_{i+1}).
\tag{A19}
$$

A_{n+1}^2 is built from A_n^2 in the following way [see Fig. 9(b)]:

(i) We keep the cut: the only change is then in the polynomial

$P_{n-i-1} \rightarrow r(P_{n-i-1})$.

(ii) We have an extra cut, for which the polynomial is 1: $\alpha_{n+1} \Pi_{i=0}^n \beta_n$. From (A19), we write $r(A_n^2)$

$$
\begin{aligned}
r(A_n^2)
&= c_{n+1} \alpha_n^* \left(\prod_{j=0}^{n-1} \beta_j\right) + \sum_{i=0}^{n-2} c_{n+1} \alpha_{i+1} \left(\prod_{j=0}^{i} \beta_j\right) P_{n-i-1}^* \\
&= \alpha_{n+1} \prod_{j=0}^{n} \beta_j + \sum_{i=0}^{n-1} \alpha_{i+1} \left(\prod_{j=0}^{i} \beta_j\right) r(P_{n-i-1}),
\end{aligned}
\tag{A20}
$$

i.e.,

$$
r(A_n^2) = A_{n+1}^2,
$$

and

$$
\frac{A_{n+1}^2}{P_{n+1}} = \left(\frac{A_n^2}{P_n}\right)^*.
\tag{A21}
$$

Similarly we have

$$
\frac{A_{n+1}^4}{P_{n+1}} = \left(\frac{A_n^4}{P_n}\right)^*.
\tag{A22}
$$

5. The recurrence relations for A_n^1/P_n and A_n^3/P_n

Let us finally turn to A_n^1 and A_n^3

$$
A_n^1 = \sum_{i=1}^{n} \alpha_i \left(\prod_{j=1}^{n} \beta_j\right) P_{i-1}(\beta_{i-1}, \alpha_{i-1}, \alpha_{i-1}', \ldots, \beta_0).
\tag{A23}
$$

It is straightforward to build A_{n+1}^1 from A_n^1 (see Fig. 10). (i) The product over the β's has to be taken up to $n+1$. (ii) There is an extra term: $\alpha_{n+1} \beta_{n+1} P_n$.

Using (A23), we obtain

$r(A_n^1)$

$$= c_{n+1}\alpha_n^* \beta_n^* P_{n-1} + \sum_{i=1}^{n-1} c_{n+1}\alpha_i \left(\prod_{j=1}^{n-1}\beta_j\right)\beta_n^* P_{i-1}$$

$$= \beta_{n+1} A_n^1 + \alpha_{n+1}\beta_{n+1}(\beta_n^2/c_{n+1})P_{n-1},$$

and so

$$A_{n+1}^1 = r(A_n^1) + \alpha_{n+1}\beta_{n+1}[P_n - (\beta_n^2/c_{n+1})P_{n-1}]$$
$$= r(A_n^1) + \alpha_{n+1}\beta_{n+1} P_{n+1}/c_{n+1}, \tag{A24}$$

or

$$\frac{A_{n+1}^1}{P_{n+1}} = \frac{\alpha_{n+1}\beta_{n+1}}{c_{n+1}} + \left(\frac{A_n^1}{P_n}\right)^*. \tag{A25}$$

A_n^3 follows, of course, the same law as A_n^1

$$\frac{A_n^3}{P_{n+1}} = \frac{\alpha'_{n+1}\beta_{n+1}}{c_{n+1}} + \left(\frac{A_n^3}{P_n}\right)^*. \tag{A26}$$

APPENDIX B

In this appendix we give a series of elementary relations which will be useful elsewhere.

(1) If a is a positive constant, we have

$$\int_0^\infty \exp(-a\sigma)\left(\sum_{k=0}^n C_n^k \frac{(\sigma b)^k}{k!}\right)d\sigma = \frac{1}{a}\left(1+\frac{b}{a}\right)^n, \tag{B1}$$

and

$$\int_0^\infty \exp(-a\sigma)\left(\sum_{k=0}^n C_n^k \frac{(\sigma b_1)^k}{k!}\right)\left(\sum_{k=0}^n C_n^k \frac{(\sigma b_2)^k}{k!}\right)$$
$$< \frac{1}{a}\left(1+\frac{|b_1|+|b_2|}{a}\right)^{2n}. \tag{B2}$$

Let g_1 be the function

$$g_1(x) = \frac{\exp[-ax - bx/(c+x)]}{(1+x/c)^\alpha}, \quad \alpha>1, \quad c>0.$$

The expansion at $(x=0)$ provides us with the inequality

$$\left|\frac{1}{n!}\frac{\partial^n g_1(x=0)}{\partial x^n}\right|$$
$$< \left(\frac{\alpha}{c}\right)^n \sum_{k=0}^n \left(\frac{|ac|+|b|}{\alpha}\right)^k \frac{1}{k!} C_n^k. \tag{B3}$$

(2) Let g_2 be the function

$$g_2(x) = \exp[-(a_1+a_2)x - bx/(1+x)].$$

We have

$$\left|\frac{1}{n!}\frac{\partial^n g_2(x=0)}{\partial x^n}\right|$$
$$< \exp(|a_1|) \sum_{k=0}^n (|a_2|+|b|)^k \frac{1}{k!} C_n^k. \tag{B4}$$

(3) Let g_3 be the function

$$g_3(x) = \frac{1}{x+1}\exp\left(-(a_1+a_2)x - b\frac{x}{x+1}\right).$$

We have

$$\left|\frac{1}{n!}\frac{\partial^n g_3(x=0)}{\partial x^n}\right|$$
$$< \exp(|a_1|) \sum_{k=0}^n (|a_2|+|b|)^k \frac{1}{k!} C_{n+1}^{k+1}. \tag{B5}$$

(4) *Let* g_4 be the function

$$g_4(x) = \exp[-ax - bx/(c+x)]$$

In the case

$$a>0$$
$$b>-ac,$$

we have

$$\operatorname*{Max}_{0\leqslant x<\infty} [g_4(x)] = g_4(0) = 1. \tag{B6}$$

(5) Finally, we have

$$\operatorname*{Max}_{X>\overline{X}} [\exp(-X)X^k] \leqslant \exp(-\overline{X})k^k, \quad \text{if } \overline{X}>0. \tag{B7}$$

APPENDIX C: BOUNDS ON THE KERNEL $L(\sigma,\delta,\gamma,\sigma^*,\delta^*,u)$

1. Bounds at $\gamma=0$

The kernel $\widetilde{L}(\sigma,\delta,\gamma;\sigma^*,\delta^*,u) = [r(\sigma)/r(\sigma^*)]L(\sigma,\delta,\gamma,\sigma^*,\delta^*,u)$ can be written

$$\widetilde{L}(\sigma,\delta,\gamma,\sigma^*,\delta^*,u) = c\frac{1}{\gamma+1}\exp\left(-a\gamma - b\frac{\gamma}{1+\gamma}\right), \tag{C1}$$

with

$$\begin{aligned}
c &= u^x \exp[-m^2(1-\omega)\sigma^* - m^2\sigma u^2/(1-u) \\
&\quad + \sigma t\delta(1-\delta)(1-u) - m^2\omega\sigma],\\
b &= \sigma[t\delta(1-\delta) - p_1^2\delta - p_3^2(1-\delta)](1-u),\\
a &= a_1 + a_2,\\
a_1 &= m^2\sigma u^2/(1-u),\\
a_2 &= m^2\sigma u.
\end{aligned} \tag{C1$'$}$$

Using (B5) we have

$$\left|\frac{1}{n!}\frac{\partial^n \widetilde{L}(\sigma,\delta,\gamma=0;\sigma^*,\delta^*,u)}{\partial\gamma^n}\right|$$
$$< ce^{a_1-1}\sum_{k=0}^n \frac{(a_2+|b|)^k}{k!} C_{n+1}^{k+1}.$$

Since $0\leqslant\delta\leqslant1$ and $\frac{1}{4}t - P^2 \leqslant p_i^2 < P^2$ for $i=1$ and 3, we find that

$$\begin{aligned}
|b| &< \sigma(1-u)P^2 \quad \text{if } t\geqslant0,\\
|b| &< \sigma(1-u)(P^2 - \tfrac{1}{4}t) \quad \text{if } t\leqslant0,
\end{aligned} \tag{C2}$$

and finally

$$\left|\frac{1}{n!}\frac{\partial^n \widetilde{L}(\sigma,\delta,\gamma=0;\sigma^*,\gamma^*,u)}{\partial\gamma^n}\right| < u^x D_n, \tag{C3}$$

with

$$D_n = \exp[-m^2(1-\omega)\sigma^* - (m^2\omega - \tfrac{1}{4}t)\sigma]$$
$$\times \sum_{k=0}^n \frac{(\sigma P^2)^k}{k!} C_{n+1}^{k+1} \quad \text{if } t\geqslant0, \tag{C4}$$

$$D_n = \exp[-m^2(1-\omega)\sigma^* - (m^2\omega - \tfrac{1}{4}t)\sigma]$$
$$\times \sum_{k=0}^n \frac{(\sigma P^2)^k}{k!} C_{n+1}^{k+1} \quad \text{if } t\geqslant0, \tag{C4}$$

If, in the case $n=0$, we use (B4) instead of (B5), we obtain

$$\widetilde{L}(\sigma,\delta,\gamma;\sigma^*,\delta^*,u) < u^x D, \tag{C5}$$

with

$$D=\frac{u^x}{1+\gamma}\exp[-m^2(1-\omega)\sigma^*-(m^2\omega-\tfrac14 t)\sigma].\tag{C6}$$

Let us define:

$$\tilde{j}_x^{(n,l)}(\sigma,\delta;\sigma^*,\delta^*)=\int_0^U du\left(\frac{\sigma}{\sigma^*}u\right)^l\frac{\partial^{n-l}\tilde{L}(\sigma,\delta,0;\sigma^*,\delta^*,u)}{\partial\gamma^n}.$$

Using inequalities (C3), (IV.2), and (IV.3), we obtain

$$\left|\frac{1}{(n-l)!}\tilde{j}^{(n,l)}(\sigma,\delta;\sigma^*,\delta^*)\right|<\frac{1}{x+l+1}D_{n-l}.\tag{C7}$$

2. Bound for $0\leqslant\gamma<\infty$

The kernel $\tilde{L}(\sigma,\delta,\gamma;\sigma^*,\delta^*,u)=[r'(\sigma,\gamma)/r'(\sigma^*,\gamma^*)]\,L(\sigma,\delta,\gamma;\sigma^*,\delta^*,u)$ can be written

$$\tilde{\tilde{L}}(\sigma,\delta,\gamma,\sigma^*,\delta^*,u)=c\frac{1}{1+\gamma}\exp\left(-\tilde{\tilde{a}}\gamma-b\frac{\gamma}{1+\gamma}\right),\tag{C8}$$

with

$$\tilde{\tilde{a}}=a+m^2\sigma\omega'(1-u),\tag{C8$'$}$$

and a, b, c defined by Eq. (C1′). Using the bound (C2) we see that, if ω' verifies the inequality

$$\begin{aligned}&\omega'>\frac{P^2}{m^2}\quad\text{if } 0\leqslant t<4m^2,\\ &\omega'>\frac{P^2-\tfrac14 t}{m^2}\quad\text{if } t<0,\end{aligned}\tag{C9}$$

then the bound (B6) can be used and we find

$$\operatorname*{Max}_{0\leqslant\gamma<\infty}\tilde{\tilde{L}}(\sigma,\delta,\gamma,\sigma^*,\delta^*,u)=c\leqslant u^xD_0.\tag{C10}$$

APPENDIX D: BOUNDS ON THE FIRST TERM $\tilde{F}_1(\sigma,\delta,\gamma)$

1. Bounds for $\gamma=0$

In this subsection, we give bounds on the modified first term [see Eq. II.23)]

$$\tilde{F}_1(\sigma,\delta,\gamma)=r(\sigma)F_1(\sigma,\delta,\gamma)=\exp(-m^2\sigma\omega)F_1(\sigma,\delta,\gamma).\tag{D1}$$

The function $\tilde{F}_1$ can be written

$$\tilde{F}_1(\sigma,\delta,\gamma)=\int_0^\infty d\gamma_0 f(\gamma_0,\delta,\sigma)\,g(\gamma_0,\gamma,\sigma),\tag{D2}$$

with

$$f(\gamma_0,\delta,\sigma)=\exp[-\sigma A(\gamma_0,\delta)]\frac{\gamma_0^x}{(1+\gamma_0)^{x+2}},\tag{D3}$$

$$A(\gamma_0,\delta)=m^2\omega+m^2\gamma_0-\frac{t\delta(1-\delta)+p_2^2\delta\gamma_0+p_4^2(1-\delta)\gamma_0}{1+\gamma_0},\tag{D4}$$

$$g(\gamma_0,\gamma,\sigma)=\frac{1}{[1+\gamma/(1+\gamma_0)]^{x+2}}\times\exp\left(+\frac{\sigma b(\gamma_0,\delta)}{1+\gamma_0}\frac{\gamma}{1+\gamma_0+\gamma}\right),\tag{D5}$$

$$b(\gamma_0,\delta)=[p_1^2\delta+p_3^2(1-\delta)](1+\gamma_0)-[t\delta(1-\delta)+p_2^2\delta\gamma_0+p_4^2(1-\delta)\gamma_0].\tag{D6}$$

We recall that we are working under conditions (IV.4)–(IV.6).

We defined a constant P^2 [see Eq. (IV.10)] which verifies [see Eq. (IV.11)]

$$\begin{aligned}&\tfrac14 t-P^2<p_i^2<P^2,\\ &P^2>m^2+\tfrac14 t\quad\text{if } t\geqslant0,\\ &P^2>m^2\quad\text{if } t\leqslant0.\end{aligned}\tag{D7}$$

Bounds on the function $A(\gamma,\delta)$

(i) If $0\leqslant t<4m^2$

$$A(\gamma,\delta)\geqslant A(\gamma)=m^2(\omega-1)-P^2+m^2(1+\gamma)+(P^2-\tfrac14 t)/(1+\gamma)\tag{D8}$$

$$A(\gamma)\geqslant\bar{A}=m^2(\omega-1)-P^2+2\sqrt{m^2(P^2-\tfrac14 t)}.\tag{D9}$$

$\bar{A}$ also verifies

$$m^2\omega-\tfrac14 t\geqslant\bar{A}.\tag{D10}$$

(2) If $t\leqslant0$

$$A(\gamma,\delta)\geqslant A(\gamma)=m^2(\omega-1)-P^2+m^2(1+\gamma)+P^2/(1+\gamma)\tag{D8$'$}$$

$$A(\gamma)\geqslant\bar{A}=m^2(\omega-1)-P^2+2\sqrt{m^2P^2}.\tag{D9$'$}$$

$\bar{A}$ also verifies

$$m^2\omega\geqslant\bar{A}.\tag{D10$'$}$$

Bounds on the nth derivative of $\tilde{F}_1$

The nth derivative of $\tilde{F}_1$ with respect to γ is the integral of the nth derivative of g

$$\frac{1}{n!}\frac{\partial^n\tilde{F}_1(\sigma,\delta,\gamma)}{\partial\gamma^n}=\int_0^\infty d\gamma_0 f(\gamma_0,\delta)\frac{1}{n!}\frac{\partial^n g(\gamma_0,\gamma,\delta)}{\partial\gamma^n}.\tag{D11}$$

Inequality (B3) can be used, and we find

$$\left|\frac{1}{n!}\frac{\partial^n g(\gamma_0,\gamma=0,\delta)}{\partial\gamma^n}\right|<(x+2)^n\sum_{k=0}^n\left(\frac{1}{x+2}\frac{\sigma|b(\gamma_0,\delta)|}{1+\gamma_0}\right)^k\frac{1}{k!}C_n^k.\tag{D13}$$

As the p_i^2 ($i=1,2,3,4$) verify the inequality (D7), we have

$$\begin{aligned}&|b(\gamma,\delta)|<b(\gamma)=(2P^2-\tfrac14 t)(1+\gamma)-(P^2-\tfrac14 t)\quad\text{if } 0\leqslant t<4m^2,\\ &|b(\gamma,\delta)|<b(\gamma)=(2P^2-\tfrac14 t)(1+\gamma)-P^2\quad\text{if } t\leqslant0.\end{aligned}\tag{D14}$$

Using these inequalities and (D8) or (D8′) we easily show that

$$\frac{|b(\gamma_0,\delta)|}{1+\gamma_0}<\bar{B}A(\gamma_0),$$

where $\bar{B}$ is a constant.

Using inequality (D8) or (D8′) we find

$$f(\gamma_0,\delta,\sigma)<\frac{\gamma_0^x}{(1+\gamma_0)^{x+2}}\exp[-\sigma A(\gamma_0)].$$

Putting these two last inequalities in Eq. (D13) and using (D11) we obtain

$$\left|\frac{1}{n!}\frac{\partial^n \widetilde{F}_1(\sigma,\delta,\gamma=0)}{\partial\gamma^n}\right| < (x+2)^n \int_0^\infty d\gamma_0 \frac{\gamma_0^x}{(1+\gamma_0)^{x+2}}$$
$$\times \exp[-\sigma A(\gamma_0)] \sum_{k=0}^{n} \left(\frac{\overline{B}}{x+2}\sigma A(\gamma_0)\right)^k \frac{1}{k!} C_n^k. \quad \text{(D15)}$$

Using (B7), it is possible to perform the integral and finally we have

$$\left|\frac{1}{n!}\frac{\partial^n \widetilde{F}_1(\sigma,\delta,\gamma=0)}{\partial\gamma^n}\right| < \overline{C}_n \exp(-\sigma\overline{A}), \quad \text{(D16)}$$

with

$$\overline{C}_n = \frac{(x+2)^n}{x+1} \sum_{k=0}^{n} \left(\overline{B}\frac{k}{x+2}\right)^k \frac{1}{k!} C_n^k. \quad \text{(D17)}$$

2. Bound for $0 \leqslant \gamma < \infty$

The modified first term $\widetilde{\widetilde{F}}_1(\sigma,\delta,\gamma) = r'(\sigma,\gamma)F_1(\sigma,\delta,\gamma)$ can be written

$$\widetilde{\widetilde{F}}_1(\sigma,\delta,\gamma) = \int_0^\infty d\gamma_0 f(\gamma_0,\delta,\sigma)\, \widetilde{\widetilde{g}}(\gamma_0,\gamma,\sigma), \quad \text{(D18)}$$

with

$$\widetilde{\widetilde{g}}(\gamma_0,\gamma,\sigma) = \frac{1}{[1+\gamma/(1+\gamma_0)]^{x+2}}$$
$$\times \exp\left(-\sigma m^2\omega'\gamma + \frac{\sigma b(\gamma_0,\gamma)}{1+\gamma_0}\frac{\gamma}{1+\gamma+\gamma_0}\right), \quad \text{(D19)}$$

and f and b defined by Eqs. (D3) and (D6).

If ω' verifies the inequality (C9), the bound (B6) can be used and we find

$$\operatorname*{Max}_{0<\gamma<\infty} \widetilde{\widetilde{F}}_1(\sigma,\delta,\gamma) < \frac{1}{x+1}\exp(-\sigma\overline{A}). \quad \text{(D20)}$$

[1] J. C. Polkinghorne, J. Math. Phys. **4**, 503 (1963); **5**, 431 (1964); R. J. Eden, P. V. Landshoff, D. I. Olive, and J. C. Polkinghorne, *The Analytic S-Matrix* (Cambridge U. P., Cambridge, 1966).
[2] O. I. Zav'Yalov, J. Exptl. Theoret. Phys. **47**, 1099 (1964) [Sov. Phys. JETP **20**, 736 (1965)]; O. I. Zav'Yalov and B. M. Stepanov, Yad. Fiz. **1**, 922 (1965) [Sov. J. Nucl. Phys. **1**, 658 (1965)].
[3] M. C. Bergère and C. Gilain, J. Math. Phys. **19**, 1495 (1978); M. C. Bergère and C. de Calan, Phys. Rev. D **20**, 2047 (1979).
[4] G. C. Wick, Phys. Rev. **96**, 1124 (1954).
[5] R. E. Cutkosky, Phys. Rev. **96**, 1135 (1954).
[6] N. Nakanishi, Prog. Theor. Phys. Suppl. **43**, 1 (1969).
[7] G. Domokos and P. Suranyi, Nucl. Phys. **54**, 529 (1964).
[8] S. G. Mikhlin, *Integral Equations* (Pergamon, New York, 1957); F. G. Tricomi, *Integral Equations* (Interscience, New York, 1957).
[9] I. J. Muzinich and H. S. Tsao, Phys. Rev. D **11**, 2203 (1975).
[10] C. Gilain, thesis, Universite Parissud Centre d'Orsay, (1981).

Representations of supergroups [a)]

A. Baha Balantekin
Wright Nuclear Structure Laboratory, Physics Department, Yale University, New Haven, Connecticut 06520
Itzhak Bars
J. W. Gibbs Laboratory, Physics Department, Yale University, New Haven, Connecticut 06520

(Received 19 January 1981; accepted for publication 20 March 1981)

An explicit construction of representations of supergroups is given in terms of direct products of covariant and contravariant fundamental representations. The rules of supersymmetrization are characterized by extended Young supertableaux. This constructive approach leads to explicit transformation properties of higher representations as well as to closed explicit formulas for characters from which other invariants such as dimensions and eigenvalues of all Casimir operators can be calculated. We have applied this approach so far to the supergroups SU(N/M), OSP($N/2M$), P(N), for which we have obtained all the representations constructible as direct products of the fundamental (defining) representations. An argument is presented toward the irreducibility of all these representations.

PACS numbers: 11.30.Pb, 02.20.Qs

I. INTRODUCTION

Supergroups and superalgebras[1] have been proposed to construct models in various areas of physics.[2] The first evidence of the existence of supersymmetry in nature has come with the supersymmetric model of nuclei[3] based on the SU (6/4) supergroup where the bosonic and fermionic levels of five different nuclei are placed in the same supermultiplet. In this work certain aspects of the representation theory of supergroups were used as developed by us in Ref. 4. Independent investigations by other authors[5,6] partially overlap with ours.

The basic technique of Ref. 4. is a symmetrization–antisymmetrization procedure (supersymmetrization) associated with Young tableaux generalized to supergroups (supertableaux). The main observation was that known results for the characters of ordinary groups can be generalized to supergroups by a simple replacement of traces by supertraces *in the fundamental representation*. Once the supercharacters are obtained, it is straightforward to derive many properties of the supergroups including dimension formulas and eigen values of Casimir operators. Using this technique in Ref. 4 we derived character and dimension formulas for certain representations of U(N/M) and OSP($N/2M$) type supergroups. Formulas for the quadratic Casimir operators were included along with a general procedure for calculating higher order Casimir operators from the characters.

The representations of U(N/M) supergroups studied in Ref. 4 are the ones which can be constructed using only covariant *or* contravariant basis vectors of either class I *or* class II representations. An extension of that work is presented in Sec. II, where we discuss the representations obtained by taking tensor products of both covariant *and* contravariant basis vectors of the same class, as well as tensor products of class I with class II representations.

For the ordinary SU(N) group all representations can be written in terms of only covariant tensors, but for the supergroup SU(N/M) a contravariant tensor cannot be rewritten in terms of covariant ones, so that the representations we discuss here are not included in our work of Ref. 4. We give character and dimension formulas for such representations of supergroups. A further set of representations for SU(N/M) is obtained by taking direct products of class I and class II representations which differ from each other by interchanging bosons and fermions. These products turn out to give new representations of SU(N/M). Character and dimension formulas are provided for these representations as well.

In Sec. III we apply similar techniques to the P(N) type supergroups. We find that the structure of either class I or II representations for P(N) are formally similar to those for $Sp(2N)$, just as we showed in Ref. 4 that OSP($N/2M$) had the same structure as O(N). We calculate the characters of mixed class I and II representations of these supergroups, from which other quantities such as dimensions and Casimir operator eigenvalues can be computed. The supergroup $Q(N)$ is the only "classical" supergroup whose representations remain to be investigated.

We have thus derived the characters and the dimensions for all representations that can be constructed from direct products of fundamental class I and II representations of all supergroups [except $Q(N)$]. This leaves out certain non integral representations[7] of SU($N/1$) which cannot be obtained from the fundamental representation.

II. REPRESENTATIONS OF U (N/M)

The fundamental (defining) representation of the supergroup U(N/M) has two bases,

$$\xi_A = \begin{pmatrix} \phi_a \\ \psi_\alpha \end{pmatrix} \text{ and } \tilde{\xi}_A = \begin{pmatrix} \psi_a \\ \phi_\alpha \end{pmatrix},$$

which we call class I and class II, respectively. In ξ_A ϕ_a, $a = 1,\cdots,N$ are bosonic and $\psi_\alpha, \alpha = 1,\cdots,M$ are fermionic. For this representation the degree of the index A is $g(a) = 0$ and $g(\alpha) = 1$. On the other hand, in $\tilde{\xi}_A$ the components ψ_a,

[a)]Research supported in part by the U.S. Department of Energy under Contracts Nos. DE-ACO2-76ERO374 and EY-76-C-02-3075.

 © 1981 American Institute of Physics

$a = 1,\cdots,N$ are fermionic and ϕ_α, $\alpha = 1,\cdots,M$ are bosonic; we then have $g(a) = 1$ and $g(\alpha) = 0$. Both of these basis vectors transform with the same supergroup element

$$\xi'_A = \mathscr{U}^B_A \xi_B \text{ and } \tilde{\xi}'_A = \mathscr{U}^B_A \tilde{\xi}_B, \tag{2.1}$$

with

$$\mathscr{U} = \begin{pmatrix} \mathscr{A} & \mathscr{B} \\ \mathscr{C} & \mathscr{D} \end{pmatrix} = \exp i \begin{pmatrix} H_N & \theta \\ \theta^\dagger & H_M \end{pmatrix}. \tag{2.2}$$

Note that in contrast with Ref. 4 we are distinguishing between lower and upper indices. Here the $N \times N$ matrices $(\mathscr{A}, H_n)$ and the $M \times M$ matrices $(\mathscr{D}, H_M)$ have bosonic elements while the $N \times M$ matrices $(\mathscr{B}, \theta)$ and the $M \times N$ matrices $(\mathscr{C}, \theta^\dagger)$ have fermionic elements. The unitarity condition is implemented by requiring that H_N and H_M are Hermitian and $\theta^\dagger$ is the Hermitian conjugate of θ. Hermitian conjugation is defined to interchange the order of fermions in a given product. The supertrace, which is defined as

$$\text{Str } \mathscr{U} = \sum_A (-1)^{g(A)} \mathscr{U}^A_A, \tag{2.3}$$

is an invariant of the supergroup; so is the superdeterminant

$$\text{Sdet } \mathscr{U} = \exp[\text{Str}(\ln \mathscr{U})]. \tag{2.4}$$

The condition Sdet $\mathscr{U} = 1$ is identical to $\text{Tr } H_N = \text{Tr} H_M$.

A. Covariant and contravariant class I representations

For class I (and similarly for class II) representations the contravariant basis vector ξ^A is defined as the Hermitian conjugate of the covariant basis vector,

$$\xi^A = (\xi^\dagger)_A. \tag{2.5}$$

The Hermitian conjugate of the matrix $\mathscr{U}$ of Eq. (2.2) is

$$(\mathscr{U}^\dagger)^A_B = (\mathscr{U}^*)^B_A. \tag{2.6}$$

Following Eqs. (2.5) and (2.6) one can show that the basis vector ξ^A transform as

$$\xi'_A = \xi^B \mathscr{U}^{\dagger A}_B. \tag{2.7}$$

For the supergroup $U(N/M)$ the only invariant tensor is the Kronecker delta δ^C_B, with one lower and one upper index. Unlike $SU(N)$, there is no completely "antisymmetric" invariant tensor since the superdeterminant is not a finite polynomial. We can write

$$\mathscr{U}^{A'}_A \delta_{A'}{}^{B'} \mathscr{U}^{\dagger B}_{B'} = \delta^B_A. \tag{2.8}$$

This implies that $\xi^A \delta^B_A \xi_B$ is an invariant. If $\mathscr{U}^B_A$ and $\mathscr{U}^{\dagger D}_C$ are interchanged in a product one picks up some minus signs from the fermionic components. The general rule as given in Ref. 4 is

$$\mathscr{U}^B_A \mathscr{U}^{\dagger D}_C = (-1)^{[g(A)-g(B)]\cdot[g(C)-g(D)]} \mathscr{U}^{\dagger D}_C \mathscr{U}^B_A. \tag{2.9}$$

Using this property and Eq. (2.8) it can be shown that the left inverse is equal to the right inverse, i.e.,

$$\mathscr{U}^{\dagger A'}_A \delta^{B'}_{A'} \mathscr{U}^B_{B'} = \delta^B_A. \tag{2.10}$$

The supersymmetrized and superantisymmetrized tensor products of two covariant (and similarly contravariant) vectors belonging to the same class are given by

$$\xi_{(AB)} = \xi^{(1)}_A \xi^{(2)}_B + \xi^{(2)}_A \xi^{(1)}_B, \tag{2.11}$$

$$\xi_{[A,B]} = \xi^{(1)}_A \xi^{(2)}_B - \xi^{(2)}_A \xi^{(1)}_B. \tag{2.12}$$

Note that the order of the indices is kept the same while the wavefunctions are interchanged. Higher order supersymmetrized or superantisymmetrized tensor products follow the same rule. These tensors form irreducible bases since we have no invariant, covariant or contravariant tensors. We have studied such representations in Ref. 4. The character of the class-I representation obtained by using covariant bases and associated with a single row supertableau containing n boxes [as in Fig. 1(a)] was given as[4]

$$\mathscr{K}_n = H_n \equiv \oint_C \frac{dz}{2\pi i} \frac{z^{-n-1}}{\text{Sdet}(1 - z\mathscr{U})}, \tag{2.13}$$

where $\mathscr{U}$ is the fundamental representation given in Eq. (2.2) and C is a contour around the origin. The character of the representation $(n_1, n_2, \cdots)$ associated with the supertableau [as in Fig. 1(b)] with n_1 boxes in the first row, n_2 boxes in the second row, etc., was given as[4]

$$\mathscr{K}_{(n_1,n_2,\cdots)} = \det(H_{n_j+i-j}) = \begin{vmatrix} H_{n_1} & H_{n_2-1} & H_{n_3-2}\cdots \\ H_{n_1+1} & H_{n_2} & H_{n_3-1}\cdots \\ H_{n_1+2} & H_{n_2+1} & H_{n_3}\cdots \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{vmatrix}, \tag{2.14}$$

where we have shown the ijth element of the matrix whose determinant is calculated. The rank of the determinant is equal to the number of nontrivial rows in the supertableau.

Representations obtained by using contravariant bases are associated with the supertableaux which are mirror images of the covariant supertableaux, and are indicated by

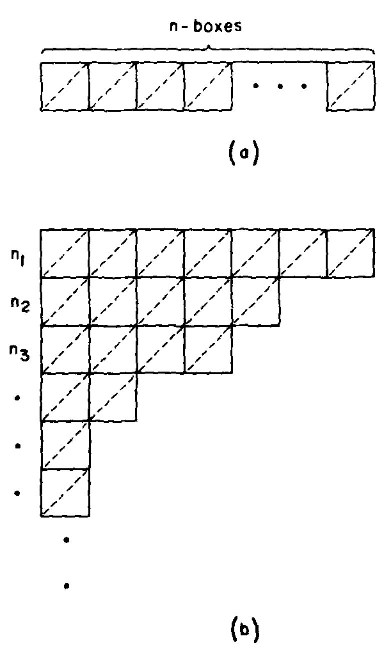


FIG. 1. Supertableaux corresponding to the representations (a) $(\cdots,0,0;n,0,\cdots)$ and (b) $(\cdots,0;n_1,n_2,n_3,\cdots)$ of the supergroup $SU(N/M)$ which are constructed only covariant basis vectors.

dotted boxes [as in Fig. 2(a)]. They are required to satisfy $n_1 \geqslant n_2 \geqslant n_3 \geqslant \cdots$. These contravariant supertableaux correspond to new representations which cannot be obtained from the covariant ones, unlike the ordinary SU(N) representations. This is due to the lack of an invariant completely superantisymmetric tensor. Note that even though there exists an invariant superdeterminant it does not correspond to an invariant supertensor.

If we replace $\mathscr{U}$ by $\mathscr{U}^\dagger$ in the character formulas of covariant representations, we obtain the character formula of contravariant representations. This follows from Eq. (2.5). For example, the character of the representation in Fig. 2(b) with n dotted boxes is obtained from (2.13) as

$$\dot{\mathscr{X}}_n = \dot{H}_{-n} = \oint_C \frac{dz}{2\pi i} \frac{z^{-n-1}}{\mathrm{Sdet}(1 - z\mathscr{U}^\dagger)}. \tag{2.15}$$

It is convenient to *define* $\dot{H}_{-n}$ with a negative index $(-n)$, where the dot on $\dot{H}_{-n}$ reminds us of the contravariant representation which uses $\mathscr{U}^\dagger$ instead of $\mathscr{U}$. Note that according to Eqs. (2.13) and (2.15) $H_n = 0 = \dot{H}_{-n}$ if n is negative while $H_0 = \dot{H}_0 = 1$.

The character of the representation $(-n_1, -n_2, -n_3, \cdots)$ obtained by using contravariant basis vectors is given by

$$\dot{\mathscr{X}}_{(-n_1, -n_2, \cdots)} = \det(\dot{H}_{-n_i - i + j}) = \begin{vmatrix} \vdots & \vdots & \vdots \\ \cdots \dot{H}_{-n_3} & \dot{H}_{-n_2 - 1} & \dot{H}_{-n_1 - 2} \\ \cdots \dot{H}_{-n_3 + 1} & \dot{H}_{-n_2} & \dot{H}_{-n_1 - 1} \\ \cdots \dot{H}_{-n_3 + 2} & \dot{H}_{-n_2 + 1} & \dot{H}_{-n_1} \end{vmatrix}, \tag{2.16}$$

in which $\mathscr{U}^\dagger$ appears instead of $\mathscr{U}$ and otherwise it is formally identical to Eq. (2.14). This result follows with the methods of Ref. 4.

B. Mixed covariant–contravariant class I representations

We now consider representations obtained by taking tensor products of both covariant and contravariant bases of the same class. Since we have an invariant mixed tensor (Kroenecker delta) such tensor products are not irreducible. They contain invariant subspaces which we need to eliminate by demanding that the supertrace with respect to any two pairs of covariant and contravariant indices must vanish. We will study some examples below. Representations obtained in this way are associated with the mixed supertableaux in Fig. 3 (a), where n_i and $-m_j$ indicate the number of the boxes, respectively. We denote such representations as

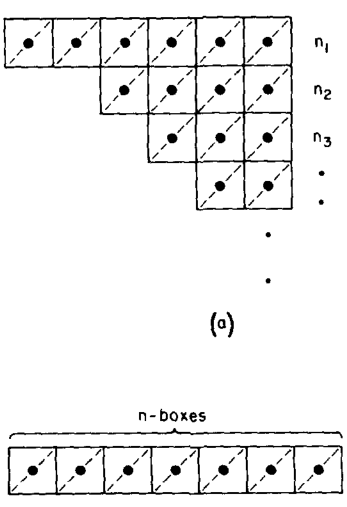


FIG. 2. Supertableaux corresponding to the representations (a) $(\cdots, -n_3, -n_2, -n_1; 0,0,\cdots)$ and (b) $(\cdots,0, -n;0,0,\cdots)$ of the supergroup SU(N/M) which are constructed using only contravariant basis vectors.

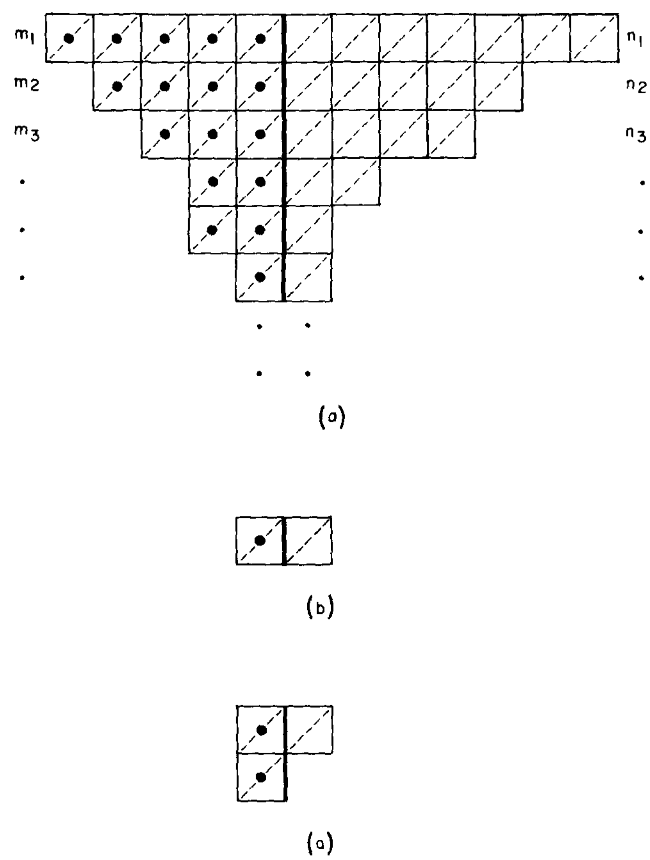


FIG. 3. Supertableaux corresponding to the representations (a) $(\cdots, -m_3, -m_2, -m_1; n_1, n_2, n_3, \cdots)$, (b) $(\cdots,0, -1;1,0,\cdots)$, and (c) $(\cdots,0, -1, -1;1,0,\cdots)$ of the supergroup SU(N/M).

$(\cdots, -m_j, \cdots, -m_2, -m_1; n_1, n_2, \cdots, n_i, \cdots)$. For example, $(-1;1)$ is the representation obtained from one contravariant and one covariant basis vector; similarly, $(-2;1,1)$ is the representation obtained from four basis vectors: two covariant ones which are antisymmetrized and two contravariant ones which are symmetrized. In both cases the invariant subspaces are subtracted. We write the covariant indices on the left and the contravariant indices on the right as in $\psi_{ABC\cdots}{}^{DEF\cdots}$. As a first example we will explicitly construct the class I representation $(-1;1)$ represented by the supertableau in Fig. 3(b). This is the adjoint representation. The basis ξ_A^B of this representation is constructed as

$$\xi_A^B = \xi_A^{(1)}\xi^{(2)B} - [\delta_A^B/\eta(N-M)]\xi_C^{(1)}\xi^{(2)C}(-1)^{g(C)}, \quad (2.17)$$

where the second term insures that the supertrace of ξ_A^B is zero, i.e., $(-1)^{g(A)}\xi_A^A = 0$. Note that the second term is invariant under the supertransformations since $(-1)^{g(C)}\xi_C^{(1)}\xi^{(2)C} = \xi^{(2)C}\xi_C^{(1)}$ is invariant according to Eqs. (2.1), (2.7), and (2.8). The factor $\eta = +1(-1)$ for class I, (II) representations. Under the supergroup SU(N/M), ξ_A^B transforms into

$$\xi'^B_A = \mathscr{U}_{AB'}^{BA'}\xi_{A'}^{B'}, \quad (2.18)$$

where the supergroup element $\mathscr{U}_{AB'}^{BA'}$ in this representation is obtained by transforming $\xi_A^{(1)}$ and $\xi^{(2)B}$ according to Eqs. (2.1) and (2.7). In obtaining $\mathscr{U}_{AB'}^{BA'}$ the order of the factors $\mathscr{U}_A^{A'}\xi_{A'}^{(1)}\xi^{(2)B'}\mathscr{U}_{B'}^{\dagger B}$ is important important because $\mathscr{U}$'s and $\mathscr{U}^\dagger$'s must be shifted to the left of the expression. This shifting produces certain minus signs and yields

$$\mathscr{U}_{AB'}^{BA'} = (-1)^{[g(A')-g(B')][g(B)-g(B')]}\mathscr{U}_A^{A'}\mathscr{U}_{B'}^{\dagger B} - \delta_A^B(-1)^{g(A')}\delta_{B'}^{A'}/\eta(N-M). \quad (2.19)$$

The character of the representaton $(-1;1)$ is the supertrace of this matrix. To take the supertrace we let $A' = A, B' = B$, multiply the whole expression by $(-1)^{g(A)}(-1)^{g(B)}$, and sum over A and B. We obtain

$$\mathscr{K}_{(-1;1)} = (\mathrm{Str}\,\mathscr{U}^\dagger)(\mathrm{Str}\,\mathscr{U}) - 1. \quad (2.20)$$

From Eqs. (2.13) and (2.15) we have $H_1 = \mathrm{Str}\,\mathscr{U}$ and $H_{-1} = \mathrm{Str}\,\mathscr{U}^\dagger$. Therefore, (2.20) can be rewritten as the determinant

$$\mathscr{K}_{(-1;1)} = \begin{vmatrix} H_{-1} & H_0 \\ H_0 & H_1 \end{vmatrix}. \quad (2.21)$$

As a second example we will study the representation $(-1,-1;1)$ associated with the supertableau in Fig. 3(c). The basis of this representation is

$$\xi_C^{(A,B)} = \xi_C\xi^{(A,B)} - \frac{1}{[\eta(N-M)-1]}[\delta_C^A\xi_D\xi^{(D,B)}(-1)^{g(D)} - (-1)^{g(A)g(B)}\delta_C^B\xi_D\xi^{(D,A)}(-1)^{g(D)}], \quad (2.22)$$

where we use

$$\xi^{(A,B)} = \xi^{(1)A}\xi^{(2)B} - \xi^{(2)A}\xi^{(1)B},$$

which is the basis of the representation $(-1,-1)$. Note that $\xi_C^{(A,B)}$ is supertraceless in the sense of $(-1)^{g(D)}\xi_D^{(D,B)} = 0 = (-1)^{g(D)}(-1)^{g(D)\cdot g(A)}\xi_D^{(A,D)}$. The second expression follows from the fermionic properties of some of the components and is consistent with the interchange rule $\xi^{(A,B)} = (-1)^{g(A)\cdot g(B)}\xi^{(B,A)}$ as given in Ref. 4.

Under the supergroup $\xi^{(A,B)}$ transforms into

$$\xi'^{(A,B)} = \xi^{(A',B')}\mathscr{U}^{\dagger}_{(A',B')}{}^{(A,B)}, \quad (2.23)$$

where the transformation matrix can be constructed according to the methods of Ref. (4) as

$$\mathscr{U}^{\dagger}_{(A',B')}{}^{(A,B)} = \tfrac{1}{2}[(-1)^{g(B')[g(A')-g(A)]}\mathscr{U}_{A'}^{\dagger A}\mathscr{U}_{B'}^{\dagger B} - (-1)^{g(A)g(B)}(-1)^{g(B')[g(A')-g(B)]}\mathscr{U}_{A'}^{\dagger B}\mathscr{U}^{\dagger}_{B'}{}^{A}] \quad (2.24)$$

Using Eqs. (2.15) and (2.16) the character of this representation is

$$\dot{\mathscr{K}}_{(-1,-1)} = \begin{vmatrix} \dot H_{-1} & \dot H_{-2} \\ \dot H_0 & \dot H_{-1} \end{vmatrix}. \quad (2.25)$$

On the other hand, the basis of the representation $(-1,-1;1)$ transforms into

$$\xi'^{(A,B)}_C = \xi_{C'}{}^{(A',B')}\mathscr{U}^{\dagger}_{(A',B')C}{}^{C'(A,B)}. \quad (2.26)$$

where, after shifting all $\mathscr{U}$ and $\mathscr{U}^\dagger$'s to the right according to Ref. 4, we obtain

$$\mathscr{U}^{\dagger}_{(A',B')C}{}^{C'(A,B)} = \{(-1)^{[g(C)-g(C')][g(A')-g(B')-g(C')]}$$

$$\mathscr{U}_C^{C'}\mathscr{U}^{\dagger}_{(A',B')}{}^{(A,B)} - \frac{1}{2\eta(N-M)-2}[(-1)^{g(C')}\delta_C^A\delta_{A'}^{C'}\mathscr{U}^{\dagger}_{B'}{}^{B}$$
$$-(-1)^{g(C')+g(A')\cdot g(B')}\delta_C^A\delta_{B'}^{C'}\mathscr{U}^{\dagger}_{A'}{}^{B}$$
$$-(-1)^{g(C')+g(A)\cdot g(B)}\delta_C^B\delta_{A'}{}^{C'}\mathscr{U}^{\dagger}_{B'}{}^{A}$$
$$+(-1)^{g(C')+g(A)\cdot g(B)+g(A')\cdot g(B')}\delta_C^B\delta_{B'}^{C'}\mathscr{U}^{\dagger}_{A'}{}^{A}]\}. \quad (2.27)$$

The character of this representation is obtained by setting $A = A', B = B', C = C'$, multiplying by $(-1)^{g(A)+g(B)+g(C)}$, and summing over A,B,C. We obtain

$$\mathscr{K}_{(-1,-1;1)} = \mathscr{K}_1\dot{\mathscr{K}}_{(-1,-1)} - \dot{\mathscr{K}}_{-1} \quad (2.28)$$

which, using Eq. (2.25), can be written as

$$\mathscr{K}_{(-1,-1;1)} = \begin{vmatrix} \dot H_{-1} & \dot H_{-2} & 0 \\ \dot H_0 & \dot H_{-1} & H_0 \\ 0 & \dot H_0 & H_1 \end{vmatrix}. \quad (2.29)$$

We perform similar calculations for other mixed covariant–contravariant class I representations of U(N/M). We find the general rule that for the representation

$$(-m_k, -m_{k-1}, \cdots, -m_2, -m_1; n_1, n_2, \cdots, n_{i-1}, n_i)$$

the formulas (2.21) and (2.29) can be generalized as

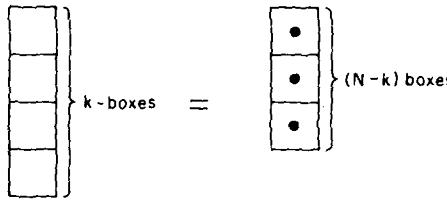


FIG. 4. For the group SU(N) the rule to replace a column of undotted boxes by a column of dotted boxes.

$$
\mathscr{K}_{-m_k,\cdots,-m_1;n_1,\cdots,n_l} = \begin{vmatrix} H_{-m_k} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \dot{H}_{-m_2} & \dot{H}_{-m_1-1} & H_{n_1-2} & H_{n_2-3} & \cdot & \cdot \\ \cdot & \cdot & \dot{H}_{-m_2+1} & \dot{H}_{-m_1} & H_{n_1-1} & H_{n_2-2} & \cdot & \cdot \\ \cdot & \cdot & \dot{H}_{-m_2+2} & \dot{H}_{-m_1+1} & H_{n_1} & H_{n_2-1} & \cdot & \cdot \\ \cdot & \cdot & \dot{H}_{-m_2+3} & \dot{H}_{-m_1+2} & H_{n_1+1} & H_{n_2} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & H_{n_l} \end{vmatrix}. \tag{2.30}
$$

The rank of this determinant is determined by the number of nontrivial rows in the generalized supertableau of Fig. 3(a). The structure of this determinant can be remembered as follows. On the diagonal we start with the obvious subscript and we increase the index going downward or decrease going upward. We must substitute $H_n = 0 = \dot{H}_{-n}$ if $n < 0$,f which are constraints that automatically follow from Eqs. (2.13) and (2.15). This result unifies all class I representations.

The dimension of the representation is now easily calculated by computing the character of the matrix $(-1)^{g(A)}\delta_A^B = \mathscr{I}_A^B$, or

$$
\mathscr{I} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.
$$

In Ref. 4 we have already calculated

$$
H_n(\mathscr{I}) = D_n = \sum_{k=0}^{n} \binom{M}{k}\binom{-N}{n-k}
$$

and now we note that $\dot{H}_{-n}(\mathscr{I}) = D_n$ has the identical value for any SU(N/M). By substituting these expressions in Eq. (2.30) we obtain the dimension formula which generalizes our results in Ref. 4. The same remarks apply to the eigenvalues of Casmir operators.

If we apply the same procedures to the ordinary SU(N) group, we will end up with the same formal expressions as Eq. (2.30) for the analogous mixed Young tableaux, except that we will be substituting traces instead of supertraces. This formula for the character of SU(N) is unfamiliar and, as far as we known, it might be new in the literature. Of course, for SU(N), if we replace $U^\dagger$ by a product of $(N-1)$ U's which are appropriately symmetrized, our formula does reduce to the familiar expressions.[4,8] For SU(N) the general rule[9] for changing from purely covariant notation to mixed covariant–contravariant notation is to replace any column with k undotted boxes by a column of N–k dotted boxes as in Fig. 4. The new resulting Young tableau should be of the form of Fig. 1(b). For SU(N) our new formula correctly reproduces the known expressions for characters, dimensions, and Casimir invariants. The new approach is very useful in those cases for which the usual Young tableau contains long columns which could be rewritten conveniently in terms of a few dotted boxes. Then all expressions and calculations simplify by many orders of magnitude.

We must again emphasize that for the supergroup SU(N/M) the replacement of dotted boxes by undotted boxes is not allowed.

C. Direct products of class I and class II fundamental representations

Before concluding this section we want to remark that similar methods can be employed to study the representations constructed by taking direct products of *both* class I *and* class II basis vectors. We recall that the class II fundamental basis vector differs form that of class I by an interchange of the grades (bosons⇆fermions). We denote covariant (contravariant) class I and II basis vectors by ξ_A (ξ^A) and $\tilde{\xi}_{\tilde{A}}$ ($\tilde{\xi}^{\tilde{A}}$), respectively. In the rest of this section we will put tilde signs on class II indices. For the grades of A and $\tilde{A}$ we have the relation

$$
g(A) = g(\tilde{A}) + 1, \ \bmod 2. \tag{2.31}
$$

We recall the transformation property of the basis vectors

$$
\xi_A \to \mathscr{U}_A^B \xi_B, \quad \tilde{\xi}_{\tilde{A}} \to \mathscr{U}_{\tilde{A}}^{\tilde{B}} \tilde{\xi}_{\tilde{B}},
$$

where

$$
\mathscr{U}_A^B = \mathscr{U}_{\tilde{A}}^{\tilde{B}}. \tag{2.32}
$$

Now, let us consider the supersymmetrized tensor product $\xi_{A\tilde{B}}$ of one class I and one class II covariant basis vector

$$
\xi_{A\tilde{B}} = \xi_A^{(1)}\xi_{\tilde{B}}^{(2)} + (-1)^{g(A)\cdot g(B)}\xi_{\tilde{B}}^{(1)}\xi_A^{(2)}. \tag{2.33}
$$

The same form was obtained from the supersymmetrization[4] of two class I vectors after placing the wavefunctions in the same order. Note that the sign is $(-1)^{g(A)\cdot g(B)}$ and *not* $(-1)^{g(\tilde{A})\cdot g(B)}$. The elements of this basis have their grades interchanged (bosons⇆fermions) in relation to the corresponding purely class I case. Equation (2.33) forms the basis of a new representation. Indeed, repeating steps analogous to those in the corresponding class I case,[4] we find that ξ_{AB} transform as

$$
\xi_{A\tilde{B}} \to \xi'_{A\tilde{B}} = \mathscr{U}_{A\tilde{B}}^{A'\tilde{B}'}\xi_{A'\tilde{B}'}, \tag{2.34}
$$

where

$$
\begin{aligned}
\mathscr{U}_{A\tilde{B}}^{A'\tilde{B}'} = \tfrac{1}{2}\big[&(-1)^{g(A')\cdot[g(\tilde{B}) - g(\tilde{B}')]}\,\mathscr{U}_A^{A'}\mathscr{U}_{\tilde{B}}^{\tilde{B}'} \\
&+ (-1)^{g(A)\cdot g(B)}(-1)^{g(A')\cdot[g(\tilde{A}) - g(\tilde{B}')]}\,\mathscr{U}_B^{A'}\mathscr{U}_{\tilde{A}}^{\tilde{B}'}.
\end{aligned} \tag{2.35}
$$

We see that the result is formally identical to the corresponding purely class I case[4] [cf. also Eq. (3.22) of this paper]. Therefore, even though the two bases have their grades interchanged, they transform with identical matrices. This property follows from an analogous property of the fundamental class I and II representations [Eq. (2.32)]. From this result it follows that the character of the representation (2.35) is formally the same as the corresponding pure class I representation. The only difference is that in one case we have to sum over indices $\tilde{A}$ with the appropriate grade $g(\tilde{A})$ as opposed to summing over indices A with grade $g(A)$. When the dust settles, the two characters can differ from each other only by an overall minus sign. Finally, to calculate the dimensions we must compute the characters of $\mathscr{I}^{B}_{A} = (-1)^{g(A)}\delta^{B}_{A}$ or $\mathscr{I}^{\tilde{B}}_{\tilde{A}} = (-1)^{g(\tilde{A})}\delta^{\tilde{B}}_{\tilde{A}}$ for the appropriate representation. Clearly the two representations have the same dimension with the number of bosonic and fermionic components interchanged.

We find that the remarks of the previous paragraph, which apply explicitly to the example above, are actually more general. In this paper we will discuss one more example explicitly. Let us consider the representation constructed by taking the tensor product of one class I covariant basis ξ_A and one class II contravariant basis $\tilde{\xi}^{\tilde{B}}$. The basis $\xi^{\tilde{B}}_{A}$ of this representation is

$$\xi^{\tilde{B}}_{A} = \xi^{(1)}_{A}\tilde{\xi}^{(2)\tilde{B}} - \frac{\delta^{\tilde{B}}_{A}}{(N-M)}\xi^{(1)}_{C}\xi^{(2)\tilde{C}}. \tag{2.36}$$

Note again the formal similarity between this basis and the purely class I case of Eq. (2.17). The apparent difference in the second term can be explained if we note that for pure class I we could write the invariant subspace as $\xi^{(1)}_{C}\xi^{(2)C}(-1)^{g(C)\cdot g(C)}$ while for the new basis we would write $\xi^{(1)}_{C}\xi^{(2)\tilde{C}}(-1)^{g(C)\cdot g(\tilde{C})}$. We then remark that $(-1)^{g(C)\cdot g(C)} = (-1)^{g(C)}$ while $(-1)^{g(C)\cdot g(\tilde{C})} = 1$. Note that the basis ξ^{B}_{A}, constructed from two basis vectors of the same class, of Eq. (2.17), is *supertraceless*, while the basis $\xi^{\tilde{B}}_{A}$ of Eq. (2.36) is *traceless*. The transformation of the new basis is obtained by steps similar to those that led to Eq. (2.19). We find

$$\xi^{\tilde{B}}_{A} \rightarrow \xi'^{\tilde{B}}_{A} = \mathscr{U}^{\tilde{B}A'}_{A\tilde{B}'}\xi^{\tilde{B}'}_{A'}, \tag{2.37}$$

where $\mathscr{U}^{BA'}_{AB'}$ is again identical to Eq. (2.19) after taking Eqs. (2.18) and (2.37) into account. Using the symmetrization and antisymmetrization procedure and taking away traces, one can similarly construct the bases of higher representations from the direct products of fundamental class I and II representations.

One final remark before we conclude this section is that quadratic and higher-order Casimir invariants for the representations discussed here can be obtained using the methods of Sec. V of Ref. 4.

III. REPRESENTATIONS OF *P(N)* TYPE SUPERGROUPS

In the first part of this section we will briefly review the character formula for the representations of the Sp($2N$) group, which is related to P(N) type supergroups. The representations of the Sp($2N$) group are labeled by a partition into N parts: $\{n_1, n_2, \cdots, n_N\}$, where $n_1 \geqslant n_2 \geqslant \cdots \geqslant n_N \geqslant 0$. The character of the irreducible representation corresponding to this partition is[8]

$$X_{(\{n_1,n_2,\cdots n_N\})} = \tfrac{1}{2}\det\left(h_{n_i+i-j} + h_{n_i-i-j+2}\right), \tag{3.1}$$

where h_n is given by[4]

$$h_n = \oint \frac{dz}{2\pi i}\frac{z^{-n-1}}{\det(1-zS)} \tag{3.2a}$$

$$= \sum_{k_1,k_2,\cdots,k_n}\prod_{l=1}^{n}\frac{1}{k_l!}\left(\frac{\mathrm{Tr}S^l}{l}\right)^{k_l}\delta(n-k_1-2k_2-\cdots-nk_n). \tag{3.2b}$$

In the above expression the matrix S is the fundamental representation of Sp($2N$).

The antisymmetric tensor C_{ab} given as

$$(C_{ab}) = \begin{pmatrix} 0 & I_N \\ -I_N & 0 \end{pmatrix} \tag{3.3}$$

is the only invariant of the Sp($2N$) group. Hence *symmetrized* tensor products of the fundamental representation form irreducible representations of Sp($2N$). However, *antisymmetrized* tensor products of the fundamental representation contain invariant subspaces due to the existence of an invariant antisymmetric tensor. In particular, Eq. (3.1) gives the character of the representation $\{n,0,0,\cdots,0\}$ as h_n, which is the character of the completely symmetrized tensor product of n fundamental representations. To illustrate this point, we consider the basis ϕ_a of the fundamental representation which, under group, transforms into $S_{ab}\phi_b$. The symmetrized product ϕ_{ab} of two such bases, given as

$$\phi_{(ab)} = \phi^{(1)}_{a}\phi^{(2)}_{b} + \phi^{(2)}_{a}\phi^{(1)}_{b}, \tag{3.4}$$

transforms into

$$\phi'_{(ab)} = S_{(ab),(a'b')}\phi_{(a'b')}, \tag{3.5}$$

where

$$S_{(ab),(a'b')} = \tfrac{1}{2}[S_{aa'}S_{bb'} + S_{ab'}S_{ba'}]. \tag{3.6}$$

The character of this representation is then

$$\chi_{(2,0,\ldots)} = \tfrac{1}{2}[(\mathrm{Tr}S)^2 + \mathrm{Tr}S^2], \tag{3.7}$$

which is equal to h_2 by Eq. (3.2b). In the same way one can show that the character of the representation which transforms the completely symmetrized product of n ϕ_a's is h_n, in agreement with Eq. (3.1).

On the other hand, the antisymmetrized tensor product $\phi_{[a,b]}$ of two basis vectors, given as

$$\phi_{[a,b]} = \phi^{(1)}_{a}\phi^{(2)}_{b} - \phi^{(2)}_{a}\phi^{(1)}_{b}, \tag{3.8}$$

contains an invariant subspace $C_{ab}\phi_{[a,b]}$. Therefore, the basis ϕ_{ab} of the representation $(\{1,1,0,\ldots\})$ can be written as

$$\phi_{ab} = \phi_{[a,b]} - (C_{ab}/2N)C_{ef}\phi_{[e,f]}. \tag{3.9}$$

The basis (3.9) transforms as $\phi'_{ab} = S_{[a,b],[a',b']}\phi_{a'b'}$ with

$$S_{[a,b],[a',b']} = \tfrac{1}{2}[S_{aa'}S_{bb'} - S_{ab'}S_{ba'}] - C_{ab}C_{a'b'}/2N. \tag{3.10}$$

The trace of the matrix (3.10) is the character of this representation:

$$\chi_{(\{1,1,0,\ldots\})} = \tfrac{1}{2}[(\mathrm{Tr}S)^2 - \mathrm{Tr}S^2] - 1. \tag{3.11}$$

At this point we can express the above results in terms of the elementary symmetric functions a_n's defined as

$$a_n = \oint \frac{dz}{2\pi i}\frac{\det(1-zS)}{z^{n+1}}, \tag{3.12}$$

which can be rewritten in terms of h_m's as an $n \times n$

determinant,[4]

$$a_n = \det(h_{j-i+1}). \qquad (3.13)$$

The function $a_n(S)$ is the character of the representation which transforms the completely antisymmetrized tensor product of n ϕ_a's [without taking away any traces, of. Eq. (3.8)] Using Eq. (3.13), Eq. (3.11) can be rewritten as

$$\chi_{(1,1,0,\dots)} = a_2 - 1 = a_2 - a_0. \qquad (3.14)$$

Alternatively, Eqs. (3.1) and (3.13) give the character of the completely antisymmetric representation $\{1,1,\dots,n_p = 1, n_{p+1} = 0,\dots,0\}$ of Sp($2N$) as

$$\chi_{(1,1,\dots,0)} = \tfrac{1}{2}\det(h_{1+i-j} + h_{3-i-j}) = a_p - a_{p-2}, \qquad (3.14a)$$

i.e., a_{p-2} is the contribution coming from the invariant subspace which is subtracted from a_p, the character of the reducible, completely antisymmetric representation. In particular, for $p = 2$ one gets $\chi_{(1,1,0,\dots)} = a_2 - a_0$, which is the character of the representation (3.9) as shown above. Similarly, using the same symmetrization–antisymmetrization procedure, one can construct higher representations and verify that their characters are given by Eq. (3.1). In this procedure one starts with tensors symmetrized according to the rules of the Young tableaux and then takes away "traces" by contracting with the antisymmetric tensor C_{ab}. The matrix which transforms this basis gives an irreducible representation of Sp($2N$) and its character agrees with the previous results of Eq. (3.1).

Now we can extend the above results to the P(N) type supergroups. We denote the basis for the fundamental (defining) representation by ψ_A which contains N bosons and N fermions. Under the supergroup, ψ_A transforms into $\psi'_A = \mathscr{M}_{AB}\psi_B$, where $\mathscr{M}$ is the element of P(N) in the $2N$ dimensional fundamental representation. $\mathscr{M}$ is of the form

$$\mathscr{M} = \begin{pmatrix} \mathscr{A} & \mathscr{B} \\ \mathscr{C} & \mathscr{D} \end{pmatrix}, \qquad (3.15)$$

where $N \times N$ matrices $\mathscr{A}$ and $\mathscr{D}$ have commuting elements and $\mathscr{B}$ and $\mathscr{C}$ have anticommuting elements. The p conjugate of the matrix $\mathscr{M}$ is defined as[10]

$$\mathscr{M}^p = \begin{pmatrix} \mathscr{D}^T & -\mathscr{B}^T \\ \mathscr{C}^T & \mathscr{A}^T \end{pmatrix}. \qquad (3.16)$$

For P(N) type supergroups the group element $\mathscr{M}$ should satisfy the condition

$$\mathscr{M}\mathscr{M}^p = 1 \qquad (3.17)$$

together with Sdet $\mathscr{M} = 1$.

For matrices with fermionic elements the ordinary matrix transposition $(\mathscr{M}_1\mathscr{M}_2)^T$ is not equal to $\mathscr{M}_2^T\mathscr{M}_1^T$. However, supertransposition, defined as

$$\mathscr{M}^{ST} = \begin{pmatrix} \mathscr{A}^T & -\mathscr{C}^T \\ \mathscr{B}^T & \mathscr{D}^T \end{pmatrix}, \qquad (3.18)$$

satisfies $(\mathscr{M}_1\mathscr{M}_2)^{ST} = \mathscr{M}_2^{ST}\mathscr{M}_1^{ST}$. One can show that $\mathscr{M}^P$, defined by Eq. (3.16), can be rewritten as

$$\mathscr{M}^P = -C\mathscr{M}^{ST}C, \qquad (3.19)$$

where C is given by Eq. (3.3). Using (3.19) condition (3.17) takes the form

$$\mathscr{M}C\mathscr{M}^{ST} = C \qquad (3.20)$$

i.e., the antisymmetric tensor C_{AB} is an invariant of the P(N) supergroup. Hence one should eliminate invariant subspaces from "antisymmetrized" [in the sense of Eq. (2.12)] tensor products of ψ_A's.

The symmetrized tensor product of two basis vectors,

$$\psi_{(AB)} = \psi_A^{(1)}\psi_B^{(2)} + \psi_A^{(2)}\psi_B^{(1)}, \qquad (3.21)$$

transforms with the supergroup element

$$\mathscr{M}_{(AB),(A'B')} = \tfrac{1}{2}\{(-1)^{g(A')[g(B)-g(B')]}\mathscr{M}_{AA'}\mathscr{M}_{BB'} + (-1)^{g(A)g(B)}(-1)^{g(A')[g(A)-g(B)]}\mathscr{M}_{BA'}\mathscr{M}_{AB'}\}, \qquad (3.22)$$

The character $\mathscr{K}$ of this representation is then given by the supertrace of (3.22)

$$\mathscr{K} = \tfrac{1}{2}[(\mathrm{Str}\,\mathscr{M})^2 + \mathrm{Str}\,\mathscr{M}^2]. \qquad (3.23)$$

We repeat the same calculation for the symmetrized products of more than two vectors. The characters of such representations are expressed by the same formula as of Sp ($2N$), except traces are replaced by supertraces.

On the other hand, the antisymmetrized tensor product of two vectors,

$$\psi_{(A,B)} = \psi_A^{(1)}\psi_B^{(2)} - \psi_A^{(2)}\psi_B^{(1)}, \qquad (3.24)$$

is no longer irreducible. But $\psi_{[A,B]}$, defined as

$$\psi_{[A,B]} = \psi_{(A,B)} - (C_{AB}/2N)C_{CD}\psi_{(C,D)}, \qquad (3.25)$$

is. Under the supergroup $\psi_{[A,B]}$ transforms with

$$\mathscr{M}_{[A,B],[A',B']} = \tfrac{1}{2}\{(-1)^{g(A')[g(B)-g(B')]}\mathscr{M}_{AA'}\mathscr{M}_{BB'} - (-1)^{g(A)g(B)}(-1)^{g(A')[g(A)-g(B)]}\mathscr{M}_{BA'}\mathscr{M}_{AB'} - C_{AB}C_{A'B'}/2N\}. \qquad (3.26)$$

Upon calculating the supertrace we find the character to be

$$\mathscr{K}_{(1,1,\dots)} = \tfrac{1}{2}[(\mathrm{Str}\,\mathscr{M})^2 - \mathrm{Str}\,\mathscr{M}^2] - 1, \qquad (3.27)$$

which is formally the same as Eq. (3.14) for Sp($2N$), except traces are replaced by supertraces.

Similar calculations for other representations corresponding to various partitions are repeated both for Sp($2N$) and P(N). For the representations corresponding to the same partition, characters are given by the same formula, except that traces (or determinants) are replaced by supertraces (superdeterminants). We conclude that the character of the irreducible representation of the P(N) supergroups corresponding to the partition $\{n_1,n_2,\cdots\}$ should be

$$\mathscr{K}_{(n_1,n_2,\dots)} = \tfrac{1}{2}\det(H_{n_j+i-j} + H_{n_j-i-j+2}), \qquad (3.28)$$

where

$$H_n = \oint \frac{dz}{2\pi i}\,\frac{z^{-n-1}}{\mathrm{Sdet}(1 - z\mathscr{M})}. \qquad (3.29)$$

The dimensions of these representations, $D_{(n_1,n_2,\dots)}$, can again be found calculating the character of the matrix $\mathscr{J}$ of Eq. (2.33a). One gets

$$D_{(n_1,n_2,\dots)} = \tfrac{1}{2}\det(D_{n_j+i-j} + D_{n_j-i-j+2}), \qquad (3.30)$$

where

$$D_n = H_n(\mathscr{J}) = \oint \frac{dz}{2\pi i} z^{-n-1}\left(\frac{1+z}{1-z}\right)^N = \sum_{k=0}^{n}\binom{N}{k}\binom{-N}{n-k}. \qquad (3.31)$$

IV. DISCUSSION

This article, together with previous work,[4] exhausts all the finite dimensional representations of SU(N/M), OSP($N/2M$), and P(N) type supergroups which can be obtained from the direct product of fundamental representations using the symmetrization–antisymmetrization technique. This leaves out certain nonintegral representations of SU($N/1$) which cannot be obtained from the fundamental representation as well as the representations of $Q(N)$ type supergroups. Through our supertableau techniques we discovered the fact that as we generalize from Lie groups to supergroups the characters and other invariants of *higher representations* are obtained from those of ordinary Lie groups by replacing traces and determinants of the *fundamental representation* appearing in the invariants of the theory with supertraces and superdeterminants. To use this trick we first need to write the characters and other invariants of the ordinary Lie groups in terms of traces of various powers of the fundamental representation.[4] In this process we have found what seems to be a new expression for the characters of the ordinary group SU(N) [Eq. (2.30)]. A descripton of the representations of superalgebras is given in Ref. 6 in terms of highest weights. Our approach and our mathematical expressions are different and very convenient for physical applications.

A rigorous proof of irreducibility of our representations remains to be given. However, we want to outline some arguments towards a proof of irreducibility of those representations. For simplicity, we will consider only the representations of SU(N/M) obtained from class I covariant bases of fundamental representation. An extension of these arguments to mixed cases (covariant–contravariant, class I–class II) is straightforward.

We consider the decomposition SU(N/M)$\supset$SU(N) $\times$SU(M)$\times U(1)$ of the representations in question. For simplicity we assume $N > M$. The U(1) quantum numbers of the fundamental representation are assigned as

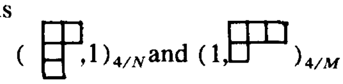
$$\begin{pmatrix} \mathscr{A} & 0 \\ 0 & \mathscr{D} \end{pmatrix}\begin{pmatrix} \phi_a \\ \psi_\alpha \end{pmatrix},$$

where $\mathscr{A} = (1/N)I_N$, $\mathscr{D} = (1/M)I_M$, with $I_N(I_M)$ being the $N\times N$ $(M\times M)$ identity matrix.

Let us assume that the representation we want to consider is associated with a supertableau containing k boxes. This representation will be reduced to SU(N)$\times$SU(M)$\times U(1)$ representations, which we will symbolically denote as $\Sigma \oplus (\mathrm{SU}(N), \mathrm{SU}(M))_{\mathrm{U(1)\text{-}quantum\ number}}$. An example is given in Fig. 5. The total number of boxes of SU(N) and SU(M) Young tableaux will be k in each of these direct product representations. Successive components in this decomposition will have one less (more) box in the SU(N) (SU(M)) tableau. They can be obtained by starting from the first (nonvanishing) representation which contains the largest number of SU(N) boxes and by applying to it a *supergenerator* which can be thought of as a step-down operator for SU(N) boxes and a step-up operator for SU(M) boxes. Each time the supergenerator acts, it moves the U(1) quantum number by $(-1/N + 1/M)$. This is because each SU(N) box gets $1/N$ for the U(1) quantum number while each SU(M) box gets $1/M$.

If the given representation were reducible into smaller representations *described by fewer boxes*, it would be impossible to find the same SU(N) and SU(N) content with the same U(1) quantum numbers. This is because SU(N) and SU(M) representations, occuring in the decomposition, automatically receive their U(1) quantum numbers from the number of boxes they contain: A component that contains n boxes for SU(N) and m boxes SU(M) (such that $m + n = k$) receives a U(1) quantum number equal to $n/N + m/M$. *Therefore, it is impossible to decompose any one of our representations into representations containing fewer boxes.*

There remains to consider the possibility of reducibility to other representations associated with supertableaux, each containing the *same* number of boxes as the supertableau of the orginal representation in question. We consider the decomposition of the original representation into SU(N) $\times$SU(M)$\times$U(1) and concentrate on the components with the minimum or the maximum value of the U(1) quantum number. In the example of Fig. 5 these are the representations



We call these the maximal and the minimal representations, respectively. They contain the maximum and the minimum values of the U(1) quantum number. These extremal representations are easily obtained in our approach. For a class I representation, if all the fermions in the fundamental representation ξ_A are set equal to zero the supertableau reduces to the SU(N) tableau of the same shape which describes the maximal representation. On the other hand, if all bosons are set equal to zero, we obtain the minimal representation, the Young tableau of which is obtained from the supertableau by reflecting it along the diagonal. If either the maximal or the minimal (or both) representations were unique, then it is impossible to reduce the original representations into smaller representations associated with supertableaux containing the same number of boxes as the original one. This is because the whole representation can be constructed by applying the supergenerators on either the maximal or the minimal representation.

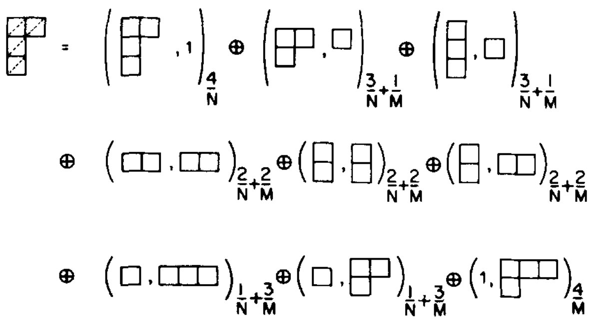


FIG. 5. The decomposition of the representation (···,0,0;2,1,1,0,···) of the supergroup SU(N/M) into the represenatations of SU(N)$\times$SU(M)$\times$U(1). In each parenthesis the first Young tableau corresponds to the SU(N) and the second tableau to the SU(M) representation. The U(1) quantum number of each direct product representation appears outside the parenthesis at the lower right hand corner.

What is left is to find out if all of our representations contain a unique maximal or minimal representation. In our construction both maximal representations are clearly unique for supertableaux with less than $N+1$ columns and less than $M+1$ rows. Therefore, *these representations are irreducible*. The representations containing $M+1$ or more rows but less than $N+1$ columns have a clearly unique maximal representation, and similarly there is a unique minimal representation with the roles of N and M interchanged. Thus, *these are also irreducible representations*. There remain the representations containing $N+1$ or more rows and columns. For such representations, if we set all fermions in ξ_A equal to zero the SU(N) Young tableau of the same shape vanishes. Similarly if all bosons in ξ_A are set equal to zero the SU(M) reflected Young tableaux vanishes. Therefore, the maximal and minimal representations cannot be directly obtained from the supertableau. We must now try to determine them by decomposing the superrepresentation to SU(N)$\times$SU(M)$\times$U(1) and finding the components with maximum and minimum values of the U(1) quantum number. If the maximal representations obtained in this way are unique then the superrepresentation will be irreducible. In each case in which we are able to perform these operations we found that indeed the maximal representations were unique. Thus, we believe that all the representations we discussed (including mixed cases) satisfy this property. However, at this point we do not have a rigorous proof of this last point. When this is established it will complete the proof of irreducibility of all the representations we have discussed.

[1]P. G. O. Freund and I. Kaplansky, J. Math. Phys. **17**, 228 (1976); V. G. Kac, Commun. Math. Phys. **53**, 31 (1977). V. Rittenberg, "A Guide to Lie Superalgebras" in *Lecture Notes in Physics* 79, edited by P. Kramer and A. Rieckers (Springer-Verlag, Berlin, 1978); M. Scheunert, *The Theory of Lie Superalgebras, An Introduction*; *Lecture Notes in Mathematics*, **716** (Springer-Verlag, Berlin, 1979).

[2]L. Corwin, Y. Ne'eman, and S. Stenberg, Rev. Mod. Phys. **47**, 573 (1975); *Supergravity*, P. VanNieuwenhuizen and D. Z. Freedman, edited by (North-Holland, Amsterdam, 1979).

[3]A. B. Balantekin, I. Bars, and F. Iachello, Yale preprint; for a preliminary version of this work see I. Bars, *Dynamical Supersymmetry* SU(6/4) *in Nuclei*, Invited talk presented at (a) Vth Johns Hopkins Workshop on Current Problems in High Energy Particle Theory, Bad Honeff, Germany, June 1980; (b) IXth International Colloquium on Group Theoretical Methods in Physics, Cocoyac, Mexico, June 1980. Also, Proceedings of the workshop "Interacting Bose–Fermi Systems in Nuclei", Erice, Sicily, June 1980, edited by F. Iachello, to be published by Plenum, New York.

[4]A. B. Balantekin and I. Bars, Yale preprint YTP 80-06, May 1980; J. Math. Phys. **22**, 1149 (1981).

[5]Y. Ne'eman and S. Sternberg, Proc. Natl. Acad. Sci. (USA) **77**, 3127 (1980); P. H. Dondi and P. D. Jarvis, Z. Phys. C. **4**, 201 (1980); P. H. Dondi and P. D. Jarvis, University of Tasmania preprint, May 1980.

[6]V. G. Kac, "Representation of Classical Lie Superalgebras" in *Differential Geometrical Methods in Mathematical Physics*, edited by K. Bleuler, H. R. Petry, and A. Reetz (Springer-Verlag, Berlin, 1978).

[7]M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 155 (1977); M. Marcu, *ibid.* **21**, 1277 (1980).

[8]H. Weyl, *The Classical Groups* (Princeton U. P., Princeton, NJ, 1946).

[9]R. C. King, J. Math. Phys. **11**, 280 (1970); J. Abramsky and R. C. King, Nuovo Cimento, A**67**, 153 (1970); N. El Samra and R. C. King, J. Phys. A **12**, 2305 (1979).

[10]V. G. Kac, *Some Remarks on Lie Superalgebra Theory*, MIT preprint (unpublished); see also reference on the entry 1.

Incompressible fluid turbulence at large Reynolds numbers: Theoretical basis for the t^{-1} decay law and the form of the longitudinal correlation function

Gerald Rosen[a)]
Department of Physics, Drexel University, Philadelphia, Pennsylvania 19104

(Received 24 November 1980; accepted for publication 18 March 1981)

Approximately valid for large values of the time t, a formal solution to the Hopf Φ equation is obtained here as an asymptotic power series in t^{-1}. This approximate solution is directly applicable to grid-generated isotropic homogeneous turbulence at large Reynolds numbers during the initial (inertial-force dominated) period of decay; thus, the solution accounts for the observed t^{-1} decay law and the fact that the longitudinal correlation function f is independent of t. It is observed that the longitudinal correlation function measured by Frenkiel, Klebanoff, and Huang is consistent with the theoretical asymptotic behavior $f \doteq (\text{const}) r^{-3}$ as $r \to \infty$ and fitted by the expression $f = [1 + 0.770(r/M)]^{-3}$, where M is the grid mesh length and the separation distance r is greater than the Taylor microscale $(10\nu t)^{1/2}$. Interestingly enough, this form for the longitudinal correlation function is shown to be derivable from a variational principle.

PACS numbers: 47.25.Cg, 47.10 + g, 47.30. + s

I. INTRODUCTION

Let $u = [u_1(x,t), u_2(x,t), u_3(x,t)]$ denote the velocity field of an incompressible fluid governed by the Navier–Stokes equation

$$\frac{\partial u}{\partial t} = -u\cdot\nabla u + \nu\nabla^2 u - \rho^{-1}\nabla p, \tag{1}$$

in which ν, ρ are positive constants. For boundary-free flow with $x = (x_1, x_2, x_3)$ in R_3, the incompressibility condition $\nabla\cdot u = 0$ can be used to eliminate the pressure term from (1); the resulting integro-differential equation

$$\frac{\partial u}{\partial t} = -(u\cdot\nabla u)^{\rm tr} + \nu\nabla^2 u \equiv Q(u) \tag{2}$$

features the *transverse* (solenoidal) part of the inertial term, where for any vector field in R_3,

$$\begin{aligned} v^{\rm tr}(x) &\equiv v(x) - \nabla\{\nabla^{-2}[\nabla\cdot v(x)]\} \\ &\equiv v(x) + \frac{1}{4\pi}\nabla\int\frac{\nabla\cdot v(x')}{|x-x'|}\,d^3x'. \end{aligned} \tag{3}$$

A statistical state of incompressible fluid turbulence is described by a Gibbsian ensemble of solenoidal velocity fields that evolve dynamically according to (2). All equal-time multipoint velocity correlation tensors are contained in the complex-valued Fourier transform of the probability measure, the *Hopf characteristic functional*[1]

$$\begin{aligned} \Phi(y,t) \equiv 1 &+ i\int\langle u_j(x',t)\rangle y_j(x')\,d^3x' \\ &- \tfrac{1}{2}\int\langle u_j(x',t)u_k(x'',t)\rangle y_j(x')y_k(x'')\,d^3x'\,d^3x'' \\ &- \frac{i}{6}\int\langle u_j(x',t)u_k(x'',t)u_l(x''',t)\rangle \\ &\times y_j(x')y_k(x'')y_l(x''')\,d^3x'\,d^3x''\,d^3x''' + \cdots. \end{aligned} \tag{4}$$

In Eq. (4), the real parameter field $y = [y_1(x), y_2(x), y_3(x)]$ is required to be continuous, differentiable and in $L^2(R_3)$ (i.e., $\|y\|^2 \equiv \int y\cdot y\,d^3x < \infty$) but is otherwise arbitrary and disposable. Since the correlation tensors inherit the solenoidal quality of u, the characteristic functional depends exclusively on the transverse part of y: $\Phi[y,t] \equiv \Phi[y^{\rm tr},t]$. The reality and non-negativity of the normalized probability measure implies that $\Phi[y,t]^* \equiv \Phi[-y,t]$ and $|\Phi[y,t]| \leqslant 1$. Furthermore, since all u satisfy Eq. (2), it follows that Φ satisfies the time-evolution equation derived by Hopf[1]

$$\begin{aligned} \frac{\partial\Phi}{\partial t} &= i\int y_j(x)Q_j\left(\frac{\delta}{i\delta y(x)}\right)d^3x\,\Phi \\ &= \int\left(iy_j^{\rm tr}(x)\frac{\delta}{\delta y_k(x)}\nabla_k\frac{\delta\Phi}{\delta y_j(x)} \right. \\ &\quad \left. + \nu y_j(x)\nabla^2\frac{\delta\Phi}{\delta y_j(x)}\right)d^3x, \end{aligned} \tag{5}$$

in which $\delta/\delta y_j(x)$ denotes the Volterra functional derivative[2] with respect to $y_j(x)$.

The Φ equation [Eq. (5)] puts the theoretical problem of incompressible fluid turbulence in a nutshell: *To determine all experimentally measurable velocity correlation tensors embodied by definition in Eq.* (4), *obtain the physically relevant solution to the single functional differential Eq.* (5).

II. FORMAL ASYMPTOTIC SERIES SOLUTION TO Φ EQUATION

Suppose that the desired solution to Eq. (5) depends parametrically on the constant velocity parameter U and the constant length parameter M, quantities associated with the turbulence-generating mechanism. Then for values of the time such that $t \gg (M/U)$, the characteristic functional is given to order $t^{-(N-1)}$ (units $U = M = 1$) by the asymptotic power series in t^{-1}

$$\Phi = 1 + \sum_{n=0}^{N-2} n!\,t^{-(n+1)}(\Omega_\nu)^{N-n-2}\Gamma + O(t^{-N}), \tag{6}$$

a)This work was supported by NASA grant NAG1-110.

 © 1981 American Institute of Physics

in which there appears the linear operator

$$\Omega_\nu \equiv -i\int y_j^{\rm tr}(x)\frac{\delta}{\delta y_k(x)}\nabla_k\frac{\delta}{\delta y_j(x)}\,d^3x - \nu\int y_j(x)\nabla^2\frac{\delta}{\delta y_j(x)}\,d^3x, \tag{7}$$

and where the functional $\Gamma = \Gamma(y) \equiv \Gamma(y^{\rm tr})$ is independent of t and satisfies the condition

$$(\Omega_\nu)^{N-1}\Gamma = 0. \tag{8}$$

To see this, observe that Eq. (5) becomes

$$\frac{\partial\Phi}{\partial t} + \Omega_\nu\Phi = 0 \tag{9}$$

in terms of the definition in Eq. (7). The left-hand side of Eq. (9) works out to yield

$$\frac{\partial\Phi}{\partial t} + \Omega_\nu\Phi = O(t^{-N}) \tag{10}$$

by direct substitution of Eq. (6) and use of the condition in Eq. (8). Hence, by virtue of Eq. (10), the Φ equation [Eq. (9)] is satisfied to order $t^{-(N-1)}$ by the asymptotic series in Eq. (6), thereby confirming validity of the latter asymptotic series in powers of t^{-1}.

From Eq. (6) one obtains the N-term approximation for the characteristic functional

$$\Phi \cong 1 + \sum_{n=0}^{N-2} n!\,t^{-(n+1)}(\Omega_\nu)^{N-n-2}\Gamma \tag{11}$$

for sufficiently large values of t. The optimum choice for the disposable integer N depends on the solution to Eq. (8) for Γ and the practical computational complexities that enter for large values of N. For t ranging from 20 to 40 [as is customary in wind and water tunnel measurements for the initial period of decay,[3-5] where $(t + 10)$ is the approximate distance downstream from the turbulence-generating grid in units $U = M = 1$], the numerical coefficient $n!\,t^{-(n+1)}$ is of order 10^{-5} for $n = 3$, and thus the asymptotic series in Eq. (11) with $N = 4$ would ordinarily be quite adequate in practical applications.

A suitable functional $\Gamma = \Gamma(y)$ must be prescribed in accordance with Eq. (8) in order to make Eq. (11) an explicitly useful approximate solution. Consider an *ansatz* of the $N^{\rm th}$ order polynomial form

$$\Gamma = \sum_{M=1}^{N}\frac{(i)^M}{M!}\int \gamma^{(M)}_{j_1\cdots j_M}(x_{(1)},\dots,x_{(M)})y_{j_1}(x_{(1)})\cdots \times y_{j_M}(x_{(M)})\,d^3x_{(1)}\cdots d^3x_{(M)}, \tag{12}$$

in which the $\gamma^{(M)}$'s are totally symmetric and solenoidal tensors. Because the first integral operator in Eq. (7) reduces the homogeneity order of a functional of y by one while the second integral operator leaves the homogeneity order unchanged, Eq. (12) produces another Nth order polynomial in y on the left-hand side of Eq. (8), and thus Eq. (8) provides precisely N tensorial conditions on the N $\gamma^{(M)}$'s. In the important special case treated below, the latter N conditions are all satisfied identically.

III. APPROXIMATE SOLUTION FOR GRID-GENERATED ISOTROPIC HOMOGENEOUS TURBULENCE AT LARGE REYNOLDS NUMBERS

In cases for which the turbulence is homogeneous and isotropic in space, the probability measure over the Gibbsian ensemble is invariant under translations and rotations of the spatial coordinates in a Galilean frame for which the mean velocity vanishes, $\langle u(x,t)\rangle = 0$. Moreover, isotropy, homogeneity, and incompressibility imply that the quadratic velocity correlation tensor has the generic form[3]

$$\begin{aligned}\langle u_j(x',t)u_k(x'',t)\rangle &= u^2(t)\left[\left(f + \frac{r}{2}\frac{\partial f}{\partial r}\right)\delta_{jk} - \frac{1}{2r}\frac{\partial f}{\partial r}x_jx_k\right]\\ &= \tfrac{1}{2}u^2(t)\left(\delta_{jk} - \frac{\partial^2}{\partial x_j\,\partial x_k}\nabla^{-2}\right)\left(3f + r\frac{\partial f}{\partial r}\right),\end{aligned} \tag{13}$$

where $u^2(t) \equiv \langle u_1(x,t)^2\rangle$, $x_j \equiv x_j' - x_j''$, $r \equiv (x_kx_k)^{1/2}$, and the *longitudinal correlation function* $f = f(r,t)$ is normalized to give $f(0,t) \equiv 1$. The final member of Eq. (13) equals the second because

$$-\nabla^{-2}\left(3f + r\frac{\partial f}{\partial r}\right) = \int_r^\infty sf(s,t)\,ds, \tag{14}$$

as verified by applying $\nabla^2 = r^{-2}(\partial/\partial r)r^2(\partial/\partial r)$ to both sides. Hence, the quadratic term in Eq. (4) is expressible as

$$\Phi^{(2)} = -\tfrac{1}{4}u^2(t)\int\left(3f + r\frac{\partial f}{\partial r}\right)y^{\rm tr}(x')\cdot y(x'')\,d^3x'\,d^3x'' \tag{15}$$

by substituting the final member of Eq. (13) into Eq. (4). Experiments give the empirical t^{-1} decay law[3]

$$u^2(t) = UM/at \tag{16}$$

for the initial period of grid-generated isotropic homogeneous turbulence at large Reynolds numbers, where U denotes the mean fluid flow speed, M denotes the grid mesh length, and a is an absolute numerical constant (of order 10^2, depending on the Reynolds number UM/ν). In view of Eqs. (15) and (16), the characteristic functional Eq. (4) becomes

$$\Phi(y,t) = 1 - \frac{UM}{4at}\int\left(3f + r\frac{\partial f}{\partial r}\right)y^{\rm tr}(x')\cdot y(x'')\,d^3x'\,d^3x'' + O(y^3). \tag{17}$$

Now although the semilinear integro-differential Navier-Stokes equation [Eq. (2)] is parabolic (with the highest-order x derivatives of u appearing in the viscous-force term $\nu\nabla^2 u$), the Hopf equation [Eq. (5)] is "hyperbolic" in the sense that the highest-order functional derivatives of Φ appear in the inertial-force term. Thus, owing to the subordinate differential structure of the viscous-force term, Eq. (5) can be expected to admit physically relevant solutions which are continuous and analytic in the parameter ν about $\nu = 0$. It follows that the inviscid ($\nu = 0$) specialization of Eq. (11) ought to apply to inertial-force dominated varieties of turbulence at large Reynolds numbers. Consider, in particular, grid-generated isotropic homogeneous turbulence during the initial period of decay and at large grid Reynolds numbers, i.e., such that $\nu < UM \times 10^{-4}$. Then Eq. (17) and the $\nu \to 0$ limit of Eq. (11) are in natural correspondence if f is independent of t and the terms of order to t^{-1} are equal, viz.,

$$(\Omega_0)^{N-2}\Gamma = -\frac{UM}{4a}\int\left(3f + r\frac{\partial f}{\partial r}\right)y^{\rm tr}(x')\cdot y(x'')\,d^3x'd^3x''. \qquad (18)$$

Indeed, experiments have shown that f is independent of t during the initial period of decay at large Reynolds numbers [see Appendix A, Eq. (A5)]. Furthermore, Eq. (18) is consistent with Eq. (8), as seen by applying

$$\Omega_0 = -i\int y_j^{\rm tr}(x)\frac{\delta}{\delta y_k(x)}\nabla_k\frac{\delta}{\delta y_j(x)}\,d^3x \qquad (19)$$

to both sides of Eq. (18) and observing that

$$\frac{\delta}{\delta y_k(x)}\nabla_k\frac{\delta}{\delta y_j(x)}\int\left(3f + r\frac{\partial f}{\partial r}\right)y^{\rm tr}(x')\cdot y(x'')\,d^3x'd^3x''$$

$$= 2\int\left[\left[\nabla_k\left(3f + r\frac{\partial f}{\partial r}\right)_{r=|x-x'|}\right]\right]\delta_{jk}^{\rm tr}(x-x')\,d^3x' \equiv 0, \qquad (20)$$

where $\delta_{jk}^{\rm tr}(x) \equiv \delta_{jk}\delta(x) - \nabla_j\nabla_k\nabla^{-2}\delta(x)$ is the transverse part of the three-dimensional Dirac function. Because the integral operator [Eq. (19)] reduces the homogeneity order of a functional of y by 1, the operator $(\Omega_0)^{N-2}$ reduces the homogeneity-order by $(N-2)$. Thus, the functional of homogeneity-order two on the right-hand side of Eq. (18) is obtainable from the single-term specialization of Eq. (12)

$$\Gamma = \frac{(i)^N}{N!}\int\gamma^{(N)}_{j_1\cdots j_N}(x_{(1)},\dots,x_{(N)}) \times y_{j_1}(x_{(1)})\cdots y_{j_N}(x_{(N)})\,d^3x_{(1)}\cdots d^3x_{(N)}. \qquad (21)$$

Here $\gamma^{(N)}$ is invariant under the Euclidean group of spatial rotations and translations, in addition to being totally symmetric and solenoidal. By comparison with Eq. (4) one obtains

$$\langle u_{j_1}(x_{(1)})\cdots u_{j_N}(x_{(N)})\rangle = (N-2)!\,t^{-(N-1)}\gamma^{(N)}_{j_1\cdots j_N}(x_{(1)},\dots,x_{(N)}) \qquad (22)$$

from the term in Eq. (11) with $n = N-2$. Moreover, the n-point velocity correlation tensor for $n < N$ is proportional to $t^{-(n-1)}$ times a (contracted, confluenced, and differentiated) concomitant of $\gamma^{(N)}$, according to Eqs. (4), (11), (19), and (21). That the triple velocity correlation tensor decays as t^{-2} during the initial period is consistent with the Kármán-Howarth equation for f independent of t (see Appendix B). In fact, the connection between the n-point and $(n+1)$-point velocity correlation tensors [implied by the iterated operator $(\Omega_\nu)^{N-n-2}$ in Eq. (11)] can be viewed as a suitably generalized version of the Kármán-Howarth connection-equation between the quadratic and triple velocity correlation tensors.

Hence, the experimentally confirmed solution for grid-generated isotropic homogeneous turbulence is given by Eq. (11) with $\nu = 0$ for the initial period of decay at large grid Reynolds numbers. In turn, this provides a theoretical explanation for the observed t^{-1} decay law [Eq. (16)] and the fact that f is independent of t. The N-point tensor $\gamma^{(N)}$ in Eq. (21) remains disposable (modulo required symmetry), and thus an additional statistical condition is required to fix the form of the longitudinal correlation function f in Eq. (17).

IV. VARIATIONAL PRINCIPLE FOR THE LONGITUDINAL CORRELATION FUNCTION

Consider the functional

$$\mathfrak{B}(f) \equiv \int_0^\infty\left[\tfrac{1}{2}\left(\frac{df}{dr}\right)^2 + \alpha M^{-2}f^\alpha\right]dr \qquad (23)$$

on the non-negative longitudinal correlation function $f = f(r)$, where α in Eq. (23) denotes an adjustable positive numerical constant and M is the grid mesh-length constant. During the initial period of decay at large Reynolds numbers, the form of f is such that $\mathfrak{B}(f)$ defined by Eq. (23) is a minimum. That is, the observed longitudinal correlation function is derivable from the variational principle

$$\delta\mathfrak{B}(f) \equiv \mathfrak{B}(f+\delta f) - \mathfrak{B}(f) = 0, \qquad (24)$$

in which δf vanishes at $r = 0$ and ∞ but is an otherwise arbitrary continuous small variation, and admissible f satisfy the normalization condition $f(0) = 1$ and the generally required asymptotic behavior [see (A4) in Appendix A]

$$f(r) \doteq Ir^{-3} \quad \text{as } r\to\infty \qquad (25)$$

with I denoting a positive constant.

Proof of Eq. (24): The Euler-Lagrange equation equivalent to Eq. (24) follows from Eq. (23) as

$$\frac{-d^2f}{dr^2} + \alpha^2M^{-2}f^{\alpha-1} = 0 \qquad (26)$$

and admits the first integral

$$\left(\frac{df}{dr}\right)^2 = 2\alpha M^{-2}f^\alpha \qquad (27)$$

with the constant of integration equal to zero by virtue of the boundary condition $f(\infty) = 0$. Substitution of the required asymptotic form [Eq. (25)] into Eq. (27) shows that $\alpha = \frac{8}{3}$, and hence the square root of Eq. (27) is

$$\frac{df}{dr} = \pm(4/\sqrt{3}\,M)f^{4/3}. \qquad (28)$$

With the minus sign required in order to satisfy both boundary conditions, the admissible solution to Eq. (28) is

$$f = [1 + (4/3\sqrt{3})(r/M)]^{-3} = [1 + (0.7698)(r/M)]^{-3}, \qquad (29)$$

a result in close agreement with experimental data [see (A5) and the discussion in Appendix A].

Therefore, the variational principle in Eq. (24), with $\alpha = \frac{8}{3}$ in the integral in Eq. (23), is confirmed by experimental observation. One would expect that this variational principle arises physically from the (nondissipative) transfer of energy to higher wave-number components of the flow by the nonlinear inertial force, and this is indeed a matter for future investigation.

APPENDIX A: FORM OF THE LONGITUDINAL CORRELATION FUNCTION DURING THE INITIAL PERIOD OF DECAY

The existence of a normalized probability measure which varies continuously over velocity fields implies that

TABLE I. Comparison of experimental values for the longitudinal correlation function [Ref. 4, Fig. 2, $f = R(r/U)$] with values given by Eq. (A5).

r/M	0	0.10	0.20	0.30	0.40	0.60	1.00	1.60	2.00	2.40	2.80	3.20
f(Ref. 6)	1	0.80	0.65	0.52	0.45	0.32	0.19	0.09	0.06	0.04	0.03	.02–.03
f[Eq. (A5)]	1	0.800	0.651	0.536	0.447	0.320	0.180	0.090	0.061	0.043	0.032	0.024

quantity (4) satisfies the Fourier interference inequality[1,6] $|\Phi[y,t]| < 1$ for all admissible y such that $y^{tr} \neq 0$ (or equivalently, such that $|\text{curl}\, y| \neq 0$). Clearly, this requires Eq. (15) to be finite and negative definite in y^{tr}, since $[\Phi(y,t) - 1 - \Phi^{(2)}] = O(\|y^{tr}\|^3)$ for parameter fields with $\|y^{tr}\|$ sufficiently small in magnitude. The r dependence in f involves a physical decay distance L, which is either the mesh length M or the Taylor microscale[3] λ in the case of grid-generated turbulence. Hence, for a parameter field that is quasiconstant over the distance L [i.e., $(\nabla y^{tr}:\nabla y^{tr}) \ll L^{-2} y^{tr}\cdot y^{tr}$ for all x], Eq. (15) becomes

$$\Phi^{(2)} \cong -\pi u^2(t) I(t) \int y^{tr}(x'')\cdot y(x'')\, d^3x'', \tag{A1}$$

in which

$$\begin{aligned} I(t) &\equiv \frac{1}{4\pi}\int\left(3f + r\frac{\partial f}{\partial r}\right) d^3x = \int_0^\infty \left(3f + r\frac{\partial f}{\partial r}\right) r^2 dr \\ &= \int_0^\infty \frac{\partial}{\partial r}(r^3 f)\, dr = \lim_{r\to\infty} [r^3 f(r,t)]. \end{aligned} \tag{A2}$$

The latter quantity (A2) must be finite and positive in (A1) since

$$\int y^{tr}(x'') \cdot y(x'')\, d^3x'' = \|y^{tr}\|^2 \tag{A3}$$

is positive definite in y^{tr}. Therefore, the final member of (A2) requires the asymptotic form

$$f(r,t) \doteq I(t) r^{-3} \text{ as } r \to \infty \tag{A4}$$

with $I(t)$ finite and positive.

An immediate consequence of the general result (A4) is that the formal Loitsyansky invariant $\int_0^\infty r^4 f(r,t)\, dr$ is a divergent integral, as conjectured many years ago by Birkhoff.[7] From the nonexistence of the Loitsyansky invariant it follows[3,7] that the energy spectrum $E(\kappa,t)$ cannot be expanded as a power series in the wavenumber κ about $\kappa = 0$. Hence, it cannot be demonstrated by power series analysis[3] that the large eddies (i.e., small κ flow components) retain their initial amplitude and store of energy during the decay of the turbulence.

To check the consistency of (A4) with experiment, consider the initial period measurements of Frenkiel *et al.*[4] at Reynolds numbers UM/ν from 12 800 to 81 000 and typical turbulence levels $u/U \sim 0.02$ in air and water, with U denoting the mean fluid flow speed. There is a beautiful universality to this experimental data for values of r greater than the Taylor microscale $\lambda = (10\nu t)^{1/2}$, since $f(\cong R(r/U)$ by Taylor's equivalence approximation[5] with $R(h)$ as shown in Fig. 2 of Ref. 6] is observed to be independent of t and to depend exclusively on the dimensionless geometrical ratio (r/M). Table I shows that

$$f(r,t) = [1 + 0.770(r/M)]^{-3} \tag{A5}$$

is a valid representation of the data for $r > \lambda$ ($\sim 0.05M$). Clearly, the theoretical result (A4) is consistent with formula (A5) for the initial period at large Reynolds numbers,[8] and in this case one finds that $I(t) = 2.19\, M^3$. As shown in Sec. IV, the form (A5) is derivable from Eq. (23) and the variational principle Eq. (24), in combination with (A4) and the normalization condition $f(0,t) = 1$.

TABLE II. Comparison of experimental estimates for $-k$ [Ref. 4, Fig. 3, wind tunnel data having $t \cong 38.5\,(M/U)$ and $a = 74.1$] with values given by Eq. (B5).

r/M	0	0.30	0.60	0.80	1.00	1.20	1.40	1.60	1.80
$-k$ (Ref. 6)	0	0.037	0.047	0.048	0.047	0.044	0.036	0.030	0.026
$-k$ [Eq. (B5)]	0	0.045	0.053	0.053	0.050	0.047	0.043	0.040	0.037

APPENDIX B: CONSISTENCY WITH THE KÁRMÁN-HOWARTH EQUATION

For values of r greater than the Taylor microscale λ, the viscous-force term is relatively negligible and thus the Kármán-Howarth equation[3,7,9] becomes

$$\frac{\partial}{\partial t}[u^2(t) f] \cong K(r,t). \tag{B1}$$

Here $u^2(t)$ and f are as defined in Eq. (13), and $K(r,t)$ appears in the contracted and confluenced triple velocity correlation tensor

$$\langle u_k(x',t) u_j(x'',t) u_k(x'',t)\rangle \equiv \tfrac{1}{2} K(r,t)(x'_j - x''_j), \tag{B2}$$

in which $r \equiv |x' - x''|$. By virtue of the initial period decay law [Eq. (16)] and f being independent of t, (B1) implies that

$$K(r,t) \cong -UMf/at^2 \tag{B3}$$

for $r > \lambda = (10\nu t)^{1/2}$. Hence, the triple velocity correlation tensor (B2) decays as t^{-2} during the initial period at large Reynolds numbers.

The relation (B3) can be checked against experimental data. Introducing the auxiliary scalar function $k = k(r,t)$ which satisfies (see Ref. 3, pp. 53 and 100)

$$K(r,t) = [u^2(t)]^{3/2}\left(\frac{\partial k}{\partial r} + \frac{4}{r}k\right) \tag{B4}$$

and making use of Eqs. (16), (B3), and (A5), one obtains

$$\begin{aligned} -k(r,t) &\cong \left(\frac{a}{UMt}\right)^{1/2} r^{-4} \int_0^r s^4 f(s)\, ds \\ &\cong \frac{1}{5}\left(\frac{aM}{Ut}\right)^{1/2} \left(\frac{r}{M}\right)[1 + 0.770(r/M)]^{-3}. \end{aligned} \tag{B5}$$

In view of the experimental definition[4] and Taylor's equivalence approximation,[5] the left-hand side of (B5) is roughly equal to the third-order correlation function in time,

$$-k(r,t) \simeq \mathscr{R}^{2,1}(h) \quad \text{for} \quad r = Uh, \tag{B6}$$

although with an accuracy significantly less than that in the Taylor approximation $f \cong R(r/U)$ above (A5) [because the third-order correlation function on the right-hand side of (B6) is the *difference* between two experimentally measured quantities of the same sign and nearly-equal magnitudes]. Table II shows the $-k$ values given by (B6) and wind tunnel measurements [Fig. 3 of Ref. 4, with $t \cong 38.5(M/U)$ and $a = 74.1$] along with the corresponding theoretical values given by the final member of (B5). In Table II, the approximate agreement to two places beyond the decimal point is within the expected accuracy. Hence, (B3) is corroborated provisionally by experiment.

[1]E. Hopf, J. Ratl. Mech. Anal. **1**, 87 (1952); E. Hopf and E. W. Titt, J. Ratl. Mech. Anal. **2**, 587 (1953); G. Rosen, Phys. Fluids **3**, 519, 525 (1960); I. Hosokawa, J. Math. Phys. **8**, 221 (1967).

[2]For example, G. Rosen, in *Formulations of Classical and Quantum Dynamical Theory* (Academic, New York, 1969), pp. 96–100.

[3]G. K. Batchelor, in *Homogeneous Turbulence* (Cambridge University, New York, 1960), pp. 46–53, 88–92, 99–100, 134–135, 150–151.

[4]F. N. Frenkiel, P. S. Klebanoff, and T. T. Huang, Phys. Fluids **22**, 1606 (1979).

[5]G. I. Taylor, Proc. R. Soc. London, Ser. A **164**, 478 (1938)

[6]B. V. Gnedenko and A. N. Kolmogorov, in *Limit Distributions for Sums of Independent Random Variables* (Addison-Wesley, Cambridge, Mass., 1954), pp. 45–61.

[7]G. Birkhoff, Commun. Pure App. Math. **7**, 19 (1954).

[8]Manifest in the Frenkiel-Klebanoff-Huang data for $r \gg \lambda$, the general asymptotic form (A4) is expected to show up at correspondingly larger values of r in experiments that feature larger values of λ [e.g., S. C. Ling and T. T. Huang, Phys. Fluids **13**, 2912 (1970)].

[9]T. Kármán and L. Howarth, Proc. R. Soc. London, Ser. A **164**, 192 (1938).

Convex covariant entropy density, symmetric conservative form, and shock waves in relativistic magnetohydrodynamics

Tommaso Ruggeri and Alberto Strumia
Istituto di Matematica Applicata, Universitá di Bologna, Via Vallescura 2, 40136 Bologna, Italy

(Received 3 December 1980; accepted for publication 20 February 1981)

The system of conservation laws governing the relativistic magnetohydrodynamics (MHD) is shown to possess a covariant entropy density which is a convex function of suitable field variables. Therefore, the results of a general theory developed in a previous paper hold and in particular: (a) there exists a *main field* such that the system exhibits a conservative symmetric hyperbolic form, in the sense of Friedrichs, and therefore the local Cauchy problem is well posed in a Sobolev space H^s $(s \geqslant 4)$; (b) the entropy increases across a shock wave front; (c) the shock propagation velocities do not exceed the speed of light; (d) the jump of thermodynamic entropy determines the jumps of each field variable.

PACS numbers: 52.30. + N, 52.60. + h, 04.20.Me, 52.25. Kn

1. THE EQUATIONS OF RELATIVISTIC MAGNETOHYDRODYNAMICS

Let (R^4, g) be a given Minkowskian space–time, where g is the flat metric, x a point belonging to R^4 and x^α $(\alpha = 0,1,2,3)$ pseudocartesian coordinates of x. We denote by $g_{\alpha\beta}$ the components of $g \equiv \mathrm{diag}(+1, -1, -1, -1)$.

The system of equations governing a perfectly conductor-relativistic plasma form a set of covariant laws (see, e.g., Ref. 1):

$$\partial_\alpha T^{\alpha\beta} = 0, \quad \text{(energy-momentum conservation)}, \tag{1.1}$$

$$\partial_\alpha (r u^\alpha) = 0, \quad \text{(matter conservation)}, \tag{1.2}$$

$$\partial_\alpha (u^\alpha B^\beta - u^\beta B^\alpha) = 0, \quad \text{(Maxwell equations)}, \tag{1.3}$$

where $\partial_\alpha = \partial / \partial x^\alpha$. The energy-momentum tensor has components

$$T^{\alpha\beta} = T^{\alpha\beta}_{\text{fluid}} + T^{\alpha\beta}_{\text{mag}}, \tag{1.4}$$

with

$$T^{\alpha\beta}_{\text{fluid}} = r f u^\alpha u^\beta - p g^{\alpha\beta}, \quad \text{(fluid energy-momentum tensor)}, \tag{1.5}$$

$$T^{\alpha\beta}_{\text{mag}} = B^2 (u^\alpha u^\beta - \tfrac{1}{2} g^{\alpha\beta}) - B^\alpha B^\beta, \quad \text{(magnetic coupling)}. \tag{1.6}$$

As usual r is the rest matter density, f the index of the fluid,

$$rf = \rho + p, \tag{1.7}$$

ρ is the proper energy density, p the pressure, and u^α the unit 4-velocity oriented towards the future, so that

$$u_\alpha u^\alpha = 1, \tag{1.8}$$

$B^\alpha = (\mu)^{1/2} H^\alpha$, with $\mu > 0$ the (constant) magnetic permeability, and H^α the proper magnetic field 4-vector (spacelike). Furthermore,

$$B^2 = -B^\alpha B_\alpha > 0, \tag{1.9}$$

$$u^\alpha B_\alpha = 0, \tag{1.10}$$

and the speed of light is equal to unity.

Now we must point out that the compact form (1.3) of the covariant Maxwell equations prevents us from following directly the way shown in Ref. 2, because of the presence of the 4-vector B^α, the components of which are not independent variables because of the constraints (1.10). We shall overcome the difficulty without losing the explicit covariance of the equations by employing an orthonormalized tetrad formed by constant congruences. Let $\{\xi^\alpha, \zeta^\alpha_I\}$, $I = 1,2,3$ be our set of congruences such that

$$\begin{aligned} &g_{\alpha\beta}\xi^\alpha \xi^\beta = 1, \quad \text{(timelike)},\\ &g_{\alpha\beta}\zeta^\alpha_I \zeta^\beta_J = -\delta_{IJ}, \quad \text{(spacelike)},\\ &g_{\alpha\beta}\xi^\alpha \zeta^\beta_I = 0,\\ &\partial_\beta \xi^\alpha = 0,\\ &\partial_\beta \zeta^\alpha_I = 0 \quad (I,J = 1,2,3). \end{aligned} \tag{1.11}$$

For any 4-vector V^α we may define the invariant scalar components with respect to the congruences through the relations

$$v = V_\alpha \xi^\alpha, \quad v_I = V_\alpha \zeta^\alpha_I, \quad V^\alpha = v \xi^\alpha + v^I \zeta^\alpha_I, \tag{1.12}$$

(it is easily found that $v_I = -v^I$).

Taking into account the definition (1.12) and introducing the operators

$$\partial_T = \xi^\alpha \partial_\alpha, \quad \partial_I = \zeta^\alpha_I \partial_\alpha, \quad T = \xi_\alpha x^\alpha, \quad X_I = \zeta^\alpha_I x_\alpha, \tag{1.13}$$

we can separate the components of (1.3) without losing covariance. When we project (1.3) on ξ_α we gain the divergence equation

$$\partial_I (u^I b - b^I u) = 0. \tag{1.14}$$

This is not a propagation equation since it does not involve time, but it represents a constraint that we shall verify holds at any time T if it is fulfilled at $T = 0$. Therefore (1.14) will not be taken into account when we consider the hyperbolic system of wave equations.

While projecting (1.3) on ζ^I_α we reach

$$\partial_T (u b^I - u^I b) + \partial_J (u^J b^I - u^I b^J) = 0, \tag{1.15}$$

or equivalently,

$$\partial_\alpha M^{I\alpha} = 0, \quad M^{I\alpha} = u^\alpha b^I - u^I B^\alpha. \tag{1.16}$$

Applying ∂_I to (1.15) we obtain,

 © 1981 American Institute of Physics

$$\partial_T\{\partial_I(bu^I - ub^I)\} = 0,$$

i.e., the condition that ensures us that Eq. (1.14) is fulfilled at any time T if it holds at $T = 0$. The procedure adopted is a covariant extension of that followed when one considers the divergence and curl equations of electrodynamics in a special frame. The previous procedure has been performed in special relativity for the sake of simplicity and clarity, but it could be of interest to investigate under which assumptions it may be extended to general relativity.

Now the hyperbolic system (1.1), (1.2), and (1.16) is a set of eight independent wave equations for eight independent unknowns, in the conservative form

$$\partial_\alpha \mathbf{F}^\alpha = \mathbf{f}; \quad \mathbf{F}^\alpha \equiv \begin{vmatrix} T^{\alpha\beta} \\ ru^\alpha \\ M^{I\alpha} \end{vmatrix}, \quad \mathbf{f} = 0, \quad (\beta = 0,1,2,3; I = 1,2,3). \tag{1.17}$$

Now following Ref. 2 we may choose as a field variable

$$\mathbf{U} = \mathbf{F}^\alpha \xi_\alpha \equiv \begin{vmatrix} T^{\alpha\beta}\xi_\alpha \\ ru \\ ub^I - u^I b \end{vmatrix}, \tag{1.18}$$

the components of which are eight independent variables, unknowns of (1.17).

The adiabatic condition $u^\alpha \partial_\alpha S = 0$ arises as a consequence of (1.17) and the first principle of thermodynamics:

$$r\,df = r\theta dS + dp, \tag{1.19}$$

and through (1.2) we have the supplementary conservation law

$$\partial_\alpha h^\alpha = g, \quad h^\alpha = -rSu^\alpha, \quad g = 0, \tag{1.20}$$

S being the entropy per mass unit of the plasma and θ the absolute temperature.

2. OUTLINES OF THE GENERAL THEORY

In the previous section we have seen that the hyperbolic system governing relativistic MHD is a set of covariant conservative forms (1.17) possessing a supplementary conservation law (1.20). In Ref. 2 a general theory for such systems has been developed, starting from some works by Lax and Friedrichs,[3] Friedrichs,[4] Boillat,[5] Boillat and Ruggeri.[6] We give a brief outline of the results of that theory. A supplementary conservation law (1.20) for a system of type (1.17) will exist if and only if suitable compatibility conditions hold, i.e., there exists a vector $\mathbf{U}'$ such that

$$\mathbf{U}'\cdot\delta\mathbf{F}^\alpha = \mathbf{U}'\cdot\nabla\mathbf{F}^\alpha\delta\mathbf{U} \equiv \delta h^\alpha \Rightarrow \mathbf{U}'\cdot\delta\mathbf{U} = \delta h \Longleftrightarrow \mathbf{U}' = \nabla h, \tag{2.1}$$

$$(\nabla = \partial/\partial\mathbf{U}),$$

$\delta\mathbf{U}$ denoting a generic variation of $\mathbf{U} = \mathbf{F}^\alpha\xi_\alpha$ and

$$h = h^\alpha \xi_\alpha. \tag{2.2}$$

In Ref. 2 it is assumed that there exists at least one timelike covector ξ_α, independent of the field, such that h is a convex function of $\mathbf{U}$ in a convex domain $D \subseteq R^N$. Under this assumption the following have been proved:

(1) the mapping $\mathbf{U} \longleftrightarrow \mathbf{U}'$ is globally univalent on D. It follows that $\mathbf{U}'$ can be chosen as a field vector;

(2) there exists a 4-vector h'^α defined as

$$h'^\alpha = \mathbf{U}'\cdot\mathbf{F}^\alpha - h^\alpha, \tag{2.3}$$

such that $\mathbf{F}^\alpha = \nabla' h'^\alpha$, $(\nabla' = \partial/\partial\mathbf{U}')$, that is to say, the system (1.17) becomes symmetric hyperbolic and preserves the conservative form, when $\mathbf{U}'$ is chosen as field variable, i.e.,

$$L\,\mathbf{U}' = \mathbf{f}, \quad L = \nabla'\nabla' h'^\alpha \partial_\alpha. \tag{2.4}$$

Then the Cauchy problem is well posed in Sobolev space H^s, $(s \geqslant 4)$ in a neighborhood of the initial manifold.[7] Moreover it has been pointed out that for physical systems $\mathbf{U}'$ seems to be a privileged field not only from a mathematical point of view, but also for physical reasons. This is why we called it the *main field*.[2] h'^α, considered as a function of $\mathbf{U}'$, generates the differential operator L and therfore we called it the *4-vector generating function* of the system.

(3) If Γ is a noncharacteristic shock manifold in space–time, of the Cartesian equation $\Phi(x^\alpha) = 0$, $\Phi \in C^2$, the function

$$\eta = [h^\alpha]\Phi_\alpha, \tag{2.5}$$

($\Phi_\alpha = \partial_\alpha\Phi$ and [ ] denoting the jump) defined on Γ is nonvanishing. In fluid dynamics this circumstance is equivalent to the increasing of the thermodynamic entropy across the shock.

(4) If η is known as a function of the unperturbed field $\mathbf{U}_*$ and Φ_α, then the jump of each component of $\mathbf{U}'$ is determined.

(5) If the characteristic manifolds are time- or lightlike also, the shock manifolds are such.

The aim of the present paper is to prove the convexity of the function h for relativistic MHD, a condition that is sufficient to ensure that the five important properties mentioned hold.

3. CONVEXITY OF THE COVARIANT DENSITY h IN MHD

Our goal is to prove the convexity of

$$h = h^\alpha\xi_\alpha = -rSu, \tag{3.1}$$

as a function of the field $\mathbf{U}$ defined by (1.18). The first step is to evaluate the main field $\mathbf{U}'$. Taking into account the first principle of thermodynamics (1.19) it has been shown elsewhere[2] that for the fluid

$$\theta\delta(-rSu) = -u_\alpha\delta(T^{\alpha\beta}_{\text{fluid}}\xi_\beta) + (G+1)\delta(ru), \tag{3.2}$$

G being the free enthalpy,

$$G = f - \theta S - 1, \qquad dG = -Sd\theta + (1/r)dp. \tag{3.3}$$

From (1.4) and (3.2) we have

$$\theta\delta(-rSu) = -u_\alpha\delta(T^{\alpha\beta}\xi_\beta) + (G+1)\delta(ru) + u_\alpha\delta(T^{\alpha\beta}_{\text{mag}}\xi_\beta). \tag{3.4}$$

But from (1.6), (1.8), (1.9), and (1.10) one obtains

$$u_\alpha(T^{\alpha\beta}_{\text{mag}}\xi_\beta) = -B_\alpha\delta(uB^\alpha - u^\alpha b), \tag{3.5}$$

and through the decomposition (1.12),

$$u_\alpha\delta(T^{\alpha\beta}_{\text{mag}}\xi_\beta) = -b_I\delta(ub^I - u^I b). \tag{3.6}$$

On introducing (3.6) into (3.5) and taking account of the compatibility conditions (2.1) and of (1.18) we arrive at

$$\mathbf{U}' \equiv \frac{1}{\theta}\begin{vmatrix} -u_\alpha \\ G+1 \\ -b_I \end{vmatrix}, \quad \alpha = 0,1,2,3; \quad I = 1,2,3. \qquad (3.7)$$

It is remarkable that in this case the components of $\mathbf{U}'$ are physical observables of major importance. In fact the first block of five variables is the same as the fluid and individuates the velocity of the plasma, the thermodynamic quantities $1/\theta$ and $(G+1)/\theta$. In continuum mechanics $1/\theta$ plays an important role, which Müller called *coldness*[8]: it appears as an *integrating factor* in order that the rhs in (3.4) is an exact differential. The *thermal potential* $(G+1)/\theta$ is also a quantity of prime importance, e.g., in the kinetic theory of the simple relativistic gas, as illustrated in the paper by W. Isräel.[9] It was first introduced by Landau and Liftshitz to generalize the Fourier equation for a relativistic fluid conductor of heat.[10] The three variables in the second block represent the components of the magnetic field on the space platform and are also observables of the system.

The last step is to prove the convexity of $h = h^\alpha \xi_\alpha$, i.e., that there exists at least one covector ξ_α such that

$$Q = \delta^2 h = \delta\mathbf{U}\cdot\nabla\nabla h\cdot\delta\mathbf{U} > 0, \quad \forall\delta\mathbf{U}\neq 0. \qquad (3.8)$$

Since with the field choice (1.18), $\mathbf{U}' = \nabla' h$ for (2.1), then (3.8) is equal to

$$Q = \delta\mathbf{U}'\cdot\delta\mathbf{U} > 0, \quad \forall\delta\mathbf{U}\neq 0. \qquad (3.9)$$

From direct computation we have

$$Q = \delta(-u_\alpha/\theta)\delta(T^{\alpha\beta}_{\text{fluid}}\xi_\beta + T^{\alpha\beta}_{\text{mag}}\xi_\beta) + \delta\{(G+1)/\theta\}\delta(ru) + \delta(-b_I/\theta)\delta(ub^I - u^I b).$$

Let Q_{fluid} be the quadratic form corresponding to (3.9) for the uncharged fluid.[2] Then (3.9) looks like

$$Q = Q_{\text{fluid}} - \{\delta u_\alpha \delta(T^{\alpha\beta}_{\text{mag}}\xi_\beta) + \delta b_I \delta(ub^I - u^I b)\}/\theta + \delta\theta\{u_\alpha\delta(T^{\alpha\beta}_{\text{mag}}\xi_\beta) + b_I\delta(ub^I - u^I b)\}/\theta^2.$$

The last term in the rhs vanishes due to (3.6); taking into account (1.8) and (1.10) after some calculations which do not involve conceptual trouble, one obtains

$$\theta Q = \theta Q_{\text{fluid}} - uB^2\delta u^\alpha\delta u_\alpha - u\delta B^\alpha\delta B_\alpha - 2B^\alpha\delta B_\alpha\delta u + 2b\delta u^\alpha\delta B_\alpha. \qquad (3.10)$$

Since Q, Q_{fluid} are covariant scalars and we are interested in the signature of Q, which is independent of the frame, we shall put ourselves in the rest frame of the fluid $\mathscr{S}$. In $\mathscr{S}$, $\{u^\alpha\}\equiv\{1,\vec{0}\}$ and $\delta u^0 = 0$, since $u^\alpha\delta u_\alpha = 0$ for (1.8). Then $u = \xi^0$, $\delta u^\alpha\delta u_\alpha = -(\delta\vec{u})^2$, $B^0 = 0$ for (1.10) and $\delta B^0 = \vec{B}\cdot\delta\vec{u}$, $B^2 = \vec{B}^2$, $\delta u = -\vec{\xi}\cdot\delta\vec{u}$, $\delta b = -\vec{\xi}\cdot\delta\vec{B}$, $b = B^\alpha\xi_\alpha = -\vec{B}\cdot\vec{\xi}$.

Therefore in $\mathscr{S}$ Eq. (3.10) becomes

$$\theta(Q - Q_{\text{fluid}}) = uB^2(\delta\vec{u})^2 - u(\vec{B}\cdot\delta\vec{u})^2 - 2b\delta\vec{u}\cdot\delta\vec{B} - 2(\vec{B}\cdot\delta\vec{B})(\vec{\xi}\cdot\delta\vec{u}) + u(\delta\vec{B})^2. \qquad (3.11)$$

If $\vec{B} = 0$ and $\delta\vec{B}\neq 0$ the rhs is positive. Then there remains to be examined the case $\vec{B}\neq 0$. We have to study the quadratic form of the vector $(\delta\vec{u},\delta\vec{B})$ of the coefficient matrix

$$\mathscr{A} \equiv \left\| \begin{matrix} u(B^2 I - \vec{B}\otimes\vec{B}) & -(bI + \vec{\xi}\otimes\vec{B}) \\ -(bI + B\otimes\vec{\xi}) & uI \end{matrix} \right\|,$$

where I denotes the 3×3 identity matrix and $\otimes$ the tensor product. The matrix is symmetric and its six real eigenvalues are easily evaluable. They are $\omega_1 = u$, $\omega_2 = B^2(u-1)$; ω_3 and ω_4 are the roots of $\omega^2 - u(B^2+1)\omega + B^2 = 0$; ω_5 and ω_6 are the roots of $\omega^2 - u(B^2+1)\omega + u^2B^2 - b^2 = 0$. Since $u\geqslant 1$ it follows that ω_1, ω_2, ω_3, and ω_4 are positive. Also ω_5 and ω_6 are positive since $u^2B^2 - b^2 = u^2B^2 - (\vec{B}\cdot\vec{\xi})^2 \geqslant B^2(u^2 - \vec{\xi}^2) > 0$ for $1 = \xi^\alpha\xi_\alpha = \xi_0^2 - \vec{\xi}^2 = u^2 - \vec{\xi}^2$. Then $\mathscr{A}$ is positive definite and the rhs in (3.11) becomes positive. It follows that $Q\geqslant Q_{\text{fluid}}$ and the conditions implying convexity in the fluid case are enough also for MHD.

The conditions are as follows:

(i) $-G = -G(\theta,p)$ must be convex, i.e.,

$$G_{\theta\theta} < 0, \qquad (3.12)$$

$$J = D(G_\theta, G_p)/D(\theta,p) = G_{\theta\theta}G_{pp} - (G_{\theta p}) > 0, \qquad (3.13)$$

where $G_\theta = (\partial G/\partial\theta)_p$, $G_p = (\partial G/\partial p)_\theta$. These conditions are usual in thermodynamics (see, e.g., Ref. 11).

(ii) The sound velocity must be smaller than that of light in vacuo

$$(\partial p/\partial\rho)_S < 1. \qquad (3.14)$$

Here we want to point out the fact that (3.13) is not independent since it follows from (3.12) and (3.14).[12] In fact from (1.19) and (3.3) there result

$$dG = -S\,d\theta + V\,dp, \quad V = 1/r,$$
$$dp = r\theta\,dS - fr^2\,dV, \qquad (3.15)$$

Then,

$$\left(\frac{\partial V}{\partial p}\right)_S = \frac{D(V,S)}{D(p,S)} = -\frac{D(G_p,G_\theta)}{D(p,S)} = -\frac{D(G_p,G_\theta)}{D(p,\theta)}\frac{D(p,\theta)}{D(p,S)} = J/G_{\theta\theta},$$

$$\left(\frac{\partial p}{\partial\rho}\right)_S = \frac{D(p,S)}{D(\rho,S)} = \frac{D(p,S)}{D(V,S)}\frac{D(V,S)}{D(\rho,S)} = -fr^2\left(\frac{\partial V}{\partial p}\right)_S = -fr^2 J/G_{\theta\theta}.$$

Therefore if $G_{\theta\theta} < 0$ and $(\partial p/\partial\rho)_S > 1$, we have $J > 0$.

Concluding, if (3.12) and (3.14) hold, $h = -rSu^\alpha\xi_\alpha$ is a convex function of $\mathbf{U}$ in any convex domain $D\subseteq R$,[8] for any unit timelike covector ξ_α oriented toward the future. In Ref. 4 the author gives an equivalent proof, under stronger assumptions, of the convexity of energy density. He considers the energy conservation law instead of entropy as the supplementary equation. We point out that even if it is always possible to interchange the roles of the equations, the physics of the shocks requires that the supplementary equation is the one that does not fulfil the Rankine–Hugoniot equations and generates the function η, increasing across the shock.

4. CONSEQUENCES OF CONVEXITY; CONCLUSION

As a consequence of the previous proof the system of MHD possesses the five properties stated in Sec. 2. In particular,

(1) The system of MHD is a conservative symmetric

hyperbolic system in the main field $\mathbf{U}'$ given by (3.7), with a 4-vector generating function (2.3) of the form

$$h'^{\alpha} = \{(p + \tfrac{1}{2}B^2)u^{\alpha} + b(bu^{\alpha} - uB^{\alpha})\}/\theta,$$

and the local Cauchy problem is well posed in H^s. The Legendre conjugate density
$h' = h'^{\alpha}\xi_{\alpha} = \mathbf{U}'\cdot\mathbf{U} - h$ is $h' = u(p + \tfrac{1}{2}B^2)/\theta$.

(2) $[S] > 0$ across the shock manifold when $u^{\alpha}_{*}\Phi_{\alpha} < 0$, u^{α}_{*} being the 4-velocity of the unperturbed plasma.

(3) The knowledge of $[S]$ as a function of the unperturbed field $\mathbf{U}_{*}$ and the shock normal Φ_{α} are enough to determine completely the shock.

(4) The propagation velocity of MHD shocks never exceeds the speed of light.

The results (2), (3), and (4) are exactly the same as for the fluid since h has the same expression (for proof see Ref. 2).

The result (4) had already been proved in a different way by Lichnerowicz[1] under the assumptions $\tau_p < 0$ and $\tau_{pp} > 0$, with $\tau = \tau(p,S) = f/r$. The former hypothesis is the same as (3.14) while the latter replaces (3.12).

ACKNOWLEDGMENT

This work was supported by the CNR (Gruppo Nazionale per la Fisica Matematica).

[1] A. Lichnerowicz, J. Math. Phys. **17** (12), 2135 (1976); *Relativistic Fluid Dynamics*, first C.I.M.E session (1970), edited by C. Cattaneo (Cremonese, Roma, 1971).

[2] T. Ruggeri and A. Strumia, Ann. Inst. Henri Poincaré **34**, 65 (1981). See also Atti 5 Conv. A.I.M.E.T.A. **1** Meccanica Generale, 225 (1980) and T. Ruggeri "Entropy principle and main field for a nonlinear covariant system," to appear in the lecture notes of the first C.I.M.E session on "Wave Propagation," Bressanone (1980).

[3] K. O. Friedrichs and P. D. Lax, Proc. Natl. Acad. Sci. U.S.A. **68**, 1686 (1971).

[4] K. O. Friedrichs, Commun. Pure Appl. Math. **27**, 749 (1974); **31**, 123 (1978).

[5] G. Boillat, C. R. Acad. Sci. Paris A **278**, 909 (1974); A **283**, 409 (1976); lectures to appear in the lecture notes of the first C.I.M.E. session on "Wave Propagation" Bressanone (1980).

[6] G. Boillat and T. Ruggeri, C. R. Acad. Sci. Paris A **289**, 257 (1979); Acta Mechanica **35**, 271 (1980).

[7] A. Fischer and D. P. Marsden, Commun. Math. Phys. **28**, 1(1972), "General relativity, partial differential equations and dynamical systems," in *Proceedings of the Symposium on Pure Mathematics*, edited by C. Spencer (Am. Math. Soc., Providence, RI, 1973), Vol. 23, p. 309.

[8] I. Müller, Arch. Rat. Mech. Anal. **40**, 1(1971); *Trends in application of pure mathematics to mechanics*, edited by H. Zorski (Pitman, London, **281**, (1979), Vol. II; see also the work by I. Shih Liu, Arch. Rat. Mech. Anal. **46**, 131 (1972).

[9] W. Isräel, J. Math. Phys. **4**, 1163 (1963).

[10] L. Landau and E. Liftshitz, *Fluid Mechanics* (Addison-Wesley, Reading, Mass., 1959), p. 505.

[11] D. Ter Haar and H. Wergeland, *Elements of thermodynamics* (Addison-Wesley, Reading, Mass., 1966).

[12] This was noticed during a discussion with G. Boillat to whom we are particularly indebted.

On black holes in magnetic universes

William A. Hiscock
Physics Department,[a] *Yale University, New Haven, Connecticut 06520*
and Institute for Theoretical Physics, University of California, Santa Barbara, California 93106

(Received 7 November 1980; accepted for publication 3 April 1981)

The magnetized black hole solutions discovered by Ernst are studied. It is shown that no static magnetic-universe Kerr–Newman black holes exist if either a, the Kerr angular momentum parameter, or e, the electric charge parameter, is nonzero. Robinson's identity is used to prove that the Schwarzschild–Melvin black hole solution is the unique static, axisymmetric black hole solution of the sourceless Einstein–Maxwell equations which asymptotically resembles Melvin's magnetic universe. This may be viewed as a generalization of Israel's theorem, in which one extra assumption (axisymmetry) is required, but the boundary conditions at infinity are somewhat relaxed.

PACS numbers: 97.60.Lf, 98.80.Dr, 04.20.Jb

I. INTRODUCTION

There are two distinct reasons for studying black hole solutions in magnetic universes. First, the effects of a cosmological magnetic field on the electromagnetic and metric field structure near the black hole's horizon are of possible interest to astrophysicists seeking models of black holes in astrophysical environments. A cosmological magnetic field can represent (near the hole) the externally generated magnetic field that an astrophysical black hole is likely to be immersed in, perhaps driven by currents in an accretion disk. Secondly, black holes in magnetic universes are of interest because the solutions are not asymptotically flat. Virtually all of our intuitions concerning black holes come from either the known exact asymptotically flat solutions, or the general theorems of Hawking, Carter, Israel, *et al.*, most of which assume asymptotic flatness as a boundary condition. Since we have no guarantee that the universe we live in is asymptotically flat, it seems worthwhile to attempt to find reasonable non-asymptotically-flat black hole solutions to the Einstein equations, and to attempt to extend the general theorems (or to find counterexamples to them), i.e., to study the generic properties of black holes in non-asymptotically-flat spacetimes.

In this paper, I shall explore some of the uniqueness properties of the family of magnetized black hole solutions discovered by Ernst[1] and Ernst and Wild.[2] These solutions are generated by applying a Harrison-type transformation[3,4] to the usual asymptotically flat black hole solutions of the Einstein–Maxwell equations. The transformed solutions asymptotically resemble stationary cylindrically symmetric magnetic universes. One effect of the transformation is to turn the three-parameter Kerr–Newman solution[5] into a four-parameter Kerr–Newman-magnetic universe solution. Since the new parameter is simply the asymptotic cosmological magnetic field strength, which, set equal to zero, gives back the ordinary Kerr–Newman solution, these metrics may be regarded as a generalization of the ordinary solutions.

Section II of this paper reviews the derivation of the Ernst solutions and sets forth several simple but necessary results converning the global nature of the solutions. An explicit prescription is given for extending the local metric forms given by Ernst and Wild[1,2] to a global form which has a regular symmetry axis (i.e., no cone singularities). Also, Ernst and Wild[1,2] have previously noted that the transformed black hole solutions asymptotically "resemble Melvin's magnetic universe" (MMU).[6–9] This statement is made somewhat more mathematically precise by showing that the transformed Schwarzschild metric does, in a mathematically meaningful way, asymptotically approach MMU, whereas if a or e is nonzero, the asymptotic region appears to be some sort of stationary, cylindrically symmetric electromagnetic cosmology, but not precisely MMU.

Ernst has previously pointed out that placing a nonzero electric charge on a black hole in a magnetic universe leads to frame-dragging effects, due to the $\mathbf{E}\times\mathbf{B}$ circulating momentum flux in the stress–energy tensor. In Sec. III I show that given a nonzero value of e, the electric charge parameter for the black hole, there is no way to adjust a, the Kerr angular momentum parameter, so as to yield a static solution wherein the two sorts of frame dragging would exactly cancel. Thus, the only static magnetic universe black hole solution obtainable by Harrison-transforming an asymptotically flat black hole solution is the Schwarzschild–Melvin solution.

The awkwardness of the previous sentence leads us to an obvious question: Are there other stationary axisymmetric magnetic universe black holes, or does the unique set of asymptotically flat black hole solutions transform into a unique set of magnetic universe black hole solutions? In Sec. IV I take a small step towards answering this query by using Robinson's identity[10] to prove that the Schwarzschild–Melvin solution is the unique static axisymmetric asymptotically Melvin's magnetic universe black hole solution to the Einstein–Maxwell equations. This uniqueness theorem also represents the first attempt to extend the black hole uniqueness theorems of Israel, Carter, Robinson, *et al.*[11] by relaxing the boundary conditions at infinity.

II. REVIEW AND PRELIMINARIES

The Ernst solutions are generated by means of a Harri-

[a]Present address.

 © 1981 American Institute of Physics

son-type transformation[3,4] applied to the asymptotically flat Kerr–Newman black hole metrics. The Kerr–Newman metric may be written in the form

$$ds^2 = \Sigma\left(\frac{-\Delta}{A}dt^2 + \frac{dr^2}{\Delta} + d\theta^2\right) + \frac{A\sin^2\theta}{\Sigma}\left(d\phi - \frac{(2Mr - e^2)a}{A}dt\right)^2, \quad (1)$$

where

$$\Delta = r^2 - 2Mr + a^2 + e^2, \quad (2)$$

$$\Sigma = r^2 + a^2\cos^2\theta, \quad (3)$$

$$A = (r^2 + a^2)^2 - \Delta a^2\sin^2\theta, \quad (4)$$

M is the mass of the black hole, $a = J/M$ its specific angular momentum, and e is its electric charge. The Kerr–Newman metric may also be written in the form

$$ds^2 = f^{-1}(-2P^{-2}\,d\zeta d\zeta^* + \rho^2\,dt^2) - f(d\phi - \omega\,dt)^2 \quad (5)$$

by identifying

$$d\zeta = 2^{-1/2}\left(\frac{dr}{\Delta^{1/2}} + id\theta\right), \quad (6)$$

$$\rho = \Delta^{1/2}\sin\theta, \quad (7)$$

$$P = (A^{1/2}\sin\theta)^{-1}, \quad (8)$$

$$f = -A\sin^2\theta/\Sigma, \quad (9)$$

$$\omega = a(2Mr - e^2)/A. \quad (10)$$

The transformation to a magnetic universe is accomplished by replacing f and ω in Eq. (5) with new metric functions f' and ω' which are defined by the following equations:

$$f' = |\Lambda|^{-2}f, \quad (11)$$

$$\nabla\omega' = |\Lambda|^2\nabla\omega + \rho f^{-1}(\Lambda^*\nabla\Lambda - \Lambda\nabla\Lambda^*), \quad (12)$$

where

$$\Lambda = 1 + B\Phi - \tfrac{1}{4}B\mathscr{E}, \quad (13)$$

B is a constant representing the strength of the cosmological magnetic field, and Φ and $\mathscr{E}$ are the complex Ernst potentials[12,13] for Kerr–Newman metric. In the Boyer–Lindquist coordinate system of Eq. (1) the Ernst potentials are

$$\Phi = e\frac{a - ir\cos\theta}{r + ia\cos\theta}, \quad (14)$$

$$\mathscr{E} = -\left(r^2 + a^2 - a\frac{2Ma + i(2Mr - e^2)\cos\theta}{r + ia\cos\theta}\right)\sin^2\theta - \left(4Ma + ie^2\cos\theta\right)\frac{a - ir\cos\theta}{r + ia\cos\theta}. \quad (15)$$

The gradient operator in Eq. (12) is defined by

$$\nabla = \Delta^{1/2}\frac{\partial}{\partial r} + i\frac{\partial}{\partial\theta}. \quad (16)$$

Corresponding to the new metric functions f' and ω' are new complex Ernst potentials for the transformed spacetime. The new complex gravitational potential is given by

$$\mathscr{E}' = \Lambda^{-1}\mathscr{E}, \quad (17)$$

and the new complex electromagnetic potential is given by

$$\Phi' = \Lambda^{-1}(\Phi - \tfrac{1}{2}B\mathscr{E}). \quad (18)$$

The electromagnetic field of the new spacetime may be found from Eq. (18) and the definitions in Ref. 13.

Since the magnetizing Harrison transformation is a local transformation of the complex Ernst potentials, one cannot *a priori* assume that the range of the coordinates in the transformed spacetime is the same as in the original spacetime. The "local value of π" for a small circle about the symmetry axis of the magnetized Kerr–Newman spacetime may be found by expanding $g_{\theta\theta}$ and $g_{\phi\phi}$ in powers of θ [or $(\pi - \theta)$] near $\theta = 0$ [or $(\pi - \theta) = 0$] using Eqs. (5)–(15) in the magnetized metric. In order for the axis to be regular (i.e., free of cone singularities), it is necessary that the angular coordinate ϕ of Eqs. (5) and (11) have $\phi = 0$ identified with $\phi = 2\pi F$ (not simply 2π), where

$$F = \left[1 + \tfrac{3}{2}B^2e^2 + 2MaeB^3 + \left(\frac{e^4}{16} + M^2a^2\right)B^4\right]^{1/2}. \quad (19)$$

Alternatively, one can define a new angular coordinate $\psi = \phi/F$ which runs from zero to 2π.

As Ernst pointed out in his original paper,[1] the magnetizing transformation applied to Minkowski space (i.e., with $M = a = e = 0$) yields the cylindrically symmetric electromagnetic solution discovered by Bonnor[6,7] and Melvin,[8] and studied thoroughly by Melvin[9] and Thorne,[14] known as "Melvin's magnetic universe" (hereafter, MMU). Its metric is

$$ds^2 = \Lambda^2(-dt^2 + d\rho^2 + dz^2) + \Lambda^{-2}\rho^2\,d\phi^2, \quad (20)$$

with

$$\Lambda = 1 + \tfrac{1}{4}B^2\rho^2. \quad (21)$$

The components of the magnetic field, in an orthonormal frame, are

$$H_{\hat z} = \Lambda^{-2}B, \quad H_{\hat\rho} = H_{\hat\phi} = 0. \quad (22)$$

The magnetized Schwarzschild solution is the easiest of the magnetic black hole solutions to study the global properties of, since it is static. Its metric may be written

$$ds^2 = \Lambda^2\left[-\left(1 - \frac{2M}{r}\right)dt^2 + \left(1 - \frac{2M}{r}\right)^{-1}dr^2 + r^2d\theta^2\right] + \Lambda^{-2}r^2\sin^2\theta\,d\phi^2, \quad (23)$$

the orthonormal components of the magnetic field are

$$H_{\hat r} = \Lambda^{-2}B\cos\theta, \quad H_{\hat\theta} = -\Lambda^{-2}B\left(1 - \frac{2M}{r}\right)^{1/2}\sin\theta, \quad (24)$$

and

$$\Lambda = 1 + \tfrac{1}{4}B^2r^2\sin^2\theta. \quad (25)$$

The global structure of the magnetized Schwarzschild solution is evident from Eq. (23). Any $\phi =$ const slice through the spacetime gives a three-dimensional spacetime whose metric is exactly the Schwarzschild metric multiplied by a simple conformal factor. Its radial null geodesics can be identified with those of the regular Schwarzschild metric, and its causal structure and Penrose diagram are the same as the $B = 0$ Schwarzschild case. Thus we see that the magnetized Schwarzschild solution is, in terms of the usual definitions, a black hole, with an event horizon and trapped surfaces.

It is not so easy to determine the causal structure of the other magnetized black hole solutions, as they are only stationary. The complexity of the metrics has prevented me from analyzing their global structure; it is natural to assume that they represent black holes in magnetic universes, but it should be emphasized that it has not been shown that the magnetized Kerr–Newman solution with a or e nonzero is in fact a black hole, despite everyone in the literature (including this paper) naming it as such.

It is easy to see that the magnetized Schwarzschild solution is, in a meaningful sense, "asymptotically Melvin's magnetic universe." The MMU metric may be written in a spherical-type coordinate system as

$$ds^2 = \Lambda^2[-dt^2 + dr^2 + r^2\,d\theta^2] + \Lambda^{-2}r^2\sin^2\theta\,d\phi^2, \tag{26}$$

where

$$r^2 = \rho^2 + z^2, \quad \tan\theta = z/\rho,$$

and Λ is given by Eq. (25). Comparing the spherical form of the MMU metric in Eq. (26) with the magnetized Schwarzschild metric, it is obvious that the magnetized Schwarzschild metric approaches the Melvin form as $r \to \infty$. The magnetic field components also asymptotically approach the MMU values of Eq. (22). Thus, in the same spirit in which one talks about the Kerr–Newman–de Sitter metric, it is meaningful to call the magnetized Schwarzschild solution, described by Eqs. (23)–(25), the "Schwarzschild–Melvin" solution.

The obvious next question to ask is whether the more general magnetized Kerr–Newman metrics, with a or e nonzero, are in any sense "asymptotically Melvin's magnetic universe."

An easy way to see that the magnetized Kerr–Newman solutions do not globally approach MMU is to evaluate the electric field on the symmetry axis. From Eq. (4.1) of Ref. 2 we find

$$E_{\hat r}(\theta = 0) = (\Lambda\Lambda^*)^{-2}B^2\{\tfrac{3}{2}e - 2MaB + \tfrac{1}{4}e^3B^2 - \tfrac{1}{2}e^2MaB^3 - \left(\frac{e^5}{32} - \frac{M^2a^2e}{2}\right)B^4\} + \mathcal{O}(r^{-1}), \tag{27}$$

$$E_{\hat\theta}(\theta = 0) = 0,$$

where $\Lambda\Lambda^* = F^2$. Thus the locally measured electric field on the symmetry axis in general approaches a constant, nonzero value far from the black hole. Since MMU has zero electric field, the magnetized Kerr–Newman solution cannot globally (i.e., for all θ) approach MMU.

III. NO MAGNETIZED ELECTRICALLY CHARGED STATIC BLACK HOLES

Thus far we have seen that there exists at least one static magnetized black hole solution, namely, the Schwarzschild–Melvin solution. One of the most fascinating aspects of the magnetized Kerr–Newman black hole solutions is that there are two independent sources of rotation and frame-dragging in these solutions. The magnetized electrically charged Reissner–Nordström solution[1] ($a = 0$, $e \neq 0$) is stationary, not static, since the nonzero Poynting vector $\mathbf{E}\times\mathbf{B}$ shows that there is angular momentum in the electromagnetic field. The magnetized Kerr metric[2] ($a \neq 0$, $e = 0$) is also stationary.

Considering the magnetized Kerr–Newman solutions, we are led to the tantalizing question of whether there might be some special values of a and e (where both are nonzero) such that $\mathbf{E}\times\mathbf{B}$ type frame-dragging exactly cancels the a type frame-dragging, leaving a static magnetized Kerr–Newman black hole with both a and e nonzero. In other words, are there any static Harrison-transformed Kerr–Newman black holes with both a and e nonzero?

Such a solution would have $\nabla\omega' = 0$, where ω' is defined by Eq. (12). Since the Reissner–Nordström ($a = 0$) and Kerr ($e = 0$) magnetized solutions are not static, it is clear that both a and e must be nonzero in the hypothetical static solution, which leads to horrendous complexity in the expression for $\nabla\omega'$. In order to determine whether $a \neq 0$, $e \neq 0$ static magnetized solutions exist, I have computed $\nabla\omega'$ for the magnetized Kerr–Newman metric. I will not, however, inflict the entire expression on either the journal or the readers (I estimate it would fill two to three journal pages). For the purposes of this section, it is sufficient to note that

$$\frac{\partial\omega'}{\partial\theta} = \frac{\Delta}{A^2\Sigma^2}\left\{\left[-\frac{B^4Ma\cos\theta\sin\theta\,(3 - \cos^2\theta)}{2} - B^3e\cos\theta\sin\theta\,\right]r^{11} + \mathcal{O}(r^{10}) + \cdots + \mathcal{O}(r^0)\right\}. \tag{28}$$

If there exist $a \neq 0$, $e \neq 0$ static magnetized black holes, the coefficient of r^{11} in Eq. (28) must vanish for all θ (this term will dominate ω' at large r). There are clearly no nonzero values of e and a which will accomplish this. Thus, *the Schwarzschild–Melvin solution is unique among the Harrison-transformed magnetized Kerr–Newman black hole solutions in being static with no naked singularities.*

If we make a further assumption, that the only stationary asymptotically cylindrical magnetic universe black hole solutions are the transformed Kerr–Newman solutions, then the above proven uniqueness property has an extremely interesting consequence. A civilization living near an electrically charged black hole in the magnetic universe could extract an infinite amount of energy from it via the Penrose process.[15]

If the hole is electrically charged, then either a, e, or both are nonzero. So long as the civilization is careful not to discharge the black hole, it can never evolve into the $a = 0$, $e = 0$ static Schwarzschild–Melvin solution. As the civilization uses the Penrose process to extract energy from the hole, a and e will evolve, and the shape and size of the ergosphere will change. So long as the black hole maintains its charge, the ergosurface can never retract onto the event horizon, and there will always be a finite ergosphere from which energy can be extracted. Where does this energy come from? It seems likely that it is extracted from the cosmological magnetic field by the charged, rotating black hole. Thus, a charged rotating black hole may act as a catalyst for extracting energy from what was, in its absence, a static uniform magnetic field.

IV. UNIQUENESS OF THE SCHWARZSCHILD–MELVIN SOLUTION

So far we have seen two ways in which the Schwarzschild–Melvin black hole is unique among the magnetized

Kerr–Newman solutions. First, in Sec. II, we saw that only the magnetized Schwarzschild solution asymptotically resembles Melvin's magnetic universe. Secondly, in Sec. III, we saw that the Schwarzschild–Melvin black hole is the unique static magnetized Kerr–Newman black hole solution.

In this section a different sort of uniqueness is proven for the Schwarzschild–Melvin solution. I will show that the Schwarzschild–Melvin black hole solution is the only asymptotically MMU static axisymmetric black hole solution to the electrovac Einstein–Maxwell equations.

The difference between the result of Sec. III and the result of this section is the set of spacetimes within which uniqueness is proven. In Sec. III the set of spacetimes studied was the Harrison-transformed Kerr–Newman metrics. In this section, the set of spacetimes considered consists of *all* static, axisymmetric, asymptotically Melvin's magnetic universe, electrovac black hole solutions to the Einstein–Maxwell equations.

A precise statement of the theorem is as follows:

Theorem: The only spacetime which

(1) is static,

(2) axisymmetric,

(3) possesses a regular event horizon (with topology $S^2\times R$) and axis,

(4) asymptotically approaches the solution known as Melvin's magnetic universe,

(5) and which satisfies the sourceless ($j^\alpha=0$) Einstein–Maxwell equations is the Schwarzschild–Melvin solution.

The method of proof is as follows. I first establish a lemma proving that any static magnetic spacetime must have zero electric field. The Lagrangian density for this problem may then be put into a form identical to the Lagrangian density for the stationary axisymmetric vacuum Einstein equations. The divergence identity discovered and used by Robinson[10] to prove the uniqueness of the Kerr black hole is then applied to the problem at hand (with different boundary conditions than in the asymptotically flat case), yielding the desired result. I will rely heavily on the techniques, formalism, and terminology of the asymptotically flat black hole uniqueness theorems.[11,16,17]

A few words about the assumptions of the theorem are perhaps in order. Since I assume the spacetime is static, the assumption of axisymmetry is necessary; Hawking's proof that "stationary black holes are axisymmetric"[18] fails in this static limit. In the asymptotically flat case this is got around by the method of proving Israel's theorem[19,20]: One shows that all static asymptotically flat electrovac black holes are spherically symmetric, then applies Birkhoff's theorem to yield the uniqueness of the Reissner–Nordström black hole. In the present case the spacetime is asymptotically cylindrical; there is no possibility of the black hole spacetime being spherically symmetric, and hence the static-implies-spherical theorems developed by Israel for black holes are useless here.

Hawking's proof that the topology of the event horizon must be spherical[18] does hold here. Assumptions (3) and (4) of the theorem simply define the boundary conditions for the spacetimes of interest.

The spacetime's metric may be written in the form

$$ds^2 = -V\,dt^2 + 2W\,d\phi\,dt + X\,d\phi^2 + g_{AB}\,dx^A dx^B, \tag{29}$$

where the coordinates t and ϕ are defined uniquely (up to additive constants) by the time translation (k^α) and axisymmetry (m^α) Killing vector fields,

$$t_{,\alpha}k^\alpha = \phi_{,\alpha}m^\alpha = 1, \tag{30}$$

$$t_{,\alpha}m^\alpha = \phi_{,\alpha}k^\alpha = 0, \tag{31}$$

$$t_{,[\alpha}k_\beta m_{\gamma]} = \phi_{,[\alpha}k_\beta m_{\gamma]} = 0, \tag{32}$$

A and B run from 2 to 3, and V, X, W, and g_{AB} are functions only of x^2 and x^3.

The vector potential for the electromagnetic field may be written

$$A = \Upsilon\,dt + \Psi\,d\phi, \tag{33}$$

where Υ and Ψ are again only functions of x^2 and x^3.

In Ref. 16 it is shown that the specification of the four functions X, W, Υ, Ψ is sufficient to completely determine the spacetime. In our case, where the spacetime is assumed to be static, we already know that

$$W = k_\alpha m^\alpha = 0. \tag{34}$$

Thus, the spacetimes satisfying Assumptions (1)–(5) of the theorem are completely specified by giving the three functions X, Υ, and Ψ, which are related through the Einstein–Maxwell equations.

Reference 16 defines an electromagnetic field to be static if

$$F_{[\alpha\beta}k_{\gamma]} = 0. \tag{35}$$

This definition requires that all magnetic fields vanish (as seen by the static observers). This is clearly an unacceptable definition for our purposes, where there is a cosmological magnetic field. I will insist only that the electromagnetic field be sourceless,

$$j_\alpha = F_\alpha{}^\beta{}_{;\beta} = 0 \tag{36}$$

and that the field be only a function of x^2 and x^3, i.e.,

$$\Upsilon_{,t} = \Upsilon_{,\phi} = \Psi_{,t} = \Psi_{,\phi} = 0. \tag{37}$$

It is now necessary to show that the assumption that the metric is static actually requires that $\Upsilon = \text{const}$ (no electric fields).

Lemma: A spacetime satisfying Assumptions (1)–(5) has $\Upsilon = \text{const}$ and hence no electric fields as seen by a static observer.

Proof: Since the metric is static, the rotation vector of the time translation Killing vector must vanish

$$\xi_\alpha = \tfrac{1}{2}\epsilon_{\alpha\beta\gamma\delta}k^\beta k^{\gamma;\delta} = 0. \tag{38}$$

Differentiation of Eq. (38), combined with the usual Killing vector–Ricci tensor identity, yields

$$\xi_{[\alpha;\beta]} = \tfrac{1}{2}\epsilon_{\alpha\beta\gamma\delta}k^\gamma R^{\delta\sigma}k_\sigma = 0 \tag{39}$$

since $\xi_\alpha = 0$. Using the fact that $R^\alpha{}_\alpha = 0$ for an electromagnetic field, and noting that $k^\gamma = \delta_t{}^\gamma$, Eq. (39) implies that

$$T^{tt} = 0, \tag{40}$$

where $i = 1,2,3$, (ϕ, x^2, x^3).

It is convenient at this point to switch to orthonormal frame components defined by the frame forms

$$\omega^{\hat{0}} = V^{1/2}\, dt, \tag{41}$$

$$\omega^{\hat{1}} = X^{1/2}\, d\phi, \tag{42}$$

$$\omega^{\hat{2}} = (g_{22})^{1/2}\, dx^2, \tag{43}$$

$$\omega^{X3} = (g_{33})^{1/2}\, dx^3, \tag{44}$$

where I have assumed that g_{23} is chosen to be zero [always possible by Sec. (10) of Ref. 16]. Equation (40) is then

$$T^{\hat{0}\hat{i}} = 0. \tag{45}$$

Since the stress–energy tensor is given by

$$T^{\hat{\alpha}\hat{\beta}} = \frac{1}{4\pi}\left(F^{\hat{\alpha}\hat{\mu}}F^{\hat{\beta}}{}_{\hat{\mu}} - \tfrac{1}{4}\eta^{\hat{\alpha}\hat{\beta}}F_{\hat{\mu}\hat{\nu}}F^{\hat{\mu}\hat{\nu}}\right), \tag{46}$$

this implies that

$$(i=1)\quad F^{\hat{0}\hat{2}}F^{\hat{1}}{}_{\hat{2}} + F^{\hat{0}\hat{3}}F^{\hat{1}}{}_{\hat{3}} = 0, \tag{47}$$

$$(i=2)\quad F^{\hat{0}\hat{1}}F^{\hat{2}}{}_{\hat{1}} + F^{\hat{0}\hat{3}}F^{\hat{2}}{}_{\hat{3}} = 0, \tag{48}$$

$$(i=3)\quad F^{\hat{0}\hat{1}}F^{\hat{3}}{}_{\hat{1}} + F^{\hat{0}\hat{2}}F^{\hat{3}}{}_{\hat{2}} = 0, \tag{49}$$

Equations (48) and (49) are trivially satisfied since $F_{\hat{0}\hat{1}} = F_{\hat{2}\hat{3}} = 0$ under our assumptions. Equation (47) implies that

$$F_{\hat{0}\hat{2}} = hF_{\hat{1}\hat{3}}, \tag{50}$$

$$F_{\hat{0}\hat{3}} = hF_{\hat{2}\hat{1}}, \tag{51}$$

where $h = h(x^2, x^3)$. The boundary conditions at infinity [Assumption (4)] tells us that $F_{\hat{1}\hat{3}}$ and $F_{\hat{2}\hat{1}}$ are nonzero. If $h = 0$, then the lemma follows immediately; hence, in what follows, I assume $h \neq 0$. Equations (50) and (51) take an interesting form when written in terms of the dual of the Maxwell tensor,

$$^*F_{\hat{\alpha}\hat{\beta}} = \tfrac{1}{2}F^{\hat{\mu}\hat{\nu}}\epsilon_{\hat{\mu}\hat{\nu}\hat{\alpha}\hat{\beta}}, \tag{52}$$

namely,

$$F_{\hat{0}\hat{3}} = h\, {}^*F_{\hat{0}\hat{3}}, \tag{53}$$

$$F_{\hat{0}\hat{2}} = h\, {}^*F_{\hat{0}\hat{2}}, \tag{54}$$

$$F_{\hat{2}\hat{1}} = -h^{-1}\, {}^*F_{\hat{2}\hat{1}}, \tag{55}$$

$$F_{\hat{1}\hat{3}} = -h^{-1}\, {}^*F_{13}. \tag{56}$$

Maxwell's equations in a source-free region are

$$F_{\hat{\alpha}}{}^{\hat{\beta}}{}_{;\hat{\beta}} = 0, \tag{57}$$

$$^*F_{\hat{\alpha}}{}^{\hat{\beta}}{}_{;\hat{\beta}} = 0. \tag{58}$$

Setting $\alpha = 0$ in Eq. (57) and applying Eqs. (53), (54), and (58), one finds

$$^*F_{\hat{0}}{}^{\hat{\alpha}}h_{,\hat{\alpha}} = 0, \tag{59}$$

while $\alpha = 1$ yields

$$^*F_{\hat{1}}{}^{\hat{\alpha}}h_{,\hat{\alpha}} = 0. \tag{60}$$

The implication of Eqs. (59) and (60) is most easily seen by noting that $^*F_{\hat{0}}{}^{\hat{\alpha}} = B^{\hat{\alpha}}$ and $F^{\hat{1}\hat{\alpha}} = \epsilon^{\hat{1}\hat{\alpha}\hat{\beta}}E_{\hat{\beta}}$, the local magnetic and electric field vectors. Also noting that Eqs. (50) and (51) imply that

$$E_{\hat{\alpha}} = hB_{\hat{\alpha}}, \tag{61}$$

we see that Eqs. (59) and (60) may be written in the form

$$B^{\hat{i}}\nabla_{\hat{i}}h = 0, \tag{62}$$

$$\epsilon^{\hat{i}\hat{j}\hat{k}}B_{\hat{i}}\nabla_{\hat{j}}h = 0. \tag{63}$$

Obviously, the only solution compatible with Eqs. (62), (63) and the fact that $h = h(x^2, x^3)$ is

$$\nabla_{\hat{i}}h = 0, \tag{64}$$

which implies that $h = \text{const}$. Since Assumption (4) requires the electric field to vanish at infinity, this implies that $h = 0$. Alternatively, one may always perform a duality rotation, i.e.,

$$F'_{\alpha\beta} = \cos\Theta F_{\alpha\beta} + \sin\Theta\, {}^*F_{\alpha\beta}$$

through an angle $\Theta = \cot^{-1}(h)$ which will reduce the electric field to zero everywhere. ■

Since the potential Υ is always adjustable by an additive constant, we may now take $\Upsilon = 0$ without loss of generality. Thus, the spacetimes of interest are completely determined by the two functions X, Ψ.

Following Carter,[16] we note that the Lagrangian density for the Einstein–Maxwell equations in this case takes the simple form

$$\mathscr{L} = \frac{|\nabla X|^2}{2X^2} + \frac{2|\nabla\Psi|^2}{X}, \tag{65}$$

where the Lagrangian integral to be varied is

$$I = \int \mathscr{L}\, d\lambda\, d\mu, \tag{66}$$

and λ and μ are x^2 and x^3, constructed according to the prescription of Ref. 16, with 2-space metric

$$ds_{\text{II}}^2 = g_{AB}\, dx^A dx^B = \Xi\left(\frac{d\lambda^2}{\lambda^2 - M^2} + \frac{d\mu^2}{1-\mu^2}\right). \tag{67}$$

Here the mass parameter of the black hole, M (a positive constant) fixes the scale of the manifold, and the coordinates range over $M < \lambda < \infty$, $-1 < \mu < +1$. The conformal factor Ξ is a well-behaved function which is nonzero everywhere, including on the axis.

Introduction of a new function, $Z = X^{1/2}$, yields the new Lagrangian density

$$\mathscr{L} = \frac{|\nabla Z|^2 + |\nabla\Psi|^2}{Z^2}, \tag{68}$$

which is formally identical to the Lagrangian density for the asymptotically flat Kerr uniqueness problem.

At this point, one follows exactly the steps of Ref. 10 in proving the uniqueness of the Kerr black hole, with only the boundary conditions altered.

The relevant boundary conditions are as follows. As $\mu \to \pm 1$ (the axisymmetry axis), X and Ψ are well-behaved and go like

$$X = \mathscr{O}(1-\mu^2), \tag{69}$$

$$X^{-1}X_{,\mu} = -2\mu(1-\mu^2)^{-1} + \mathscr{O}(1), \tag{70}$$

$$\Psi_{,\lambda} = \mathscr{O}(1-\mu^2), \tag{71}$$

$$\Psi_{,\mu} = \mathscr{O}(1). \tag{72}$$

As $\lambda \to M$ (the event horizon), X and Ψ are well-behaved functions such that

$$X = \mathscr{O}(1), \quad X^{-1} = \mathscr{O}(1),$$

$$\Psi_{,\mu} = \mathscr{O}(1), \quad \Psi_{,\lambda} = \mathscr{O}(1). \tag{73}$$

The main difference between the case at hand and the more familiar asymptotically flat case is in the boundary conditions at infinity. At infinity, we insist that X and Ψ be well-behaved functions with asymptotic behavior given by

$$\rho^{-2}X = (1 + \tfrac{1}{4}B^2\rho^2)^2[1 + \mathscr{O}(\lambda^{-1})]. \tag{74}$$

$$\Psi = \frac{2}{B}\left(1 + \frac{1}{4B^2\rho^2}\right)^{-1}[1 + \mathscr{O}(\lambda^{-1})], \tag{75}$$

where the asymptotic cylindrical coordinate $\rho^2 = \lambda^2(1 - \mu^2)$ has been introduced to more easily display the asymptotically cylindrical nature of the spacetime.

Robinson's identity [Eq.(6) of Ref. 10], relating the Schwarzschild–Melvin solution (X_1, Ψ_1) to a second hypothetical solution (X_2, Ψ_2) with the same values of M and B, may now be integrated over the two-dimensional manifold of Eq. (67). The application of Stoke's theorem and the boundary conditions [Eqs. (69)–(75)] shows that the boundary integral vanishes, just as in the asymptotically flat case. Exactly as in Ref. 10, the Einstein equations may then be manipulated, and the boundary conditions applied again, to yield the desired result,

$$\Psi_2 = \Psi_1, \quad X_2 = X_1, \tag{76}$$

and so the theorem is proven.

ACKNOWLEDGMENTS

This work was supported by the NSF under Grant No. PHY79-16482 at Yale and No. PHY77-27084 at Santa Barbara.

[1] F. J. Ernst, J. Math. Phys. **17**, 54 (1976).
[2] F. J. Ernst and W. J. Wild, J. Math. Phys. **17**, 182 (1976).
[3] B. K. Harrison, J. Math. Phys. **9**, 1744 (1968).
[4] W. Kinnersley, J. Math. Phys. **14**, 651 (1973).
[5] See, for example, C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973), Chap. 33.
[6] W. B. Bonnor, Proc. Roy. Soc. London Ser. A **66**, 145 (1953).
[7] W. B. Bonnor, Proc. Roy. Soc. London Ser. A **67**, 225 (1954).
[8] M. A. Melvin, Phys. Lett. **8**, 65 (1964).
[9] M. A. Melvin, Phys. Rev. **139**, B225 (1965).
[10] D. C. Robinson, Phys. Rev. Lett. **34**, 905 (1975).
[11] For a recent review, see B. Carter's article in *General Relativity: An Einstein centenary survey*, edited by S. W. Hawking and W. Israel (Cambridge U. P., Cambridge, 1979).
[12] F. J. Ernst, Phys. Rev. **168**, 1415 (1968).
[13] F. J. Ernst, J. Math. Phys. **15**, 1409 (1974).
[14] K. S. Thorne, Phys. Rev. **139**, B244 (1965).
[15] R. Penrose, Nuovo Cimento **1**(special number), 252 (1969).
[16] B. Carter, in *Black Holes*, edited by C. DeWitt and B. S. DeWitt (Gordon and Breach, New York, 1973).
[17] D. C. Robinson, Phys. Rev. D **10**, 458 (1974).
[18] S. W. Hawking, Commun. Math. Phys. **25**, 152 (1972).
[19] W. Israel, Phys. Rev. **164**, 1776 (1967).
[20] W. Israel, Commun. Math. Phys. **8**, 245 (1968).