Generalized Grassmann algebras with applications to Fermi systems

Y. Ohnuki
Department of Physics, Nagoya University, Nagoya 464, Japan
S. Kamefuchi
Institute of Physics, University of Tsukuba, Ibaraki 305, Japan

(Received 10 September 1979; accepted for publication 26 November 1979)

Generalized Grassmann numbers x_i $(i = 1,2,\cdots,n)$ are defined as those satisfying the relations $x_i x_j = \eta_{ij} x_j x_i$ with $\eta_{ij} = -(+ \text{ or } -)$ for $i = j$ $(i \neq j)$. Ordinary Grassmann numbers correspond to a special case $\eta_{ij} = -$ for all i, j. Mathematical properties of such numbers are discussed in detail, and it is found that most of the results known for the ordinary case can naturally be extended to the general case. Applications are made to a description of general Fermi systems where the commutation relations belong to an arbitrary anomalous case.

I. INTRODUCTION

Recently, Grassmann algebras[1] have been extensively utilized in the description of Fermi systems. This is due to their specific property such that Grassmann numbers x_i can play the role of, so to speak, the classical counterparts of quantum Fermi operators a_i or $a_i^\dagger$ satisfying the commutation relations of the normal case, i.e., $[a_i, a_j^\dagger]_+ = \delta_{ij}$ and $[a_i, a_j]_+ = 0$ with $i, j = 1,2,\cdots,n$. On the other hand, it is also known[2] that Fermi operators a_i and $a_i^\dagger$ may alternatively satisfy the commutation relations of anomalous cases which are specified by a set of *relative signatures* $\eta_{ij} = \eta_{ji}$:

$$[a_i, a_j^\dagger]_{-\eta_{ij}} = \delta_{ij}, \quad [a_i, a_j]_{-\eta_{ij}} = 0, \tag{1.1}$$

with $\eta_{ij} = -(- \text{ or } +)$ for $i = j$ $(i \neq j)$. Corresponding to this, we may thus consider new numbers x_i such that the commutation relations are given by

$$[x_i, x_j]_{-\eta_{ij}} = 0 \text{ or } x_i x_j = \eta_{ij} x_j x_i . \tag{1.2}$$

In what follows x_i's of this kind will be called *generalized Grassmann numbers*. They form a *generalized Grassmann algebra*. Evidently, ordinary Grassmann numbers correspond to a particular case such that all $\eta_{ij} = -$. The purpose of the present paper is to study mathematical properties of these numbers and to discuss some possible applications to Fermi systems. The adjectives "Grassmann" and "generalized Grassmann" will hereafter be abbreviated to "G" and "gG," respectively.

One of the most important properties of gG numbers consists in the relationship between different x_i's specified by η_{ij}'s. Throughout the following discussion we shall always work within the framework of a given set of η_{ij}'s, namely, we shall not be concerned simultaneously with other gG numbers having different sets of relative signatures η'_{ij}, η''_{ij}, $\cdots$. The x_i's satisfying Eq. (1.2) are said to form a *gG vector*: $x = (x_1, x_2, \cdots, x_n)$. We shall consider below a number of such vectors $x = (x_1, x_2, \cdots, x_n)$, $x' = (x'_1, x'_2, \cdots, x'_n), \cdots$, which are all subject to

$$x_i x_j = \eta_{ij} x_j x_i, \; x'_i x'_j = \eta_{ij} x'_j x'_i, \; x_i x'_j = \eta_{ij} x'_j x_i \; , \cdots . \tag{1.3}$$

The present paper is arranged as follows: In Sec. II we study algebraic properties of gG numbers. A number of basic lemmas and theorems are proved concerning gG matrices and gG determinants. Section III is devoted to a discussion of "analytical" properties, including differentiation, integration, Fourier transformations, and δ functions. Transformations of gG numbers are discussed in Sec. IV. Lastly, in Sec. V, the algebras are applied to general Fermi systems where the operators a_i and $a_i^\dagger$ satisfy the commutation relations with an arbitrary set of η_{ij}'s [Eq. (1.1)]. Coherent states are explicitly constructed for $a_i^\dagger$'s as well as for a_i's. It is then found that the quantum mechanics of Fermi systems can be formulated in what we shall call the *gG representation*.

II. ALGEBRAIC PROPERTIES: gG MATRICES, gG DETERMINANTS

Let us begin by introducing *gG matrices* $M, M', \cdots$. A gG matrix M is defined as one whose ij element M_{ij} satisfies

$$M_{ij} x_k = \eta_{ik} \eta_{jk} x_k M_{ij} \; , \tag{2.1}$$

$$M_{ij} M'_{kl} = \eta_{ik} \eta_{il} \eta_{jk} \eta_{jl} M'_{kl} M_{ij} \; .$$

The quantity M_{ij} has the same commutation property as that of $x_i x_j$ when commuted with other quantities. More generally, any quantities $T^{i_1 i_2 \cdots}_{j_1 j_2 \cdots}$ which have the same commutation property as that of $x_{i_1} x_{i_2} \cdots x_{j_1} x_{j_2} \cdots$ when commuted with other quantities will hereafter be referred to as *gG tensors*. The product of M and M' is defined as usual by $(MM')_{ij} = M_{ik} M'_{kj}$, where and in the following the summation convention is applied only to those indices that are repeated in factors other than η_{ij}'s. It is clear that MM' also is a gG matrix.

A set of n numbers $\xi = (\xi_1, \xi_2, \cdots, \xi_n)$ with $\xi_i = M_{ij} x_j$ is a gG vector. Forming a product of such ξ_i's and considering the property $(x_i)^2 = 0$ (no summation over i), we define the *gG determinant* corresponding to M as follows[3]:

$$(M_{1j_1} x_{j_1})(M_{2j_2} x_{j_2}) \cdots (M_{nj_n} x_{j_n}) \equiv \det M \cdot x_1 x_2 \cdots x_n \; . \tag{2.2}$$

From the definition it immediately follows that

$$\det(MM') = \det M \cdot \det M', \tag{2.3}$$

$$[\det M, \; x_i] = [\det M, \; M'_{ij}] = 0.$$

The transpose M^T of M, defined by $(M^T)_{ij} \equiv M_{ji}$, is a gG matrix. However, relations such as

 0022-2488/80/040601-08$1.00

$(MM')^T = M'^T M^T$, $\det M^T = \det M$, do not necessarily hold. Thus, we introduce another gG matrix $\tilde{M}$ related to M by

$$\tilde{M}_{ij} \equiv -\eta_{ij}(M^T)_{ij}, \tag{2.4}$$

for which there holds the following:

Theorem 1:

$$\widetilde{(MM')} = \tilde{M}'\tilde{M}, \qquad \det\tilde{M} = \det M. \tag{2.5}$$

Proof: Since Eq. $(2.5)_1$ follows immediately from the definition (2.4), let us prove Eq. $(2.5)_2$ only.[4] By virtue of $(x_i)^2 = 0$ we can also write Eq. (2.2) in the form

$$\begin{aligned}\det M \cdot x_1 x_2 \cdots x_n &= \sum_{\tau\in S_n} (M_{1,\tau 1}x_{\tau 1})(M_{2,\tau 2}x_{\tau 2})\cdots(M_{n,\tau n}x_{\tau n}) \\ &= \sum_{\tau\in S_n} \eta(\tau)(M_{\tau^{-1}1,1}x_1)(M_{\tau^{-1}2,2}x_2)\cdots(M_{\tau^{-1}n,n}x_n),\end{aligned} \tag{2.6}$$

where τ stands for an element $\binom{1\ 2\cdots n}{\tau 1\ \tau 2\cdots\tau n}$ of the symmetric group S_n of order $n!$, and $\eta(\tau)$ is the signature function defined by

$$x_1 x_2 \cdots x_n \equiv \eta(\tau) x_{\tau^{-1}1} x_{\tau^{-1}2} \cdots x_{\tau^{-1}n}. \tag{2.7}$$

Similarly, for $\tilde{M}$ we have

$$\det\tilde{M}\cdot x_1 x_2\cdots x_n = \sum_{\tau\in S_n} \eta(\tau)(\tilde{M}_{\tau^{-1}1,1}x_1)(\tilde{M}_{\tau^{-1}2,2}x_2)\cdots(\tilde{M}_{\tau^{-1}n,n}x_n). \tag{2.8}$$

On the right-hand side of Eq. (2.8) we collect all x_i's at the right end, use Eq. (2.7) there, and then return them to the original positions, i.e.,

$$\begin{aligned}\text{Eq. (2.8)} &= \sum_{\tau\in S_n} \eta(\tau)\prod_{i<j}(\eta_{i,\tau^{-1}j}\eta_{ij})\tilde{M}_{\tau^{-1}1,1} \\ &\quad\times\tilde{M}_{\tau^{-1}2,2}\cdots\tilde{M}_{\tau^{-1}n,n}x_1 x_2\cdots x_n \\ &= \sum_{\tau\in S_n}\prod_{i<j}(\eta_{i,\tau^{-1}j}\eta_{ij})\tilde{M}_{\tau^{-1}1,1} \\ &\quad\times\tilde{M}_{\tau^{-1}2,2}\cdots\tilde{M}_{\tau^{-1}n,n}x_{\tau^{-1}1}x_{\tau^{-1}2}\cdots x_{\tau^{-1}n} \\ &= \sum_{\tau\in S_n}\prod_{i<j}(\eta_{i,\tau^{-1}j}\eta_{ij})\prod_{i>j}(\eta_{\tau^{-1}i,\tau^{-1}j}\eta_{i,\tau^{-1}j}) \\ &\quad\times(\tilde{M}_{\tau^{-1}1,1}x_{\tau^{-1}1}) \\ &\quad\times(\tilde{M}_{\tau^{-1}2,2}x_{\tau^{-1}2})\cdots(\tilde{M}_{\tau^{-1}n,n}x_{\tau^{-1}n}) \\ &= \sum_{\tau\in S_n}\prod_{i\neq j}(\eta_{i,\tau^{-1}j})(\tilde{M}_{\tau^{-1}1,1}x_{\tau^{-1}1}) \\ &\quad\times(\tilde{M}_{\tau^{-1}2,2}x_{\tau^{-1}2})\cdots(\tilde{M}_{\tau^{-1}n,n}x_{\tau^{-1}n}),\end{aligned} \tag{2.8'}$$

where use is made of the relation $\Pi_{i<j}\eta_{ij} = \Pi_{i>j}\eta_{\tau^{-1}i,\tau^{-1}j}$. In Eq. (2.8') we replace τ^{-1} by τ, use Eq. (2.4) to obtain $\tilde{M}_{\tau i,i} = -\eta_{i,\tau i}M_{i,\tau i}$, and then take account of the relation $\Pi_{i,j}\eta_{i,\tau j} = \Pi_{i,j}\eta_{ij} = (-1)^n$. The result is

$$\det\tilde{M}\cdot x_1 x_2\cdots x_n = \sum_{\tau\in S_n}(M_{1,\tau 1}x_{\tau 1})(M_{2,\tau 2}x_{\tau 2})\cdots(M_{n,\tau n}x_{\tau n}), \tag{2.8''}$$

which, according to Eq. (2.6), equals $\det M\cdot x_1 x_2\cdots x_n$. Q.E.D.

Further, we have the following:

Theorem 2: Given a real parameter s, there holds the relation

$$\det[\exp(sM)] = \exp(s\,\mathrm{Tr}M), \tag{2.9}$$

where the equality means that when expanded as power series in s, the terms of the same degree in s on both sides are equal to each other, and $\mathrm{Tr}M \equiv \Sigma_i M_{ii}$.

We shall first prove the following lemma, which is well known for ordinary determinants:

Lemma 1: Let K be a gG matrix such that the elements K_{kj} for the k th row are given as

$$K_{kj} = M_{kj} + L_{kj} \quad (j = 1,2,\cdots,n), \tag{2.10}$$

and the elements K_{ij} $(i\neq k)$ for other rows are given as

$$K_{ij} = M_{ij} = L_{ij}. \tag{2.10'}$$

Then, there holds the relation

$$\det K = \det M + \det L. \tag{2.11}$$

Proof:

$$\begin{aligned}\det K\cdot x_1 x_2\cdots x_n &= (K_{1j_1}x_{j_1})\cdots(K_{k-1,j_{k-1}}x_{j_{k-1}})\{(M_{kj_k}+L_{kj_k})x_{j_k}\} \\ &\quad\times(K_{k+1,j_{k+1}}x_{j_{k+1}})\cdots(K_{nj_n}x_{j_n}) \\ &= (K_{1j_1}x_{j_1})\cdots(K_{k-1,j_{k-1}}x_{j_{k-1}})(M_{kj_k}x_{j_k}) \\ &\quad\times(K_{k+1,j_{k+1}}x_{j_{k+1}})\cdots(K_{nj_n}x_{j_n}) \\ &\quad+(K_{1j_1}x_{j_1})\cdots(K_{k-1,j_{k-1}}x_{j_{k-1}})(L_{kj_k}x_{j_k}) \\ &\quad\times(K_{k+1,j_{k+1}}x_{j_{k+1}})\cdots(k_{nj_n}x_{j_n}) \\ &= (\det M + \det L)x_1 x_2\cdots x_n.\end{aligned} \tag{2.12}$$

Q.E.D.

Let $K(s)$ be a gG matrix whose elements are all differentiable with respect to a real parameter s. By virtue of the above lemma we can then easily obtain $d/ds\cdot(\det K(s))$: First construct a matrix $M_{(k)}(s)$ from $K(s)$ by making a replacement of the elements of the k th row such as $K(s)_{kj} \to d/ds\cdot K(s)_{kj}$ $(j = 1,2,\cdots,n)$, compute the determinant of the resulting matrix, and then sum the result over all k. That is to say,

$$\frac{d}{ds}\det K(s) = \sum_{k=1}^{n}\det M_{(k)}(s), \tag{2.13}$$

where

$$M_{(k)}(s)_{ij} = (1-\delta_{ik})K(s)_{ij} + \delta_{ik}\frac{d}{ds}K(s)_{ij}. \tag{2.14}$$

Let us now put $K(s) = \exp(sM)$, where M is a gG matrix independent of s. Then, $M_{(k)}$ is obtained by use of Eq. (2.14) as

$$\begin{aligned}M_{(k)}(s)_{ij} &= (1-\delta_{ik})K(s)_{ij} + \delta_{ik}\{MK(s)\}_{ij} \\ &= \{G_{(k)}K(s)\}_{ij},\end{aligned} \tag{2.15}$$

with

$$G_{(k)ij} \equiv (1-\delta_{ik})\delta_{ij} + \delta_{ik}M_{ij}. \tag{2.16}$$

Here, $G_{(k)}$ is a gG matrix such that the elements of the k th row are given by M_{kj} $(j = 1,2,\cdots,n)$, the diagonal elements are unity, except for the kk element being M_{kk}, and all other elements are zero. Equation (2.2) therefore gives $\det G_{(k)}$

$= M_{kk}$. Then, from Eqs. (2.15) and $(2.3)_1$ we find that $\det M_{(k)}(s) = M_{kk}\det K(s)$, which when substituted in Eq. (2.13) gives

$$\frac{d}{ds}\det K(s) = \mathrm{Tr}M\det K(s). \tag{2.17}$$

On the other hand, Eq. $(2.3)_2$ shows that $\det K(s)$ may be treated as if it were a c number. Thus, by integrating Eq. (2.17) under the condition $K(0) = 1$ we arrive at $\det K(s) = \exp(s\,\mathrm{Tr}M)$, thereby completing the proof of Theorem 2.

Replacing $\exp M$ by M in Eq. (2.9) with $s = 1$, we obtain $\det M = \exp(\mathrm{Tr}\log M)$. This formula is convenient for formal operations, but the meaning of $\log M$ is not clear for the case of general M's.[5]

Let us now turn to the questions of under what conditions the inverse M^{-1} of a gG matrix M exists, and of whether M^{-1}, if it exists, is uniquely determined. We start our discussion by proving the following:

Lemma 2: For a gG matrix M, there holds the relation

$$\begin{aligned} &M_{ij}\frac{\partial}{\partial M_{kj}}\det M \\ &= \left(\frac{\partial}{\partial M_{ji}}\det M\right)M_{jk} \\ &= \delta_{ik}\det M. \end{aligned} \tag{2.18}$$

Here the operator $\partial/\partial M_{kj}$ means left differentiation, and the left derivative $\partial/\partial M_{kj}\cdot\det M$ is obtained as follows: Since $\det M$ is expressible as a homogeneous polynomial in the matrix elements, the derivative in question is given as a sum of the derivatives of the respective monomials. Any of such monomials contains in it at most one M_{kj}. If the monomial does not contain M_{kj}, its left derivative is zero, of course. On the other hand, if it contains one M_{kj}, we shift the factor M_{kj} to the leftmost position by use of Eq. $(2.1)_2$ with $M = M'$, and then drop the M_{kj} to obtain the derivative. Later we shall also use the right derivative $\bar\partial/\partial M_{ji}\cdot\det M$. In this case the procedure differs from the above one in that the factor M_{ji}, if contained in a monomial, is to be shifted to the rightmost position and then dropped there.

The proof of the above lemma proceeds as follows: We begin by considering the first part thereof, i.e., $M_{ij}\partial/\partial M_{kj}\cdot\det M = \delta_{ik}\det M$. By use of the relation $(M_{ij}x_j)(M_{i'j'}x_{j'}) = \eta_{ii'}(M_{i'j'}x_{j'})(M_{ij}x_j)$, which follows from Eq. (2.1), we shift the factor $(M_{kj}x_j)$ contained in the left-hand side of Eq. (2.2) to the leftmost position and then perform the operation $\partial/\partial M_{kj}$. In this way we find that

$$\begin{aligned} \left(\frac{\partial}{\partial M_{kj}}\det M\right)x_1x_2\cdots x_n &= \eta_{1k}\eta_{2k}\cdots\eta_{k-1,k}\,x_j(M_{1j_1}x_{j_1}) \\ &\cdots(M_{k-1,j_{k-1}}x_{j_{k-1}})(M_{k+1,j_{k+1}}x_{j_{k+1}})\cdots(M_{nj_n}x_{j_n}), \end{aligned} \tag{2.19}$$

which, when multiplied from the left by M_{ij} and summed over j, results in

$$\begin{aligned} &\left(M_{ij}\frac{\partial}{\partial M_{kj}}\det M\right)x_1x_2\cdots x_n = \eta_{1k}\eta_{2k}\cdots\eta_{k-1,k} \\ &\times(M_{ij}x_j)(M_{1j_1}x_{j_1})\cdots(M_{k-1,j_{k-1}}x_{j_{k-1}}) \\ &\times(M_{k+1,j_{k+1}}x_{j_{k+1}})\cdots(M_{nj_n}x_{j_n}). \end{aligned} \tag{2.20}$$

If $i = k$, we shift the factor $(M_{ij}x_j)$ on the right-hand side to the position in between $(M_{k-1,j_{k-1}}x_{j_{k-1}})$ and $(M_{k+1,j_{k+1}}x_{j_{k+1}})$, thereby obtaining the left-hand side of Eq. (2.2), i.e., $\det M\cdot x_1x_2\cdots x_n$. On the other hand, if $i \neq k$, the right-hand side of Eq. (2.20) contains the factor $(M_{ij}x_j)$ twice, and hence must vanish owing to $(M_{ij}x_j)^2 = 0$. This completes the proof of the first part of the lemma.

Next we consider the second part of the lemma, i.e., $(\partial/M_{ji}\cdot\det M)M_{jk} = \delta_{ik}\det M$. Using Eqs. (2.3), (2.4), and $(2.5)_2$, we find that

$$\begin{aligned} x_1x_2\cdots x_n\cdot\det M &= \det\widetilde{M}\cdot x_1x_2\cdots x_n \\ &= (\widetilde{M}_{1j_1}x_{j_1})(\widetilde{M}_{2j_2}x_{j_2})\cdots(\widetilde{M}_{nj_n}x_{j_n}) \\ &= (x_{j_1}M_{j_11})(x_{j_2}M_{j_22})\cdots(x_{j_n}M_{j_nn}). \end{aligned} \tag{2.21}$$

On the right-hand side of Eq. (2.21) we shift the factor (x_jM_{ji}) to the rightmost position, perform the operation $\bar\partial/\partial M_{ji}$, multiply M_{jk} from the right, and sum the resulting expression over j. Then, by repeating the same arguments as above we arrive at

$$\left(\frac{\bar\partial}{\partial M_{ji}}\det M\right)M_{jk} = \delta_{ik}\det M. \tag{2.22}$$

For $\det M$, however, the right and left derivatives with respect to an arbitrary matrix element are equal:

$$\frac{\partial}{\partial M_{ji}}\det M = \frac{\bar\partial}{\partial M_{ji}}\det M. \tag{2.23}$$

To prove this we note that $\det M$ is given as a sum of terms such as $M_{\sigma1,\tau1}M_{\sigma2,\tau2}\cdots M_{\sigma n,\tau n}(\sigma,\tau\in S_n)$ apart from a sign, where $2n$ indices $\sigma1,\tau1,\sigma2,\tau2,\cdots,\sigma n,\tau n$ contain every index twice. This means that any term thereof commutes with M_{ji} owing to Eq. $(2.1)_2$. Thus, if we write a term containing M_{ji} as $M_{ji}T$, then we have $M_{ji}T = TM_{ji}$, which leads to Eq. (2.23). Combining Eq. (2.22) with (2.23), we find that the second part of the lemma is true. Q.E.D.

The quantities defined by

$$\Delta_{ij} \equiv \frac{\partial}{\partial M_{ji}}\det M, \tag{2.24}$$

now provides us with a gG matrix Δ, and Eq. (2.18) can then be written as $\Delta\cdot M = M\cdot\Delta = \det M$. This implies that if $(\det M)^{-1}$ exists, then the inverse matrix M^{-1} of M is uniquely given as $M^{-1} = (\det M)^{-1}\Delta$. Conversely, if M^{-1} exists, then $M^{-1}M = 1$ yields $\det(M^{-1})\det M = 1$, thereby implying the existence of $(\det M)^{-1}$. Summarizing, we obtain the following:

Theorem 3: The inverse matrix M^{-1} of a gG matrix M exists if and only if $(\det M)^{-1}$ exists.

We conclude the present section by making a remark on more general cases. The above discussion has been confined to those gG matrices (and determinants) which combine gG vectors having the same relative signatures η_{ij}. It is possible, however, to remove such a restriction and to generalize gG matrices. Introducing a gG matrix such as

$$\xi_\kappa = M_{\kappa i}x_i \quad (i,\kappa = 1,2,\cdots,n), \tag{2.25}$$

where

$$\begin{aligned} &x_ix_j = \eta_{ij}x_jx_i, \quad \xi_\kappa\xi_\rho = \eta'_{\kappa\rho}\xi_\rho\xi_\kappa, \\ &x_i\xi_\kappa = \eta''_{i\kappa}\xi_\kappa x_i,\ \eta_{ii} = \eta'_{\kappa\kappa} = -1 \text{ (no summation over } i,\kappa), \text{ and} \end{aligned}$$

$$M_{\kappa i}M_{\rho j} = \eta_{ij}\eta'_{\kappa\rho}\eta''_{i\rho}\eta''_{j\kappa}M_{\rho j}M_{\kappa i}, \tag{2.26}$$

we define the corresponding determinant by

$$\xi_1\xi_2\cdots\xi_n \equiv \det M \cdot x_1x_2\cdots x_n . \tag{2.27}$$

In this case $\det M$ does not necessarily commute with x_i. On the basis of such matrices it is possible to introduce Klein transformations[6] for gG vectors. The problem will be discussed elsewhere.

III. ANALYTICAL PROPERTIES: DIFFERENTIATION, INTEGRATION

Regarding a gG vector $x = (x_1,x_2,\cdots,x_n)$ as a kind of variables, to be referred to hereafter as gG variables, we consider a function $f(x)$. Since $(x_i)^2 = 0, f(x)$ is always expressible as a polynomial in x_i, where the coefficients may consist of complex numbers, other gG numbers, matrix elements of gG matrices, etc.

As for differentiation we employ again left differentiation. The procedure to obtain the left derivative $\partial/\partial x_i \cdot f(x)$ is extremely simple here: For any monomial $m(x)$ contains x_i, at most, once. It then follows that the operators $\partial/\partial x_i$ are subject to the commutation relation $[\partial/\partial x_i, \partial/\partial x_j]_{-\eta_{ij}} = 0$.

For integration we adopt, as usual, the formulas[1,7]

$$\int x_j dx_j \equiv i, \quad \int dx_j \equiv 0 \quad (j = 1,2,\cdots,n). \tag{3.1}$$

These are further generalized to the case of multiple integrals:

$$\begin{aligned} &\int x_1x_2\cdots x_n d^n x = (i)^n , \\ &\int x_{j_1}x_{j_2}\cdots x_{j_r} d^n x = 0 \quad (0 \leqslant r < n). \end{aligned} \tag{3.2}$$

The linearity relation is also assumed:

$$\begin{aligned} &\int (\alpha_1 f_1(x) + \alpha_2 f_2(x)) d^n x \\ &\quad = \alpha_1 \int f_1(x) d^n x + \alpha_2 \int f_2(x) d^n x, \end{aligned} \tag{3.3}$$

where α_1 and α_2 are quantities independent of x, and $f_1(x)$ and $f_2(x)$ are arbitrary functions of x.

It is to be remarked here that since the notion of "magnitude" cannot be associated with gG variables themselves, the integration considered above should not be regarded as an infinite sum of infinitesimal quantities, but rather as a kind of linear mapping. It is often convenient to write

$$d^n x \equiv dx_n dx_{n-1} \cdots dx_1. \tag{3.4}$$

When combined with the symbol $\int$, $d^n x$ thus means a succession of mappings such that we perform firstly a mapping with respect to x_n in a way specified by Eq. (3.1) with $j = n$, secondly a similar mapping with respect to x_{n-1}, $\cdots$, and lastly a similar mapping with respect to x_1. Thus, if the order of dx_i's is changed in Eq. (3.4), this will necessitate a corresponding change in the order of mappings. The consistency with the definition of integration then requires that[8]

$$dx_i dx_j = \eta_{ij} dx_j dx_i \quad (i \neq j). \tag{3.5}$$

Further, it is also consistent to treat dx_i's as if they were gG numbers satisfying

$$dx_i x_j = \eta_{ij} x_j dx_i . \tag{3.6}$$

More generally, if α is a gG number such that $x_i\alpha = \eta\alpha x_i$, then we may put

$$dx_i \alpha = \eta\alpha dx_i . \tag{3.7}$$

Relations (3.5)–(3.7) will prove useful for practical purposes.

The δ function for a gG variable x_j belonging to a gG vector $x = (x_1,x_2,\cdots x_n)$ is defined by[7]

$$\delta(x_j - x_j') \equiv \frac{1}{i}(x_j - x_j') , \tag{3.8}$$

where x_j' belongs to a gG vector $x' = (x_1',x_2',\cdots,x_n')$ which is independent of x. From the definition of integration it is easy to check that for an arbitrary function $f(x_j)$, which is always written in the form $\alpha + \beta x_j$, there holds the relation

$$\int f(x_j)\delta(x_j - x_j')dx_j = f(x_j') . \tag{3.9}$$

The δ function is also written in the form of a Fourier integral[7]

$$\delta(x_j - x_j') = \int \exp[-(x_j - x_j')x_j'']dx_j'' . \tag{3.10}$$

These are easily generalized to the case of n gG variables. The definition

$$\delta(x - x') \equiv \delta(x_1 - x_1')\delta(x_2 - x_2')\cdots\delta(x_n - x_n'), \tag{3.11}$$

immediately leads us to

$$\int f(x)\delta(x - x')d^n x = f(x') \tag{3.12}$$

for an arbitrary function $f(x)$ of a gG vector x. From Eq. (3.10) its Fourier integral is found to be

$$\delta(x - x') = \int \exp[-(\overline{x - x'}\cdot x'')]\widetilde{d^n x''}, \tag{3.13}$$

where $(x\cdot x') \equiv \Sigma_{i=1}^n x_i x_i'$ and $\widetilde{d^n x} \equiv dx_1dx_2\cdots dx_n$.

Substituting Eq. (3.13) in (3.12), we find

$$\begin{aligned} f(x') &= \int\int f(x)\exp[-(\overline{x - x'}\cdot x'')]\widetilde{d^n x''}d^n x \\ &= (-1)^n \int\int f(x)\exp(x''\cdot x)d^n x \exp(x'\cdot x'')\widetilde{d^n x''}, \end{aligned} \tag{3.14}$$

where use is made of $\widetilde{d^n x''}d^n x = (\Pi_{i,j}\eta_{ij})d^n x\widetilde{d^n x''}$ [cf. Eq. (3.5) and the footnote[8]] and $\Pi_{i,j}\eta_{ij} = (-1)^n$. Performing a change of variables $x_j'' \to -x_j''$ $(j = 1,2,\cdots,n)$ in Eq. (3.14), we arrive at what corresponds to Fourier transformations.

Theorem 4: Let $\tilde{f}(x')$ be the Fourier transform of a function $f(x)$:

$$\tilde{f}(x') = \int f(x)\exp[-(x'\cdot x)]d^n x . \tag{3.15}$$

Then the inverse transformation is given by

$$f(x) = \int \tilde{f}(x')\exp[-(x\cdot x')]\widetilde{d^n x'} . \tag{3.16}$$

It is of interest to note that in contrast with the case of real variables the above integrals exist for any functions $f(x)$.

IV. TRANSFORMATIONS OF gG VARIABLES

In the present section we take up the problem of transformations or changes of gG variables. Here we are especially interested in the question of how the measure $d^n x$ is transformed under a change of variables $x = (x_1, x_2, \cdots, x_n) \rightarrow x' = (x_1', x_2', \cdots, x_n')$. Now the new variables x' can generally be expressed as a polynomial in x, such as

$$x_i' = \lambda_i + x_j \lambda_{ji} + \sum_{r=2}^{n} x_{j_1} x_{j_2} \cdots x_{j_r} \lambda_{j_1 j_2 \cdots j_r i} , \tag{4.1}$$

where the coefficients λ_i, λ_{ji}, and $\lambda_{j_1 j_2 \cdots j_r i}$ are gG tensors independent of x. As before we are concerned here with such changes as $x \rightarrow x'$ with x and x' being gG vectors sharing a fixed set of η_{ij}'s. Let us now denote by λ the gG matrix whose ji element is given by λ_{ji}, i.e., $\lambda = \|\lambda_{ji}\|$, and assume that λ^{-1} exists. According to Theorem 3, this assumption means the existence of $(\det\lambda)^{-1}$, the significance of which will be made clear by Theorem 5 below. For the sake of brevity let us write the third term in Eq. (4.1) as $Q_i(x)$:

$$Q_i(x) \equiv \sum_{r=2}^{n} x_{j_1} x_{j_2} \cdots x_{j_r} \lambda_{j_1 j_2 \cdots j_r i} . \tag{4.2}$$

Thus, Eq. (4.1) is rewritten as

$$x_i' = \lambda_i + x_j \lambda_{ji} + Q_i(x) . \tag{4.1'}$$

In this case the following theorem holds:

Theorem 5: The inverse transformation of Eq. (4.1') exists if and only if λ^{-1} exists.

Proof: Let us first assume that λ^{-1} exists. Then, multiplying both sides of Eq. (4.1') by $(\lambda^{-1})_{ik}$ from the right and summing the result over i, we find

$$x_k = (x_i' - \lambda_i)(\lambda^{-1})_{ik} - Q_i(x)(\lambda^{-1})_{ik} , \tag{4.3}$$

which can be solved for x by iteration. After repeating the iteration, at most, n times, we can express x as a polynomial in x'. Clearly, the inverse transformation thus obtained is unique. This proves the sufficiency part of the condition. Next we assume that the inverse transformation exists. It can then be written as

$$x_i = \lambda_i' + x_j' \lambda_{ji}' + Q_i'(x') , \tag{4.4}$$

or equivalently

$$= \lambda_i'' + x_j'' \lambda_{ji}'' + Q_i''(x'') , \tag{4.4'}$$

where $x_i'' \equiv x_i' - \lambda_i$ and $Q_i'(x')$ $[Q_i''(x'')]$ is a polynomial in x' (x'') consisting of terms of the second or higher degrees. Substituting x' given by Eq. (4.1'), i.e., $x_j'' = x_k \lambda_{kj} + Q_j(x)$, in the right-hand side of Eq. (4.4'), we obtain an identity in x such as

$$x_i = \lambda_i'' + x_k \lambda_{kj} \lambda_{ji}'' + Q_i'''(x) , \tag{4.5}$$

where $Q_i'''(x)$ is a polynomial in x of the same kind as the Q_i's introduced above. It therefore follows from Eq. (4.5) that $\lambda_i'' = 0$, $Q_i'''(x) = 0$, and

$$\lambda_{kj} \lambda_{ji}'' = \delta_{ki} , \tag{4.6}$$

which implies that $(\det\lambda)^{-1} = \det\lambda''$, namely, that λ^{-1} exists owing to Theorem 3. This proves the necessity part of the condition given above. Q.E.D.

The reason why we have assumed above the existence of λ^{-1} is now clear: By making this assumption we are confining ourselves to those transformations whose inverses exist. In this case all gG vectors $x = (x_1, x_2, \cdots, x_n)$, $x' = (x_1', x_2', \cdots, x_n'), \cdots$ which are connected with each other by such transformations are placed, so to speak, on an equal footing. Accordingly, any basic formulas valid for x are required to be so for any other x's. For example, Eq. (3.2) should hold true in the same form but with x_i and $d^n x$ being replaced by x_i' and $d^n x'$, respectively. In order to ensure such invariance of the integral formulas we have to require the measure $d^n x$ to satisfy Eq. (4.7) below.

Theorem 6: The necessary and sufficient condition for Eq. (3.2) to remain invariant under a change of variables $x = (x_1, x_2, \cdots, x_n) \rightarrow x' = (x_1', x_2', \cdots, x_n')$ is that

$$d^n x' = (\det J)^{-1} d^n x , \tag{4.7}$$

where J is the gG matrix whose ji element is given by $\partial / \partial x_j \cdot x_i'$, i.e.,

$$J = \|\partial x_i' / \partial x_j\| . \tag{4.8}$$

Proof: The proof is performed in essentially the same way as in the case of G variables.[7] Thus, we show first that in the transformation (4.1') the inverse of $\det J$ exists. Taking the left derivative thereof, we find that J is given in matrix form as

$$J = \lambda \left(1 + \lambda^{-1} \left\| \frac{\partial}{\partial x_j} Q_i(x) \right\| \right), \tag{4.9}$$

by virtue of the existence of λ^{-1} being assumed. Here every matrix element $\partial / \partial x_j \cdot Q_i(x)$ of $\|\partial / \partial x_j \cdot Q_i(x)\|$ is given as a polynomial in x consisting of terms of the first or higher degrees, whence J^{-1} is obtained as

$$\begin{aligned} J^{-1} &= \left(1 + \lambda^{-1} \left\| \frac{\partial}{\partial x_j} Q_i(x) \right\| \right)^{-1} \lambda^{-1} \\ &= \sum_{m=0}^{n} \left(-\lambda^{-1} \left\| \frac{\partial}{\partial x_j} Q_i(x) \right\| \right)^m \lambda^{-1} , \end{aligned} \tag{4.10}$$

which guarantees the existence of J^{-1} and hence that of $(\det J)^{-1}$ owing to Theorem 3.

Next we consider two transformations $x_i' = f_i(x)$ and $x_i' = g_i(x)$, specified by functions f and g. In the following the former, for example, will be referred to simply as the transformation f. When the transformation g is applied to x, i.e., $x_i' = g_i(x)$, and then followed by the transformation f, i.e., $x_i'' = f_i(x')$, let us denote the resulting transformation $x_i'' = f_i[g(x)]$ by $x_i'' = (fg)_i(x)$ and call it the transformation fg. In this case it is clear from the definition of left differentiation that

$$\frac{\partial x_i''}{\partial x_j} = \frac{\partial x_k'}{\partial x_j} \frac{\partial x_i''}{\partial x_k'} , \tag{4.11}$$

whence

$$\det \left\| \frac{\partial x_i''}{\partial x_j} \right\| = \det \left\| \frac{\partial x_k'}{\partial x_j} \right\| \det \left\| \frac{\partial x_i''}{\partial x_k'} \right\| . \tag{4.12}$$

The above result implies that if Theorem 6 is true for the transformations f and g separately, then it is also true for the transformation fg. [Notice that $(\det J)^{-1} d^n x = d^n x (\det J)^{-1}$ on account of Eq. (2.3) and (3.7)]. Thus, to prove the theorem for a general transformation (4.1') we can proceed as

follows: We first express Eq. (4.1') as a succession of some basic transformations and then prove the theorem for each of such transformations.

In view of this we consider $(n+1)$ transformations $f_{(1)}, f_{(2)},\cdots, f_{(n)}, f_{(n+1)}$ such that

$$f_{(k)i}(x)\equiv x_i + \delta_{ik}\widetilde{Q}_k(x) \quad (k=1,2,\cdots,n),$$
$$f_{(n+1)i}(x)\equiv\lambda_i + x_j\lambda_{ji}\,, \tag{4.13}$$

where $\widetilde{Q}_k(x)$ is a polynomial in x consisting of terms of the second or higher degrees. Let us now focus our attention on the transformation $x_i' = (f_{(n+1)}\, f_{(n)} f_{(n-1)}\cdots f_{(2)} f_{(1)})_i(x)$. A simple calculation shows that

$$x_i' = \lambda_i + x_j\lambda_{ji} + \widetilde{Q}_j(f_{(j-1)}\, f_{(j-2)}\cdots f_{(1)}(x))\lambda_{ji}\,, \tag{4.14}$$

where $\widetilde{Q}_j(f_{(j-1)}\, f_{(j-2)}\cdots f_{(1)}(x))|_{j=1}\equiv\widetilde{Q}_1(x)$. It is now evident that Eq. (4.14) coincides with Eq. (4.1'), provided $Q_i(x)$'s in Eq. (4.1') or $\widetilde{Q}_k(x)$'s in Eq. (4.13) are chosen to be

$$\begin{aligned} Q_i(x) &= \widetilde{Q}_j(f_{(j-1)}\, f_{(j-2)}\cdots f_{(1)}(x))\lambda_{ji}\,,\\ \widetilde{Q}_i(x) &= Q_j[(f_{(i-1)}\, f_{(i-2)}\cdots f_{(1)})^{-1}(x)](\lambda^{-1})_{ji}\,. \end{aligned} \tag{4.15}$$

This means that by successively applying the transformations (4.13) with $\widetilde{Q}_i(x)$'s being chosen in an appropriate manner we can realize an arbitrary transformation.

We now show that our theorem holds true for each of the transformations (4.13). Let us first consider $f_{(n+1)}$. In this case we have

$$x_1'x_2'\cdots x_n' = (x_{j_1}\lambda_{j_1 1})(x_{j_2}\lambda_{j_2 2})\cdots(x_{j_n}\lambda_{j_n n}) + P_{(n-1)}(x)\,, \tag{4.16}$$

where $P_{(l)}(x)$ stands for a polynomial in x of degree not higher than l. Substituting $x_j\lambda_{ji} = \tilde{\lambda}_{ij}x_j$ in Eq. (4.16) and using Eqs. (2.2) and $(2.5)_2$, we can rewrite Eq. (4.16) as

$$x_1'x_2'\cdots x_n' = \det\lambda\cdot x_1x_2\cdots x_n + P_{(n-1)}(x). \tag{4.16'}$$

On the other hand, we obviously have

$$x_{j_1}'x_{j_2}'\cdots x_{j_r}' = P_{(r)}(x) \quad (0\leqslant r<n). \tag{4.17}$$

Noticing the fact that $\det J = \det\lambda$ for the transformation $f_{(n+1)}$, we find that the necessary and sufficient condition concerned is given precisely by Eq. (4.7).

Next we consider the transformation $f_{(k)}$ $(k=1,2,\cdots,n)$. In this case $\widetilde{Q}_k(x)$ is written as a sum of two terms:

$$\widetilde{Q}_k(x) = x_k U_{(k)}(x) + V_{(k)}(x)\,, \tag{4.18}$$

where $U_{(k)}(x)$ and $V_{(k)}(x)$ do not contain x_k. Noticing the fact that $U_{(k)}(x)$ commutes with any x_i because of $\widetilde{Q}_k(x)x_i = \eta_{ik}x_i\widetilde{Q}_k(x)$, we find from Eq. $(4.13)_1$ that

$$\begin{aligned} x_1'x_2'\cdots x_n' = {}& (1+U_{(k)}(x))x_1x_2\cdots x_n\\ &+ x_1x_2\cdots x_{k-1}V_{(k)}(x)x_{k+1}x_{k+2}\cdots x_n\,. \end{aligned} \tag{4.19}$$

Here the polynomial $V_{(k)}(x)$ does not contain terms of the zeroth and first degrees and moreover does not contain x_k. Hence, the second term on the right-hand side of Eq. (4.19) must vanish [in fact, such terms can be nonvanishing only when $V_{(k)}(x)$ is of the form $\alpha+\beta x_k$]. Further the term containing $U_{(k)}(x)$ in the above vanishes as well, but is retained there for later convenience. At any rate, Eq. (4.19) is rewritten in the form

$$x_1'x_2'\cdots x_n' = (1+U_{(k)}(x))x_1x_2\cdots x_n\,. \tag{4.19'}$$

As for $x_{j_1}'x_{j_2}'\cdots x_{j_r}'(0\leqslant r<n)$ we may divide the possibilities in the following way: case (i) in which the monomial contains x_k; case (ii) in which the monomial does not contain x_k. In case (i) let us put $j_1 = k$ for simplicity. The following results are then obtained:

$$x_{j_1}'x_{j_2}'\cdots x_{j_r}' = \begin{cases} (1+U_{(k)}(x))x_{j_1}x_{j_2}\cdots x_{j_r} + V_{(k)}(x)x_{j_2}x_{j_3}\cdots x_{j_r}, \\ \qquad\qquad \text{for case (i)},\\ x_{j_1}x_{j_2}\cdots x_{j_r} \quad (j_l\neq k,\ l=1,2,\cdots,r<n), \\ \qquad\qquad \text{for case (ii)}, \end{cases} \tag{4.20}$$

where it is to be noticed that the right-hand sides, except for the first term of case (i), do not contain x_k.

On the other hand, it is easy to see that for the transformation $f_{(k)}$ $(k=1,2,\ldots,n)$

$$\det J = 1 + U_{(k)}(x). \tag{4.21}$$

From Eqs. (4.19'), (4.20), and (4.21) we thus find that the necessary and sufficient condition concerned is given again by Eq. (4.7) for the present case. Q.E.D.

V. APPLICATIONS TO FERMI SYSTEMS: COHERENT STATES, gG REPRESENTATION

In the present section we shall generalize the state-vector space $\mathscr{H}$ for a system described by Fermi operators a_i and $a_i^\dagger$ $(i=1,2,\ldots,n)$ to $\mathscr{H}_{\mathrm{G}}$ by introducing gG numbers. Here the Fermi operators a_i and $a_i^\dagger$ are assumed to satisfy the commutation relations (1.1) with an arbitrary set of η_{ij}'s. Our generalization consists in employing, together with complex numbers, gG numbers or more generally gG tensors like elements of gG matrices as coefficients of basis vectors $a_{j_1}^\dagger a_{j_2}^\dagger\cdots a_{j_r}^\dagger|0\rangle$ $(r=0,1,\cdots,n)$ which span the state-vector space. Since gG numbers do not form a field, the resulting space $\mathscr{H}_{\mathrm{G}}$ is neither a linear nor a Hilbert space in the usual sense of the word. The dual space $\mathscr{H}^*$ of $\mathscr{H}$ is also genealized to $\mathscr{H}_{\mathrm{G}}{}^*$ in such a way that gG as well as complex numbers are allowed to be coefficients of basis vectors $\langle 0|a_{j_r}a_{j_{r-1}}\cdots a_{j_1}$.

For definiteness let us further introduce the following assumptions: Firstly, as far as the commutation relations with gG numbers or tensors are concerned, the operators a_i and $a_i^\dagger$ behave in the same way as gG numbers x_i. Thus, for example,

$$\begin{aligned} &\hat{a}_i x_j = \eta_{ij}x_j\hat{a}_i\,,\\ &\hat{a}_i T^{j_1j_2\cdots}_{k_1k_2\cdots} = \eta_{ij_1}\eta_{ij_2}\cdots\eta_{ik_1}\eta_{ik_2}\cdots T^{j_1j_2\cdots}_{k_1k_2\cdots}\hat{a}_i\,, \end{aligned} \tag{5.1}$$

where $\hat{a}_i$ stands for a_i or $a_i^\dagger$, and $T^{j_1j_2\cdots}_{k_1k_2\cdots}$ for a gG tensor. Secondly, the vacuum state $|0\rangle$ has commutation properties such as

$$\begin{aligned} &x_i|0\rangle = |0\rangle x_i, \quad T^{j_1j_2\cdots}_{k_1k_2\cdots}|0\rangle = |0\rangle T^{j_1j_2\cdots}_{k_1k_2\cdots},\\ &\langle 0|x_i = x_i\langle 0|, \quad \langle 0|T^{j_1j_2\cdots}_{k_1k_2\cdots} = T^{j_1j_2\cdots}_{k_1k_2\cdots}\langle 0|. \end{aligned} \tag{5.2}$$

Lastly, in view of the fact that the right-hand sides of Eqs. (3.1) and (3.2) are c numbers, we must require in accordance with Eqs. (5.1) and (5.2) that

$$\hat{a}_i dx_j = \eta_{ij} dx_j \hat{a}_i \,, \qquad dx_i|0\rangle = |0\rangle dx_i, \quad \langle 0|dx_i = dx_i\langle 0|. \tag{5.3}$$

On the basis of Eqs. (1.1), (5.1), and (5.2) we can now compute inner products between state vectors of $\mathscr{H}_G$ and those of $\mathscr{H}_G{}^*$. For example, the inner product of $\langle 0|a_i \in \mathscr{H}_G{}^*$ and $M_{jk}a_l^\dagger|0\rangle \in \mathscr{H}_G$ is found to be $\eta_{jl}\eta_{kl}\delta_{il}M_{jk}$, which is not a c number. In fact, state vectors of $\mathscr{H}_G$ are not, in general, those susceptible of the usual probability interpretation. In this sense they are not directly connected with observational facts. As will be seen below, however, it is possible to arrive at physical results through the intermediary of such formal objects.

Let us now introduce the state vectors $|x\rangle$ and $\langle x|$ which will play an important role in the following discussion:[9]

$$|x\rangle \equiv \exp(a^\dagger \cdot x)|0\rangle, \quad \langle x| \equiv \langle 0|\delta(x-a), \tag{5.4}$$

where $(a^\dagger \cdot x) \equiv \Sigma_{i=1}^n a_i^\dagger x_i$, and

$$\delta(x-a) \equiv \delta(x_1-a_1)\delta(x_2-a_2)\cdots\delta(x_n-a_n), \qquad \delta(x_j-a_j) \equiv \frac{1}{i}(x_j-a_j)\,. \tag{5.5}$$

In this respect we notice that the following relations hold:

$$\exp(-a_i^\dagger x_i)a_j \exp(a_i^\dagger x_i) = a_j + \delta_{ij}x_j \,,$$

$$\delta(x_i-a_i)a_i = \delta(x_i-a_i)x_i, \quad \text{(no summation over } i\text{)}. \tag{5.6}$$

With the help of Eq. (5.6) we can show that $|x\rangle$ and $\langle x|$ are the eigenstates, or coherent states, of a_i with eigenvalue x_i being a gG number:

$$a_i|x\rangle = x_i|x\rangle, \qquad \langle x|a_i = \langle x|x_i \,. \tag{5.7}$$

This corresponds to the situation for the case of Bose operators that the annihilation operators have eigenstates, called coherent states, where their eigenvalues are regarded as the classical counterparts of the operators. It is worth noticing, in the present case, that not only ket but also bra vectors can be eigenstates of a_i's. This is due to the fact that a_i's are bounded operators, contrary to the case of Bose operators.

One of the most characteristic features of the above coherent states is stated as the following:

Theorem 7: The coherent states for the operators a_i satisfy the orthocompleteness relation:[10]

$$\langle x|x'\rangle = \delta(x-x'), \qquad \int |x\rangle\langle x|\, d^n x = 1. \tag{5.8}$$

Proof: The first equation follows directly from Eqs. (5.4) and (5.7). The proof of the second equation proceeds as follows: For the state vectors $|k\rangle$ $(k = 0,1,\cdots,n)$ defined by

$$|k\rangle \equiv a_k^\dagger a_{k-1}^\dagger \cdots a_2^\dagger a_1^\dagger |0\rangle, \tag{5.9}$$

there holds the relation

$$\begin{aligned}\int \exp(a_k^\dagger x_k)|k-1\rangle\langle k-1|\delta(x_k-a_k)dx_k &= (-i)\int (1+a_k^\dagger x_k)|k-1\rangle\langle k-1|(x_k-a_k)\,dx_k \\ &= (-i)\Big\{\int |k-1\rangle\langle k-1|\, x_k\, dx_k \\ &\quad + \int a_k^\dagger |k-1\rangle\langle k-1|\, a_k x_k\, dx_k\Big\} \\ &= |k-1\rangle\langle k-1| + |k\rangle\langle k|, \end{aligned} \tag{5.10}$$

where use is made of Eq. (3.1) and $a_k^\dagger x_k|k-1\rangle\langle k-1| = a_k^\dagger|k-1\rangle\langle k-1|\, x_k$, which follows from Eqs. (5.1) and (5.2). By mathematical induction Eq. (5.10) leads us eventually to

$$\int |x\rangle\langle x|\, d^n x = \sum_{\substack{\text{all possible}\\ \text{states}}} |k\rangle\langle k| = 1. \tag{5.11}$$

Q.E.D.

Next let us construct eigenstates of the operators $a_i^\dagger$. To distinguish them from the eigenstates of a_i's we shall use, as eigenvalues, gG vectors with asterisk such as $x^* = (x_1^*, x_2^*, \cdots, x_n^*)$. For the state vectors defined by

$$|x^*\rangle \equiv \tilde{\delta}(a^\dagger - x^*)|0\rangle, \qquad \langle x^*| \equiv \langle 0|\exp(x^*\cdot a), \tag{5.12}$$

where

$$\tilde{\delta}(a^\dagger - x^*) \equiv \delta(a_n^\dagger - x_n{}^*)\delta(a_{n-1}^\dagger - x_{n-1}{}^*)\cdots\delta(a_1^\dagger - x_1{}^*), \tag{5.13}$$

we can prove in a similar manner that

$$a_i^\dagger|x^*\rangle = x_i^*|x^*\rangle, \qquad \langle x^*|a_i^\dagger = \langle x^*|x_i^* \,. \tag{5.14}$$

Further, what corresponds to Theorem 7 is now stated as follows:

Theorem 7′: The coherent states for the operators $a_i^\dagger$ satisfy the orthocompleteness relation:

$$\langle x^*|x^{*\prime}\rangle = \tilde{\delta}(x^* - x^{*\prime}), \qquad \int d^n x^*|x^*\rangle\langle x^*| = 1, \tag{5.8′}$$

where $d^n x^* \equiv dx_1{}^* dx_2{}^* \cdots dx_n{}^*$.

The operation of attaching an * or * operation which appears above may be regarded as a generalization of Hermitian conjugation. Thus, we may assume rules such that $(\alpha x_i + \beta x_j)^* = x_i{}^*\alpha^* + x_j{}^*\beta^*, (x_i a_j)^* = a_j{}^\dagger x_i{}^*$, $(dx_i\, dx_j)^* = dx_j{}^* dx_i{}^*$, $(|0\rangle)^* = \langle 0|, (\ )^{**} = (\ )$, and in particular

$$(\delta(x-x'))^* = \tilde{\delta}(x'^* - x^*), \qquad \Big(\int x_j\, dx_j\Big)^* = \int dx_j{}^* x_j{}^* = -\int x_j{}^*\, dx_j{}^* \,. \tag{5.15}$$

Incidentally, the reason why "i" was introduced in Eq. $(3.1)_1$ is now clear: The integral formulas have been so defined as to remain invariant under * operation. It is also easy to check that in each of the following pairs the two members connected by $\sim$ are the adjoint of each other: $(5.4)_1 \sim (5.12)_2$;

$(5.4)_2 \sim (5.12)_1$; $(5.7)_1 \sim (5.14)_2$; $(5.7)_2 \sim (5.14)_1$; $(5.8)_1 \sim (5.8')_1$; $(5.8)_2 \sim (5.8')_2$.

In view of this let us agree in general that given a gG vector $x = (x_1, x_2, \cdots, x_n)$ another gG vector $x^* = (x_1^*, x_2^*, \cdots, x_n^*)$ can always be so chosen as to be in a one-to-one correspondence[11] with x. Our formalism as a whole will then be invariant under * operation. Whether such x and x^* are connected by a transformation such as Eq. (4.1), they will be regarded hereafter as independent quantities. This is in accord with the fact that the corresponding operators a_i and $a_i^\dagger$, which are related through Hermitian conjugation, are independent of each other.

When the products are formed between Eqs. $(5.8)_2$ and $(5.8')_2$, the completeness relation is expressed in different forms:

$$\int\int |x^*\rangle\langle x| \exp(x^*\cdot x)\, d^n x^* d^n x = 1,$$
$$\int\int |x\rangle\langle x^*| \exp(x\cdot x^*)\, d^n x d^n x^* = 1, \tag{5.16}$$

where $(x^*\cdot x) = \sum_{i=1}^n x_i^* x_i$, and use is made of the relations

$$\langle x^*|x\rangle = \exp(x^*\cdot x),$$
$$\langle x|x^*\rangle = \exp(x\cdot x^*). \tag{5.17}$$

On the basis of the above results the theory of Fermi operators given in $\mathscr{H}$ can be transcribed in the language of $\mathscr{H}_G$ or of the gG representation. By using the coherent states we define, for a given state vector $|\rangle$ in $\mathscr{H}$, the *gG representative* or *gG wavefunction* by

$$\psi(x^*) \equiv \langle x^*|\rangle, \tag{5.18}$$

and its adjoint by

$$\bar{\psi}(x) \equiv \langle|x\rangle. \tag{5.19}$$

In particular, the vacuum state $|0\rangle$ or $\langle 0|$ is represented by

$$\psi_0(x^*) = \bar{\psi}_0(x) = 1. \tag{5.20}$$

From the completeness relation it is then clear that the description by gG wavefunctions is equivalent with the one by state vectors $|\rangle$: For example, the use of Eqs. $(5.16)_2$ and (5.18) gives us immediately

$$|\rangle = \int\int |x\rangle\psi(x^*) \exp(x\cdot x^*) d^n x d^n x^*. \tag{5.21}$$

Let $\psi_1(x^*)$ and $\bar{\psi}_2(x)$ be the gG wavefunctions for $|1\rangle$ and $|2\rangle$, respectively. Then, by sandwiching both sides of Eq. $(5.16)_2$ with $\langle 2|$ and $|1\rangle$ we obtain

$$\langle 2|1\rangle = \int\int \bar{\psi}_2(x)\psi_1(x^*) \exp(x\cdot x^*) d^n x d^n x^*. \tag{5.22}$$

This also shows that the normalization condition for $\psi(x^*)$ is given by

$$\int\int \bar{\psi}(x)\psi(x^*) \exp(x\cdot x^*) d^n x d^n x^* = 1. \tag{5.23}$$

As for observables $F(a^\dagger, a)$ in general, which we assume for simplicity to be normal-ordered, we can show similarly that there exists the following one-to-one correspondence with $F(x^*, x)$:

$$\langle x^*|F(a^\dagger, a)|x\rangle = F(x^*, x) \exp(x^*\cdot x),$$
$$F(a^\dagger, a) = \int\cdots\int |x'\rangle F(x'^*, x)\langle x^*| \times \exp[(x'\cdot x'^*) + (x'^*\cdot x) + (x\cdot x^*)] \times d^n x' d^n x'^* d^n x d^n x^*. \tag{5.24}$$

Further, the matrix element $\langle 2|F(a^\dagger, a)|1\rangle$ is expressed in terms only of gG quantities as

$$\langle 2|F(a^\dagger, a)|1\rangle = \int\cdots\int \bar{\psi}_2(x') F(x'^*, x)\psi_1(x^*) \times \exp[(x'\cdot x'^*) + (x'^*\cdot x) + (x\cdot x^*)] \times d^n x' d^n x'^* d^n x d^n x^*. \tag{5.25}$$

The Schrödinger equation with the normal-ordered Hamiltonian $H(a^\dagger, a)$ then takes the form:

$$i\frac{\partial\psi(x^*)}{\partial t} = \int\int H(x^*, x')\psi(x'^*) \exp[(x^*\cdot x') + (x'\cdot x'^*)] \times d^n x' d^n x'^*. \tag{5.26}$$

We have thus found that the system under consideration can be completely described in terms of gG quantities, i.e., in the gG representation. As in our previous paper[7] the path-integral method can be easily formulated in this representation. In the paper that immediately follows[12] it will be shown that when suitably modified, most of the above results also apply to para-Fermi systems.

[1] See, for example, F.A. Berezin, *The Method of Second Quantization* (Academic, New York, 1966); Theor. Math. Phys. **6**, 194 (1971); F.A. Berezin and M.S. Marinov, Ann. Phys. (N.Y.) **104**, 336 (1977).

[2] See, for example, H. Umezawa, J. Podolanski and S. Oneda, Proc. Phys. Soc. London Sect. A **68**, 503 (1955).

[3] For the special case of G algebras our definitions are slightly different from those employed by other authors. For example, our determinants for this case correspond to the inverses of those used by R. Arnowitt, P. Nath, and B. Zumino, Phys. Lett. B **56**, 81 (1975). Compare Theorem 2 below.

[4] Throughout the present paper the ith equation in Eq. (2.5), for example, will be quoted as Eq. $(2.5)_i$.

[5] For this reason it does not seem appropriate to define det M in general by exp(Tr log$M$) as, for example, in Ref. 3. On the contrary, our definition (2.2) always provides us, for any given M, with det$M$ as a homogeneous polynomial in M_{ij}'s.

[6] O. Klein, J. Phys. (U.S.S.R.) **9**, 1 (1938); G. Lüders, Z. Naturforsch. **139**, 254 (1958).

[7] Y. Ohnuki and T. Kashiwa, Prog. Theor. Phys. **60**, 548 (1978).

[8] For the case of multiple integrals such as $\int\int\cdots d^n x d^n x'$, Eq. (3.5) must be generalized to $dx_i dx_j' = \eta_{ij} dx_j' dx_i$ for all i, j.

[9] Most of the formulas given in this section are straightforward generalizations of those given previously for the case of ordinary Fermi systems.[7]

[10] Here, by orthogonality we do not mean that the inner product vanishes for $x \neq x'$, but only that it becomes a δ function.

[11] When two complex numbers c and c' are given, it is possible to decide, by comparing their real and imaginary parts, whether they are complex conjugates of each other. For the case of gG vectors, on the contrary, we have no corresponding way of deciding whether given x and x' are the adjoint of each other. In this sense, * operation is a very formal one. Indeed, whether x is the adjoint of x' or not is merely a matter of definition.

[12] Y. Ohnuki and S. Kamefuchi, J. Math. Phys. **21**, 609 (1980).

Para-Grassmann algebras with applications to para-Fermi systems

Y. Ohnuki
Department of Physics, Nagoya University, Nagoya 464, Japan

S. Kamefuchi
Institute of Physics, University of Tsukuba, Ibaraki 305, Japan

(Received 10 September 1979; accepted for publication 26 November 1979)

Grassmann algebras of the usual kind are generalized to what are to be called para-Grassmann algebras. Para-Grassmann numbers are defined as those satisfying the trilinear commutation relations that resemble the para-Fermi commutation relations. Basic mathematical properties of these algebras are studied in detail on the basis of generalized Grassmann algebras discussed in the preceding paper. Applications are made to description of para-Fermi systems, and it is found that such systems can completely be described in what we call the para-Grassmann representation.

I. INTRODUCTION

Just as Fermi operators are generalized to para-Fermi operators,[1] a Grassmann algebra of the usual kind[2] is generalized to what is to be called a *para-Grassmann algebra*.[3] The purpose of the present paper is to discuss some of the basic properties of such algebras and to apply the results to description of para-Fermi systems. In what follws the adjectives "para-Fermi" and "para-Grassmann" will be abbreviated to "pF" and "pG," respectively.

A set of independent numbers $\xi_1, \xi_2, \cdots, \xi_n$ are called pG *numbers* to form a pG *algebra* of order $p(=1,2,\ldots)$, when they satisfy the following relations:

$$[\xi_i[\xi_j,\xi_k]] = 0,$$
$$[\xi_{i_1},\xi_{i_2},\cdots,\xi_{i_m}]_+ = 0, \quad \text{only for } m \geqslant p+1, \tag{1.1}$$

where $[A, B, \cdots, Z]_+$ denotes the completely symmetrized product of $AB\cdots Z$, and i's, j, $k = 1,2,\cdots,n$. For the cases $p = 1$ and $p = 2$, Eqs. (1.1) are simplified, respectively, to

$$[\xi_i,\xi_j]_+ = 0, \quad \text{for } p = 1,$$
$$\langle\xi_i,\xi_j,\xi_k\rangle_+ = 0, \quad \text{for } p = 2, \tag{1.2}$$

where $\langle A, B, C\rangle_+ \equiv ABC + CBA$. Thus, the case $p = 1$ corresponds to a Grassmann algebra of the usual kind.

It is often convenient to use what we may call the *Green representation*:[1]

$$\xi_i = \sum_{\alpha=1}^{p} \xi_i^{(\alpha)}, \tag{1.3}$$

with

$$[\xi_i^{(\alpha)},\xi_j^{(\alpha)}]_+ = 0,$$
$$[\xi_i^{(\alpha)},\xi_j^{(\beta)}]_- = 0 \quad (\alpha \neq \beta). \tag{1.4}$$

Let us call $\xi_i^{(\alpha)}$ the Green component of ξ_i with Green index α. For the present algebra there may exist some other types of representations. Throughout the present paper, however, we shall exclusively be concerned, for a physical reason, with the one represented by Eqs. (1.3) and (1.4).

In the preceding paper[4] we have made a detailed study on *generalized Grassmann* ("gG") numbers x_i. They also correspond to a kind of generalization of ordinary Grassmann ("G") numbers. As shown there, we may regard G and gG numbers as the classical counterparts of Fermi operators satisfying the commutation relations, respectively, of the normal case and of a general, normal or anomalous, case.[5] Since the algebra of $\xi_i^{(\alpha)}$'s is a special case of gG algebras x_i, the following discussion will heavily lean on the results of (I).

More precisely speaking, the correspondence $\xi_i^{(\alpha)} \longleftrightarrow x_i$ between $\xi_i^{(\alpha)}$'s and x_i's of (I) is given as follows: A pair of suffixes (i,α) of $\xi_i^{(\alpha)}$ is taken to correspond to the subscript i of x_i, and the relative signatures η introduced through $\xi_i^{(\alpha)}\xi_j^{(\beta)} \equiv \eta_{(i,\alpha)(j,\beta)}\xi_j^{(\beta)}\xi_i^{(\alpha)}$ are such that

$$\eta_{(i,\alpha)(j,\beta)} = \eta_{\alpha\beta} = \begin{cases} -, & \text{for } \alpha = \beta, \\ +, & \text{for } \alpha \neq \beta. \end{cases} \tag{1.5}$$

Alternatively, it is also possible to identify the suffixes α and i of $\xi_i^{(\alpha)}$, respectively, with the subscript i and the label by which to distinguish n gG vectors $x, x', \cdots x^{(n)}$. In the following we shall make use of both kinds of correspondence.

For the pG numbers or variables ξ_j $(j = 1,2,\cdots,n)$ it is possible to define "analytical" operations such as differentiation and integration. In addition to these numbers we may consider a finite number of "constants" μ_s $(s = 1,2,\ldots)$ which also are pG numbers of order p but not subject to the above-mentioned operations for ξ_i's. The Green components $\mu_s^{(\alpha)}$ for which $\mu_s = \sum_{\alpha=1}^{p}\mu_s^{(\alpha)}$ satisfy the following relations:

$$[\mu_s^{(\alpha)},\mu_{s'}^{(\beta)}]_{-\eta_{\alpha\beta}} = [\mu_s^{(\alpha)},\xi_j^{(\beta)}]_{-\eta_{\alpha\beta}} = 0. \tag{1.6}$$

In connection with the Green representation let us also remark the following: When working with a pG algebra we should be interested only in those functions in ξ_i's and μ_s's which do not explicitly depend on Green indices: For one thing the Green components are merely mathematical auxiliaries, and for another only those quantities of the above-mentioned kind are defined, in our physical applications, to be of physical meaning.

The present paper is arranged as follows: In Sec. II analytical operations such as differentiation and integration are introduced. Some basic formulas for these operations are summarized as two lemmas. Further mathematical properties are discussed in Sec. III. A general theorem is proved concerning a change of variables in integrals, and the delta functions and Fourier transformation are defined. Section IV deals with applications to pF systems. Coherent states for

 0022-2488/80/040609-08$1.00

pF operators a_i and $a_i^\dagger$ $(i = 1,2,\cdots,n)$ are explicitly constructed, and what we call the pG *representation* is defined on the basis of such states. It is shown that pF systems can be described completely in this representation.

II. ANALYTICAL OPERATIONS

A. Differentiation

As in (I) we shall be concerned mostly with left differentiation. We define left differentiation with respect to ξ_i by the sum of left differentiations $\partial/\partial\xi_i^{(\alpha)}$ with respect to the Green components $\xi_i^{(\alpha)}$:

$$\frac{\partial}{\partial\xi_i} \equiv \sum_{\alpha=1}^{p} \frac{\partial}{\partial\xi_i^{(\alpha)}} . \tag{2.1}$$

Here, as for the operation $\partial/\partial\xi_i^{(\alpha)}$, we adopt the same definition as the one given in (I) (cf. Sec. III thereof). The definition also requires that

$$\frac{\partial}{\partial\xi_i^{(\alpha)}} \frac{\partial}{\partial\xi_j^{(\beta)}} = \eta_{\alpha\beta} \frac{\partial}{\partial\xi_j^{(\beta)}} \frac{\partial}{\partial\xi_i^{(\alpha)}} . \tag{2.2}$$

Right differentiation $\bar{\partial}/\partial\xi_i$ is similarly defined in terms of the Green components $\bar{\partial}/\partial\xi_i^{(\alpha)}$.

Let us consider a function $f(\xi_1, \xi_2, \cdots \xi_n)$ of $\xi_1, \xi_2, \cdots, \xi_n$, and differentiate it with respect to a specific variable ξ_i. The function f may also contain other pG numbers μ_s. For the sake of convenience we hereafter denote the ξ_i under consideration by ξ and all other pG numbers by ζ_r, ρ_s,... Thus, we write $f(\xi,\zeta_r,\zeta_{r'},\cdots,\rho_s, \rho_{s'},\ldots) \equiv f(\xi)$. At this point we notice that the so-called decomposition theorem,[6] proved originally for pF operators, is applicable as well to the present case owing to the similarity between Eqs. (1.1) and the pF commutation relations[1,6,7]: Actually, the situation is much simpler here because the right-hand side of Eqs. (1.1) vanishes, so that no contracted terms are needed. Thus, adapting the theorem to the present case, we obtain the following:

Theorem 1: An arbitrary function of ξ and other pG number can be expressed as a linear combination of terms of the standard form, i.e.,

$$[\xi,\zeta_1][\xi,\zeta_2]\cdots[\xi,\zeta_q][\xi,\xi,\cdots,\xi,\rho_1,\rho_2,\cdots,\rho_s]_{+,r+s} \times[\ ,\ ][\ ,\ ]\cdots[\ ,\ ] \quad (q+r\leqslant p,\ r+s\leqslant p), \tag{2.3}$$

where the subscript attached to the bracket $[\ ,\cdots]_+$ denotes the number of pG numbers contained therein, and $[\ ,\ ]$'s represent commutators of pG numbers other than ξ.

It is to be emphasized here that the theorem is a consequence only of Eqs. (1.1). The factors $[\ ,\ ][\ ,\ ]\cdots[\ ,\ ]$ are quite independetn of ξ and commute with ξ and $\partial/\partial\xi$. They may therefore be omitted often in the following discussion.

We now show that the result of differentiating an arbitrary function $f(\xi)$, i.e., $\partial/\partial\xi \cdot f(\xi)$, can be expressed again as a function only of ξ, ζ's, and ρ's, namely, in a form which does not explicitly depend on Green indices. To this end we note, first of all, that

$$\frac{\partial}{\partial\xi}\{[\xi,\zeta]g(\xi)\} = \frac{\partial}{\partial\xi}[\xi,\zeta]\cdot g(\xi) + [\xi,\zeta]\frac{\partial g(\xi)}{\partial\xi}, \tag{2.4}$$

$$\frac{\partial}{\partial\xi}[\xi,\zeta] = 2\zeta,$$

which follow directly from the definition. Thanks to Eqs. (2.4) and (2.3) it suffices to show that $\partial/\partial\xi\cdot[\xi,\xi,\cdots,\xi,\rho_1,\rho_2,\cdots,\rho_s]_{+,r+s}$ is a function only of pG numbers. In so doing we make use of the following relations:

$$\begin{aligned}\frac{\partial}{\partial\xi^{(\alpha)}}\{(\xi+\rho)g(\xi)\} &= \sum_\beta \frac{\partial}{\partial\xi^{(\alpha)}}\{(\xi^{(\beta)}+\rho^{(\beta)})g(\xi)\} \\ &= \sum_\beta \left\{\delta_{\alpha\beta}g(\xi) + \eta_{\alpha\beta}(\xi^{(\beta)}+\rho^{(\beta)})\frac{\partial g(\xi)}{\partial\xi^{(\alpha)}}\right\} \\ &= g(\xi) + \sum_\beta (1-2\delta_{\alpha\beta})(\xi^{(\beta)}+\rho^{(\beta)})\frac{\partial g(\xi)}{\partial\xi^{(\alpha)}} \\ &= g(\xi) + (\xi+\rho)\frac{\partial g(\xi)}{\partial\xi^{(\alpha)}} - 2(\xi^{(\alpha)}+\rho^{(\alpha)})\frac{\partial g(\xi)}{\partial\xi^{(\alpha)}} .\end{aligned} \tag{2.5}$$

Firstly, summing Eq. (2.5) over α, and secondly, multiplying Eq. (2.5) from the left by $(\xi^{(\alpha)}+\rho^{(\alpha)})$ and then summing the resulting equation over α, we obtain

$$\frac{\partial}{\partial\xi}\{(\xi+\rho)g(\xi)\} = pg(\xi) + (\xi+\rho)\frac{\partial g(\xi)}{\partial\xi} - 2Dg(\xi) \tag{2.6}$$

and

$$D\{(\xi+\rho)g(\xi)\} = (\xi+\rho)g(\xi) + (\xi+\rho)Dg(\xi), \tag{2.7}$$

respectively, where $D \equiv \Sigma_\alpha(\xi^{(\alpha)}+\rho^{(\alpha)})\cdot\partial/\partial\xi^{(\alpha)}$. When $g(\xi) = (\xi+\rho)^{r-1}$ is substituted in Eq. (2.7), it is found that $Dg(\xi) = (r-1)(\xi+\rho)^{r-1}$. Substitution of the last expression in Eq. (2.6) then provides

$$\frac{\partial}{\partial\xi}(\xi+\rho)^r = (p+2-2r)(\xi+\rho)^{r-1} + (\xi+\rho)\frac{\partial(\xi+\rho)^{r-1}}{\partial\xi}, \tag{2.8}$$

whence

$$\frac{\partial}{\partial\xi}(\xi+\rho)^r = r(p+1-r)(\xi+\rho)^{r-1} . \tag{2.9}$$

Making a change $r \to r+s$ and noticing the relation $(\xi+\rho)^k = [(\xi+\rho),(\xi+\rho),\cdots,(\xi+\rho)]_{+,k}/k!$, we rewrite Eq. (2.9) as

$$\frac{1}{(r+s)!}\frac{\partial}{\partial\xi}[(\xi+\rho),(\xi+\rho),\cdots,(\xi+\rho)]_{+,r+s} = \frac{(r+s)(p+1-r-s)}{(r+s-1)!}[(\xi+\rho),(\xi+\rho),\cdots,(\xi+\rho)]_{+,r+s-1} . \tag{2.10}$$

Picking out of both sides the terms containing s ρ's, we obtain

$$\frac{1}{(r+s)!}\binom{r+s}{s}\frac{\partial}{\partial\xi}\left[\xi,\xi,\cdots,\xi,\ \underset{r}{\rho},\underset{s}{\rho},\cdots,\rho\right]_{+,r+s} = \frac{(r+s)(p+1-r-s)}{(r+s-1)!}\binom{r+s-1}{s}\left[\xi,\xi,\cdots,\xi,\ \underset{r-1}{\rho},\underset{s}{\rho},\cdots,\rho\right]_{+,r+s-1} , \tag{2.11}$$

i.e.,

$$\frac{\partial}{\partial\xi}\left[\xi,\xi,\cdots,\xi,\ \underset{r}{}\ \rho,\rho,\cdots,\rho\ \underset{s}{}\right]_{+,r+s}=r(r+s)(p+1-r-s)\left[\xi,\xi,\cdots,\xi,\ \underset{r-1}{}\ \rho,\rho,\cdots,\rho\ \underset{s}{}\right]_{+,r+s-1}. \tag{2.11'}$$

Putting $\rho=\rho_1+\rho_2+\cdots+\rho_s$ on both sides of Eq. (2.11') and picking out the terms containing each ρ_i $(i=1,2,\cdots,s)$ linearly, we obtain the differential formula for $[\xi,\cdots,\xi,\rho_1,\cdots,\rho_s]_{+,r+s}$. This and Eq. (2.4) may be summarized as the following:

Lemma 1: The differential formulas for terms of the standard form (2.3) are given as

$$\frac{\partial}{\partial\xi}\{[\xi,\zeta]g(\xi)\}=2\zeta g(\xi)+[\xi,\zeta]\frac{\partial g(\xi)}{\partial\xi},$$

$$\frac{\partial}{\partial\xi}[\xi,\xi,\cdots,\xi,\rho_1,\rho_2,\cdots,\rho_s]_{+,r+s}=r(r+s)(p+1-r-s)[\xi,\xi,\cdots,\xi,\rho_1,\rho_2,\cdots,\rho_s]_{+,r+s-1}. \tag{2.12}$$

We have thus proved that the left derivative $\partial/\partial\xi\cdot f(\xi)$ of an arbitrary function $f(\xi)$ is again a function of pG numbers. It is obvious that the same is also true of right differentiation.

B. Integration

Evidently, integration of $f(\xi_1,\xi_2,\cdots,\xi_n)$ with respect to the variables of $\xi_1,\xi_2,\cdots,\xi_n$ is to be defined as a successive integration of f with respect to the individual variables ξ_i performed in a fixed order. Therefore, we have only to define the integration of f with respect to a single variable. Again let us denote such a specific variable by ξ, other pG variables by ζ's or ρ's, and a general function of these variables by $f(\xi)$. In accordance with Eq. (I.3.2) we now adopt the following formulas expressed in terms of the Green components:

$$\int\xi^{(1)}\xi^{(2)}\cdots\xi^{(p)}\,d^p\xi=(i)^p,$$

$$\int\xi^{(\alpha_1)}\xi^{(\alpha_2)}\cdots\xi^{(\alpha_r)}\,d^p\xi=0\quad(0\leqslant r<p), \tag{2.13}$$

with

$$d^p\xi\equiv d\xi^{(1)}d\xi^{(2)}\cdots d\xi^{(p)}, \tag{2.14}$$

where $d\xi^{(\alpha)}\,d\xi^{(\beta)}=d\xi^{(\beta)}d\xi^{(\alpha)}$ $(\alpha\neq\beta)$ is assumed. Further, if $d\xi$ is defined as $d\xi\equiv\Sigma_{\alpha=1}^{p}d\xi^{(\alpha)}$, then $d^p\xi=(d\xi)^p/p!$. Similarly, we can rewrite Eqs. (2.13) in the form

$$\int\xi^p\,d^p\xi=p!\,(i)^p,$$

$$\int\xi^r\,d^p\xi=0\quad(0\leqslant r<p). \tag{2.13'}$$

Noticing that right differentiation of pth order with respect to ξ can be written as

$$\frac{\bar\partial^p}{\partial\xi^p}=p!\,\frac{\bar\partial^p}{\partial\xi^{(1)}\partial\xi^{(2)}\cdots\partial\xi^{(p)}}, \tag{2.15}$$

we can easily realize that

$$\int f(\xi)\,d^p\xi=\frac{i^p}{p!}\frac{\bar\partial^p}{\partial\xi^p}f(\xi). \tag{2.16}$$

Since, as already noticed above, the right-hand side is expressible in a form which does not explicitly depend on Green indices, we can thereby conclude that the result of integrating an arbitrary function $f(\xi)$, i.e., $\int f(\xi)\,d^p\xi$ is again a function only of ξ and other pG numbers. Needless to say, $f(\xi)$ here may also contain pG numbers μ_s's other than ξ_i's.

More explicitly we can prove the following:

Lemma 2: The integral formula for terms of the standard form (2.3) is given as

$$\int[\xi,\xi_1][\xi,\xi_2]\cdots[\xi,\xi_{p-r}][\xi,\xi,\cdots,\xi,\rho_1,\rho_2,\cdots\rho_s]_{+,r+s}\,d^p\xi$$

$$=\frac{(-1)^{p-r}2^{p-r-s}i^p r!(r+s)!}{[(p-r-s)!]^2}\sum_{\tau\in S_{p-r}}[\rho_1,\zeta_{\tau 1}][\rho_2,\zeta_{\tau 2}]\cdots[\rho_s,\zeta_{\tau s}]$$

$$\times[\zeta_{\tau(s+1)},\zeta_{\tau(s+2)},\cdots,\zeta_{\tau(p-r)}]_{+,p-r-s}\qquad(r+s\leqslant p). \tag{2.17}$$

The lemma is proved as follows: Taking an arbitrary complex number c, we consider

$$([\xi,\zeta]+c\xi)^p=\left\{\sum_{\alpha=1}^{p}\xi^{(\alpha)}(2\zeta^{(\alpha)}+c)\right\}^p=p!\xi^{(1)}\xi^{(2)}\cdots\xi^{(p)}(2\zeta^{(1)}+c)(2\zeta^{(2)}+c)\cdots(2\zeta^{(p)}+c). \tag{2.18}$$

Picking out of both sides the terms containing c^r, we obtain

$$\frac{p!}{(p-r)!r!}[\xi,\zeta]^{p-r}\xi^r=(-2)^{p-r}\frac{p!}{(p-r)!}\zeta^{p-r}\xi^{(1)}\xi^{(2)}\cdots\xi^{(p)}, \tag{2.18'}$$

which yields, on account of Eq. $(2.13)_1$,

$$\int[\xi,\zeta]^{p-r}\xi^r\,d^p\xi=(-2)^{p-r}i^p\frac{r!}{(p-r)!}[\zeta,\zeta,\cdots,\zeta]_{+,p-r}. \tag{2.19}$$

Putting $\zeta=\zeta_1+\zeta_2+\cdots+\zeta_{p-r}$ in Eq. (2.19) and picking out the terms containing each ζ_i $(i=1,2,\cdots,p-r)$ linearly, we obtain

$$\int[\xi,\xi_1][\xi,\xi_2]\cdots[\xi,\xi_{p-r}]\xi^r\,d^p\xi=(-2)^{p-r}i^p\frac{r!}{(p-r)!}[\zeta_1,\zeta_2,\cdots,\zeta_{p-r}]_+. \tag{2.19'}$$

Next let us consider the integral

$$\int [\xi,\zeta_1][\xi,\zeta_2]\cdots[\xi,\zeta_{p-r}](\xi+\rho)^{r+s}d^p\xi$$

$$= \frac{1}{(r+s)!}\int [\xi,\zeta_1][\xi,\zeta_2]\cdots[\xi,\zeta_{p-r}][\xi+\rho,\xi+\rho,\cdots\xi+\rho]_{+,r+s}d^p\xi$$

$$= \frac{1}{r!\,s!}\int [\xi,\zeta_1][\xi,\zeta_2]\cdots[\xi,\zeta_{p-r}]\Big[\underset{r}{\xi},\underset{s}{\xi},\cdots\xi,\rho,\rho,\cdots,\rho\Big]_{+,r+s}d^p\xi\,, \tag{2.20}$$

where, in the last step, use is made of Eq. (2.13). We then perform a change of variable $\xi \to \xi - \rho$ here. In this case Theorem 6 of (I) assures us that Eq. (2.20) remains true even when ξ is replaced by $\xi - \rho$ in the first line thereof. Thus, we have the third line of Eq. (2.20)

$$= \int ([\xi,\zeta_1]-[\rho,\zeta_1])([\xi,\zeta_2]-[\rho,\zeta_2])\cdots([\xi,\zeta_{p-r}]-[\rho,\zeta_{p-r}])\xi^{r+s}d^p\xi$$

$$= \frac{(-1)^s}{(p-r-s)!s!}\sum_{\tau\in S_p}{}_{,}\,[\rho,\zeta_{\tau 1}][\rho,\zeta_{\tau 2}]\cdots[\rho,\zeta_{\tau s}]\int[\xi,\zeta_{\tau(s+1)}][\xi,\zeta_{\tau(s+2)}]\cdots[\xi,\zeta_{\tau(p-r)}]\xi^{r+s}d^p\xi\,. \tag{2.20'}$$

The integral in the last line of Eq. (2.20') can be computed by use of Eq. (2.19') with r being replaced by $r+s$. Then, Eqs. (2.20') leads us to

$$\int [\xi,\zeta_1][\xi,\zeta_2]\cdots[\xi,\zeta_{p-r}]\Big[\underset{r}{\xi},\underset{s}{\xi},\cdots,\xi,\rho,\rho,\cdots,\rho\Big]_{+,r+s}d^p\xi$$

$$= \frac{(-1)^{p-r}2^{p-r-s}i^p r!(r+s)!}{[(p-r-s)!]^2}\sum_{\tau\in S_p}{}_{,}\,[\rho,\zeta_{\tau 1}][\rho,\zeta_{\tau 2}]\cdots[\rho,\zeta_{\tau s}][\zeta_{\tau(s+1)},\zeta_{\tau(s+2)},\cdots,\zeta_{\tau(p-r)}]_{+,p-r-s}\,. \tag{2.21}$$

Putting here $\rho = \rho_1 + \rho_2 + \cdots + \rho_s$ and picking out the terms containing each ρ_i $(i = 1,2,\cdots,s)$ linearly, we arrive at Eq. (2.17).

It is to be remarked incidentally that when the integrand consists only of brackets of the type $[\xi,\zeta_i]$, the necessary integral formula is obtained, as a special case, from Eq. (2.17) with $r = s = 0$, which agrees with Eq. (2.19') with $r = 0$. The repeated use of Eq. (2.17) thus enables us to carry out any multiple integrals with respect to $\xi_1, \xi_2, \cdots, \xi_n$. In what follows we shall denote the measure of such integrals by

$$(d\xi)_n \equiv d^p\xi_n d^p\xi_{n-1}\cdots d^p\xi_1\,. \tag{2.22}$$

III. FURTHER MATHEMATICAL PROPERTIES

A. Change of variables in integrals

Let us now consider transformations or changes of variables in pG integrals such as

$$\xi_i \to \xi_i' = \xi_i'(\xi,\mu)\,, \tag{3.1}$$

where ξ and μ stand for ξ_j $(j = 1,2,\cdots,n)$ and μ_s $(s = 1,2,...)$, respectively. Here we limit ourselves to transformations such that the new variables ξ_i' become again pG numbers of the same order p. By virtue of Theorem 1 we then find that the right-hand side of Eq. (3.1) should consist of terms of the standard form such as $[\rho_1,\rho_2][\rho_3,\rho_4]\cdots[\rho_{r-1},\rho_r]\zeta$, where ρ's and ζ stand for ξ_j's or μ_s's. Thus, Eq. (3.1) can be written explicitly as

$$\xi_i' = \mu_i + a_{ij}\xi_j + \tfrac{1}{2}b_{ij,s's''}[\xi_j,\mu_{s'}']\mu_{s''}'' + Q_i(\xi)\,. \tag{3.1'}$$

Here a_{ij}'s and $b_{ij,s's''}$'s are polynomials in $[\mu,\mu']$'s and hence commute with any other quantities, $Q_i(\xi)$'s are polynomials in ξ_j's and μ_s's, without constant and linear terms in ξ_j's, and the summation convention over repeated indices is implied. In terms of the Green components Eq. (3.1') reads

$$\xi_i'^{(\alpha)} = \mu_i^{(\alpha)} + \xi_j^{(\beta)}(a_{ij}\delta_{\alpha\beta} + b_{ij,s's''}\mu_{s'}'^{(\beta)}\mu_{s''}''^{(\alpha)}) + Q_i^{(\alpha)}(\xi)\,. \tag{3.1''}$$

Under the change of variables (3.1) the measure $(d\xi)_n$ undergoes the change dictated by Theorem 6 of (I). In the present case the matrix J is such that its $(j\beta)(i\alpha)$ element is given by $\partial/\partial\xi_j^{(\beta)}\cdot\xi_i'^{(\alpha)}$: $J \equiv \|\partial/\partial\xi_j^{(\beta)}\cdot\xi_i'^{(\alpha)}\|$. Considering the fact that as far as its dependence on a specific variable, say $\xi_j^{(\beta)}$, is concerned, $Q_i^{(\alpha)}(\xi)$ has a form similar to that of the second term in Eq. (3.1''), we obtain the matrix elements as follows

$$\frac{\partial\xi_i'^{(\alpha)}}{\partial\xi_j^{(\beta)}} = a_{ij}\delta_{\alpha\beta} + b_{ij,s's''}\mu_{s'}'^{(\beta)}\mu_{s''}''^{(\alpha)} + \frac{\partial}{\partial\xi_j^{(\beta)}}Q_i^{(\alpha)}(\xi)$$
$$\equiv A_{ij}\delta_{\alpha\beta} + B_{ij,s's''}\xi_{s'}'^{(\beta)}\xi_{s''}''^{(\alpha)}\,, \tag{3.2}$$

where A_{ij}'s and $B_{ij,s's''}$'s are polynomials in $[\zeta,\zeta']$'s.

Let $\kappa_j = \Sigma_{\alpha=1}^p \kappa_j^{(\alpha)}$ $(j = 1,2,\cdots,n)$ be a set of pG numbers of order p. Then the determinant is found from Eq. (I.2.2) as

$$\prod_{\beta=1}^{p}\left(\frac{\partial\xi_i'^{(\alpha)}}{\partial\xi_1^{(\beta)}}\kappa_i^{(\alpha)}\right)\prod_{\beta=1}^{p}\left(\frac{\partial\xi_i'^{(\alpha)}}{\xi_2^{(\beta)}}\kappa_i^{(\alpha)}\right)\cdots\prod_{\beta=1}^{p}\left(\frac{\partial\xi_i'^{(\alpha)}}{\partial\xi_n^{(\beta)}}\kappa_i^{(\alpha)}\right)$$
$$= \det\left(\frac{\partial\xi_i'^{(\alpha)}}{\partial\xi_j^{(\beta)}}\right)\cdot\left(\prod_{\alpha=1}^{p}\kappa_1^{(\alpha)}\right)\left(\prod_{\alpha=1}^{p}\kappa_2^{(\alpha)}\right)\cdots\left(\prod_{\alpha=1}^{p}\kappa_n^{(\alpha)}\right). \tag{3.3}$$

Here each factor on the left-hand side can be computed by use of Eq. (3.2) as follows:

$$\prod_{\beta=1}^{p}\left(\frac{\partial\xi_i'^{(\alpha)}}{\partial\xi_j^{(\beta)}}\kappa_i^{(\alpha)}\right) = \prod_{\beta=1}^{p}(A_{ij}\kappa_i^{(\beta)} + B_{ij,s's''}\xi_{s'}'^{(\beta)}\xi_{s''}''^{(\alpha)}\kappa_i^{(\alpha)})$$
$$= \prod_{\beta=1}^{p}(A_{ij}\kappa_i^{(\beta)} + \tfrac{1}{2}B_{ij,s's''}\xi_{s'}'^{(\beta)}[\zeta_{s''}'',\kappa_i])$$
$$= \frac{1}{p!}(A_{ij}\kappa_i + \tfrac{1}{2}B_{ij,s's''}\xi_{s'}'[\zeta_{s''}'',\kappa_i])^p\,. \tag{3.4}$$

We have thus found that Eqs. (3.4) and hence the left-hand side of Eq. (3.3) are expressed in terms only of pG numbers. The determinant is then obtained from Eq. (3.3) with the help of Eqs. (2.13) and (2.22):

$$\det\left(\frac{\partial \xi_i'^{(\alpha)}}{\partial \xi_j^{(\beta)}}\right) = \frac{1}{i^{pn}} \int \cdots \int [\text{left hand side of Eqs. (3.4)}]\,(d\kappa)_n\,. \tag{3.5}$$

Thus, according to Lemma 2, this determinant also is expressible in terms only of pG numbers. To indicate this property explicitly let us write the determinant as $\det J \equiv J(\xi',\xi)$. Then, from Eqs. (I.2.3) and (I.4.12) we obtain

$$[J(\xi',\xi),\xi_i] = 0\,, \qquad J(\xi'',\xi) = J(\xi'',\xi')\,J(\xi',\xi)\,. \tag{3.6}$$

Further, as can be seen from the first half of the proof of Theorem 6 of (I) and from Theorem 5 of (I), $J(\xi',\xi)^{-1}$ exists if and only if the inverse transformation exists for Eq. (3.1). On the other hand, we can show by observing Eqs. (3.2) and by arguing in the same way as in the derivation of Eq. (I.4.10) that the existence of $J(\xi',\xi)^{-1}$ is guaranteed by that of $[\det(a_{ij}\delta_{\alpha\beta})]^{-1} = [\det(a_{ij})]^{-p}$, i.e., of $[\det(a_{ij})]^{-1}$.

On the basis of Theorem 6 of (I) we can summarize the above results as follows:

Theorem 2: For a transformation of pG variables (3.1) or (3.1'),

(i) the determinant $J(\xi',\xi) \equiv \det\|\partial/\partial\xi_j^{(\beta)}, \xi_i'^{(\alpha)}\|$ is expressible in terms only of pG numbers, and commutes with any other quantities;

(ii) $J(\xi',\xi)^{-1}$ and the inverse transformation of Eq. (3.1) exist if and only if $[\det(a_{ij})]^{-1}$ exists;

(iii) under the condition of (ii) the measure $(d\xi)_n$ in pG integrals is subject to the change

$$(d\xi')_n = J(\xi',\xi)^{-1}(d\xi)_n\,. \tag{3.7}$$

B. Delta functions and Fourier transformation

The δ functions and Fourier transformation can be introduced in the same way as in (I). By suitably modifying Eqs. (I.3.11) and (I.3.8) we define the δ function for a single variable ξ_j by

$$\delta(\xi_j - \xi'_j) \equiv \frac{1}{i^p}\prod_{\alpha=1}^{p}(\xi_j^{(\alpha)} - \xi_j'^{(\alpha)}) = \frac{1}{i^p p!}(\xi_j - \xi'_j)^p\,. \tag{3.8}$$

It is easy to check that this function has the expected property

$$\int f(\xi_1,\xi_2,\cdots,\xi_j,\cdots,\xi_n)\,\delta(\xi_j - \xi'_j)\,d^p\xi_j = f(\xi_1,\xi_2,\cdots,\xi'_j,\cdots,\xi_n)\,. \tag{3.9}$$

From Eq. (I.3.13) the Fourier integral of this function is obtained as

$$\delta(\xi_j - \xi'_j) = \int \exp\left[-(\overline{\xi_j - \xi'_j}\cdot\xi''_j)\right] d^p\xi''_j\,, \tag{3.10}$$

where $(\xi_j\cdot\xi'_j) \equiv \Sigma_{\alpha=1}^{p}\xi_j^{(\alpha)}\xi_j'^{(\alpha)} = \frac{1}{2}[\xi_j,\xi'_j]$.

The operation of Fourier transformation can be performed on the basis of Theorem 4 of (I). Thus, for a function f of n pG variables we obtain the following formulas:

$$f(\xi_1,\xi_2,\cdots,\xi_n) = \int\cdots\int \tilde{f}(\xi_1^*,\xi_2^*,\cdots,\xi_n^*)\exp[-(\xi\cdot\xi^*)]\,(d\xi^*)_n\,, \tag{3.11}$$

$$\tilde{f}(\xi_1^*,\xi_2^*,\cdots,\xi_n^*) = \int\cdots\int f(\xi_1,\xi_2,\cdots,\xi_n)\exp[-(\xi^*\cdot\xi)]\,(d\xi)_n\,,$$

where ξ_i^*'s are another set of pG variables, $(\xi\cdot\xi^*) \equiv \Sigma_{i=1}^{n}(\xi_i\cdot\xi_i^*)$, and $(d\xi^*)_n \equiv d^p\xi_1^* d^p\xi_2^*\cdots d^p\xi_n^*$. It is to be noticed that for any function f its Fourier transform exists and is uniquely determined. For the δ function of n variables defined by

$$\delta^n(\xi - \xi') \equiv \delta(\xi_1 - \xi'_1)\delta(\xi_2 - \xi'_2)\cdots\delta(\xi_n - \xi'_n)\,, \tag{3.12}$$

the Fourier decomposition is obtained from Eq. (3.10) as

$$\delta^n(\xi - \xi') = \int\cdots\int \exp\left[-(\overline{\xi - \xi'}\cdot\xi^*)\right](d\xi^*)_n\,. \tag{3.13}$$

The pG variables ξ_i^* ($i = 1,2,\cdots,n$) have been introduced above as those independent of the ξ_i's. In the following, however, we shall assume that ξ_i's and ξ_i^*'s always appear in pairs, and regard the latter as the * conjugates of the former and *vice versa*. As explained in Sec. V of (I), the * operation plays the role that corresponds to a generalization of Hermitian conjugation in the usual theories.

IV. APPLICATIONS TO pF SYSTEMS

A. Coherent states

The physical object to be discussed in the present section is a pF system,[1] i.e., one described by a set of pF operators $\hat{a}_i$ ($i = 1,2,\cdots,n$) of order p, where $\hat{a}_i$ stands for a_i or $a_i^\dagger$. With each $a_i(a_i^\dagger)$ we shall associate a pG number $\xi_i(\xi_i^*)$ of the same order and examine the possibility of describing the system in terms of such $\hat{\xi}_i$'s, where $\hat{\xi}_i$ stands for ξ_i or ξ_i^*. The method we shall adopt for this purpose is essentially the same as that of describing ordinary Fermi operators in terms of G or gG numbers. In fact, we shall closely follow the procedure given in Sec. V of (I).

We begin by assuming that the Green components of $\hat{a}_i$ and ζ satisfy

$$[\hat{a}_i^{(\alpha)},\zeta^{(\beta)}]_{-\eta_{\alpha\beta}} = 0\,, \tag{4.1}$$

where ζ stands for an arbitrary $\hat{\xi}_i$ or μ_i. Thus, the following trilinear commutation relations hold:

$$[\hat{a}_i,[\hat{a}_j,\zeta]] = 2\delta(\hat{i},\hat{j})\zeta,\quad [\zeta,[\zeta',\hat{a}_i]] = 0,\quad [\zeta,[\hat{a}_i,\hat{a}_j]] = 0\,,\quad [\hat{a}_i,[\zeta,\zeta']] = 0\,, \tag{4.2}$$

where $\delta(\hat{i},\hat{j})$ equals δ_{ij} for $\hat{a}_i = a_i$, $\hat{a}_j = a_j^\dagger$ or $\hat{a}_i = a_i^\dagger, \hat{a}_j = a_j$ and vanishes otherwise. The introduction of pG numbers into the framework of our theory naturally necessitates a corresponding generalization of the state-vector space $\mathscr{A}(\mathscr{B})$. Usually, $\mathscr{A}(\mathscr{B})$ is spanned by state vectors such as $\mathscr{M}(\hat{a}_i,\hat{a}_j,\ldots)|0\rangle$ ($\mathscr{M}(\hat{a}_i^{(\alpha)},\hat{a}_j^{(\beta)},\ldots)|0\rangle$), where $\mathscr{M}$ denotes a monomial in the operators concerned. In the present case, however, we have to allow such $\mathscr{M}$'s to contain as well ξ_i's

and μ_s's ($\xi_i^{(\alpha)}$'s and $\mu_s^{(\beta)}$'s). Let us denote hereafter the space thus enlarged by $\mathscr{A}_G(\mathscr{B}_G)$. At the same time the dual space $\mathscr{A}^*(\mathscr{B}^*)$ of $\mathscr{A}(\mathscr{B})$ has to be enlarged to $\mathscr{A}_G^*(\mathscr{B}_G^*)$ in a similar manner. Corresponding to Eqs. (I.5.2)$_1$ and (I.5.2)$_3$ we further assume for ζ's that

$$\zeta^{(\alpha)}|0\rangle = |0\rangle\zeta^{(\alpha)}, \quad \langle 0|\zeta^{(\alpha)} = \zeta^{(\alpha)}\langle 0|, \tag{4.3}$$

or equivalently

$$\zeta|0\rangle = |0\rangle\zeta, \quad \langle 0|\zeta = \zeta\langle 0|. \tag{4.3'}$$

Since $a_i\zeta|0\rangle = a_i|0\rangle\zeta = 0$ owing to (4.3')$_1$, we find that $\mathscr{A}_G$ is spanned by ket vectors such as $\mathscr{M}(a_i^\dagger, a_j^\dagger, \cdots, \zeta_k, \cdots)|0\rangle$.[8] Similarly, $\mathscr{A}_G^*$ is spanned by bra vectors such as $\langle 0|\mathscr{M}(a_i, a_j, \cdots, \zeta_k, \ldots)$. The use of Eqs. (4.2) and (4.3') enables us to compute inner products of state vectors of $\mathscr{A}_G$ and those of $\mathscr{A}_G^*$, the result being functions of ζ's. For example, the inner product of $\langle 0|[\zeta, a_i]$ and $[\zeta', a_j^\dagger]|0\rangle$ is computed as follows:

$$\begin{aligned}\langle 0|[\zeta,a_i][\zeta',a_j^\dagger]|0\rangle &= \langle 0|[[\zeta,a_i],[\zeta',a_j^\dagger]]|0\rangle \\ &= \langle 0|[\zeta,[a_i,[\zeta',a_j^\dagger]]]|0\rangle \\ &= \langle 0|[\zeta,-2\delta_{ij}\zeta']|0\rangle = 2\delta_{ij}[\zeta',\zeta].\end{aligned}$$

On the other hand, the space $\mathscr{B}_G(\mathscr{B}_G^*)$ is spanned, of course, by ket (bra) vectors such as $\mathscr{M}(a_i^{(\alpha)\dagger}, a_j^{(\beta)\dagger}, \cdots \zeta^{(\gamma)}, \ldots)|0\rangle$ ($\langle 0|\mathscr{M}(a_i^{(\alpha)}, a_j^{(\beta)}, \cdots \zeta^{(\gamma)}, \ldots)$), and their inner products are obtained as functions of ζ's ($\zeta^{(\alpha)}$'s).

Following the procedure given in Sec. V of (I), we now construct eigenstates or coherent states for the operators a_i in $\mathscr{A}_G$ or $\mathscr{A}_G^*$. The state vectors corresponding to Eq. (I.5.4) are written as

$$\begin{aligned}&|(\xi)_n\rangle \equiv \exp(a^\dagger\cdot\xi)\langle 0|, \\ &\langle(\xi)_n| \equiv \langle 0|\delta^n(\xi - a),\end{aligned} \tag{4.4}$$

where

$$(a^\dagger\cdot\xi) \equiv \sum_i (a_i^\dagger\cdot\xi_i) \equiv \tfrac{1}{2}\sum_i [a_i^\dagger, \xi_i],$$

$$\begin{aligned}\delta^n(\xi - a) &\equiv \delta(\xi_1 - a_1)\delta(\xi_2 - a_2)\cdots\delta(\xi_n - a_n) \\ &= \frac{1}{i^{pn}(p!)^n}(\xi_1 - a_1)^p(\xi_2 - a_2)^p\cdots(\xi_n - a_n)^p,\end{aligned} \tag{4.5}$$

[cf. Eq. (3.12)]. As in the case of (I), it is then easy to show that

$$\begin{aligned}&a_i^{(\alpha)}|(\xi)_n\rangle = \xi_i^{(\alpha)}|(\xi)_n\rangle, \\ &\langle(\xi)_n|a_i^{(\alpha)} = \langle(\xi)_n|\xi_i^{(\alpha)},\end{aligned} \tag{4.6}$$

and hence

$$\begin{aligned}&a_i|(\xi)_n\rangle = \xi_i|(\xi)_n\rangle, \\ &\langle(\xi)_n|a_i = \langle(\xi)_n|\xi_i.\end{aligned} \tag{4.6'}$$

Thus, ξ_i's are the eigenvalues of a_i's.

Repeating the arguments that led to Theorem 7 of (I), we find that the orthocompleteness relation now takes the form

$$\begin{aligned}&\langle(\xi)_n|(\xi')_n\rangle = \delta^n(\xi - \xi'), \\ &\int |(\xi)_n\rangle\langle(\xi)_n|(d\xi)_n = 1.\end{aligned} \tag{4.7}$$

From Eqs. (4.6) and (4.6') it is clear that $|(\xi)_n\rangle$'s ($\langle(\xi)_n|$'s) are the coherent states not only for a_i's but also for $a_i^{(\alpha)}$'s. This means that all the above relations hold both in $\mathscr{A}_G(\mathscr{A}_G^*)$ and in $\mathscr{B}_G(\mathscr{B}_G^*)$. Especially, the unit operator on the right-hand side of Eq. (4.7)$_2$ is the one in the larger space $\mathscr{B}_G$ or $\mathscr{B}_G^*$.

Eigenstates or coherent states for the operators $a_i^\dagger$ can be constructed in a similar manner. The states vectors (I.5.12) can now be written as

$$\begin{aligned}&|(\xi^*)_n\rangle \equiv \tilde{\delta}^n(a^\dagger - \xi^*)|0\rangle, \\ &\langle(\xi^*)_n| \equiv \langle 0|\exp(\xi^*\cdot a),\end{aligned} \tag{4.8}$$

where

$$\begin{aligned}\tilde{\delta}^n(a^\dagger - \xi^*) &\equiv \delta(a_n^\dagger - \xi_n^*)\delta(a_{n-1}^\dagger - \xi_{n-1}^*)\cdots\delta(a_1^\dagger - \xi_1^*) \\ &= \frac{1}{i^{pn}(p!)^n}(a_n^\dagger - \xi_n^*)^p(a_{n-1}^\dagger - \xi_{n-1}^*)^p\cdots(a_1^\dagger - \xi_1^*)^p.\end{aligned} \tag{4.9}$$

For the state vectors (4.8) the following eigenvalues equations hold:

$$\begin{aligned}&a_i^{(\alpha)\dagger}|(\xi^*)_n\rangle = \xi_i^{(\alpha)*}|(\xi^*)_n\rangle, \\ &\langle(\xi^*)_n|a_i^{(\alpha)\dagger} = \langle(\xi^*)_n|\xi_i^{(\alpha)*},\end{aligned} \tag{4.10}$$

and hence

$$\begin{aligned}&a_i^\dagger|(\xi^*)_n\rangle = \xi_i^*|(\xi^*)_n\rangle, \\ &\langle(\xi^*)_n|a_i^\dagger = \langle(\xi^*)_n|\xi_i^*.\end{aligned} \tag{4.10'}$$

That is to say, ξ_i^*'s are the eigenvalues of $a_i^\dagger$'s. The orthocompleteness relation given in Theorem 7' of (I) now reads

$$\begin{aligned}&\langle(\xi^*)_n|(\xi^{*\prime})_n\rangle = \tilde{\delta}^n(\xi^* - \xi^{*\prime}), \\ &\int (d\xi^*)_n|(\xi^*)_n\rangle\langle(\xi^*)_n| = 1,\end{aligned} \tag{4.11}$$

where the same remark as given after Eqs. (4.7) applies to the present case.

Furthermore, it is also straightforward to rewrite Eqs. (I.5.16) and (I.5.17) in a form suitable for the present case:

$$\begin{aligned}&\int\int |(\xi^*)_n\rangle\langle(\xi)_n|\exp(\xi^*\cdot\xi)(d\xi^*)_n(d\xi)_n = 1, \\ &\int\int |(\xi)_n\rangle\langle(\xi^*)_n|\exp(\xi\cdot\xi^*)(d\xi)_n(d\xi^*)_n = 1,\end{aligned} \tag{4.12}$$

and

$$\begin{aligned}&\langle(\xi^*)_n|(\xi)_n\rangle = \exp(\xi^*\cdot\xi), \\ &\langle(\xi)_n|(\xi^*)_n\rangle = \exp(\xi\cdot\xi^*).\end{aligned} \tag{4.13}$$

Thus, in contrast with the Bose or para-Bose case the pF operators $\hat{a}_i$ allow four kinds of coherent states. This is due to the fact that $\hat{a}_i$'s are bounded operators.

B. pG representation

We shall now show that the pF system concerned can completely be described in terms of pG numbers when the coherent states are employed as basis vectors in $\mathscr{A}_G$.

To do this we notice first that

$$\langle(\xi^*)_n|:[\hat{a}_i,\hat{a}_j]:|(\xi)_n\rangle = [\hat{\xi}_i,\hat{\xi}_j]\exp(\xi^*,\xi),$$
$$\langle(\xi^*)_n|:[\hat{a}_{i_1},\hat{a}_{i_2},\cdots\hat{a}_{i_m}]_+:|(\xi)_n\rangle = [\hat{\xi}_{i_1},\hat{\xi}_{i_2},\cdots\hat{\xi}_{i_m}]_+\exp(\xi^*\cdot\xi). \quad (4.14)$$

Here $\hat{\xi}_i$ on the right-hand side means ξ_i or ξ_i^* depending on whether the corresponding on $\hat{a}_i$ on the left-hand side is a_i or $a_i^\dagger$, $:[\hat{a}_i,\hat{a}_j]:\equiv[\hat{a}_i,\hat{a}_j]-\langle 0|[\hat{a}_i,\hat{a}_j]|0\rangle$, and $:[\hat{a}_{i_1},\hat{a}_{i_2},\cdots\hat{a}_{i_m}]_+:\equiv m!\Sigma'\hat{a}_{i_1}^{(\alpha_1)}\hat{a}_{i_2}^{(\alpha_2)}\cdots\hat{a}_{i_m}^{(\alpha_m)}$, with the summation Σ' extending over all different α_i's.[7] By invoking the decomposition theorem and by making use of the pF commutation relation, we can express an arbitrary operator in $\mathscr{A}$ as a linear combination of terms of the following form:[6-8]

$$F(a^\dagger,a)=[a_{i_0}^\dagger,a_{i_1'}^\dagger][a_{i_2}^\dagger,a_{i_2'}^\dagger]\cdots[a_{i_r}^\dagger,a_{i_r'}^\dagger]:[a_{j_1}^\dagger,a_{j_1'}]::[a_{j_2}^\dagger,a_{j_2'}]:\cdots:[a_{j_s}^\dagger,a_{j_s'}]:$$
$$\times[a_{k_1},a_{k_1'}][a_{k_2},a_{k_2'}]\cdots[a_{k_t},a_{k_t'}]:[a_{l_1}^\dagger,a_{l_2}^\dagger,\cdots,a_{l_u}^\dagger,a_{m_1},a_{m_2},\cdots,a_{m_v}]_+:. \quad (4.15)$$

By using Eq. (4.14) together with Eq. (4.3′) and inserting the left-hand side of Eq. $(4.12)_2$ in between the last two brackets of Eq. (4.15), we obtain the following integral representation for matrix elements of $F(a^\dagger,a)$:

$$\langle(\xi^*)_n|F(a^\dagger,a)|(\xi)_n\rangle\equiv F_G(\xi^*,\xi)\exp(\xi^*\cdot\xi), \quad (4.16)$$

with

$$F_G(\xi^*,\xi)=\int\int[\xi_{i_0}^*,\xi_{i_1'}^*][\xi_{i_2}^*,\xi_{i_2'}^*]\cdots[\xi_{i_r}^*,\xi_{i_r'}^*][\xi_{j_1}^*,\xi_{j_1'}'][\xi_{j_2}^*,\xi_{j_2'}']\cdots[\xi_{j_s}^*,\xi_{j_s'}']$$
$$\times[\xi_{k_1}',\xi_{k_1'}'][\xi_{k_2}',\xi_{k_2'}']\cdots[\xi_{k_t}',\xi_{k_t'}']$$
$$\times[\xi_{l_1}^{*\prime},\xi_{l_2}^{*\prime},\cdots\xi_{l_u}^{*\prime},\xi_{m_1},\xi_{m_2},\cdots\xi_{m_v}]_+\exp[(\xi^*\cdot\xi')+(\xi'\cdot\xi^{*\prime})+(\xi^{*\prime}\cdot\xi)+(\xi\cdot\xi^*)]\,(d\xi')_n\,(d\xi^{*\prime})_n. \quad (4.17)$$

As can be seen from here, the correspondence between the functional form of $F(a^\dagger,a)$ and that of $F_G(\xi^*,\xi)$ is not so straightforward as in the case of Eq. $(I.5.24)_1$, where $F(x^*,x)$ is obtained from $F(a^\dagger,a)$ by simply replacing $\hat{a}_i$ with $\hat{x}_i$. This is due to the fact that normal ordering is not always possible in the present case. For special types of $F(a^\dagger,a)$, however, the correspondence $F(a^\dagger,a)\longleftrightarrow F_G(\xi^*,\xi)$ is relativley simple. For example,

$$1\longleftrightarrow 1,$$
$$[a_{i_0}^\dagger,a_{i_1'}^\dagger][a_{i_2}^\dagger,a_{i_2'}^\dagger]\cdots[a_{i_r}^\dagger,a_{i_r'}^\dagger]:[a_{j_1}^\dagger,a_{j_1'}]::[a_{j_2}^\dagger,a_{j_2'}]:\cdots:[a_{j_s}^\dagger,a_{j_s'}]:[a_{k_1},a_{k_1'}][a_{k_2},a_{k_2'}]\cdots[a_{k_t},a_{k_t'}]$$
$$\longleftrightarrow[\xi_{i_0}^*,\xi_{i_1'}^*][\xi_{i_2}^*,\xi_{i_2'}^*]\cdots[\xi_{i_r}^*,\xi_{i_r'}^*][\xi_{j_1}^*,\xi_{j_1'}][\xi_{j_2}^*,\xi_{j_2'}]\cdots[\xi_{j_s}^*,\xi_{j_s'}][\xi_{k_1},\xi_{k_1'}][\xi_{k_2},\xi_{k_2'}]\cdots[\xi_{k_t},\xi_{k_t'}], \quad (4.18)$$
$$:[a_{i_1}^\dagger,a_{i_2}^\dagger,\cdots,a_{i_r}^\dagger,a_{j_1},a_{j_2},\cdots a_{j_s}]_+:\longleftrightarrow[\xi_{i_1}^*,\xi_{i_2}^*,\cdots\xi_{i_r}^*,\xi_{j_1},\xi_{j_2},\cdots\xi_{j_s}]_+.$$

Needless to say, to a general observable given by $\Sigma F(a^\dagger,a)$ there corresponds the expression $\Sigma F_G(\xi^*,\xi)$.

Conversely, when $F_G(\xi^*,\xi)$ is given, the corresponding operator $F(a^\dagger,a)$ is found from

$$F(a^\dagger,a)=\int\cdots\int|(\xi')_n\rangle F_G(\xi^{*\prime},\xi)\langle(\xi^*)_n|\exp[(\xi'\cdot\xi^{*\prime})+(\xi^{*\prime}\cdot\xi)+(\xi\cdot\xi^*)]\,(d\xi')_n(d\xi^{*\prime})_n(d\xi)_n(d\xi^*)_n, \quad (4.19)$$

which is easily derived from Eqs. (4.16) and $(4.12)_2$. The above relations (4.16) and (4.19) indicate clearly that there exists a one-to-one correspondence between $F(a^\dagger,a)$ and $F_G(\xi^*,\xi)$.

Let us now turn to state vectors. Given a state vector $|\ \rangle$ in $\mathscr{A}$, we define the corresponding pG *representative* or pG *wavefunction* by

$$\psi(\xi^*)\equiv\langle(\xi^*)_n|\ \rangle, \quad (4.20)$$

and its adjoint by

$$\bar{\psi}(\xi)\equiv\langle\ |(\xi)_n\rangle. \quad (4.21)$$

Thus, the vacuum state $|0\rangle$, in particular, is represented by

$$\psi_0(\xi^*)=\bar{\psi}_0(\xi)=1. \quad (4.22)$$

As in the case of Eq. (I.5.21), the use of Eq. $(4.12)_2$ yields

$$|\ \rangle=\int\int|(\xi)_n\rangle\psi(\xi^*)\exp(\xi\cdot\xi^*)(d\xi)_n(d\xi^*)_n, \quad (4.23)$$

thereby implying that the one-to-one correspondence exists also between $|\ \rangle$ and $\psi(\xi^*)$. The inner product of two state vectors $|1\rangle$ and $|2\rangle$ is given as

$$\langle 2|1\rangle=\int\int\bar{\psi}_2(\xi)\psi_1(\xi^*)\exp(\xi\cdot\xi^*)\,(d\xi)_n(d\xi^*)_n, \quad (4.24)$$

and hence the normalization condition for $\psi(\xi)$ takes the form

$$\int\int\bar{\psi}(\xi)\psi(\xi^*)\exp(\xi\cdot\xi^*)\,(d\xi)_n(d\xi^*)_n=1. \quad (4.25)$$

More generally, for an operator $F(a^\dagger,a)$ we find from Eqs. (4.16), $(4.12)_2$, (4.20), and (4.21) that

$$\langle 2|F(a^\dagger,a)|1\rangle = \int\int \bar{\psi}_2(\xi')F_G(\xi^{*\prime},\xi)\psi_1(\xi^*)\exp[(\xi'\cdot\xi^{*\prime}) + (\xi^{*\prime}\cdot\xi) + (\xi\cdot\xi^*)]\,(d\xi')_n(d\xi^{*\prime})_n(d\xi)_n(d\xi^*)_n\,, \tag{4.26}$$

which corresponds to Eq. (I.5.25). The Schrödinger equation for $\psi(\xi^*)$ is written in the form [cf. Eq. (I.5.26)]

$$i\frac{\partial\psi(\xi^*)}{\partial t} = \int\int H_G(\xi^*,\xi')\,\psi(\xi^{*\prime})\exp[(\xi^*\cdot\xi') + (\xi'\cdot\xi^{*\prime})]\,(d\xi')_n(d\xi^{*\prime})_n\,, \tag{4.27}$$

where $H_G(\xi^*,\xi)$ is obtained from $H(a^\dagger,a)$ according to the prescription given above. We have thus found that the quantum mechanics of pF systems can completely be described in the pG as well as ordinary representations.

Lastly, as an example of the use of the pG representation we shall show how the path integral method is formulated for pF systems. Let $U(t_f,t_i) = \exp[-iH(t_f - t_i)]$ be the time-evolution operator, where t_i and t_f denote the initial and final times, respectively, and let $\langle f|U(t_f,t_i)|i\rangle$ be the transition amplitude from an initial state $|i\rangle$ to a final state $|f\rangle$. As usual, we divide the time interval $t_f - t_i$ into N equal segments: $(t_f - t_i)/N = \Delta t = t_k - t_{k-1}$, where $t_i = t_0$, $t_f = t_N$, and $k = 1,2,\cdots,N$. For N sufficiently large we can write

$$\begin{aligned}\langle(\xi^*_{(k)})_n|U(t_k,t_{k-1})|(\xi_{(k-1)})_n\rangle &= \langle(\xi^*_{(k)})_n|(1 - i\Delta t\,H(a^\dagger,a))|(\xi_{(k-1)})_n\rangle \\ &= \exp[(\xi^*_{(k)}\cdot\xi_{(k-1)}) - i\Delta t\,H_G(\xi^*_{(k)},\xi_{(k-1)})]\,.\end{aligned} \tag{4.28}$$

On the other hand, we obtain by a repeated use of Eq. $(4.12)_2$

$$\begin{aligned}\langle f|U(t_f,t_i)|i\rangle &= \lim_{N\to\infty}\langle f|U(t_N,t_{N-1})\cdots U(t_2,t_1)U(t_1,t_0)|i\rangle \\ &= \lim_{N\to\infty}\int\cdots\int\bar{\psi}_f(\xi_{(N)})\langle(\xi^*_{(N)})_n|U(t_N,t_{N-1})|(\xi_{(N-1)})_n\rangle\cdots\langle(\xi^*_{(2)})_n|U(t_2,t_1)|(\xi_{(1)})_n\rangle \\ &\quad\times\langle(\xi^*_{(1)})_n|U(t_1,t_0)|(\xi_{(0)})_n\rangle\psi_i(\xi^*_{(0)})\exp\Big[-\sum_{k=0}^{N}(\xi^*_{(k)}\cdot\xi_{(k)})\Big]\prod_{k=0}^{N}(d\xi_{(k)})_n(d\xi^*_{(k)})_n\,,\end{aligned} \tag{4.29}$$

where $\psi_i(\xi^*)$ and $\bar{\psi}_f(\xi)$ are the pG wavefunctions for the states $|i\rangle$ and $\langle f|$, respectively. Substitution of Eq. (4.28) in (4.29) then results in

$$\langle f|U(t_f,t_i)|i\rangle = \lim_{N\to\infty}\int\cdots\int\bar{\psi}_f(\xi_{(n)})\exp\Big(\prod_{k=0}^{N}i\Delta t\,L_k\Big)\psi_i(\xi^*_{(0)})\prod_{k=0}^{N}(d\xi_{(k)})_n(d\xi^*_{(k)})_n\,, \tag{4.30}$$

where

$$L_k \equiv i\Big(\xi^*_{(k)}\cdot\frac{\xi_{(k)} - \xi_{(k-1)}}{\Delta t}\Big) - H_G(\xi^*_{(k)},\xi_{(k-1)})\,, \tag{4.31}$$

with

$$\xi_{(-1)} = H_G(\xi^*_{(0)},\xi_{(-1)}) = 0\,,\ \text{i.e.,}\ L_0 = i(\xi^*_{(0)}\cdot\xi_{(0)})/\Delta t\,. \tag{4.32}$$

Thus, the usual path integral method[9] for the case $p = 1$ has been generalized to the case $p \geqslant 2$. In carrying out the above path integral it is convenient, however, to employ the Green representation where the integration variables are gG numbers. As has been discussed before,[10] this is equivalent to reinterpreting pG oscillators as Fermi oscillators with an internal (not necessarily hidden) degree of freedom.

Before closing the present paper let us add a few words about the case of para-Bose systems. Obviously, the above method cannot be extended to the para-Bose case in a straightforward manner. In particular, we encounter a difficulty in defining integrals for those variables which are to play the role of eigenvalues of para-Bose operators. The problem still remains to be investigated.

[1]H.S. Green, Phys. Rev. **90**, 270 (1953).

[2]See, for example, F.A. Berezin, *The Method of Second Quantization* (Academic, New York, 1966); Theor. Math. Phys. **6**, 194 (1971); F.A. Berezin and M.S. Marinov, Ann. Phys. (N.Y.) **104**, 336 (1977).

[3]Preliminary results were reported in M. Omote and S. Kamefuchi, Lett. Nuovo Cimento **24**, 345 (1979). To our knowledge the earliest work that considers this kind of generalization is J.L. Martin, Proc. R. Soc. (London) Ser. A **251**, 543 (1959), and the first author who introduced the term "para-Grassmann" is A.J. Kálnay, Rep. Math. Phys. **9**, 9 (1976).

[4]Y. Ohnuki and S. Kamefuchi, J. Math. Phys. **21**, 601 (1980), to be referred to hereafter as (I). The present paper employs mostly the same notation and conventions as in (I). Equation (3.2) of (I), for example, will be quoted as (I.3.2).

[5]See, for example, Ref. 2 of (I).

[6]Theorem 1 in Y. Ohnuki and S. Kamefuchi, Phys. Rev. **170**, 1279 (1968).

[7]Y. Ohnuki and S. Kamefuchi, Prog. Theor. Phys. **54**, 1860 (1975).

[8]Y. Ohnuki and S. Kamefuchi, *Quantum Field Theory and Parastatistics*, Soryushiron Kenkyu (Kyoto) **55**, (1977), special issue.

[9]Y. Ohnuki and T. Kashiwa, Prog. Theor. Phys. **60**, 548 (1978).

[10]K. Drühl, R. Haag, and J.E. Roberts, Commun. Math. Phys. **18**, 204 (1970); Y. Ohnuki and S. Kamefuchi, Prog. Theor. Phys. **50**, 258 (1973).

A Galileian formulation of spin. II. Explicit realizations [a)]

J. A. Brooke
Department of Mathematics, University of Alberta, Edmonton, Alberta T6G 2G1

(Received 8 August 1979; accepted for publication 9 November 1979)

This paper is the follow-up of an earlier one. It reviews some of the lesser known ideas involved in the representation theory of Clifford algebras and applies these ideas in computing explicit realizations of the spin groups and the Lie algebras appearing in the previous paper.

1. INTRODUCTION

This paper is a continuation of an earlier article,[1] hereinafter referred to as I. Whereas I is concerned with a definition of the spin group associated with the homogeneous Galilei group and also with a number of related matters, part II deals with the more down-to-earth problem of representing (by matrices over possibly noncommutative fields) the Clifford algebras and spin groups of I. Thus, in particular, the fundamental space of Galilei spinors is identified.

As noted briefly above, and in I, in order to deal effectively with Clifford algebras corresponding to real orthogonal spaces of arbitrary dimension, one needs concepts general enough to be applicable when dealing with noncommutative fields (in fact the quaternions). The reason for this is that these higher dimensional Clifford algebras turn out to be isomorphic to matrix algebras over $\mathbf{R}$, $\mathbf{C}$, and $\mathbf{H}$ (reals, complexes, and quaternions, respectively).

It is not the intent to enter into the subtleties mentioned in any more depth than is necessary to provide a systematic computation of the spin groups at hand: namely the de Sitter and Galilei spin groups. However a review of basic notions concerning $\mathbf{R}$-, $\mathbf{C}$-, $\mathbf{H}$-linear spaces (e.g., semilinear maps, correlations, antiinvolutions, correlated spaces) is provided. Several general results on conjugations in Clifford algebras, useful for computation of spin groups, are given. For this background material, Porteous[2] is used extensively as a reference. Indeed, the basic results required, including theorems of this nature are merely stated here without proof, as proofs may be located in Porteous.

After these preliminaries, we calculate the de Sitter and Galilei Clifford algebras and spin groups in their matrix representations. Matrix analogues of results in I are given and we end the discussion with the matrix representations of the relevant Lie algebras.

2. ALGEBRAIC PRELIMINARIES

In this section $\mathbf{K}$ will denote either $\mathbf{R}$, $\mathbf{C}$, or $\mathbf{H}$ where $\mathbf{H}$ is as usual, the real associative algebra generated by a basis $\{1, i, j, k\}$ subject to the relations: $i^2 = j^2 = k^2 = -1 = ijk$.

$\mathbf{A}$ will denote some $\mathbf{K}^s = \mathbf{K}\times\mathbf{K}\times\cdots\times\mathbf{K}$ (s times), together with the usual direct product ring structure (written $\mathbf{A} = {}^s\mathbf{K}$), in addition to:

$$(k_1, k_2, \cdots, k_s) + (k'_1, k'_2, \cdots, k'_s) = (k_1 + k'_1, k_2 + k'_2, \cdots, k_s + k'_s), \tag{2.1a}$$

$$\left.\begin{aligned}(k_1, k_2, \cdots, k_s)\cdot k' &= (k_1 k', k_2 k', \cdots, k_s k')\\ k\cdot(k'_1, k'_2, \cdots, k'_s) &= (kk'_1, kk'_2, \cdots, kk'_s)\end{aligned}\right\}, \tag{2.1b}$$

$$\left.\begin{aligned}a\cdot(kk') &= (a\cdot k)\cdot k'\\ (kk')\cdot a &= k\cdot(k'\cdot a)\end{aligned}\right\}, \tag{2.1c}$$

for all $k, k' \in \mathbf{K}$ and all $a \in \mathbf{A}$. $\mathbf{A}$ is then a $\mathbf{K}$-*linear algebra*.

A *right* (left) *linear space over* $\mathbf{A}$ or a *right* (left) $\mathbf{A}$-*linear space*, is just a right (left) $\mathbf{A}$-module where the ring $\mathbf{A}$ is not necessarily commutative or with unit. The standard example of a right $\mathbf{A}$-linear space is $\mathbf{A}^n = \{(a_1, a_2, \cdots, a_n)\}$ with the usual right $\mathbf{A}$-module defining axioms.

An $\mathbf{R}$-linear mapping t of right or left $\mathbf{A}$-linear spaces X, Y is said to be *semilinear over* $\mathbf{A}$ if there is an automorphism (a) or antiautomorphism (aa) ψ: $\mathbf{A}\to\mathbf{A}$ such that for all $x \in X$, $\lambda \in \mathbf{A}$:

(a) $t(x\cdot\lambda) = t(x)\cdot\psi(\lambda)$, t is right semilinear, (2.2a)

or

$$t(\lambda\cdot x) = \psi(\lambda)\cdot t(x), \quad t \text{ is left semilinear}, \tag{2.2b}$$

when ψ is an automorphism;

(aa) $t(x\cdot\lambda) = \psi(\lambda)\cdot t(x)$, t is right-to-left semilinear, (2.2c)

or

$$t(\lambda\cdot x) = t(x)\cdot\psi(\lambda), \quad t \text{ is left-to-right semilinear}, \tag{2.2d}$$

when ψ is an antiautomorphism.

Unless $t = 0$, ψ is uniquely determined by t. One often says that t is $\mathbf{A}^{\psi}$-linear, the semilinearity being determined to an extent by ψ.

It is helpful to keep in mind the following examples which indicate some possibilities (Ref. 2, p. 199): Here $\mathbf{A} = \mathbf{H} = \langle 1, i, j, k\rangle$, $X = Y = \mathbf{H}$; $t: X\to Y$ and ψ: $\mathbf{A}\to\mathbf{A}$ as follows

(i) $t(x) = x$; $\psi(\lambda) = \lambda$,

(ii) $t(x) = ax$, $0 \neq a \in \mathbf{A}$; $\psi(\lambda) = \lambda$,

(iii) $t(x) = xb$, $0 \neq b \in \mathbf{A}$; $\psi(\lambda) = b^{-1}\lambda b$,

(iv) $t(x) = axb$, $0 \neq a, b \in \mathbf{A}$; $\psi(\lambda) = b^{-1}\lambda b$,

(v) $t(x) = jxj^{-1}$; $\psi(\lambda) = j\lambda j^{-1}$.

Examples (i) to (v) are all of right semilinear mappings,

a)Supported in part by a National Research Council of Canada Postgraduate Scholarship.

 0022-2488/80/040617-05$1.00

whereas the following ones are right-to-left semilinear: $\mathbb{A} = \mathbb{H}$, $X = Y = \mathbb{H}^2$

(vi) $t(x, y) = (\bar{x}, \bar{y})$; $\psi(\lambda) = \bar{\lambda}$,

(vii) $t(x, y) = (\bar{y}, \bar{x})$; $\psi(\lambda) = \bar{\lambda}$.

For X, Y both right (left) A-linear (i.e., ψ is the identity automorphism of A), $L(X, Y)$ denotes the family of A-linear mappings which is itself generally not an A-linear space (right or left). The *dual space* $X^L = L(X, \mathbb{A})$ of a right (left) A-linear space has the structure of a left (right) A-linear space; for $t \in L(X, \mathbb{A})$, $x \in X$, $\lambda \in \mathbb{A}$ we have $(\lambda \cdot t)(x) = \lambda \cdot t(x)$ [respectively $(t \cdot \lambda)(x) = t(x) \cdot \lambda$]. If $t : X \to Y$ is $\mathbb{A}^\psi$-linear, then for all $\gamma \in Y^L$, $\psi^{-1}\gamma t \in X^L$. (This is fairly straightforward and entails checking the four possible cases of semilinearity. The map $t^L : Y^L \to X^L$ by $t^L(\gamma) = \psi^{-1}\gamma t$ is called the *dual* of t and is $\mathbb{A}^{\psi^{-1}}$-linear (again one checks case by case).

A *correlation* on a right A-linear space X is an A-semilinear map $\xi : X \to X^L$. The map $X \times X \to \mathbb{A}$, $(x, y) \to \xi(x) \cdot y$ is called the *product* induced by the correlation (it is analogous to inner product) and the map $X \to \mathbb{A}$, $x \to \xi(x) \cdot x$ is the *form* induced by the correlation (it is analogous to quadratic form). Although $\mathbb{R}$-bilinear, the product is generally not A-bilinear (recall Hermitian inner products which are sesquilinear). An $\mathbb{A}^\psi$ correlation $\xi : X \to X^L$ is *symmetric* or *skew* if for all $a, b \in X$,

$$\xi(b) \cdot a = \begin{cases} \psi(\xi(a) \cdot b) & \text{symmetric case,} \\ -\psi(\xi(a) \cdot b) & \text{skew case.} \end{cases} \tag{2.3}$$

A symmetric product ξ over $\mathbb{C}^\psi$, $\mathbb{H}^\psi$ where ψ is the conjugation antiautomorphism is said to be *Hermitian*. A *reflexive* correlation ξ has the property that for all $a, b \in X$, $\xi(b) \cdot a = 0 \Longleftrightarrow \xi(a) \cdot b = 0$. Symmetric and skew correlations are reflexive. An invertible correlation is said to be *nondegenerate*.

The following results (proofs to be found in Ref. 2, pp. 207–10) are important for later material.

Proposition 2.1: Let ξ, η be nondegenerate $\mathbb{A}^\psi$-correlations on finite dimensional right A-linear spaces X, Y respectively and let $t : X \to Y$ be A-linear. Then there is a unique A-linear map $t^* : Y \to X$ such that for all $a \in X$, $b \in Y$, $\eta(b) \cdot t(a) = \xi(t^*(b)) \cdot a$. (In fact $t^* = \xi^{-1} t^L \eta$ and is called the *adjoint* of t with respect to ξ, η. When $Y = X$ and $\eta = \xi$, the adjoint t^* with respect to ξ is denoted by t^ξ and t is self- or skew-adjoint if $t^\xi = t$ or $-t$, respectively.) □

Proposition 2.2: Let X be a finite dimensional right A-linear space. Then any antiinvolution of the real algebra $\mathrm{End}(X) = L(X, X)$ is representable as the adjoint antiinvolution induced by a nondegenerate reflexive correlation on X. □

Finally, one last notion. An *$\mathbb{A}^\psi$-correlated space* (X, ξ) is a right A-linear space X with an $\mathbb{A}^\psi$-correlation ξ on it, and (X, ξ) will be said to be *nondegenerate*, *reflexive*, *symmetric*, or *skew* if ξ has the respective property. It is *totally isotropic* if ξ is zero and *neutral* if X is the direct sum of two totally isotropic subspaces.

Two examples:

(i) $X = \mathbb{A}^2$ right A-linear, with the $\mathbb{A}^\psi$-sesquilinear product (ψ is an antiinvolution: $\psi^2 =$ identity and ψ is an antiautomorphism):

$$\mathbb{A}^2 \times \mathbb{A}^2 \to \mathbb{A} \text{ by } ((a, b), (a', b')) \mapsto \psi(b) \cdot a' + \psi(a) \cdot b',$$

is a symmetric, neutral, nondegenerate $\mathbb{A}^\psi$-correlated space called the *standard hyperbolic plane* $\mathbb{A}^\psi_{\mathrm{hyp}}$.

(ii) $X = \mathbb{A}^2$ with product (again ψ is an antiinvolution)

$$\mathbb{A}^2 \times \mathbb{A}^2 \to \mathbb{A} \text{ by } ((a, b), (a', b')) \mapsto \psi(b) \cdot a' - \psi(a) \cdot b',$$

is a skew, neutral, nondegenerate $\mathbb{A}^\psi$-correlated space called the *standard symplectic plane* $\mathbb{A}^\psi_{\mathrm{sp}}$.

3. CLIFFORD ALGEBRAS AS ENDOMORPHISM ALGEBRAS

We shall denote the real endomorphism algebra $\mathrm{End}(\mathbb{A}^m) = L(\mathbb{A}^m, \mathbb{A}^m)$ by $\mathbb{A}(m)$, with $\mathbb{A}^m$ right A-linear. As in I, $\mathbb{R}^{r,p,q}(n = p + q + r)$ denotes the space $\mathbb{R}^n$ with bilinear form $\mathrm{diag}(0, \cdots, 0, -1, \cdots, -1, 1, \cdots, 1)$ containing r zeros, p minus ones, and q plus ones; $\mathbb{R}^{0,p,q}$ is abbreviated to $\mathbb{R}^{p,q}$. The universal Clifford algebra associated with $\mathbb{R}^{r,p,q}$ is denoted by $\mathbb{R}_{r,p,q}$, with $\mathbb{R}_{p,q}$ as the abbreviation for $\mathbb{R}_{0,p,q}$. We have the following results (Ref. 2, pp, 248-9):

$$\text{(i)}\ \mathbb{R}_{p+1,q} \cong \mathbb{R}_{q+1,p}\ ; \tag{3.1a}$$

$$\text{(ii)}\ \mathbb{R}_{p,q+4} \cong \mathbb{R}_{p,q} \otimes \mathbb{R}_{0,4} \cong \mathbb{R}_{p,q} \otimes \mathbb{H}(2)\ ; \tag{3.1b}$$

$$\text{(iii)}\ \mathbb{R}_{p,q+8} \cong \mathbb{R}_{p,q} \otimes \mathbb{R}(16) \cong \mathbb{R}_{p,q}(16). \tag{3.1c}$$

Every $\mathbb{R}_{p,q}$ is of the form $\mathbb{A}(m)$ for $\mathbb{A}$ one of $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$, ${}^2\mathbb{R}$, ${}^2\mathbb{H}$, and $\mathbb{A}^m$ is called the *real spinor space* of $\mathbb{R}^{p,q}$ and it elements are the $\mathbb{R}^{p,q}$ *spinors*.

For the usual examples of physical interest we have the following (recall from I, (2.7): $\mathbb{R}^+_{r,p,q} \cong \mathbb{R}_{r,p,q-1} \cong \mathbb{R}_{r,q,p-1}$ for even Clifford algebras):

$$\text{(i)}\ \mathbb{R}^{p,q} = \mathbb{R}^{0,3} \text{ (Euclidean 3-space)}, \tag{3.2}$$

$$\mathbb{R}_{0,3} = {}^2\mathbb{H};\ \mathbb{R}^+_{0,3} = \mathbb{R}_{0,2} = \mathbb{H}\ ;$$

$$\text{(ii)}\ \mathbb{R}^{p,q} = \mathbb{R}^{1,3} \text{ (Minkowski space–time)}, \tag{3.3}$$

$$\mathbb{R}_{1,3} = \mathbb{H}(2);\ \mathbb{R}^+_{1,3} = \mathbb{R}_{1,2} = \mathbb{C}(2)\ ;$$

$$\text{(iii)}\ \mathbb{R}^{p,q} = \mathbb{R}^{1,4} \text{ (related to de Sitter space–time)}, \tag{3.4}$$

$$\mathbb{R}_{1,4} = {}^2\mathbb{H}(2);\ \mathbb{R}^+_{1,4} = \mathbb{R}_{1,3} = \mathbb{H}(2)\ .$$

As noted at some length in I, Spin (Galilei) is a stability subgroup of Spin (de Sitter) in addition to the inclusion $\mathbb{R}_{1,0,3} \subset \mathbb{R}_{0,1,4} = \mathbb{R}_{1,4}$ of universal Clifford algebras. This enables us to realize Spin (Galilei) by first representing Spin (de Sitter) and making a suitable restriction, thereby requiring only the theory for *nondegenerate* orthogonal spaces.

To tie things together, the ensuing results are essential.

A correlation on the spinor space induces an antiinvolution (the adjoint) on its endomorphism algebra (i.e., the Clifford algebra) and conversely by Prop. 2.1, Prop. 2.2, any antiinvolution of the Clifford algebra is induced by a symmetric or skew correlation on spinor space. It is necessary, therefore, to know the appropriate correlation. To this end we have (see Ref. 2, pp. 265–70):

Proposition 3.1: Conjugation on $\mathbb{R}_{0,n}$ is adjoint induced by the standard positive-definite correlation on spinor space. □

Proposition 3.2: For $\mathbf{R}_{p,q} = \mathrm{End}(\mathbf{A}^m) = \mathbf{A}(m)$, with $p > 0$ and $(p, q) \neq (1, 0)$, conjugation is the adjoint on $\mathbf{A}(m)$ induced by a neutral semilinear correlation on $\mathbf{A}^m$. □

4. SPIN (DE SITTER) AND SPIN (GALILEI) CALCULATIONS

First the de Sitter calculation.

As noted in Sec. 3., $\mathbf{R}_{1,4} = {}^2\mathbb{H}(2) = \mathbf{H}(2) \oplus \mathbf{H}(2)$ and $\mathbf{R}^+_{1,4} = \mathbf{H}(2)$. Let us put $F_a = f_a \oplus (-f_a)$, $0 \leqslant a \leqslant 4$, where $f_a \in \mathbf{H}(2)$ are the following:

$$f_0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad f_4 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad f_1 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix},$$
$$f_2 = \begin{pmatrix} 0 & j \\ j & 0 \end{pmatrix}, \quad f_3 = \begin{pmatrix} 0 & k \\ k & 0 \end{pmatrix}. \tag{4.1}$$

For a similar but not identical choice see Refs. 3 and 4. One readily verifies that:

(i) $f_a f_b + f_b f_a = -2g_{ab} \cdot \mathbf{1}_2$; $(g_{ab}) = \mathrm{diag}(-1,1,1,1,1)$, (4.2a)

$$F_a F_b + F_b F_a = -2g_{ab} \cdot \mathbf{1}_2 \oplus \mathbf{1}_2 = -2g_{ab} \cdot \mathbf{1}_4; \tag{4.2b}$$

(ii) $\mathbf{H}(2) = \langle f_0, f_1, f_2, f_3, f_4 \rangle = \mathbb{R}$-algebra generated by $\{f_a\}$ is of $\mathbb{R}$-dimension 16 and that $\mathbf{H}(2) \oplus \mathbf{H}(2)$ is of $\mathbb{R}$-dimension 32

(iii) $$\mathbf{R}^+_{1,4} = \langle F_a F_b : a \leqslant b \rangle \cong \langle f_a f_b : a \leqslant b \rangle = \langle f_0, f_1, f_2, f_3, f_4 \rangle = \mathbb{H}(2), \tag{4.2c}$$

the second last equality holding because $f_0 f_1 f_2 f_3 f_4 = 1$, since $\mathbb{H}(2)$ is a *nonuniversal* Clifford algebra for $\mathbf{R}^{1,4}$.

In view of Prop. 3.2, the correlations on $\mathbf{H}^2$ are of interest and to the examples of Sec. 2, we add another. Let ψ be an antiinvolution of $\mathbf{H}$ and consider the symmetric product $\mathbf{H}^2 \times \mathbf{H}^2 \to \mathbf{H}$ by $((a, b), (a', b')) \mapsto \psi(a)a' - \psi(b)b'$; let's denote this product by $\langle \cdot, \cdot \rangle_-$.

Summarizing our three examples of correlations, we denote the induced products by $\langle \cdot, \cdot \rangle_{\substack{\mathrm{sp.} \\ \mathrm{hyp.} \\ -}}$, according as

$$\langle (a, b), (a', b') \rangle = \begin{cases} \psi(b)a' - \psi(a)b' \\ \psi(b)a' + \psi(a)b', \\ \psi(a)a' - \psi(b)b' \end{cases} \tag{4.3}$$

where (a, b), $(a', b') \in \mathbf{H}^2$; ψ is an $\mathbb{H}$ antiinvolution. Now, up to isomorphism, there are exactly two antiinvolutions on $\mathbb{H}$: conjugation ($\psi(a) = \bar{a}$) and reversion with respect to j ($\psi(a) = j\bar{a}j^{-1}$). For a proof, see Ref. 2, p. 181.

We compute the adjoints associated with $\langle \cdot, \cdot \rangle_{\substack{\mathrm{sp} \\ \mathrm{hyp.} \\ -}}$. If $t = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathbb{H}(2)$, then $t^* = \begin{pmatrix} \kappa & \lambda \\ \mu & \nu \end{pmatrix}$ is determined by $\langle t^*(a,b), (a',b') \rangle = \langle (a,b), t(a',b') \rangle$, which in the symplectic and hyperbolic cases means:

$$\langle t^*(a, b), (a', b') \rangle_{\substack{\mathrm{sp.} \\ \mathrm{hyp.}}} = \langle (\kappa a + \lambda b, \mu a + \nu b), (a', b') \rangle_{\substack{\mathrm{sp.} \\ \mathrm{hyp.}}}$$
$$= \psi(\mu a + \nu b)a' \mp \psi(\kappa a + \lambda b)b'$$
$$= \psi(\mu)\psi(a)a' + \psi(a)(\mp\psi(\kappa))b' + \psi(b)\psi(\nu)a' + \psi(b)(\mp\psi(\lambda))b', \tag{4.4}$$

but

$$\langle (a, b), t(a', b') \rangle_{\substack{\mathrm{sp.} \\ \mathrm{hyp.}}} = \langle (a, b), (\alpha a' + \beta b', \gamma a' + \delta b') \rangle_{\substack{\mathrm{sp.} \\ \mathrm{hyp.}}}$$
$$= \psi(b)(\alpha a' + \beta b') \mp \psi(a)(\gamma a' + \delta b')$$
$$= \psi(b)\alpha a' + \psi(b)\beta b' + \psi(a)(\mp\gamma)a' + \psi(a)(\mp\delta)b', \tag{4.5}$$

so

$$\kappa = \psi(\delta), \quad \lambda = \mp\psi(\beta), \quad \mu = \mp\psi(\gamma), \quad \nu = \psi(\alpha), \tag{4.6}$$

hence

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}^*_{\substack{\mathrm{sp.} \\ \mathrm{hyp.}}} = \begin{pmatrix} \psi(\delta) & \mp\psi(\beta) \\ \mp\psi(\gamma) & \psi(\alpha) \end{pmatrix}. \tag{4.7}$$

Turning to $\langle \cdot, \cdot \rangle_-$:

$$\langle t^*(a, b), (a', b') \rangle_- = \langle (\kappa a + \lambda b, \mu a + \nu b), (a', b') \rangle_-$$
$$= \psi(\kappa a + \lambda b)a' - \psi(\mu a + \nu b)b'$$
$$= \psi(a)\psi(\kappa)a' + \psi(a)(-\psi(\mu))b' + \psi(b)\psi(\lambda)a' + \psi(b)(-\psi(\nu))b', \tag{4.8}$$

$$\langle (a, b), t(a', b') \rangle_- = \langle (a, b), (\alpha a' + \beta b', \gamma a' + \delta b') \rangle_-$$
$$= \psi(a)(\alpha a' + \beta b') - \psi(b)(\gamma a' + \delta b')$$
$$= \psi(a)\alpha a' + \psi(a)\beta b' + \psi(b)(-\gamma)a' + \psi(b)(-\delta)b', \tag{4.9}$$

so

$$\kappa = \psi(\alpha), \quad \lambda = -\psi(\gamma), \quad \mu = -\psi(\beta), \quad \nu = \psi(\delta), \tag{4.10}$$

hence

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}^*_- = \begin{pmatrix} \psi(\alpha) & -\psi(\gamma) \\ -\psi(\beta) & \psi(\delta) \end{pmatrix}. \tag{4.11}$$

Two of the seemingly better known representations in the de Sitter case will now be derived and their connection indicated. For a related discussion see Refs. 3 and 5.

With respect to the reversion antiinvolution on $\mathbb{H}$, $a \mapsto j\bar{a}j^{-1}$ ($1 \mapsto 1$, $i \mapsto i$, $j \mapsto -j$, $k \mapsto k$), the symplectic correlation defines the adjoint on $\mathbf{H}(2)$ by the formulas:

$$f_0^* = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}^* = \begin{pmatrix} \psi(-1) & 0 \\ 0 & \psi(1) \end{pmatrix} = -f_0, \tag{4.12a}$$

$$f_1^* = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}^* = \begin{pmatrix} 0 & -\psi(i) \\ -\psi(i) & 0 \end{pmatrix} = -f_1, \tag{4.12b}$$

$$f_2^* = \begin{pmatrix} 0 & j \\ j & 0 \end{pmatrix}^* = \begin{pmatrix} 0 & -\psi(j) \\ -\psi(j) & 0 \end{pmatrix} = f_2, \tag{4.12c}$$

$$f_3^* = -f_3, \tag{4.12d}$$

$$f_4^* = -f_4. \tag{4.12e}$$

Now $\mathbf{R}^+_{1,4} = \mathbf{R}_{1,3} = \langle f_0, f_1, f_3, f_4 \rangle$ and the adjoint so defined is conjugation on $\mathbf{R}^+_{1,4}$; moreover $N_{\mathrm{sp.}}(s) = \bar{s}s = s^*s$, $s \in \mathbb{H}(2)$, defines the norm obtained from the symplectic correlation with the result that [for Spin$^+$(1,4), see I, p. 956, (5.4)-(5.5)]:

$\mathrm{Spin}^+(1,4)$

$$= \left\{ \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathbf{H}(2): \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}^* \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right\},$$

$$= \left\{ \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} : \psi(\alpha)\delta - \psi(\gamma)\beta = 1, \quad \psi(\delta)\beta = \psi(\beta)\delta, \right.$$

$$\left. \psi(\gamma)\alpha = \psi(\alpha)\gamma \right\},$$

and $\psi(\alpha) = j\bar{\alpha}j^{-1}$.

Next, with respect to the conjugation antiinvolution on $\mathbf{H}$ the symmetric, neutral correlation (with product $\langle \cdot,\cdot \rangle_-$) induces an adjoint on $\mathbf{H}(2)$ defined by:

$$f_0^* = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}^* = \begin{pmatrix} \bar{1} & 0 \\ 0 & -\bar{1} \end{pmatrix} = f_0, \tag{4.13a}$$

$$f_1^* = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}^* = \begin{pmatrix} 0 & -\bar{i} \\ -\bar{i} & 0 \end{pmatrix} = f_1, \tag{4.13b}$$

$$f_2^* = f_2, \tag{4.13c}$$

$$f_3^* = f_3, \tag{4.13d}$$

$$f_4^* = f_4. \tag{4.13e}$$

An easy check shows that $\{f_\alpha f_4 : \alpha = 0,1,2,3\}$ is an orthonormal basis of $\mathbf{R}_{1,3}$, so $\mathbf{R}_{1,4}^+ = \mathbf{R}_{1,3} = \langle f_0 f_4, f_1 f_4, f_2 f_4, f_3 f_4 \rangle$. Furthermore $(f_\alpha f_4)^* = f_4^* f_\alpha^* = f_4 f_\alpha = -f_\alpha f_4$ and this adjoint coincides with conjugation on $\mathbf{R}_{1,4}^+$. Therefore, with $N_-(s) = s^* s$

$\mathrm{Spin}^+(1,4)$

$$= \left\{ \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathbf{H}(2): \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}^* \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right\}$$

$$= \left\{ \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} : |\alpha|^2 - |\gamma|^2 = 1 = |\delta|^2 - |\beta|^2, \ \bar{\alpha}\beta = \bar{\gamma}\delta \right\}.$$

This latter representation also yields the result:

$\mathrm{Spin}^+(1,4)$

$= \mathrm{Sp}(1,1)$

$= \{ s \in \mathbf{H}(2)$, leaving the form $\langle (a,b),(a,b) \rangle_- = |a|^2 - |b|^2$ invariant$\}$,

[for $|\alpha|^2 - |\gamma|^2 = 1 = |\delta|^2 - |\beta|^2$, $\bar{\alpha}\beta = \bar{\gamma}\delta$, iff $\bar{s}^t \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} s = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, where $s = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ and $\bar{s}^t$ = transposed conjugate of s].

Equivalence of these two definitions of $\mathrm{Spin}^+(1,4)$ follows because $N_-(CsC^{-1}) = N_{\mathrm{sp.}}(s)$ for $C = (1/\sqrt{2}) \begin{pmatrix} 1 & -j \\ -j & 1 \end{pmatrix}$ and for all $s \in \mathbf{H}(2)$. In summary we have:

Theorem 4.1: $\mathrm{Spin}^+(\text{de Sitter}) = \mathrm{Spin}^+(1,4) = \mathrm{Sp}(1,1)$

$$= \left\{ \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in GL(2;\mathbb{H}): |\alpha|^2 - |\gamma|^2 = 1 = |\delta|^2 - |\beta|^2, \right.$$

$$\left. \bar{\alpha}\beta = \bar{\gamma}\delta \right\},$$

or alternatively with $\psi(\alpha) = j\bar{\alpha}j^{-1}$,

$\mathrm{Spin}^+(\text{de Sitter})$

$$= \left\{ \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in GL(2;\mathbb{H}): \psi(\alpha)\delta - \psi(\gamma)\beta = 1, \right.$$

$$\left. \psi(\gamma)\alpha = \psi(\alpha)\gamma, \quad \psi(\beta)\delta = \psi(\delta)\beta \right\}.$$

□

Finally, the Galilei calculation.

Keeping previous notation, let

$$e_0 = (1/\sqrt{2})(f_0 + f_4), \quad e_1 = f_1, \quad e_2 = f_2, \quad e_3 = f_3, \tag{4.14a}$$

$$E_0 = e_0 \oplus (-e_0), \quad E_1 = F_1, \quad E_2 = F_2, \quad E_3 = F_3. \tag{4.14b}$$

Then

$$E_\alpha E_\beta + E_\beta E_\alpha = -2\gamma_{\alpha\beta} \cdot \mathbf{1}_2 \oplus \mathbf{1}_2 = -2\gamma_{\alpha\beta} \cdot \mathbf{1}_4, \tag{4.15a}$$

$$e_\alpha e_\beta + e_\beta e_\alpha = -2\gamma_{\alpha\beta} \cdot \mathbf{1}_2, \tag{4.15b}$$

where $(\gamma_{\alpha\beta}) = \mathrm{diag}(0,1,1,1)$ and $\alpha, \beta = 0,1,2,3$. A tedious but elementary computation shows that

$$\dim_{\mathbf{R}} \langle E_0, E_1, E_2, E_3 \rangle$$

$$= \dim \ \mathrm{span}_{\mathbf{R}} \{ 1, E_0, \cdots, E_0E_1, \cdots, E_0E_1E_2, \cdots, E_0E_1E_2E_3 \}$$

$$= 16,$$

and hence that $\{E_\alpha\}$ is an orthonormal basis for $\mathbf{R}_{1,0,3}$. Note, however (recall I Prop. 2.1), $\dim_{\mathbf{R}} \langle e_0,e_1,e_2,e_3 \rangle = 12$, and hence $\mathbf{R}_{1,0,3}^+ \neq \langle e_0,e_1,e_2,e_3 \rangle$, contrary to what might have been guessed on the strength of the de Sitter situation. Of course,

$$\mathbf{R}_{1,0,3}^+ = \mathrm{span}_{\mathbf{R}} \{ 1, E_0E_1, E_0E_2, E_0E_3, E_1E_2, E_2E_3, E_3E_1, E_0E_1E_2E_3 \}$$

$$= \langle E_0E_2, E_1E_2, E_2E_3 \rangle \cong \langle e_0e_2, e_1e_2, e_2e_3 \rangle,$$

is of dimension 8. (That $e_0e_1e_2e_3 = -e_0$, is useful in calculations.) With respect to $\langle \cdot,\cdot \rangle_-$ of the de Sitter case, we calculate adjoints:

$$e_0^* = (1/\sqrt{2})(f_0 + f_4)^* = (1/\sqrt{2})(f_0 + f_4) = e_0, \tag{4.16a}$$

$$e_1^* = e_1, \quad e_2^* = e_2, \quad e_3^* = e_3, \tag{4.16b}$$

so when $\alpha \neq \beta$, $(e_\alpha e_\beta)^* = -e_\alpha e_\beta$, and conjugation on $\mathbf{R}_{1,0,3}^+$, is given by this adjoint. With a minor notation change ($\theta^\alpha \to e_\alpha$), translating I, Thm. 3.1, we have

$$\mathrm{Spin\ (Galilei)} = \mathrm{Spin}(G_0)$$

$$= \{ a + b \cdot e : (a^0)^2 + (a^1)^2 + (a^2)^2 + (a^3)^2 = 1, \ a^0b^0 + a^1b^1 + a^2b^2 + a^3b^3 = 0 \},$$

where

$$a = a^0 1 + a^1 e_2e_3 + a^2 e_3e_1 + a^3 e_1e_2,$$

$$b = b^0 1 + b^1 e_2e_3 + b^2 e_3e_1 + b^3 e_1e_2,$$

$e = e_0e_1e_2e_3 = -e_0$, as noted previously.

In matrix form

$$e_0 = (1/\sqrt{2}) \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}, \quad e_1e_2 = \begin{pmatrix} k & 0 \\ 0 & k \end{pmatrix},$$

$$e_2e_3 = \begin{pmatrix} i & 0 \\ 0 & i \end{pmatrix}, \quad e_3e_1 = \begin{pmatrix} j & 0 \\ 0 & j \end{pmatrix},$$

$$s = a + b \cdot e = \begin{pmatrix} \alpha - (1/\sqrt{2})\beta & -(1/\sqrt{2})\beta \\ (1/\sqrt{2})\beta & \alpha + (1/\sqrt{2})\beta \end{pmatrix},$$

Since

$$s^* = \begin{pmatrix} \bar{\alpha} - (1/\sqrt{2})\bar{\beta} & -(1/\sqrt{2})\bar{\beta} \\ -(-(1/\sqrt{2})\bar{\beta}) & \bar{\alpha} + (1/\sqrt{2})\bar{\beta} \end{pmatrix},$$

we have:

Theorem 4.2:

$$\text{Spin } (G_0) = \left\{ \begin{pmatrix} \alpha - (1/\sqrt{2})\beta & -(1/\sqrt{2})\beta \\ (1/\sqrt{2})\beta & \alpha + (1/\sqrt{2})\beta \end{pmatrix} : |\alpha| = 1, \bar{\alpha}\beta + \bar{\beta}\alpha = 0 \right\},$$

moreover (I, Sec. 5), Spin $(G_0) = \{s \in \text{Spin}^* = (1,4) : se_0s^{-1} = e_0\}$ as may be seen by direct calculation. □

5. THE LIE ALGEBRAS

From I Sec. 6. and notational changes $\Sigma^a \to f_a$, $\theta^\alpha \to e_\alpha$, we write in matrix form, the generators:

$$J_S^1 = \tfrac{1}{2}f_2f_3 = \tfrac{1}{2}\begin{pmatrix} i & 0 \\ 0 & i \end{pmatrix}; \quad J_S^2 = \tfrac{1}{2}f_3f_1 = \tfrac{1}{2}\begin{pmatrix} j & 0 \\ 0 & j \end{pmatrix};$$

$$J_S^3 = \tfrac{1}{2}f_1f_2 = \tfrac{1}{2}\begin{pmatrix} k & 0 \\ 0 & k \end{pmatrix};$$

$$K_S^1 = \tfrac{1}{2}f_0f_1 = \tfrac{1}{2}\begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}; \quad K_S^2 = \tfrac{1}{2}f_0f_2 = \tfrac{1}{2}\begin{pmatrix} 0 & j \\ -j & 0 \end{pmatrix};$$

$$K_S^3 = \tfrac{1}{2}f_0f_3 = \tfrac{1}{2}\begin{pmatrix} 0 & k \\ -k & 0 \end{pmatrix};$$

$$P_S^1 = \tfrac{1}{2}f_1f_4 = \tfrac{1}{2}\begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}; \quad P_S^2 = \tfrac{1}{2}f_2f_4 = \tfrac{1}{2}\begin{pmatrix} -j & 0 \\ 0 & j \end{pmatrix};$$

$$P_S^3 = \tfrac{1}{2}f_3f_4 = \tfrac{1}{2}\begin{pmatrix} -k & 0 \\ 0 & k \end{pmatrix}; \quad E_S = \tfrac{1}{2}f_0f_4 = \tfrac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

of the de Sitter Lie algebra, and

$$J_G^1 = \tfrac{1}{2}e_2e_3 = \tfrac{1}{2}\begin{pmatrix} i & 0 \\ 0 & i \end{pmatrix}; \quad J_G^2 = \tfrac{1}{2}e_3e_1 = \tfrac{1}{2}\begin{pmatrix} j & 0 \\ 0 & j \end{pmatrix};$$

$$J_G^3 = \tfrac{1}{2}e_1e_2 = \tfrac{1}{2}\begin{pmatrix} k & 0 \\ 0 & k \end{pmatrix};$$

$$K_G^1 = \tfrac{1}{2}e_0e_1 = (1/2\sqrt{2})\begin{pmatrix} i & i \\ -i & -i \end{pmatrix};$$

$$K_G^2 = \tfrac{1}{2}e_0e_2 = \tfrac{1}{2}(1/2\sqrt{2})\begin{pmatrix} j & j \\ -j & -j \end{pmatrix};$$

$$K_G^3 = \tfrac{1}{2}e_0e_3 = (1/2\sqrt{2})\begin{pmatrix} k & k \\ -k & -k \end{pmatrix},$$

of the homogeneous Galilei Lie algebra.

The one-parameter groups within Spin (Galilei), Spin (de Sitter), corresponding to rotations (angular momentum) and boosts (inertial transformations), and in addition in the de Sitter case, to spatial translations (linear momentum) and time translations (energy), are obtained by matrix exponentiation.

Many computations are more efficiently done in matrix form rather than in abstract Clifford algebra form.

6. COMMENTS

The basic matrix representations of Spin (de Sitter) and Spin (Galilei) have been derived and their relationships outlined.

It is significant that a representation theory of arbitrary degenerate Clifford algebras ($\mathbb{R}_{r,p,q}$ with $r > 0$) has *not* been worked out. This appears to be an open problem which is complicated by the fact that $\mathbb{R}_{r,p,q}$ is in general not a semisimple algebra (yet another manifestation of the degeneracy of the Galilei "metric" $(\gamma_{\alpha\beta}) = \text{diag}(0,1,1,1)$ when $r = 1$, $p = 0$, and $q = 3$).

For those uneasy about matrices over the quaternions, the $\mathbb{H}(2)$ representations may simply be converted into $\mathbb{C}(4)$ representations by using essentially the Pauli matrices in a $\mathbb{C}(2)$ representation of $\mathbb{H}$. In fact it can be shown that there are exactly two inequivalent $\mathbb{C}(4)$ faithful representations of $\mathbb{R}_{1,0,3}$ and these are of least $\mathbb{C}$-dimension, and both obtained from the corresponding lowest dimensional $\mathbb{C}$-representations of $\mathbb{R}_{1,4}$. This indicates some of the complexity referred to in the preceding paragraph, and serves as partial justification for the "embedding approach" used throughout.

7. ACKNOWLEDGMENTS

The author is grateful to H.P. Künzle for general discussions and to J.-F. Dumais for remarks on the problem of complex representations of Clifford algebras.

[1]J.A. Brooke, J. Math. Phys. **19**, 952 (1978).
[2]I. Porteous, *Topological Geometry* (Van Nostrand, London, 1969).
[3]R. Takahashi, Bull. Soc. Math. Fr. **91**, 289–433 (1963).
[4]J. Rawnsley, Proc. Cambridge Philos. Soc. (Math. Proc.) **76**, 473–80 (1974).
[5]S. Strőm, in *Lectures in Theoretical Physics*, edited by A. Barut and W. Britten (Colorado University, Boulder, 1971), Vol. 13 and related articles therein.

A note on the symmetries of the 3j and 6j coefficients. I

K. Venkatesh
Department of Physics, University of Mysore, Manasagangotri, Mysore 570 006, India

(Received 16 March 1979; accepted for publication 14 September 1979)

It is shown that the study of the symmetries of the 3j coefficient in terms of the set of six ${}_3F_2(1)$'s derived by us introduces a six-to-one homomorphism of the 72-element group of symmetries of the 3j coefficient on to the 12 permutations of parameters of a single ${}_3F_2(1)$ series of the set. Also, the study of the symmetries of the 6*j* coefficient in terms of the set of three (four) ${}_4F_3(1)$'s derived by us introduces a three (four)-to-one homomorphism of the 144-element group of symmetries of the 6j coefficient on to the 48(36) allowed permutations of parameters of a single ${}_4F_3(1)$ series of the set.

3*j* COEFFICIENT

In a recent paper[1] we derived a set of six ${}_3F_2(1)$'s for the 3*j* coefficient and discussed the 72 symmetries by giving an explicit form of the five Regge[2] symmetries of the 3*j* coefficient. The six ${}_3F_2(1)$'s of the set are given in Appendix A. Each one of the six ${}_3F_2(1)$'s describes 12 symmetries of the 3*j* coefficient. The 12 symmetries listed in Table I of Ref. 1 get mapped on to the 12 permutations of parameters of the ${}_3F_2(1)$ series given by Eq. (A1) as follows:

ABC;DE, *ACB;ED*,
CAB;DE,
BAC;ED,
BCA;DE,
CBA;ED,
BAC;DE,
CAB;ED,
BCA;ED,
CBA;DE,
ACB;DE,
ABC;ED.

While all the six ${}_3F_2(1)$'s are necessary to describe the 72 symmetries, each describing 12 of them, the 12 permutations of parameters of a single ${}_3F_2(1)$ series of the set are sufficient to account for the 72 symmetries of the 3*j* coefficient. We have explicitly listed the 12 symmetries each described by the ${}_3F_2(1)$'s given by Eqs. (A2)–(A6). A calculation of the parameters of the ${}_3F_2(1)$ series corresponding to these symmetries clearly shows that the 12 symmetries each described by the ${}_3F_2(1)$'s given by Eqs. (A2)–(A6) get mapped on to the 12 permutations of parameters of the ${}_3F_2(1)$ series given by Eq. (A1). As an illustration of the proof, the 12 symmetries each described by the ${}_3F_2(1)$'s of Eqs. (A3) and (A4) and their mapping on to the 12 permutations of parameters of the ${}_3F_2(1)$ series given by Eq. (A1) are given in Tables I and II, respectively.

TABLE I. The list of 12 symmetries of the 3*j* coefficient described by the ${}_3F_2(1)$ series given by Eq. (A3) and their mapping on to the 12 permutations of parameters of the ${}_3F_2(1)$ series given by Eq. (A1).

$$\begin{pmatrix} \frac{(j_3+m_3)+(j_1+m_1)}{2} & j_2 & \frac{(j_3-m_3)+(j_1-m_1)}{2} \\ \frac{(j_1+m_1)-(j_3+m_3)}{2} & j_3-j_1 & \frac{(j_1-m_1)-(j_3-m_3)}{2} \end{pmatrix} \rightarrow ABC;ED$$

$$\begin{pmatrix} \frac{(j_2-m_2)+(j_3-m_3)}{2} & j_1 & \frac{(j_2+m_2)+(j_3+m_3)}{2} \\ \frac{(j_2-m_2)-(j_3-m_3)}{2} & j_3-j_2 & \frac{(j_2+m_2)-(j_3+m_3)}{2} \end{pmatrix} \rightarrow ACB;DE$$

$$\begin{pmatrix} \frac{j_1+j_3-m_2}{2} & \frac{j_1+j_2-m_3}{2} & \frac{j_2+j_3-m_1}{2} \\ \frac{(j_2-m_2)+(j_2-j_1-j_3)}{2} & \frac{(j_3-m_3)+(j_3-j_1-j_2)}{2} & \frac{(j_1-m_1)+(j_1-j_2-j_3)}{2} \end{pmatrix} \rightarrow ABC;DE$$

$$\begin{pmatrix} \frac{j_2+j_3+m_1}{2} & \frac{j_1+j_2+m_3}{2} & \frac{j_1+j_3+m_2}{2} \\ \frac{(j_1+m_1)+(j_3-j_2-j_3)}{2} & \frac{(j_3+m_3)+(j_3-j_1-j_2)}{2} & \frac{(j_2+m_2)+(j_2-j_1-j_3)}{2} \end{pmatrix} \rightarrow ACB;ED$$

 0022-2488/80/040622-08$1.00

$$\begin{pmatrix} \frac{j_1+j_3-m_2}{2} & \frac{j_2+j_3-m_1}{2} & \frac{j_1+j_2-m_3}{2} \\ \frac{(j_1+j_3-j_2)-(j_2-m_2)}{2} & \frac{(j_2+j_3-j_1)-(j_1-m_1)}{2} & \frac{(j_1+j_2-j_3)-(j_3-m_3)}{2} \end{pmatrix} \rightarrow BAC;ED$$

$$\begin{pmatrix} \frac{j_2+j_3+m_1}{2} & \frac{j_1+j_3+m_2}{2} & \frac{j_1+j_2+m_3}{2} \\ \frac{(j_2+j_3-j_1)-(j_1+m_1)}{2} & \frac{(j_1+j_3-j_2)-(j_2+m_2)}{2} & \frac{(j_1+j_2-j_3)-(j_3+m_3)}{2} \end{pmatrix} \rightarrow CAB;DE$$

$$\begin{pmatrix} \frac{(j_2-m_2)+(j_3-m_3)}{2} & \frac{(j_2+m_2)+(j_3+m_3)}{2} & j_1 \\ \frac{(j_3-m_3)-(j_2-m_2)}{2} & \frac{(j_3+m_3)-(j_2+m_2)}{2} & j_2-j_3 \end{pmatrix} \rightarrow CAB;ED$$

$$\begin{pmatrix} \frac{(j_1+m_1)+(j_3+m_3)}{2} & \frac{(j_1-m_1)+(j_3-m_3)}{2} & j_2 \\ \frac{(j_3+m_3)-(j_1+m_1)}{2} & \frac{(j_3-m_3)-(j_1-m_1)}{2} & j_1-j_3 \end{pmatrix} \rightarrow BAC;DE$$

$$\begin{pmatrix} j_3 & j_1 & j_2 \\ m_3 & m_1 & m_2 \end{pmatrix} \rightarrow BCA;DE, \quad \begin{pmatrix} j_3 & j_2 & j_1 \\ -m_3 & -m_2 & -m_1 \end{pmatrix} \rightarrow CBA;ED$$

$$\begin{pmatrix} j_3 & \frac{(j_1-m_1)+(j_2-m_2)}{2} & \frac{(j_1+m_1)+(j_2+m_2)}{2} \\ j_1-j_2 & \frac{(j_2-m_2)-(j_1-m_1)}{2} & \frac{(j_2+m_2)-(j_1+m_1)}{2} \end{pmatrix} \rightarrow BCA;ED$$

$$\begin{pmatrix} j_3 & \frac{(j_2+m_2)+(j_1+m_1)}{2} & \frac{(j_2-m_2)+(j_1-m_1)}{2} \\ j_2-j_1 & \frac{(j_1+m_1)-(j_2+m_2)}{2} & \frac{(j_1-m_1)-(j_2-m_2)}{2} \end{pmatrix} \rightarrow CBA;DE$$

Thus, the study of the symmetries of the $3j$ coefficient in terms of the set of six ${}_3F_2(1)$'s introduces a six-to-one homomorphism of the 72-element group of symmetries on to the 12 permutations of parameters of a single ${}_3F_2(1)$ series of the set. This ${}_3F_2(1)$ series is determined by the condition that the parameters D and E calculated from the parameters of

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix}$$

are greater than zero. In our discussion this series is given by Eq. (A1).

6j COEFFICIENT

In our earlier paper,[3] we derived a set of three ${}_4F_3(1)$'s for the $6j$ coefficient and discussed the 144 symmetries in terms of the allowed permutations of parameters of the ${}_4F_3(1)$ series. The three ${}_4F_3(1)$'s of the set are given in Appendix A. Each one of the three ${}_4F_3(1)$'s describes 48 symmetries of the $6j$ coefficient. The list of 48 symmetries described by the permutations of parameters of the ${}_4F_3(1)$ series given by Eq. (A7) is given in Table III. While all the three ${}_4F_3(1)$'s are necessary to describe the 144 symmetries, each describing 48 of them, the 48 allowed permutations of parameters of a single ${}_4F_3(1)$ series are sufficient to account for the 144 symmetries of the $6j$ coefficient. We have explicitly listed the 48 symmetries each described by the ${}_4F_3(1)$'s given by Eqs. (A8) and (A9). A calculation of the parameters of the ${}_4F_3(1)$ series corresponding to these symmetries clearly shows that the 48 symmetries each described by the ${}_4F_3(1)$'s given by Eqs. (A8) and (A9) get mapped on to the 48 permutations of parameters of the ${}_4F_3(1)$ series given by Eq. (A7). For example, the 48 symmetries described by the ${}_4F_3(1)$ series of Eq. (A9) and their mapping on to the permutations of parameters of the ${}_4F_3(1)$ series of Eq. (A7) are given in Table IV.

Thus, the study of the symmetries of the $6j$ coefficient in terms of the set of three ${}_4F_3(1)$'s introduces a three-to-one homomorphism of the 144-element group of symmetries on to the 48 allowed permutations of parameters of a single ${}_4F_3(1)$ series of the set. This ${}_4F_3(1)$ series is determined by the minimum of the β's[5] in the series representation for

$$\begin{Bmatrix} a & b & e \\ d & c & f \end{Bmatrix}.$$

For example, in our discussion, $\beta_0 = a + b + c + d$.

TABLE II. The list of 12 symmetries of the $3j$ coefficient described by the ${}_3F_2(1)$ series given by Eq. (A4) and their mapping on to the 12 permutations of parameters of the ${}_3F_2(1)$ series by Eq. (A1).

$$(-1)^J\begin{pmatrix} j_1 & j_2 & j_3 \\ -m_1 & -m_2 & -m_3 \end{pmatrix} \rightarrow ABC;DE, \quad (-1)^J\begin{pmatrix} j_2 & j_1 & j_3 \\ m_2 & m_1 & m_3 \end{pmatrix} \rightarrow ACB;ED$$

$$(-1)^J\begin{pmatrix} \frac{(j_2-m_2)+(j_1-m_1)}{2} & \frac{(j_2+m_2)+(j_1+m_1)}{2} & j_3 \\ \frac{(j_1-m_1)-(j_2-m_2)}{2} & \frac{(j_1+m_1)-(j_2+m_2)}{2} & j_2-j_1 \end{pmatrix} \rightarrow ABC;ED$$

$$(-1)^J\begin{pmatrix} \frac{(j_2+m_2)+(j_1+m_1)}{2} & \frac{(j_2-m_2)+(j_1-m_1)}{2} & j_3 \\ \frac{(j_2+m_2)-(j_1+m_1)}{2} & \frac{(j_2-m_2)-(j_1-m_1)}{2} & j_1-j_2 \end{pmatrix} \rightarrow ACB;DE$$

$$(-1)^J\begin{pmatrix} j_2 & \frac{(j_1-m_1)+(j_3-m_3)}{2} & \frac{(j_1+m_1)+(j_3+m_3)}{2} \\ j_1-j_3 & \frac{(j_3-m_3)-(j_1-m_1)}{2} & \frac{(j_3+m_3)-(j_1+m_1)}{2} \end{pmatrix} \rightarrow CAB;ED$$

$$(-1)^J\begin{pmatrix} j_1 & \frac{(j_2+m_2)+(j_3+m_3)}{2} & \frac{(j_2-m_2)+(j_3-m_3)}{2} \\ j_2-j_3 & \frac{(j_3+m_3)-(j_2+m_2)}{2} & \frac{(j_3-m_3)-(j_2-m_2)}{2} \end{pmatrix} \rightarrow BAC;DE$$

$$(-1)^J\begin{pmatrix} \frac{j_1+j_2-m_3}{2} & \frac{j_2+j_3-m_1}{2} & \frac{j_1+j_3-m_2}{2} \\ \frac{(j_1+j_2-j_3)-(j_3-m_3)}{2} & \frac{(j_2+j_3-j_1)-(j_1-m_1)}{2} & \frac{(j_1+j_3-j_2)-(j_2-m_2)}{2} \end{pmatrix} \rightarrow CAB;DE$$

$$(-1)^J\begin{pmatrix} \frac{j_1+j_2+m_3}{2} & \frac{j_1+j_3+m_2}{2} & \frac{j_2+j_3+m_1}{2} \\ \frac{(j_1+j_2-j_3)-(j_3+m_3)}{2} & \frac{(j_1+j_3-j_2)-(j_2+m_2)}{2} & \frac{(j_2+j_3-j_1)-(j_1+m_1)}{2} \end{pmatrix} \rightarrow BAC;ED$$

$$(-1)^J\begin{pmatrix} \frac{j_1+j_3+m_2}{2} & \frac{j_1+j_2+m_3}{2} & \frac{j_2+j_3+m_1}{2} \\ \frac{(j_2+m_2)+(j_2-j_1-j_3)}{2} & \frac{(j_3+m_3)+(j_3-j_1-j_2)}{2} & \frac{(j_1+m_1)+(j_1-j_2-j_3)}{2} \end{pmatrix} \rightarrow BCA;DE$$

$$(-1)^J\begin{pmatrix} \frac{j_2+j_3-m_1}{2} & \frac{j_1+j_2-m_3}{2} & \frac{j_1+j_3-m_2}{2} \\ \frac{(j_1-m_1)+(j_1-j_2-j_3)}{2} & \frac{(j_3-m_3)+(j_3-j_1-j_2)}{2} & \frac{(j_2-m_2)+(j_2-j_1-j_3)}{2} \end{pmatrix} \rightarrow CBA;ED$$

$$(-1)^J\begin{pmatrix} \frac{(j_2+m_2)+(j_3+m_3)}{2} & j_1 & \frac{(j_2-m_2)+(j_3-m_3)}{2} \\ \frac{(j_2+m_2)-(j_3+m_3)}{2} & j_3-j_2 & \frac{(j_2-m_2)-(j_3-m_3)}{2} \end{pmatrix} \rightarrow BCA;ED$$

$$(-1)^J\begin{pmatrix} \frac{(j_3-m_3)+(j_3-m_1)}{2} & j_2 & \frac{(j_3+m_3)+(j_1+m_1)}{2} \\ \frac{(j_1-m_1)-(j_3-m_3)}{2} & j_3-j_1 & \frac{(j_1+m_1)-(j_3+m_3)}{2} \end{pmatrix} \rightarrow CBA;DE$$

TABLE III. [a] The list of 48 symmetries (Ref. 4) of the 6*j* coefficient described by the permutations of parameters of the ${}_4F_3(1)$ series given by Eq. (A7).

$$\begin{Bmatrix} a & b & e \\ d & c & f \end{Bmatrix} \rightarrow ABCD, \qquad \begin{Bmatrix} a_1 & b_1 & e_1 \\ d_1 & c_1 & f_1 \end{Bmatrix} \rightarrow CBAD, \qquad \begin{Bmatrix} e_2 & b_2 & a_2 \\ f_2 & c_2 & d_2 \end{Bmatrix} \rightarrow DBCA$$

$$\begin{Bmatrix} b_3 & a_3 & e_3 \\ c_3 & d_3 & f_3 \end{Bmatrix} \rightarrow ABDC, \qquad \begin{Bmatrix} e_4 & a_4 & b_4 \\ f_4 & d_4 & c_4 \end{Bmatrix} \rightarrow DBAC, \qquad \begin{Bmatrix} b_5 & e_5 & a_5 \\ c_5 & f_5 & d_5 \end{Bmatrix} \rightarrow CBDA$$

$$\begin{Bmatrix} d & c & e \\ a & b & f \end{Bmatrix} \rightarrow BADC, \qquad \begin{Bmatrix} d_1 & c_1 & e_1 \\ a_1 & b_1 & f_1 \end{Bmatrix} \rightarrow DABC, \qquad \begin{Bmatrix} e_2 & c_2 & d_2 \\ f_2 & b_2 & a_2 \end{Bmatrix} \rightarrow CADB$$

$$\begin{Bmatrix} c_3 & d_3 & e_3 \\ b_3 & a_3 & f_3 \end{Bmatrix} \rightarrow BACD, \qquad \begin{Bmatrix} e_4 & d_4 & c_4 \\ f_4 & a_4 & b_4 \end{Bmatrix} \rightarrow CABD, \qquad \begin{Bmatrix} c_5 & e_5 & d_5 \\ b_5 & f_5 & a_5 \end{Bmatrix} \rightarrow DACB$$

$$\begin{Bmatrix} a & c & f \\ d & b & e \end{Bmatrix} \rightarrow CDAB, \qquad \begin{Bmatrix} a_1 & c_1 & f_1 \\ d_1 & b_1 & e_1 \end{Bmatrix} \rightarrow ADCB, \qquad \begin{Bmatrix} f_2 & c_2 & a_2 \\ e_2 & b_2 & d_2 \end{Bmatrix} \rightarrow BDAC$$

$$\begin{Bmatrix} c_3 & a_3 & f_3 \\ b_3 & d_3 & e_3 \end{Bmatrix} \rightarrow CDBA, \qquad \begin{Bmatrix} f_4 & a_4 & c_4 \\ e_4 & d_4 & b_4 \end{Bmatrix} \rightarrow BDCA, \qquad \begin{Bmatrix} c_5 & f_5 & a_5 \\ b_5 & e_5 & d_5 \end{Bmatrix} \rightarrow ADBC$$

$$\begin{Bmatrix} d & b & f \\ a & c & e \end{Bmatrix} \rightarrow DCBA, \qquad \begin{Bmatrix} d_1 & b_1 & f_1 \\ a_1 & c_1 & e_1 \end{Bmatrix} \rightarrow BCDA, \qquad \begin{Bmatrix} f_2 & b_2 & d_2 \\ e_2 & c_2 & a_2 \end{Bmatrix} \rightarrow ACBD$$

$$\begin{Bmatrix} b_3 & d_3 & f_3 \\ c_3 & a_3 & e_3 \end{Bmatrix} \rightarrow DCAB, \qquad \begin{Bmatrix} f_4 & d_4 & b_4 \\ e_4 & a_4 & c_4 \end{Bmatrix} \rightarrow ACDB, \qquad \begin{Bmatrix} b_5 & f_5 & d_5 \\ c_5 & e_5 & a_5 \end{Bmatrix} \rightarrow BCAD$$

$$\begin{Bmatrix} b & a & e \\ c & d & f \end{Bmatrix} \rightarrow ABDC, \qquad \begin{Bmatrix} b_1 & a_1 & e_1 \\ c_1 & d_1 & f_1 \end{Bmatrix} \rightarrow DBAC, \qquad \begin{Bmatrix} e_2 & a_2 & b_2 \\ f_2 & d_2 & c_2 \end{Bmatrix} \rightarrow CBDA$$

$$\begin{Bmatrix} a_3 & b_3 & e_3 \\ d_3 & c_3 & f_3 \end{Bmatrix} \rightarrow ABCD, \qquad \begin{Bmatrix} e_4 & b_4 & a_4 \\ f_4 & c_4 & d_4 \end{Bmatrix} \rightarrow CBAD, \qquad \begin{Bmatrix} a_5 & e_5 & b_5 \\ d_5 & f_5 & c_5 \end{Bmatrix} \rightarrow DBCA$$

$$\begin{Bmatrix} c & d & e \\ b & a & f \end{Bmatrix} \rightarrow BACD, \qquad \begin{Bmatrix} c_1 & d_1 & e_1 \\ b_1 & a_1 & f_1 \end{Bmatrix} \rightarrow CABD, \qquad \begin{Bmatrix} e_2 & d_2 & c_2 \\ f_2 & a_2 & b_2 \end{Bmatrix} \rightarrow DACB$$

$$\begin{Bmatrix} d_3 & c_3 & e_3 \\ a_3 & b_3 & f_3 \end{Bmatrix} \rightarrow BADC, \qquad \begin{Bmatrix} e_4 & c_4 & d_4 \\ f_4 & b_4 & a_4 \end{Bmatrix} \rightarrow DABC, \qquad \begin{Bmatrix} d_5 & e_5 & c_5 \\ a_5 & f_5 & b_5 \end{Bmatrix} \rightarrow CADB,$$

$$\begin{Bmatrix} c & a & f \\ b & d & e \end{Bmatrix} \rightarrow CDBA, \qquad \begin{Bmatrix} c_1 & a_1 & f_1 \\ b_1 & d_1 & e_1 \end{Bmatrix} \rightarrow BDCA, \qquad \begin{Bmatrix} f_2 & a_2 & c_2 \\ e_2 & d_2 & b_2 \end{Bmatrix} \rightarrow ADBC$$

$$\begin{Bmatrix} a_3 & c_3 & f_3 \\ d_3 & b_3 & e_3 \end{Bmatrix} \rightarrow CDAB, \qquad \begin{Bmatrix} f_4 & c_4 & a_4 \\ e_4 & b_4 & d_4 \end{Bmatrix} \rightarrow ADCB, \qquad \begin{Bmatrix} a_5 & f_5 & c_5 \\ d_5 & e_5 & b_5 \end{Bmatrix} \rightarrow BDAC$$

$$\begin{Bmatrix} b & d & f \\ c & a & e \end{Bmatrix} \rightarrow DCAB, \qquad \begin{Bmatrix} b_1 & d_1 & f_1 \\ c_1 & a_1 & e_1 \end{Bmatrix} \rightarrow ACDB, \qquad \begin{Bmatrix} f_2 & d_2 & b_2 \\ e_2 & a_2 & c_2 \end{Bmatrix} \rightarrow BCAD$$

$$\begin{Bmatrix} d_3 & b_3 & f_3 \\ a_3 & c_3 & e_3 \end{Bmatrix} \rightarrow DCBA, \qquad \begin{Bmatrix} f_4 & b_4 & d_4 \\ e_4 & c_4 & a_4 \end{Bmatrix} \rightarrow BCDA, \qquad \begin{Bmatrix} d_5 & f_5 & b_5 \\ a_5 & e_5 & c_5 \end{Bmatrix} \rightarrow ACBD$$

[a] The first 24 symmetries in this table correspond to the denominator parameter permutation EFG and the next 24 symmetries in this table correspond to the denominator parameter permutation EGF.

TABLE IV.[a] The list of 48 symmetries (Ref. 4) of the 6j coefficient described by the ${}_4F_3(1)$ series given by Eq. (A9) and their mapping on to the permutations of parameters of the ${}_4F_3(1)$ series given by Eq. (A7)

$$\begin{Bmatrix} b & e & a \\ c & f & d \end{Bmatrix} \rightarrow ABDC,$$

$$\begin{Bmatrix} e_3 & b_3 & a_3 \\ f_3 & c_3 & d_3 \end{Bmatrix} \rightarrow CBDA,$$

$$\begin{Bmatrix} c & f & a \\ b & e & d \end{Bmatrix} \rightarrow CDBA,$$

$$\begin{Bmatrix} f_3 & c_3 & a_3 \\ e_3 & b_3 & d_3 \end{Bmatrix} \rightarrow ADBC,$$

$$\begin{Bmatrix} b & f & d \\ c & e & a \end{Bmatrix} \rightarrow DCAB,$$

$$\begin{Bmatrix} e_3 & c_3 & d_3 \\ f_3 & b_3 & a_3 \end{Bmatrix} \rightarrow DACB,$$

$$\begin{Bmatrix} c & e & d \\ b & f & a \end{Bmatrix} \rightarrow BACD,$$

$$\begin{Bmatrix} f_3 & b_3 & d_3 \\ e_3 & c_3 & a_3 \end{Bmatrix} \rightarrow BCAD,$$

$$\begin{Bmatrix} a & e & b \\ d & f & c \end{Bmatrix} \rightarrow ABCD,$$

$$\begin{Bmatrix} e_3 & a_3 & b_3 \\ f_3 & d_3 & c_3 \end{Bmatrix} \rightarrow DBCA,$$

$$\begin{Bmatrix} d & f & b \\ a & e & c \end{Bmatrix} \rightarrow DCBA,$$

$$\begin{Bmatrix} f_3 & d_3 & b_3 \\ e_3 & a_3 & c_3 \end{Bmatrix} \rightarrow ACBD,$$

$$\begin{Bmatrix} a & f & c \\ d & e & b \end{Bmatrix} \rightarrow CDAB,$$

$$\begin{Bmatrix} e_3 & d_3 & c_3 \\ f_3 & a_3 & b_3 \end{Bmatrix} \rightarrow CADB,$$

$$\begin{Bmatrix} d & e & c \\ a & f & b \end{Bmatrix} \rightarrow BADC,$$

$$\begin{Bmatrix} f_3 & a_3 & c_3 \\ e_3 & d_3 & b_3 \end{Bmatrix} \rightarrow BDAC,$$

$$\begin{Bmatrix} b_1 & e_1 & a_1 \\ c_1 & f_1 & d_1 \end{Bmatrix} \rightarrow DBAC,$$

$$\begin{Bmatrix} a_4 & b_4 & e_4 \\ d_4 & c_4 & f_4 \end{Bmatrix} \rightarrow DBCA,$$

$$\begin{Bmatrix} c_1 & e_1 & d_1 \\ b_1 & f_1 & a_1 \end{Bmatrix} \rightarrow CABD,$$

$$\begin{Bmatrix} d_4 & b_4 & f_4 \\ a_4 & c_4 & e_4 \end{Bmatrix} \rightarrow ACBD,$$

$$\begin{Bmatrix} b_1 & f_1 & d_1 \\ c_1 & e_1 & a_1 \end{Bmatrix} \rightarrow ACDB,$$

$$\begin{Bmatrix} d_4 & c_4 & e_4 \\ a_4 & b_4 & f_4 \end{Bmatrix} \rightarrow CADB,$$

$$\begin{Bmatrix} c_1 & f_1 & a_1 \\ b_1 & e_1 & d_1 \end{Bmatrix} \rightarrow BDCA,$$

$$\begin{Bmatrix} a_4 & c_4 & f_4 \\ d_4 & b_4 & e_4 \end{Bmatrix} \rightarrow BDAC,$$

$$\begin{Bmatrix} a_1 & e_1 & b_1 \\ d_1 & f_1 & c_1 \end{Bmatrix} \rightarrow CBAD,$$

$$\begin{Bmatrix} b_4 & a_4 & e_4 \\ c_4 & d_4 & f_4 \end{Bmatrix} \rightarrow CBDA,$$

$$\begin{Bmatrix} d_1 & e_1 & c_1 \\ a_1 & f_1 & b_1 \end{Bmatrix} \rightarrow DABC,$$

$$\begin{Bmatrix} c_4 & a_4 & f_4 \\ b_4 & d_4 & e_4 \end{Bmatrix} \rightarrow ADBC,$$

$$\begin{Bmatrix} a_1 & f_1 & c_1 \\ d_1 & e_1 & b_1 \end{Bmatrix} \rightarrow ADCB,$$

$$\begin{Bmatrix} c_4 & d_4 & e_4 \\ b_4 & a_4 & f_4 \end{Bmatrix} \rightarrow DACB,$$

$$\begin{Bmatrix} d_1 & f_1 & b_1 \\ a_1 & e_1 & c_1 \end{Bmatrix} \rightarrow BCDA,$$

$$\begin{Bmatrix} b_4 & d_4 & f_4 \\ c_4 & a_4 & e_4 \end{Bmatrix} \rightarrow BCAD,$$

$$\begin{Bmatrix} a_2 & e_2 & b_2 \\ d_2 & f_2 & c_2 \end{Bmatrix} \rightarrow ABCD$$

$$\begin{Bmatrix} e_5 & a_5 & b_5 \\ f_5 & d_5 & c_5 \end{Bmatrix} \rightarrow CBAD$$

$$\begin{Bmatrix} d_2 & f_2 & b_2 \\ a_2 & e_2 & c_2 \end{Bmatrix} \rightarrow DCBA$$

$$\begin{Bmatrix} e_5 & d_5 & c_5 \\ f_5 & a_5 & b_5 \end{Bmatrix} \rightarrow DABC$$

$$\begin{Bmatrix} a_2 & f_2 & c_2 \\ d_2 & e_2 & b_2 \end{Bmatrix} \rightarrow CDAB$$

$$\begin{Bmatrix} f_5 & a_5 & c_5 \\ e_5 & d_5 & b_5 \end{Bmatrix} \rightarrow ADCB$$

$$\begin{Bmatrix} d_2 & e_2 & c_2 \\ a_2 & f_2 & b_2 \end{Bmatrix} \rightarrow BADC$$

$$\begin{Bmatrix} f_5 & d_5 & b_5 \\ e_5 & a_5 & c_5 \end{Bmatrix} \rightarrow BCDA$$

$$\begin{Bmatrix} b_2 & e_2 & a_2 \\ c_2 & f_2 & d_2 \end{Bmatrix} \rightarrow ABDC$$

$$\begin{Bmatrix} e_5 & b_5 & a_5 \\ f_5 & c_5 & d_5 \end{Bmatrix} \rightarrow DBAC$$

$$\begin{Bmatrix} c_2 & f_2 & a_2 \\ b_2 & e_2 & d_2 \end{Bmatrix} \rightarrow CDBA$$

$$\begin{Bmatrix} e_5 & c_5 & d_5 \\ f_5 & b_5 & a_5 \end{Bmatrix} \rightarrow CABD$$

$$\begin{Bmatrix} b_2 & f_2 & d_2 \\ c_2 & e_2 & a_2 \end{Bmatrix} \rightarrow DCAB,$$

$$\begin{Bmatrix} f_5 & b_5 & d_5 \\ e_5 & c_5 & a_5 \end{Bmatrix} \rightarrow ACDB$$

$$\begin{Bmatrix} c_2 & e_2 & d_2 \\ b_2 & f_2 & a_2 \end{Bmatrix} \rightarrow BACD$$

$$\begin{Bmatrix} f_5 & c_5 & a_5 \\ e_5 & b_5 & d_5 \end{Bmatrix} \rightarrow BDCA$$

[a] The first 24 symmetries in this table correspond to the denominator parameter permutation EFG and the next 24 symmetries in this table correspond to the denominator parameter permutation EGF.

We had also derived[5] an equivalent set of four ${}_4F_3(1)$'s for the $6j$ coefficient. The equivalence between the two sets of ${}_4F_3(1)$'s has also been discussed earlier.[6] The ${}_4F_3(1)$'s of set II are given in Appendix A. Each one of the four ${}_4F_3(1)$'s of set II describes 36 symmetries of the $6j$ coefficient. We have explicitly listed the 36 symmetries each described by a single ${}_4F_3(1)$ series of this set. A proof similar to that given above for the set I of ${}_4F_3(1)$'s is given for the set II of ${}_4F_3(1)$'s and examples are given in Tables V and VI. This shows that the discussion of the symmetries of the $6j$ coefficient in terms of the set II of ${}_4F_3(1)$'s introduces a four-to-one homomorphism of the 144-element group of symmetries on to the 36 allowed permutations of parameters of a single ${}_4F_3(1)$ series of set II. This ${}_4F_3(1)$ series is determined by the maximum of the α's[5] in the series representation for

$$\begin{Bmatrix} a & b & e \\ d & c & f \end{Bmatrix}.$$

We have chosen $\alpha_0 = a + b + e$ in our discussion.[7]

TABLE V.[a] The list of 36 symmetries (Ref. 4) of the $6j$ coefficient described by the permutations of parameters of the ${}_4F_3(1)$ series given by Eq. (A10).

$\begin{Bmatrix} a & b & e \\ d & c & f \end{Bmatrix} \rightarrow EFG,$	$\begin{Bmatrix} a_1 & b_1 & e_1 \\ d_1 & c_1 & f_1 \end{Bmatrix} \rightarrow FEG,$	$\begin{Bmatrix} e_2 & b_2 & a_2 \\ f_2 & c_2 & d_2 \end{Bmatrix} \rightarrow GFE$
$\begin{Bmatrix} b_3 & a_3 & e_3 \\ c_3 & d_3 & f_3 \end{Bmatrix} \rightarrow EGF,$	$\begin{Bmatrix} e_4 & a_4 & b_4 \\ f_4 & d_4 & c_4 \end{Bmatrix} \rightarrow GEF,$	$\begin{Bmatrix} b_5 & e_5 & a_5 \\ c_5 & f_5 & d_5 \end{Bmatrix} \rightarrow FGE$
$\begin{Bmatrix} b & a & e \\ c & d & f \end{Bmatrix} \rightarrow EGF,$	$\begin{Bmatrix} b_1 & a_1 & e_1 \\ c_1 & d_1 & f_1 \end{Bmatrix} \rightarrow GEF,$	$\begin{Bmatrix} e_2 & a_2 & b_2 \\ f_2 & d_2 & c_2 \end{Bmatrix} \rightarrow FGE$
$\begin{Bmatrix} a_3 & b_3 & e_3 \\ d_3 & c_3 & f_3 \end{Bmatrix} \rightarrow EFG,$	$\begin{Bmatrix} e_4 & b_4 & a_4 \\ f_4 & c_4 & d_4 \end{Bmatrix} \rightarrow FEG,$	$\begin{Bmatrix} a_5 & e_5 & b_5 \\ d_5 & f_5 & c_5 \end{Bmatrix} \rightarrow GFE$
$\begin{Bmatrix} a & e & b \\ d & f & c \end{Bmatrix} \rightarrow FEG,$	$\begin{Bmatrix} a_1 & e_1 & b_1 \\ d_1 & f_1 & c_1 \end{Bmatrix} \rightarrow EFG,$	$\begin{Bmatrix} b_2 & e_2 & a_2 \\ c_2 & f_2 & d_2 \end{Bmatrix} \rightarrow GEF$
$\begin{Bmatrix} e_3 & a_3 & b_3 \\ f_3 & d_3 & c_3 \end{Bmatrix} \rightarrow FGE,$	$\begin{Bmatrix} b_4 & a_4 & e_4 \\ c_4 & d_4 & f_4 \end{Bmatrix} \rightarrow GFE,$	$\begin{Bmatrix} e_5 & b_5 & a_5 \\ f_5 & c_5 & d_5 \end{Bmatrix} \rightarrow EGF$
$\begin{Bmatrix} e & a & b \\ f & d & c \end{Bmatrix} \rightarrow FGE,$	$\begin{Bmatrix} e_1 & a_1 & b_1 \\ f_1 & d_1 & c_1 \end{Bmatrix} \rightarrow GFE,$	$\begin{Bmatrix} b_2 & a_2 & e_2 \\ c_2 & d_2 & f_2 \end{Bmatrix} \rightarrow EGF$
$\begin{Bmatrix} a_3 & e_3 & b_3 \\ d_3 & f_3 & c_3 \end{Bmatrix} \rightarrow FEG,$	$\begin{Bmatrix} b_4 & e_4 & a_4 \\ c_4 & f_4 & d_4 \end{Bmatrix} \rightarrow EFG,$	$\begin{Bmatrix} a_5 & b_5 & e_5 \\ d_5 & c_5 & f_5 \end{Bmatrix} \rightarrow GEF$
$\begin{Bmatrix} e & b & a \\ f & c & d \end{Bmatrix} \rightarrow GFE,$	$\begin{Bmatrix} e_1 & b_1 & a_1 \\ f_1 & c_1 & d_1 \end{Bmatrix} \rightarrow FGE,$	$\begin{Bmatrix} a_2 & b_2 & e_2 \\ d_2 & c_2 & f_2 \end{Bmatrix} \rightarrow EFG$
$\begin{Bmatrix} b_3 & e_3 & a_3 \\ c_3 & f_3 & d_3 \end{Bmatrix} \rightarrow GEF,$	$\begin{Bmatrix} a_4 & e_4 & b_4 \\ d_4 & f_4 & c_4 \end{Bmatrix} \rightarrow EGF,$	$\begin{Bmatrix} b_5 & a_5 & e_5 \\ c_5 & d_5 & f_5 \end{Bmatrix} \rightarrow FEG$
$\begin{Bmatrix} b & e & a \\ c & f & d \end{Bmatrix} \rightarrow GEF,$	$\begin{Bmatrix} b_1 & e_1 & a_1 \\ c_1 & f_1 & d_1 \end{Bmatrix} \rightarrow EGF,$	$\begin{Bmatrix} a_2 & e_2 & b_2 \\ d_2 & f_2 & c_2 \end{Bmatrix} \rightarrow FEG$
$\begin{Bmatrix} e_3 & b_3 & a_3 \\ f_3 & c_3 & d_3 \end{Bmatrix} \rightarrow GFE,$	$\begin{Bmatrix} a_4 & b_4 & e_4 \\ d_4 & c_4 & f_4 \end{Bmatrix} \rightarrow FGE,$	$\begin{Bmatrix} e_5 & a_5 & b_5 \\ f_5 & d_5 & c_5 \end{Bmatrix} \rightarrow EFG$

[a]The six symmetries present in the first two rows of this table correspond to the numerator parameter permutation BCD. The subsequent five sets of two rows of symmetries in this table correspond to the numerator parameter permutations CBD, BDC, DBC, DCB, and CDB, respectively.

TABLE VI.[a] The list of 36 symmetries (Ref. 4) described by the ${}_4F_3(1)$ series given by Eq. (A12) and their mapping on to the permutations of parameters of the ${}_4F_3(1)$ series given by Eq. (A10).

$\begin{Bmatrix} b & d & f \\ c & a & e \end{Bmatrix} \rightarrow EGF,$	$\begin{Bmatrix} b_1 & d_1 & f_1 \\ c_1 & a_1 & e_1 \end{Bmatrix} \rightarrow GEF,$	$\begin{Bmatrix} e_2 & d_2 & c_2 \\ f_2 & a_2 & b_2 \end{Bmatrix} \rightarrow FGE$
$\begin{Bmatrix} a_1 & c_1 & f_1 \\ d_1 & b_1 & e_1 \end{Bmatrix} \rightarrow EFG,$	$\begin{Bmatrix} f_4 & c_4 & a_4 \\ e_4 & b_4 & d_4 \end{Bmatrix} \rightarrow FEG,$	$\begin{Bmatrix} d_5 & e_5 & c_5 \\ a_5 & f_5 & b_5 \end{Bmatrix} \rightarrow GFE$
$\begin{Bmatrix} a & c & f \\ d & b & e \end{Bmatrix} \rightarrow EFG,$	$\begin{Bmatrix} a_1 & c_1 & f_1 \\ d_1 & b_1 & e_1 \end{Bmatrix} \rightarrow FEG,$	$\begin{Bmatrix} e_2 & c_2 & d_2 \\ f_2 & b_2 & a_2 \end{Bmatrix} \rightarrow GFE$
$\begin{Bmatrix} b_3 & d_3 & f_3 \\ c_3 & a_3 & e_3 \end{Bmatrix} \rightarrow EGF,$	$\begin{Bmatrix} f_4 & d_4 & b_4 \\ e_4 & a_4 & c_4 \end{Bmatrix} \rightarrow GEF,$	$\begin{Bmatrix} c_5 & e_5 & d_5 \\ b_5 & f_5 & a_5 \end{Bmatrix} \rightarrow FGE$
$\begin{Bmatrix} e & d & c \\ f & a & b \end{Bmatrix} \rightarrow FGE,$	$\begin{Bmatrix} e_1 & d_1 & c_1 \\ f_1 & a_1 & b_1 \end{Bmatrix} \rightarrow GFE,$	$\begin{Bmatrix} b_2 & d_2 & f_2 \\ c_2 & a_2 & e_2 \end{Bmatrix} \rightarrow EGF$
$\begin{Bmatrix} a_3 & f_3 & c_3 \\ d_3 & e_3 & b_3 \end{Bmatrix} \rightarrow FEG,$	$\begin{Bmatrix} c_4 & f_4 & a_4 \\ b_4 & e_4 & d_4 \end{Bmatrix} \rightarrow EFG,$	$\begin{Bmatrix} d_5 & b_5 & f_5 \\ a_5 & c_5 & e_5 \end{Bmatrix} \rightarrow GEF$
$\begin{Bmatrix} a & f & c \\ d & e & b \end{Bmatrix} \rightarrow FEG,$	$\begin{Bmatrix} a_1 & f_1 & c_1 \\ d_1 & e_1 & b_1 \end{Bmatrix} \rightarrow EFG,$	$\begin{Bmatrix} b_2 & f_2 & d_2 \\ c_2 & e_2 & a_2 \end{Bmatrix} \rightarrow GEF$
$\begin{Bmatrix} e_3 & d_3 & c_3 \\ f_3 & a_3 & b_3 \end{Bmatrix} \rightarrow FGE,$	$\begin{Bmatrix} c_4 & d_4 & e_4 \\ b_4 & a_4 & f_4 \end{Bmatrix} \rightarrow GFE,$	$\begin{Bmatrix} f_5 & b_5 & d_5 \\ e_5 & c_5 & a_5 \end{Bmatrix} \rightarrow EGF$
$\begin{Bmatrix} b & f & d \\ c & e & a \end{Bmatrix} \rightarrow GEF,$	$\begin{Bmatrix} b_1 & f_1 & d_1 \\ c_1 & e_1 & a_1 \end{Bmatrix} \rightarrow EGF,$	$\begin{Bmatrix} a_2 & f_2 & c_2 \\ d_2 & e_2 & b_2 \end{Bmatrix} \rightarrow FEG$
$\begin{Bmatrix} e_3 & c_3 & d_3 \\ f_3 & b_3 & a_3 \end{Bmatrix} \rightarrow GFE,$	$\begin{Bmatrix} d_4 & c_4 & e_4 \\ a_4 & b_4 & f_4 \end{Bmatrix} \rightarrow FGE,$	$\begin{Bmatrix} f_5 & a_5 & c_5 \\ e_5 & d_5 & b_5 \end{Bmatrix} \rightarrow EFG$
$\begin{Bmatrix} e & c & d \\ f & b & a \end{Bmatrix} \rightarrow GFE,$	$\begin{Bmatrix} e_1 & c_1 & d_1 \\ f_1 & b_1 & a_1 \end{Bmatrix} \rightarrow FGE,$	$\begin{Bmatrix} a_2 & c_2 & f_2 \\ d_2 & b_2 & e_2 \end{Bmatrix} \rightarrow EFG$
$\begin{Bmatrix} b_3 & f_3 & d_3 \\ c_3 & e_3 & a_3 \end{Bmatrix} \rightarrow GEF,$	$\begin{Bmatrix} d_4 & f_4 & b_4 \\ a_4 & e_4 & c_4 \end{Bmatrix} \rightarrow EGF,$	$\begin{Bmatrix} c_5 & a_5 & f_5 \\ b_5 & d_5 & e_5 \end{Bmatrix} \rightarrow FEG$

[a]The six symmetries present in the first two rows of this table correspond to the numerator parameter permutation BCD. The subsequent five sets of two rows of symmetries in this table correspond to the numerator parameter permutations CBD, BDC, DBC, DCB, and CDB, respectively.

ACKNOWLEDGMENTS

The author is grateful to Dr. G. Ramachandran for his kind encouragement and to Professor B. Sanjeevaiah for providing facilities for research. The author thanks the CSIR (India) for the award of the Senior Research Fellowship. The author is thankful to the referee for his comments and suggestions.

APPENDIX A: THE SET OF GENERALIZED HYPERGEOMETRIC FUNCTIONS FOR THE 3*j* AND 6*j* COEFFICIENTS

A. Set of six ${}_3F_2(1)$'s for the 3*j* coefficient

$$A = -(j_1 + j_2 - j_3), \quad B = -(j_1 - m_1), \quad C = -(j_2 + m_2),$$
$$D = j_3 - j_2 + m_1 + 1, \quad E = j_3 - j_1 - m_2 + 1, \tag{A1}$$

$$A = -(j_1 + m_1), \quad B = -(j_1 + j_3 - j_2), \quad C = -(j_3 - m_3),$$
$$D = j_2 - j_3 - m_1 + 1, \quad E = j_2 - j_1 + m_3 + 1, \tag{A2}$$

$$A=-(j_2-m_2),\quad B=-(j_3+m_3),\quad C=-(j_2+j_3-j_1),$$
$$D=j_1-j_3+m_2+1,\quad E=j_1-j_2-m_3+1, \tag{A3}$$

$$A=-(j_1+j_2-j_3),\quad B=-(j_1+m_1),\quad C=-(j_2-m_2),$$
$$D=j_3-j_2-m_1+1,\quad E=j_3-j_1+m_2+1, \tag{A4}$$

$$A=-(j_1-m_1),\quad B=-(j_1+j_3-j_2),\quad C=-(j_3+m_3),$$
$$D=j_2-j_3+m_1+1,\quad E=j_2-j_1-m_3+1, \tag{A5}$$

and

$$A=-(j_2+m_2),\quad B=-(j_3-m_3),\quad C=-(j_2+j_3-j_1),$$
$$D=j_1-j_3-m_2+1,\quad E=j_1-j_2+m_3+1. \tag{A6}$$

B. Set of three ${}_4F_3(1)$'s for the 6j coefficient

$$A=e-a-b,\quad B=e-c-d,\quad C=f-a-c,\quad D=f-b-d,$$
$$E=-a-b-c-d-1,\quad F=e+f-a-d+1,\quad G=e+f-b-c+1, \tag{A7}$$

$$A=d-b-f,\quad B=a-b-e,\quad C=d-c-e,\quad D=a-c-f,$$
$$E=-b-c-e-f-1,\quad F=a+d-e-f+1,\quad G=a+d-b-c+1, \tag{A8}$$

and

$$A=b-a-e,\quad B=b-d-f,\quad C=c-a-f,\quad D=c-d-e,$$
$$E=-a-d-e-f-1,\quad F=b+c-e-f+1,\quad G=b+c-a-d+1. \tag{A9}$$

C. Set of four ${}_4F_3(1)$'s for the 6j coefficient

$$A=a+b+e+2,\quad B=a-c-f,\quad C=b-d-f,\quad D=e-c-d,$$
$$E=a+b-c-d+1,\quad F=a+e-d-f+1,\quad G=b+e-c-f+1, \tag{A10}$$

$$A=c+d+e+2,\quad B=c-a-f,\quad C=d-b-f,\quad D=e-a-b,$$
$$E=c+d-a-b+1,\quad F=c+e-b-f+1,\quad G=d+e-a-f+1, \tag{A11}$$

$$A=a+c+f+2,\quad B=c-d-e,\quad C=a-b-e,\quad D=f-b-d,$$
$$E=a+c-b-d+1,\quad F=a+f-d-e+1,\quad G=c+f-b-e+1, \tag{A12}$$

and

$$A=b+d+f+2,\quad B=b-a-e,\quad C=d-c-e,\quad D=f-a-c,$$
$$E=b+d-a-c+1,\quad F=b+f-c-e+1,\quad G=d+f-a-e+1. \tag{A13}$$

[1] K. Venkatesh, J. Math. Phys. **19**, 2060 (1978).
[2] T. Regge, Nuovo Cimento **10**, 544 (1958).
[3] K. Srinivasa Rao, T.S. Santhanam, and K. Venkatesh, J. Math. Phys. **16**, 1528 (1975).
[4] The subscript in the parameters indicate that the corresponding Regge symmetry is superposed on the tetrahedral symmetry. See T. Regge, Nuovo Cimento **11**, 116 (1959). For example,

$$\begin{Bmatrix} d_5 & f_5 & b_5 \\ a_5 & e_5 & c_5 \end{Bmatrix}$$

means that the Regge symmetry R5 is superposed on the tetrahedral symmetry

$$\begin{Bmatrix} d & f & b \\ a & e & c \end{Bmatrix}.$$

[5] K. Srinivasa Rao and K. Venkatesh, *Proc. Fifth Int. Colloquium on Group Theoretical Methods in Physics, Montreal*, 1976 (Academic, New York, 1977), p. 649.
[6] K. Venkatesh, J. Math. Phys. **19**, 1973 (1978).
[7] The complete listing of symmetry relations and homomorphisms have been given in K. Venkatesh, Ph.D. thesis, Mysore University, 1979, unpublished.

The most degenerate irreducible representations of the symplectic group

M. K. F. Wong
Fairfield University, Fairfield, Connecticut 06430

Hsin-Yang Yeh
Moorhead State University, Moorhead, Minnesota 56560

(Received 22 June 1979; accepted for publication 20 September 1979)

Matrix elements of the generators of the symplectic group Sp(2n) have been obtained for the most degenerate irreducible representations $(m_{2n},\dot{0})$ and $(\dot{m}_{2n})$. These are the only two cases for $n>2$ where the state labels according to the branching laws of Hegerfeldt give rise to an orthogonal set of basic vectors.

I. INTRODUCTION

It has been almost thirty years since Gel'fand and Zetlin[1] obtained the matrix elements of the generators of U(n) and O(n). However, the representation theory of the symplectic group for $n>2$ has not yet been worked out so far. The reason is that in the decomposition $\mathrm{Sp}(2n)\supset\mathrm{Sp}(2n-2)\cdots\supset\mathrm{Sp}(2)$ there is a multiplicity problem for the labeling of the states. It was shown by Hegerfeldt[2] that if one introduces the intermediate labels $m_{1,2n-1}, m_{2,2n-1}, \ldots, m_{n,2n-1}$, together with the "betweenness" conditions $m_{j+1,i+1}\leqslant m_{ji}\leqslant m_{j,i+1}$, $i=1,2,\ldots,2n-1$, $j=1,2,\ldots,[\frac{1}{2}i]$, $m_k^{2k+1}\geqslant 0, m_{ji}$, all positive integers or zero, then the states

$$|(m)\rangle = \left|\begin{array}{ccccc} m_{1,2n} & m_{2,2n} & \cdots\cdots & & m_{n,2n} \\ m_{1,2n-1} & m_{2,2n-1} & \cdots\cdots & & m_{n,2n-1} \\ & m_{1,2n-2} & \cdots\cdots & & m_{n-1,2n-2} \\ & m_{1,2n-3} & \cdots\cdots & & m_{n-1,2n-3} \\ & & m_{14} & m_{24} & \\ & & m_{13} & m_{23} & \\ & & m_{12} & & \\ & & m_{11} & & \end{array}\right\rangle , \tag{1.1}$$

will form a complete set of basis vectors.

However, it so happens that, for a general irreducible representation of Sp(2n), $n>2$, the state $|(m)\rangle$ in (1.1) is not a "pure" state. For example, in Sp(6) the states

$$\left|\begin{array}{ccccc} 3 & & 1 & & 0 \\ 3 & & 0 & & 0 \\ & 2 & & 0 & \\ & 2 & & 0 & \\ & & 2 & & \\ & & 2 & & \end{array}\right\rangle \quad \text{and} \quad \left|\begin{array}{ccccc} 3 & & 1 & & 0 \\ 2 & & 1 & & 0 \\ & 2 & & 0 & \\ & 2 & & 0 & \\ & & 2 & & \\ & & 2 & & \end{array}\right\rangle ,$$

are not orthogonal to each other. One must therefore choose a linear combination of these states in order to form an orthonormal set of basis vectors. This is the multiplicity problem which is very difficult to solve.

In view of this difficulty, we have to be content to look at the multiplicity free case first. There are two irreducible representations of Sp(2n) where an orthogonal set of basis vectors can be uniquely determined by (1.1). These are the most degenerate irreducible representations of the first kind $(m_{2n},\dot{0})$ and of the second kind $(\dot{m}_{2n})$. In this paper we obtain the matrix elements of the generators of Sp(2n) for these two representations.

The most degenerate representations of the first kind $(m_{2n},\dot{0})$ have been considered by Pajas and Rączka.[3] They used the Laplace–Beltrami operator and obtained the basis function for this irreducible representation. However, they did not use the Hergerfeldt decomposition. Consequently, they used a different set of labels, and the branching laws for these labels are not explicitly stated. In our work, we use the infinitesimal method exclusively and obtain the matrix elements of our generators in a much simpler form than Pajas and Rączka, especially for the generators $E_{\mp e_\mu \pm e_\mu}$. Let us also add that in our work we have also obtained the basis function in the form of the normalized lowering (raising) operators operating on the highest (lowest) weight.

In Sec. II, we present the representations of Sp(2) and Sp(4). These two can be completely solved because of the isomorphism between Sp(2) and SO(3) and Sp(4) and SO(5). In Secs. III and IV, we discuss the representations $(m_{2n},\dot{0})$ and

 0022-2488/80/040630-06$1.00

$(\dot{m}_{2n})$, respectively. In Sec. V, we discuss the general irreducible representations of Sp(2n). We show that the matrix elements of the generators can be obtained between some special states which are orthogonal and uniquely determined by (1.1).

II. REPRESENTATIONS OF Sp(2) AND Sp(4)

For the generators of Sp(2n), we follow the notation of Perelomov and Popov,[4] first derived by Racah.[5]

$$[G^a_b, G^c_d] = \delta^c_b G^a_d - \delta^a_d G^c_b + \epsilon^a \epsilon^b (\delta^{\bar{b}}_d G^c_{\bar{a}} - \delta^c_{\bar{a}} G^{\bar{b}}_d), \tag{2.1}$$

where

$$\epsilon^a = \begin{cases} 1 & a>0, \\ 0 \quad \text{for} & a=0, \\ -1 & a<0, \end{cases}$$

$$G^a_b = -\epsilon^a \epsilon^b G^{\bar{b}}_{\bar{a}}, \tag{2.2}$$

$$G^{a+}_b = G^b_a. \tag{2.3}$$

The indices range over $\bar{n}, \ldots, \bar{1}, 1, \ldots, n$.

It is clear from the commutation relations (2.1) that G^i_i corresponds to the diagonal generator of the Cartan subalgebra H_i. G^p_q $(q<p)$ corresponds to the lowering generator E_{-q+p}, G^q_p $(q<p)$ corresponds to the raising generator E_{q-p}, G^{-q}_p corresponds to the lowering generator E_{-q-p}, $G^q{}_{-p}$ corresponds to the raising generator E_{q+p}, G^{-p}_p corresponds to the lowering generator E_{-2p}, and $G^p{}_{-p}$ corresponds to the raising generator E_{2p}.

Thus in Sp(2), there are three generators G^1_1, $G^1{}_{-1}$, and G^{-1}_1. The commutation relations are

$$[G^1_1, \quad G^{-1}_1] = -2G^{-1}_1, \tag{2.4}$$

$$[G^1_1, \quad G^1{}_{-1}] = 2G^1{}_{-1}, \tag{2.5}$$

$$[G^1{}_{-1}, \quad G^{-1}_1] = 4G^1_1. \tag{2.6}$$

The second order Casimir invariant is

$$C_2 = 2G^1_1 G^1_1 + G^1{}_{-1} G^{-1}_1 + G^{-1}_1 G^1_1 = 2m^2_{12} + 4m_{12}. \tag{2.7}$$

The dimension of the representation is $m_{12}+1$.

We obtain the matrix elements of the generators as follows:

$$\left\langle \begin{matrix} m_{12} \\ m_{11} \end{matrix} \middle| G^1_1 \middle| \begin{matrix} m_{12} \\ m_{11} \end{matrix} \right\rangle = 2m_{11} - m_{12}, \tag{2.8}$$

$$\left\langle \begin{matrix} m_{12} \\ m_{11-1} \end{matrix} \middle| G^{-1}_1 \middle| \begin{matrix} m_{12} \\ m_{11} \end{matrix} \right\rangle = 2[m_{11}(m_{12} - m_{11} + 1)]^{1/2}, \tag{2.9}$$

$$\left\langle \begin{matrix} m_{12} \\ m_{11+1} \end{matrix} \middle| G^1{}_{-1} \middle| \begin{matrix} m_{12} \\ m_{11} \end{matrix} \right\rangle = 2[(m_{11}+1)(m_{12} - m_{11})]^{1/2}. \tag{2.10}$$

The connection between Sp(2) and SO(3) is as follows:

$$G^1_1 = 2J_z, \tag{2.11}$$

$$G^{-1}_1 = 2J_-, \tag{2.12}$$

$$G^1{}_{-1} = 2J_+, \tag{2.13}$$

$$m_{12} = 2j, \tag{2.14}$$

$$m_{11} = m + j. \tag{2.15}$$

For Sp(4) we use the fact that the Lie algebras of Sp(4) and SO(5) are identical. Using Hecht's notation,[6] we classify SO(5) $\supset$ SO(3) $\otimes$ SO(3) by the labels J_m, Λ_m, J, Λ, M_J, M_Λ with

$$J = J_m - \tfrac{1}{2}n - \tfrac{1}{2}m, \quad 0 \leqslant n \leqslant 2(J_m - \Lambda_m), \tag{2.16}$$

$$\Lambda = \Lambda_m + \tfrac{1}{2}n - \tfrac{1}{2}m, \quad 0 \leqslant m \leqslant 2\Lambda_m. \tag{2.17}$$

Then it is easily seen that

$$m_{14} = 2J_m, \tag{2.18}$$

$$m_{24} = 2\Lambda_m, \tag{2.19}$$

$$m_{13} = 2\Lambda_m + n = J_m + \Lambda_m + \Lambda - J, \tag{2.20}$$

$$m_{23} = m = J_m + \Lambda_m - J - \Lambda, \tag{2.21}$$

$$m_{12} = 2L,$$

where L is the angular momentum in the direct product

$$J \otimes \Lambda \to L, \tag{2.22}$$

and

$$m_{11} = M + L, \tag{2.23}$$

where $M = M_J + M_\Lambda$. Thus, e.g., the state

$$\left| \begin{matrix} 3 & & 1 \\ 3 & & 0 \\ & 2 & \\ & 2 & \end{matrix} \right\rangle$$

corresponds to $J_m = \frac{3}{2}$, $\Lambda_m = \frac{1}{2}$, $J = \frac{1}{2}$, $\Lambda = \frac{3}{2}$, and $L = 1$, i.e., $J \otimes \Lambda = \frac{1}{2} \otimes \frac{3}{2} \to 1$,

$$M = 1, \quad \begin{matrix} M_J = \frac{1}{2}, & M_\Lambda = \frac{1}{2} \\ \text{or } M_J = -\frac{1}{2}, & M_\Lambda = \frac{3}{2}. \end{matrix}$$

Then all the matrix elements of the generators of Sp(4) can be calculated using Hecht's results and angular momentum coupling schemes.

Note that Hecht's decomposition is Sp(4) $\supset$ Sp(2) $\otimes$ Sp(2) and is different from the Hegerfeldt decomposition (1.1). In the following two sections we use the Hegerfeldt decomposition to obtain the most degenerate representations. Therefore, the results will be different from Hecht's when applied to Sp(4).

III. THE MOST DEGENERATE IRREDUCIBLE REPRESENTATIONS OF THE FIRST KIND

It is obvious that for the irreducible representation $(m_{2n}, \dot{0})$, the states are specified by the labels $m_{1,2n} m_{1,2n-1}$, ..., $m_{1,1}$ only. We shall denote this state by

$$\left| \begin{matrix} m_{2n} & \\ m_{2n-1} & (\dot{0}) \\ \vdots & \\ m_1 & \end{matrix} \right\rangle.$$

We now follow Mickelsson[7] and denote the semimaximal vector to be

$$|\mathrm{sm}\rangle = \left| \begin{matrix} m_{2n} & \\ m_{2n-1} & (\dot{0}) \\ m_{2n-2} & \\ \max & \end{matrix} \right\rangle. \tag{3.1}$$

The diagonal generators G_i^i have the following matrix elements:

$$G_i^i|(m)\rangle = \left(\sum_{j=1}^{i} m_{j,2i-1} - \sum_{j=1}^{i-1} m_{j,2i-2} - \sum_{j=1}^{i} m_{j,2i}\right) \times |(m)\rangle = H_i|(m)\rangle. \tag{3.2}$$

There are two kinds of lowering (raising) operators. The first one lowers (raises) all the numbers from m_{2n-1} down to m_1 by 1. The second one lowers (raises) all the numbers from m_{2n-1} down to m_1 by 1. Let us denote these operators by $L_n^{-1}(R_{n-1}^{\ n})$ and $L_1^n(R_n^{\ 1})$, respectively. Then the commutation relations will be as follows:

$$[G_1^1, L_n^{-1}] = -L_n^{-1}, \quad [G_1^1, R_{n-1}^{\ n}] = R_n^{-1}, \tag{3.3}$$

$$[G_n^n, L_n^{-1}] = -L_n^{-1}, \quad [G_n^n, R_{n-1}^{\ n}] = R_{n-1}^{\ n}, \tag{3.4}$$

$$[G_{b+1}^b, L_n^{-1}(R_n^{-1})]|\mathrm{sm}\rangle = 0, \quad b = 1,2,\ldots,n-2, \tag{3.5}$$

$$[G_{n-n+1}^{n-1}, L_n^{-1}(R_n^{-1})]|\mathrm{sm}\rangle = 0, \tag{3.6}$$

$$[G_1^1, L_1^n] = -L_1^n, \quad [G_1^1, R_n^1] = R_n^1, \tag{3.7}$$

$$[G_n^n, L_1^n] = L_1^n, \quad [G_n^n, R_n^1] = -R_n^1, \tag{3.8}$$

$$[G_{b+1}^b, L_1^n(R_1^n)]|\mathrm{sm}\rangle = 0, \tag{3.9}$$

$$[G_{n-n+1}^{n-1}, L_1^n(R_n^1)]|\mathrm{sm}\rangle = 0. \tag{3.10}$$

It is easy to check that

$$L_n^{-1} = G_n^{-1}, \quad R_{n-1}^{\ n} = G_{n-1}^n, \tag{3.11}$$

$$L_1^n = G_1^n, \quad R_n^1 = G_n^1. \tag{3.12}$$

Hence a general state $|(m)\rangle$ can be obtained as follows:

$$|(m)\rangle = N(G_1^{-1})^{m_2 - m_1}\cdots(G_1^n)^{m_{2n} - m_{2n-2}} \times (G_n^{-1})^{m_{2n} - m_{2n-1}}|M\rangle, \tag{3.13}$$

where $M =$ maximal state, with $m_i = m_{2n}$ for $i = 1,2,\ldots, n-1$ and N is the normalization constant of the generators. The normalization constants are connected with the matrix elements of the generators. Below we shall obtain both the normalization constants and the matrix elements of the generators.

It is clear that for the normalization constants we only have to obtain the matrix elements of the generators G_1^n and G_n^{-n}, acting on the semimaximal state. We show below how the matrix elements of G_1^n are obtained.

Define

$$\left\langle \begin{matrix} m_{2n} \\ m_{2n-1} \\ m_{2n-2}+\alpha-1 \\ m_{2n-2}+\alpha-1 \\ \max \end{matrix} \right| (\dot{0}) \left| G_1^n \right| \left. \begin{matrix} m_{2n} \\ m_{2n-1} \\ m_{2n-2}+\alpha \\ m_{2n-2}+\alpha \\ \max \end{matrix} \right| (\dot{0}) \Bigg\rangle$$

$$= X(\alpha). \tag{3.14}$$

Now we use the commutation relation

$$[G_n^1, G_1^n] = H_1 - H_n. \tag{3.15}$$

From (3.15) we obtain

$$X^2(\alpha) - X^2(\alpha+1) = m_{2n} - 2m_{2n-1} + 2m_{2n-1} + 2\alpha. \tag{3.16}$$

We know, however,

$$X^2(m_{2n-1} - m_{2n-2} + 1) = 0. \tag{3.17}$$

Therefore, summing over (3.16) from $\alpha = 0$ to $m_{2n-1} - m_{2n-2}$, we obtain

$$X^2(0) = (m_{2n-1} - m_{2n-2} + 1)(m_{2n} - m_{2n-1} + m_{2n-2}). \tag{3.18}$$

By a similar procedure we obtain the matrix elements of G_n^{-n}

$$\left\langle \begin{matrix} m_{2n} \\ m_{2n-1}-1 \\ m_{2n-2} \\ \max \end{matrix} \right| (\dot{0}) \left| G_n^{-n} \right| \left. \begin{matrix} m_{2n} \\ m_{2n-1} \\ m_{2n-2} \\ \max \end{matrix} \right| (\dot{0}) \Bigg\rangle$$

$$= 2[(m_{2n-1} - m_{2n-2})(m_{2n} - m_{2n-1} + 1)]^{1/2}. \tag{3.19}$$

Next we show how to obtain the matrix elements of the generator G_{n-1}^n. First define

$$\left\langle \begin{matrix} m_{2n} \\ m_{2n-1} \\ m_{2n-2}+\alpha-1 \\ m_{2n-2}+\alpha-1 \\ m_{2n-4} \\ \max \end{matrix} \right| (\dot{0}) \left| G_{n-1}^n \right| \left. \begin{matrix} m_{2n} \\ m_{2n-1} \\ m_{2n-2}+\alpha \\ m_{2n-2}+\alpha \\ m_{2n-4} \\ \max \end{matrix} \right| (\dot{0}) \Bigg\rangle$$

$$= X_1(\alpha). \tag{3.20}$$

Then use the commutation relation

$$[G_{n-1}^n, G_n^{n-1}] = H_n - H_{n-1}. \tag{3.21}$$

From (3.21), we obtain

$$X_1^2(\alpha+1) - X_1^2(\alpha) = 2m_{2n-1} - m_{2n} - 2m_{2n-2} + m_{2n-4} - 2\alpha. \tag{3.22}$$

Summing over (3.22) from $\alpha = 0$ to $\alpha = m_{2n-1} - m_{2n-2}$, we obtain

$$X_1^2(0) = (m_{2n-1} - m_{2n-2} + 1)(m_{2n} - m_{2n-1} + m_{2n-2} - m_{2n-4}). \tag{3.23}$$

Next we wish to find the matrix elements of G_{n-1}^n acting on a general state, i.e.,

$$\left\langle \begin{matrix} m_{2n} \\ m_{2n-1} \\ m_{2n-2}-1 \\ m_{2n-3}-1 \\ m_{2n-4} \\ \max \end{matrix} \right| (\dot{0}) \left| G_{n-1}^n \right| \left. \begin{matrix} m_{2n} \\ m_{2n-1} \\ m_{2n-2} \\ m_{2n-3} \\ m_{2n-4} \\ \max \end{matrix} \right| (\dot{0}) \Bigg\rangle. \tag{3.24}$$

First we use the lowering operator G_{n-1}^{-n+1} on the state in (3.20) for $\alpha = 0$, $m_{2n-2} - m_{2n-3}$ times. We then have

$$\left|\begin{matrix} m_{2n} & (\dot{0}) \\ m_{2n-1} & \\ m_{2n-2} & \\ m_{2n-3} & \\ m_{2n-4} & \\ \max & \end{matrix}\right\rangle = \frac{1}{2^{m_{2n-2}-m_{2n-3}} \prod_{\alpha=m_{2n-3}+1}^{m_{2n-2}} [(\alpha - m_{2n-4})(m_{2n-2} - \alpha + 1)]^{1/2}}$$

$$\times (G_{n-1}^{-n+1})^{m_{2n-2}-m_{2n-3}} \left|\begin{matrix} m_{2n} & (\dot{0}) \\ m_{2n-1} & \\ m_{2n-2} & \\ m_{2n-2} & \\ m_{2n-4} & \\ \max & \end{matrix}\right\rangle. \tag{3.25}$$

Substituting (3.25) into (3.24) and commuting G_{n-1}^{n} over to the right, we find that the value of (3.24) is

$$[(m_{2n-1} - m_{2n-2} + 1)(m_{2n} - m_{2n-1} + m_{2n-2} - m_{2n-4})]^{1/2}$$
$$\times \prod_{\alpha=m_{2n-3}}^{m_{2n-2}-1} [(\alpha - m_{2n-4})(m_{2n-2} - \alpha)]^{1/2} \left(\prod_{\alpha=m_{2n-3}+1}^{m_{2n-2}} [(\alpha - m_{2n-4})(m_{2n-2} - \alpha + 1)]^{1/2} \right)^{-1}$$
$$= \frac{[(m_{2n-1} - m_{2n-2} + 1)(m_{2n} - m_{2n-1} + m_{2n-2} - m_{2n-4})(m_{2n-3} - m_{2n-4})]^{1/2}}{[(m_{2n-2} - m_{2n-4})]^{1/2}}. \tag{3.26}$$

Comparing our results with those of Pajas and Rączka, we find that the matrix elements of H_p and $E_{\pm 2\epsilon_p}$ can be made to agree with each other provided one makes the following identification:

$$m - \bar{m} = 2m_{2p-1} - m_{2p} - m_{2p-2}, \tag{3.27}$$

$$l = L_p - L_{p-1} = m_{2p} - m_{2p-2}. \tag{3.28}$$

However, for the other generators, Pajas and Rączka obtained a very complicated result, whereas our results, in the form of (3.26), is very simple.

IV. THE MOST DEGENERATE IRREDUCIBLE REPRESENTATIONS OF THE SECOND KIND

For the irreducible representations $(\dot{m}_{2n})$, the states are specified by the labels $m_{n,2n} m_{n,2n-1}, m_{n-1,2n-2}, m_{n-1,2n-3}, \ldots, m_{1,1}$. We shall denote this state by

$$\left|\begin{matrix} (\dot{m}_{2n}) & m_{2n} \\ & m_{2n-1} \\ & m_{2n-2} \\ & \vdots \\ & m_1 \end{matrix}\right\rangle. \tag{4.1}$$

The lowest state is

$$\left|\begin{matrix} (\dot{m}_{2n}) & m_{2n} \\ & 0 \\ & 0 \\ & \vdots \\ & 0 \end{matrix}\right\rangle.$$

To obtain the general state from the lowest state, we use the raising operators. There are two kinds of raising operators. The first one, denoted by R^{i}_{-i}, raises m_{2i-1} by 1, but does not change any other labels. The second one, denoted by R_i^{i-1}, raises both m_{2i-2}, and m_{2i-3} by 1. The commutation relations are

$$[G_i^i, R^{i}_{-i}] = 2R^{i}_{-i}. \tag{4.2}$$

R^{i}_{-i} commutes with all the generators of the group Sp$(2i-2)$.

$$[G_i^i, R_i^{i-1}] = -R_i^{i-1}, \tag{4.3}$$

$$[G_{i-1}^{i-1}, R_i^{i-1}] = R_i^{i-1}. \tag{4.4}$$

R_i^{i-1} commutes with all the generators of the group Sp($2i-4$). It is clear that

$$R_i^{i-1} = G_i^{i-1},\quad R^i{}_{-i} = G^i{}_{-i}. \tag{4.5}$$

The matrix elements of the generators are completely determined if we can find those for $G^n{}_{-n}$ and G_n^{n-1}. By a procedure similar to the one used in the previous section, we find

$$\left\langle \begin{matrix} & m_{2n} \\ (\dot m_{2n}) & m_{2n-1}-1 \\ & m_{2n-2} \\ & \vdots \end{matrix} \middle| G_n^{-n} \middle| \begin{matrix} & m_{2n} \\ (\dot m_{2n}) & m_{2n-1} \\ & m_{2n-2} \\ & \vdots \end{matrix} \right\rangle = 2[m_{2n-1}(m_{2n-2}-m_{2n-1}+1)]^{1/2}, \tag{4.6}$$

$$\left\langle \begin{matrix} & m_{2n} \\ (\dot m_{2n}) & m_{2n-1} \\ & m_{2n-2}-1 \\ & m_{2n-2}-1 \\ & m_{2n-4} \\ & \vdots \end{matrix} \middle| G_{n-1}^{n} \middle| \begin{matrix} & m_{2n} \\ (\dot m_{2n}) & m_{2n-1} \\ & m_{2n-2} \\ & m_{2n-2} \\ & m_{2n-4} \\ & \vdots \end{matrix} \right\rangle$$

$$= [(m_{2n}-m_{2n-2}+1)]^{1/2}\,[(2m_{2n}-2m_{2n-1}+m_{2n-2}-m_{2n-4})]^{1/2}. \tag{4.7}$$

Then we have

$$\left| \begin{matrix} (\dot m_{2n}) & m_{2n} \\ & m_{2n-1} \\ & m_{2n-2} \\ & m_{2n-3} \\ & m_{2n-4} \\ & \vdots \end{matrix} \right\rangle = \left(2^{m_{2n-2}-m_{2n-3}} \prod_{\alpha=m_{2n-3}+1}^{m_{2n-2}} [\alpha(m_{2n-4}-\alpha+1)]^{1/2} \right)^{-1}$$

$$\times (G_{n-1}^{-n+1})^{m_{2n-2}-m_{2n-3}} \left| \begin{matrix} (\dot m_{2n}) & m_{2n} \\ & m_{2n-1} \\ & m_{2n-2} \\ & m_{2n-2} \\ & m_{2n-4} \\ & \vdots \end{matrix} \right\rangle. \tag{4.8}$$

Now we wish to find

$$\left\langle \begin{matrix} (\dot m_{2n}) & m_{2n} \\ & m_{2n-1} \\ & m_{2n-2}-1 \\ & m_{2n-3}-1 \\ & m_{2n-4} \end{matrix} \middle| G_{n-1}^{n} \middle| \begin{matrix} (\dot m_{2n}) & m_{2n} \\ & m_{2n-1} \\ & m_{2n-2} \\ & m_{2n-3} \\ & m_{2n-4} \end{matrix} \right\rangle. \tag{4.9}$$

Substituting (4.8) into (4.9) and commuting G_{n-1}^{n} over to the right, we find that the value of (4.9) is

$$[(m_{2n}-m_{2n-2}+1)(2m_{2n}-2m_{2n-1}+m_{2n-2}-m_{2n-4})]^{1/2}$$

$$\times \prod_{\alpha=m_{2n-3}}^{m_{2n-2}-1} [(\alpha)(m_{2n-4}-\alpha+1)]^{1/2} \left(\prod_{\alpha=m_{2n-3}+1}^{m_{2n-2}} [(\alpha)(m_{2n-4}-\alpha+1)]^{1/2} \right)^{-1}$$

$$= [(m_{2n}-m_{2n-2}+1)(2m_{2n}-2m_{2n-1}+m_{2n-2}-m_{2n-4})]^{1/2}$$

$$\times [m_{2n-3}(m_{2n-4}-m_{2n-3}+1)]^{1/2}\,[m_{2n-2}(m_{2n-4}-m_{2n-2}+1)]^{-1/2}. \tag{4.10}$$

V. SPECIAL CASES FOR THE MATRIX ELEMENTS OF THE GENERATORS FOR A GENERAL IRREDUCIBLE REPRESENTATION

For a general irreducible representation of Sp($2n$), $n>2$, the states labeled by (1.1) are in general not orthogonal to each other. There are, however, special cases, where two orthogonal states can be uniquely determined. For example, the highest weight state is always uniquely determined. Therefore, the matrix elements of these generators which are primitive roots, acting on the highest weight, can always be uniquely determined. Aside from the highest weight state, there are a few other

states where the matrix elements of the generators can also be determined. We shall give some examples in Sp($2n$) to illustrate the point.

For the highest state weight we find, e.g.,

$$\left\langle \begin{matrix} m_{1,2n} & & & & \cdots & m_{n,2n} \\ m_{1,2n} & & & & \cdots & m_{n,2n} \\ & m_{1,2n} & \cdots & m_{n-2,2n} & & m_{n-1,2n}-1 \\ & m_{1,2n} & \cdots & m_{n-2,2n} & & m_{n-1,2n}-1 \end{matrix} \right| G^{n}_{n-1} \left| M \right\rangle = (m_{n-1,2n} - m_{n,2n})^{1/2}, \tag{5.1}$$

$$\left\langle \begin{matrix} m_{1,2n} & \cdots & m_{n-1,2n} & m_{n,2n} \\ m_{1,2n} & \cdots & m_{n-1,2n} & m_{n,2n}-1 \end{matrix} \right| G^{-n}_{n} \left| M \right\rangle = 2(m_{n,2n})^{1/2}. \tag{5.2}$$

For other special states, we find, e.g.,

$$\left\langle \begin{matrix} m_{1,2n} & \cdots & m_{n,2n} \\ m_{1,2n} & \cdots & m_{n,2n} \\ m_{1,2n} & \cdots & m_{n-1,2n-1}-1 \\ m_{1,2n} & \cdots & m_{n-1,2n-1}-1 \end{matrix} \right| G^{n}_{n-1} \left| \begin{matrix} m_{1,2n} & \cdots & m_{n-1,2n} & m_{n,2n} \\ m_{1,2n} & \cdots & m_{n-1,2n} & m_{n,2n} \\ m_{1,2n} & \cdots & m_{n-2,2n} & m_{n-1,2n-1} \\ m_{1,2n} & \cdots & m_{n-2,2n} & m_{n-1,2n-1} \end{matrix} \right\rangle$$
$$= [(m_{n-1,2n-1} - m_{n,2n})(m_{n-1,2n} - m_{n-1,2n-1} + 1)]^{1/2}, \tag{5.3}$$

$$\left\langle \begin{matrix} m_{1,2n} & \cdots & m_{n-1,2n} & m_{n,2n} \\ m_{1,2n} & \cdots & m_{n-1,2n} & m_{n,2n}-1 \end{matrix} \right| G^{-n}_{n} \left| \begin{matrix} m_{1,2n} & \cdots & m_{n-1,2n} & m_{n,2n} \\ m_{1,2n} & \cdots & m_{n-1,2n} & m_{n,2n-1} \end{matrix} \right\rangle$$
$$= 2[m_{n,2n-1}(m_{n,2n} - m_{n,2n-1} + 1)]^{1/2}. \tag{5.4}$$

ACKNOWLEDGMENT

We wish to thank Moorhead State University for financial support of this work.

[1]I.M. Gel'fand and M.L. Zetlin, Dokl. Akad. Nauk. SSSR **71**, 875, 1071 (1950).
[2]G.C. Hegerfeldt, J. Math. Phys. **8**, 1195 (1967). See also D.P. Zhelobenko, Usp. Mat. Nauk. **17**, 27 (1962) [Russ. Math. Surv. **17**, 1 (1962)].
[3]P. Pajas and R. Rączka, J. Math. Phys. **9**, 1188 (1968).
[4]A.M. Perelomov and V.S. Popov, Yad. Fiz. **3**, 1127 (1966) [Sov. J. Nucl. Phys. **3**, 819 (1966)].
[5]G. Racah, Ergeb. Exact. Naturw. **37**, 28 (1965).
[6]K.T. Hecht, Nucl. Phys. **63**, 177 (1965).
[7]J. Mickelsson, Rep. Math. Phys. **3**, 193 (1972).

The Lorentz group in the oscillator realization. II. Integral transforms and matrix elements of SO(2,1)

Debabrata Basu and D. Mitra
Department of Physics, Indian Institute of Technology, Kharagpur 721302, India

(Received 26 January 1979)

We use the boson realization to investigate the connection between integral transforms and matrix elements of SO(2,1) in a continuous basis for both the principal and discrete series of representations. Matrices in the SO(1,1) basis are shown to be related to an integral transform of Mellin–Barnes type.

1. INTRODUCTION

Recently there has been considerable interest in the study of a class of new realizations of the unitary irreducible representations (UIR's) of the three dimensional Lorentz Group SO(2,1). In contrast to the canonical realizations of Gel'fand[1] *et al.* and Bargmann[2] in which the group acts transitively as a group of point transformations in a function space, the group action in this construction,[3,4] is an integral transform,

$$T_g f(x) = \int K_g(x,x') f(x')\, dx'.$$

This realization which arouse out of the works of Moshinsky,[3] Wolf,[4] and coworkers on the role of generalized canonical transforms in quantum mechanics turns out to be equivalent to the boson representations used by Holman and Biedenharn[5] and others.[6]

The finite transformation matrices in a continuous noncompact basis, in this construction, will appear naturally as the appropriate integral transform of the Kernel $K(x,x')$. The object of this short note is to show that a given subgroup reduction corresponds, in general, to integral transform of a specific type. The reduction in SO(1,1) basis, for example, yields the continuous basis matrix elements of Barut and Phillips[7] as Mellin transforms of the kernel. The horocyclic basis which will be treated elsewhere likewise corresponds to Hankel transforms. Our method, which is similar to that of a previous paper,[8] (I) is quite general and covers both the principal and discrete[9] series of UIR's of the group.

2. MELLIN TRANSFORM OF THE KERNEL AND MATRIX ELEMENT

A. Principal series of representations:

Since for the principal series of UIR's the SO(1,1) subspace is doubly degenerate, the Hilbert space, in the notation of I, consists of pairs:

$$f = \begin{pmatrix} f_1(r) \\ f_2(r) \end{pmatrix}. \tag{2.1}$$

The action of an element of the group is then given by

$$T_{g(\alpha)} f(r) = \int K(r,r';\alpha) f(r')\, r'\, dr', \tag{2.2}$$

where the kernel K is a two by two matrix. The elements of K for pure space rotations are given by[6]

$$\begin{aligned} K_{12} &= \exp[i(x-x')\cot(\alpha/2)]K_{21} \\ &= -\frac{\cos\pi\nu/2}{\pi \sin\alpha/2} \exp[(i/2)(x-x')\cot\alpha/2] K_\nu(a), \\ K_{11} &= \exp[i(x+x')\cot(\alpha/2)]K_{22} \\ &= \frac{i\exp[(i/2)(x+x')\cot\alpha/2]}{\sin(\alpha/2)} [e^{-i\pi\nu/2} H_\nu^{(2)}(a) \\ &\quad - e^{i\pi\nu/2} H_\nu^{(1)}(a)], \end{aligned} \tag{2.3}$$

where

$$\nu = 2is, \quad a = (xx')^{1/2} \csc\alpha/2, \quad x = r^2. \tag{2.4}$$

The normalized SO(1,1) bases consist of the vectors

$$f_\mu = \frac{1}{\sqrt{2\pi}} \begin{pmatrix} x^{i\mu - 1/2} \\ \epsilon x^{i\mu - 1/2} \end{pmatrix}, \quad \epsilon = \pm 1, \tag{2.5}$$

which span a pair of mutually orthogonal subspaces H_ϵ satisfying

$$(f_{\mu'}^{\epsilon'}, f_\mu^{\epsilon}) = \delta_{\epsilon\epsilon'}\, \delta(\mu - \mu'). \tag{2.6}$$

The finite transformation matrices

$$D_{\mu'\mu}^{\epsilon'\epsilon}(\alpha) = (f_{\mu'}^{\epsilon'}, \quad T_{g(\alpha)} f_\mu^{\epsilon}), \tag{2.7}$$

in this basis can therefore be identified as the Mellin transforms of K.

Following I this can be easily obtained by first expanding K_{ij} in powers of x'. The series thus obtained can be recast in the form of an integral over the pure imaginary axis. Thus, for example,

$$\begin{aligned} &K_{12}(r,r',\alpha) \\ &= -\frac{\cos\pi\nu/2}{\pi \sin\alpha/2} \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} dx \left(\frac{ix'}{2} \cot\frac{\alpha}{2} \right)^{-z-1/2} \\ &\quad \times \Gamma(z+\nu/2+1/2)\Gamma(z-\nu/2+1/2)(X_\nu + X_{-\nu}), \end{aligned} \tag{2.8}$$

where

$$\begin{aligned} X_\nu &= \frac{\Gamma(-\nu)}{\Gamma(z-\nu/2+1/2)} (-ix\csc\alpha)^{\nu/2} \\ &\quad \times \exp[(1/2)x\cot\alpha/2] \\ &\quad \times F(z+\nu/2+1/2; \nu+1; -ix\csc\alpha). \end{aligned} \tag{2.9}$$

 0022-2488/80/040636-02$1.00

The Taylor expansion of $X_{\pm\nu}$ shows that the two infinite series can be identified as the sum of residues at $y = \mp\nu/2 - 1/2 - n$, $n = 0,1,2\cdots$ of the meromorphic function

$$\chi(y) = (-ix\csc\alpha)^{-y-1/2}(\cos^2\alpha/2)^{z-y} \times \frac{\Gamma(y+\nu/2+1/2)\Gamma(y-\nu/2+1/2)}{\Gamma(z+y+1)} \times {}_2F_1\left[\begin{matrix} z+\nu/2+1/2, & z-\nu/2+1/2 \\ & z+y+1 \end{matrix}; \sin^2\alpha/2\right], \quad (2.10)$$

which vanishes rapidly at $|y| \to \infty$ for $\mathrm{Re}\, y < 0$. We therefore obtain

$$X_\nu + X_{-\nu} = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} \chi(y)\, dy, \quad (2.11)$$

and

$$K_{12} = \frac{-2\cos\pi\nu/2}{\pi}\left(\frac{1}{2\pi}\right)^3 \int_{-i\infty}^{i\infty}\int_{-i\infty}^{i\infty} dz\, dy \left(\frac{ix'}{2}\right)^{-z-1/2} \times (-ix/2)^{-y-1/2}\Gamma(z+\nu/2+1/2) \times \Gamma(z-\nu/2+1/2)\Gamma(y+\nu/2+1/2) \times \Gamma(y-\nu/2+1/2)\ \sin\pi(z+y) F_1(z,y), \quad (2.12)$$

where

$$F_1(z,y) = \Gamma(-z-y)\,(\cos\alpha/2)^{z-y}\,(\sin\alpha/2)^{z+y} \times {}_2F_1\left[\begin{matrix} z+\nu/2+1/2, & z-\nu/2+1/2 \\ & z+y+1 \end{matrix}; \sin^2\alpha/2\right]. \quad (2.13)$$

Similarly

$$K_{11} = \frac{\csc\pi\nu/2}{\pi}\left(\frac{1}{2\pi}\right)^3 \int_{-i\infty}^{i\infty}\int_{-i\infty}^{i\infty} dz\, dy \left(-\frac{ix'}{2}\right)^{-z-1/2} \times \left(-\frac{ix}{2}\right)^{-y-1/2}\Gamma(z+\nu/2+1/2)\,\Gamma(z-\nu/2+1/2)\Big\{\Gamma(y+\nu/2+1/2)\Gamma(y-\nu/2+1/2) \times [e^{i\pi\nu/2}\cos\pi(z+\nu/2)\cos\pi(y+\nu/2) - e^{i\pi\nu/2} \times \cos\pi(z-\nu/2)\cos\pi(y-\nu/2)]\, F_1(z,y) - \frac{2i\pi^2\sin\pi\nu/2}{\Gamma(z+\nu/2+1/2)\Gamma(z-\nu/2+1/2)} F_2(z,y)\Big\}, \quad (2.14)$$

where

$$F_2(z,y) = \Gamma(z+y)(\cos\alpha/2)^{z-y}(-\sin\alpha/2)^{-z-y} \times {}_2F_1\left[\begin{matrix} -y+\nu/2+1/2, & -y-\nu/2+1/2 \\ & -z-y+1 \end{matrix}; \sin^2\alpha/2\right]. \quad (2.15)$$

The corresponding integral representations for K_{21} and K_{22} can be immediately written down by symmetry. The matrix elements $D^{\epsilon'\epsilon}_{\mu'\mu}$ which are simple linear combinations of the Mellin transforms of K_{ij} can then be read off from these equations.

B. Discrete representations:

Since for the discrete UIR's, the procedure is quite similar to that discussed above we present only the outlines. In this case the SO(1,1) bases consist of the single set of basis vectors

$$f_\mu(z) = \frac{1}{\sqrt{\pi}} x^{i\mu-1/2}, \quad (2.16)$$

and the kernel K is given by

$$K(x,x';\alpha) = \frac{e^{i\pi k}}{\sin\alpha/2} \times \exp[-(1/2)(x+x')\cot\alpha/2]\, J_{2k-1}(a). \quad (2.17)$$

The desired integral representation for K of Mellin Barnes type can be obtained easily by following the procedure as outlines above. Thus,

$$K(x,x=;\alpha) = -\frac{e^{i\pi/2}(\sec^2\alpha/2)^{k-1/2}}{\Gamma(2k)\sin\alpha/2}\left(\frac{1}{2\pi}\right)^2 \times \int_{-i\infty}^{i\infty}\int_{-i\infty}^{i\infty} dz'\, dz \left(\frac{ix'}{2}\cot\alpha/2\right)^{-z'-1/2} \times \left(\frac{ix}{2}\cot\alpha/2\right)^{-z-1/2}\Gamma(k+z)\Gamma(k+z')\,{}_2F_1 \times \left[\begin{matrix} k+z, & k+z' \\ & 2k \end{matrix}; \frac{1}{\cos^2\alpha/2}\right]. \quad (2.18)$$

The continuous basis matrix elements can once again be read off from this formula.

[1] I.M. Gel'fand, M.I. Graev, and N.Ya. Vilenkin, *Generalized Functions* (Academic, New York, 1966), Volume 5.
[2] V. Bargmann, Ann. Math. **48**, 568 (1947).
[3] M. Moshinsky and C. Quesne, J. Math. Phys. **12**, 1772–80 (1971); M. Moshinsky, T.H. Seligman, and K.B. Wolf, J. Math. Phys. **13**, 901 (1972); M. Moshinsky, SIAM J. Appl. Math **25**, 193 (1973); P. Kramer, Moshinsky, and T.H. Seligman *Group Theory and Its Applications*, edited by E.M. Loebl (Academic, New York, 1975), Volume III.
[4] K.B. Wolf, J. Math. Phys. **15**, 1295, 2102(1974); **17**, 601 (1976); Integral Transforms (Plenum, to be published); C.P Boyer and K.B. Wolf, J. Math. Phys. **16**, 1493, 2215 (1975).
[5] W.J. Holman and L.C. Biedenharn, Ann. Phys. **39**, 1 (1966).
[6] N. Mukunda and B. Radhakrishnan, J. Math. Phys. **14**, 254 (1973); E.C. Kalnins and W. Willer Jr., J. Math. Phys. **15**, 1728 (1974).
[7] A.O. Barut and E.C. Phillips, Commun. Math. Phys. **8**, 52 (1968).
[8] D. Basu, J. Math. Phys. **19**, 1667 (1978).
[9] A similar result has recently been obtained by C.P. Boyer and K.B. Wolf, Revis, Mex. Fis **25**, 31 (1976), for the discrete UIR's of SO(2,1). We would like to thank Professor K.B. Wolf for bringing this to our notice.

Permutational symmetry of many particle states

C. R. Sarma and G. G. Sahasrabudhe
Indian Institute of Technology, Bombay-400 076, India

(Received 30 January 1979)

The symmetry of the Nth rank tensor basis for an irreducible representation of $U(n)$ under the operations of the permutation group has been investigated. It has been found that symmetrized linear combinations of the elements of the matrix algebra of S_N lead to a tensor basis for $U(n)$ yielding the same matrix elements as the Gel'fand–Tsetlin basis for the generators $E_{i,i+1}$ of $U(n)$. Based on these developments an algorithm has been developed for directly determining the matrix elements of the generators $E_{ij}(j \neq i \pm 1)$ of $U(n)$ using a pattern calculus.

1. INTRODUCTION

Over the past few years considerable effort has gone into realizing basis states spanning the irreducible representations (IR's) of U(n).[1-8] The study of many-particle systems using these basis states leads naturally to the consideration of transformations induced in them by the generators of the permutation group S_N on the particle coordinates. In view of this, there have been attempts to realize these states by applying Young operators of S_N[7,8] to reducible tensor products of the basis orbitals spanning the fundamental representation of U(n). However, the nonhermiticity of Young idempotents used, led to nonorthogonality of the generated bases as pointed out by Baird and Biedenharn[2] and Patterson and Harter.[8] This nonorthogonality problem prevented an extensive application of such a method to many-particle states till quite recently. An alternative approach based on the generalized hook concept was used by Biedenharn and Ciftan[3] and Ciftan[4] to realize explicit forms for the semimaximal states of U(n). This approach proved relatively easier to use than the one based exclusively on the process of lowering from the highest weight state (HWS), developed by Baird and Biedenharn[1,2] and Nagel and Moshinsky.[5]

A further alternative to the use of the semimaximal states approach was suggested recently by Lezuo.[7] Using essentially idempotent elements of the associative algebra of S_N, he was able to realize a projected tensor basis for SU(3). Such a basis was found to differ at most by a phase factor from the corresponding Gel'fand basis.[6,9] In a more recent note Patterson and Harter[8,10] were able to demonstrate that such considerations could be generalized to U(n). However, to the best of our knowledge, an explicit determination of the normalization and other factors involved in such a projected tensor basis state has not been carried out.

In the recent note we have used nonstandard irreducible representations[12] to obtain a set of operators defining a matrix algebra of S_N,[11,12] and applied them to an Nth rank tensor product of single particle basis orbitals to obtain a properly normalized basis set for the IR's of U(n). The procedure for obtaining these tensor basis sets is outlined in Sec. 2. Using these basis sets it has also been shown that the matrix elements of the elementary generators $E_{i-1,i}$ of U(n) agree with those resulting from the use of Gel'fand bases.[2,6,9] In Sec. 3, we have developed a pattern calculus for calculation of matrix elements of nonelementary generators E_{ij} $(j \neq i \pm 1)$ using a parallelism with the work of Holman and Biedenharn.[13] A brief discussion of the method is presented in Sec. 4.

2. A TENSOR BASIS SET FOR THE IR'S OF U(n)

Consider an N-particle system described by an ordered set of orthonormal single particle orbitals, $\{\phi_i \mid i = 1,2,...,n\}$. These orbitals span the fundamental representation space V_n of U(n). A reducible Nth rank tensor basis set can be generated for U(n) by using these as,

$$V_n \otimes {}^N : |(N_1,N_2,...,N_n)\rangle = |\phi_1\rangle^{N_1}|\phi_2\rangle^{N_2}\cdots|\phi_n\rangle^{N_n}, \tag{1}$$

where

$$N_1,N_2,...,N_n \geqslant 0, \quad \sum_{i=1}^{n} N_i = N.$$

A standard (Young–Yamanouchi) realization of basis states of an IR of the permutation group S_N now follows if we apply the operators,[12]

$$\omega_{rs}^{[m]} = \left(\frac{f_N^{[m]}}{N!}\right)^{1/2} \sum_{P \in S_N} [P]_{rs}^{[m]} P \tag{2}$$

to a tensor product of Eq. (1). In Eq. (2) $[m] = [m_{1n},m_{2n},...,m_{nn}]$ subject ot $m_{1n} \geqslant m_{2n} \geqslant \cdots \geqslant m_{nn} \geqslant 0$ and $\Sigma_{i=1}^{n} m_{in} = N$, represents a specific IR of S_N of dimensionality $f_N^{[m]}$, $[P]_{rs}^{[m]}$ is the (r,s) matrix element for the Young orthogonal representation matrix for $P \in S_N$. For a fixed index r we can generate a set of nonzero states,

$$|rs;(N_1,N_2,...,N_n)\rangle^{[m]} = \omega_{rs}^{[m]}|(N_1,N_2,...,N_n)\rangle,$$

 0022-2488/80/040638-12$1.00

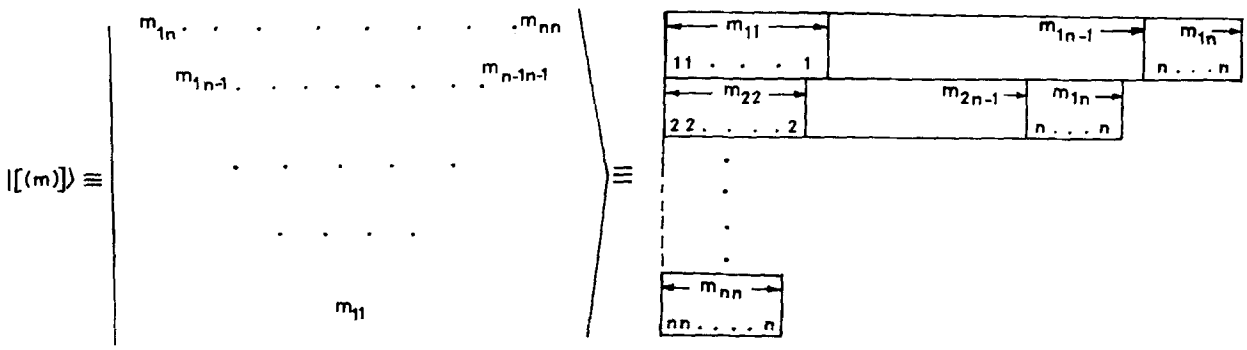
which, if chosen linearly independent, can provide a basis for the IR$[m]$ of U(n)[7,8,10,11] This set can be put into a one to one correspondence with Gel'fand bases $|[m]\rangle$ of U(n) if we identify N_i as $N_i = \Sigma_{j=1}^{i} m_{ji} - \Sigma_{j=1}^{i-1} m_{ji-1}$, where the m_{ji} satisfy the betweenness condition $m_{ji} \geqslant m_{ji-1} \geqslant m_{j+1\ i}$ for $i,j = 1,2,...,n$, subject to $m_{ji} = 0$ if $j > i$. The main problem with the procedure is choice of a linearly independent subset of $\{|rs;(N_1,...,N_n)\rangle^{[m]}\}$. The chief source of lack of linear independnece is that the Young–Yamanouchi representation of S_N[11,12] used in defining $\omega_{rs}^{[m]}$ is sequence adapted to the chain of subgroups $S_N \supset S_{N-1} \supset \cdots \supset S_1$, whereas the monomial tensor products of Eq. (1) are invariant under all $P \in S_{N_1} \otimes S_{N_2} \otimes \cdots \otimes S_{N_n}$. As pointed out by Lezuo,[5] this leads to the result,

$$|rs;(N_1,N_2,...,N_n)\rangle^{[m]} = (\text{constant})\ |rs';(N_1,N_2,...,N_n)\rangle^{[m]}$$

if $t_s^{[m]}$ and $t_{s'}^{[m]}$ are any two standard Young tableaux (SYT's),[12] related to each other through an elementary transposition of the set defined by $\mathscr{A} = \{P\,|P \in \Pi_{k=1}^{n} \otimes S_{N_k}\}$. A way out of this difficulty is to define a normalized linear combination of the operators of Eq. (2),

$$\omega_{r[(m)]}^{[m]} = N_{[(m)]}^{[m]} \sum_s a_s^{[m]} \omega_{rs}^{[m]} \tag{3}$$

which is right invariant ($\omega_{r[(m)]}^{[m]} P = \omega_{r[(m)]}^{[m]}$) under all $P \in \mathscr{A}$. The coefficients $a_s^{[m]}$ are determined solely from the symmetry requirement and $N_{[(m)]}^{[m]} = [\Sigma_s (a_s^{[m]})^2]^{-1/2}$. At this juncture it is necessary to indicate which $\omega_{rs}^{[m]}$ are to be used in the linear combination of Eq. (3). Patterson and Harter[8,9] have shown that the SYT $t_s^{[m]}$ defining $\omega_{rs}^{[m]}$ could be any of the tableaux in which the numbers i in the Weyl tableau are replaced by numbers $p_{i-1}+1, p_{i-1}+2, ..., p_i$, where $p_i = \Sigma_{j=1}^{i} m_{ji}$, so as to produce an SYT. The tableaux $t_s^{[m]}$ corresponding $\omega_{rs}^{[m]}$ of Eq. (3) are therefore determined by the Weyl pattern corresponding to a basis state of $[m]$. For convenience the Weyl tableau, and the corresponding Gel'fand pattern are shown below:



The index $[(m)]$ is used to identify the Weyl tableau (m) of the Ir$[m]$ of U(n). For a fixed index r this choice of $\omega_{r[(m)]}^{[m]}$ permits us to obtain a linearly independent basis set,

$$\{|[(m)];(N_1,N_2,...,N_n)\rangle^{[m]}\},$$

spanning the IR$[m]$ of U(n), using a subset of tensor products of Eq. (1) in which no two elements are related to each other through a $P \in S_N$.

Transformations among the elements of the above basis set can be induced using the generators E_{ij} $(i,j = 1,2,...n)$ of U(n). These E_{ij} can be defined using the matrix elements $e_{ij}(\alpha)$ $(i,j = 1,2,...,n;\alpha = 1,2,...,N)$ as,[8,9]

$$E_{ij} = \sum_{\alpha=1}^{N} e_{ij}(\alpha), \tag{4}$$

where $e_{ij}(\alpha)\phi_k(\beta) = \delta_{jk}\delta_{\alpha\beta}\phi_i(\alpha)$ so that,

$$e_{ij}(\alpha)e_{km}(\beta) = \delta_{jk}\delta_{\alpha\beta}e_{im}(\alpha). \tag{5}$$

The definition of E_{ij} in Eq. (4) and the relation given in Eq. (5) lead to the result

$$[E_{ij}, E_{kl}] = \delta_{jk}E_{jl} - \delta_{il}E_{kj}, \tag{6}$$

so that E_{ij}, $i,j = 1,2,...,n$ may be treated as the generators of U(n). Since the E_{ij} are totally symmetric in particle coordinates they commute with $\omega_{r[(m)]}^{[m]}$, and we obtain

$$E_{ij}\omega_{r[(m)]}^{[m]}\,|\,(N_1,N_2,...,N_n)\rangle = \omega_{r[(m)]}^{[m]}(\phi_1^{N_1}\cdots\phi_i^{N_i}\cdots(\phi_i\phi_j^{N_j-1})\cdots\phi_n^{N_n}), \tag{7}$$

where $(\phi_i\phi_j^{N_j-1}) = \Sigma_{k=0}^{N_j-1}\phi_j^k\phi_i\phi_j^{N_j-k-1}$ is a symmetrized sum of monomials with one ϕ_i and $N_j - 1$ ϕ_j's. Since $\omega_{r[(m)]}^{[m]}$ is right symmetric with respect to all permutations of these N_j coordinates

$$E_{ij}[(m)]\,|\,(N_1,...,N_i,...,N_j,...,N_n)\rangle = N_j[(m)]\phi_1^{N_1}\cdots\phi_i^{N_i}\cdots\phi_i\phi_j^{N_j-1}\cdots\phi_n^{N_n}. \tag{8}$$

In Eq. (8) we have, for notational convenience, replaced $\omega_{r[(m)]}^{[m]}$ by $[(m)]$. The orthonormality of basis orbitals ϕ_i leads to the result that the only nonzero matrix elements of E_{ij} are

$$\langle [(m')] \,;(N_1,...,N_i+1,...,N_j-1,...,N_n) \mid E_{ij} \mid [(m)];(N_1,...,N_i,...,N_j,...,N_n)\rangle \,.$$

Applying $E_{ij}^{\dagger} = E_{ji}$ to the bra vector in this matrix element we obtain a weighting factor N_i+1 in place of N_j as in Eq. (8). In order to resolve this ambiguity we use the geometric mean $[(N_i+1)N_j]^{1/2}$ to replace N_j in Eq. (8), if $i \neq j$. Using this replacement we express Eq. (8) as,

$$E_{ij} \mid [(m)];(N_1,...,N_i,...,N_j,...,N_n)\rangle = [(N_i+1)N_j]^{1/2} \mid [(m)];\phi_1^{N_1}\cdots\phi_i^{N_i},...,\phi_i\phi_j^{N_j-1},...,\phi_n^{N_n})\rangle$$

$$= [(N_i+1)N_j]^{1/2} \sum_{|(m')|} [(P)]^{|m|}_{|(m')|[(m)]} \mid [(m')];(N_1,...,N_i+1,...,N_j-1,...,N_n)\rangle \,. \tag{9}$$

We now describe a method of constructing the operators $\omega^{|m|}_{r|(m)|} \equiv [(m)]$ of Eq. (3) and the states $|[(m)];(N_1,...,N_n)\rangle$. We first concentrate on U(2) and then generalize the procedure to U(n). Using these basis states and Eq. (9) we will then determine the matrix elements of $E_{i-1,i}$. Consider the Gel'fand pattern and the corresponding Weyl tableau for a state belonging to the IR$[m_{12},m_{22}]$ of U(2):

$$\left| \begin{matrix} m_{12} & & m_{22} \\ & m_{11} & \end{matrix} \right\rangle \equiv \begin{array}{|c|c|}\hline \multicolumn{1}{c}{m_{11}} & \multicolumn{1}{c}{m_{12}} \\ \hline 111\ldots\ldots\ldots 11 & 22\ldots\ldots 2 \\ \hline \multicolumn{1}{c}{m_{22}} & \\ \cline{1-1} 222\ldots\ldots 22 & \end{array} \tag{10}$$

For such a state the basic tensor product of Eq. (1) is,

$$|(N_1,N_2)\rangle \equiv \phi_1^{N_1}\,\phi_2^{N_2}\,, \tag{11}$$

where $N_1 = m_{11}$, $N_2 = m_{12}+m_{22}-m_{11}$, $N_1+N_2 = m_{12}+m_{22} = N$. This tensor product is invariant under all $P \in S_{N_1} \otimes S_{N_2}$. This requires that we define $[(m)]$ as in Eq. (3) which is right invariant under all these permutations. Before taking up this general case, consider a specific example of $|(3,3)\rangle = \phi_1^3\phi_2^3$ for the IR[4,2] of U(2). Consider the Gel'fand state

$$\left| \begin{matrix} 4 & & 2 \\ & 3 & \end{matrix} \right\rangle \equiv \begin{matrix} 1 & 1 & 1 & 2 \\ 2 & 2 & & \end{matrix} \,.$$

From what has been said above regarding the $\omega^{|m|}_{rs}$ to be considered in the linear combination of Eq. (3), the relevant indices s here correspond to the lattice permutation symbols: 111122, 111212, 111221, to be written as $[(1^3)(1)(2^2)]$, $[(1^3)(2)(1)(2)]$, $[(1^3)(2^2)(1)]$. The parentheses have been introduced with a definite aim in mind. These indicate symmetrization over the particles whose row numbers are contained within the same parentheses. Obviously then, the symbols in which no parentheses contain two or more distinct row numbers denote single SYT's. Those symbols for which this is not the case are, in general, linear combinations of several SYT's, the linear combinations being totally symmetric with respect to permutations of particles within each of the parentheses. For the U(2) state under consideration, then, we write $[(1^3)(12^2)]$ for $[(m)] \equiv \omega^{|m|}_{r|(m)|}$. The symbol $[(1^3)(12^2)]$, therefore, denotes an operator totally symmetric in: (a) the first three particles, all in the first row of the SYT's and (b) the last three particles, one of which is in the first row and the other two in the second row. Using the transpositions (4,5), (5,6) on the right and the Y–Y orthogonal representation matrices, we find that

$$[(1^3)(12^2)] = (10/3)^{1/2}\{(4.5)^{-1/2}[(1^3)(1)(2^2)] + (3.4)^{-1/2}[(1^3)(2)(1)(2)] + (2.3)^{-1/2}[(1)^3(2^2)(1)]\} \tag{12}$$

is normalized and right invariant under $P \in S_3 \otimes S_3 \subset S_6$. By combining the last two terms on the right of Eq. (12) and normalizing the result we obtain a form which proves convenient for later generalization. We reexpress Eq. (12) as,

$$[(1^3)(12^2)] = \sqrt{1/6}\,[(1^3)(1)(2^2)] + \sqrt{5/6}\,[(1^3)(2)(12)], \tag{13}$$

where each operator combination on the right can be readily verified to be right symmetric under (5,6) in addition to being symmetric under interchange of the first three particles.

As a generalization of Eq. (13) to the case of $|(N_1,N_2)\rangle$ of Eq. (11), we first define a normalized operator combination,

$$[(1^{m_{11}})(1^{m_{12}-m_{11}}2^{m_{22}})] = \left[\frac{(m_{12}-m_{11})(m_{11}-m_{22}+1)}{(m_{11}+1)(m_{12}+m_{22}-m_{11})}\right]^{1/2}[(1^{m_{11}})(1)(1^{m_{12}-m_{11}-1}2^{m_{22}})]$$

$$+\left[\frac{(m_{12}+1)m_{22}}{(m_{11}+1)(m_{12}+m_{22}-m_{11})}\right]^{1/2}[(1^{m_{11}})(2)(1^{m_{12}-m_{11}}2^{m_{22}-1})] \tag{14}$$

and find that it tallies with Eq. (13) for $m_{12}=4$, $m_{22}=2$, $m_{11}=3$.

For low values of m_{11}, m_{12}, m_{22} it can be directly verified that the expansion given in Eq. (14) is right invariant under all $P \in S_{N_1} \otimes S_{N_2}$. Assuming that each term on the right-hand side of Eq. (14) is normalized and right invariant under permutations of the last N_2-1 particles, we can demonstrate that the linear combination is additionally right invariant under the transposition (N_1+1, N_1+2). In order to demonstrate this we decompose each operator combination on the right of Eq. (14) into further pairs of terms, each of which is totally symmetric under permutations of the last N_2-2 particles. We note, in this process, that Eq. (14) is valid for the further decomposition of $[(1^{m_{11}})(1)(1^{m_{12}-m_{11}-1}2^{m_{22}})]$ omitting the first $m_{11}+1$ boxes in

the first row, instead of the first m_{11} as in the original decomposition. Hence we have to replace m_{11} by $m_{11}+1$ wherever it occurs in the coefficients of Eq. (14) during the further decomposition. Similarly, the decomposition of $[(1^{m_{11}})(2)(1^{m_{12}-m_{11}}2^{m_{22}-1})]$ requires that we omit the first m_{11} boxes in the first row and the first box in the second row while doing the "city-block distance"[14] counting required for the calculation of the coefficients of Eq. (14). This means that the first column of the SYT's has to be omitted in obtaining the coefficients of the further decomposition of the second term in Eq. (14). Therefore, we replace m_{11}, m_{12}, m_{22} by $m_{11}-1, m_{12}-1, m_{22}-1$, respectively, wherever they occur in the coefficients of this expansion. With these considerations in mind we obtain

$$
\begin{aligned}
&[(1^{m_{11}})(1^{m_{12}-m_{11}}2^{m_{22}})] \\
&= \left[\frac{(m_{12}-m_{11})(m_{11}-m_{12}+1)}{(m_{11}+1)(m_{12}+m_{22}-m_{11})}\right]^{1/2}\Bigg[\left\{\frac{(m_{12}-m_{11}-1)(m_{11}-m_{22}+2)}{(m_{11}+2)(m_{12}+m_{22}-m_{11}-1)}\right\}^{1/2}[(1^{m_{11}})(1)(1)(1^{m_{12}-m_{11}-2}2^{m_{22}})] \\
&\quad+\left\{\frac{m_{22}(m_{12}+1)}{(m_{11}+2)(m_{12}+m_{22}-m_{11}-1)}\right\}^{1/2}[(1^{m_{11}})(1)(2)(1^{m_{12}-m_{11}-1}2^{m_{22}-1})]\Bigg]+\left[\frac{m_{22}(m_{12}+1)}{(m_{11}+1)(m_{12}+m_{22}-m_{11})}\right]^{1/2} \\
&\quad\times\Bigg[\left\{\frac{(m_{12}-m_{11})(m_{11}-m_{22}+1)}{m_{11}(m_{12}+m_{22}-m_{11}-1)}\right\}^{1/2}[(1^{m_{11}})(2)(1)(1^{m_{12}-m_{11}-1}2^{m_{22}-1})] \\
&\quad+\left\{\frac{(m_{22}-1)m_{12}}{m_{11}(m_{12}+m_{22}-m_{11}-1)}\right\}^{1/2}[(1^{m_{11}})(2)(2)(1^{m_{12}-m_{11}}2^{m_{22}-2})]\Bigg].
\end{aligned}
\tag{15}
$$

We find that the first and the last operator combinations on the right-hand side of Eq. (15) are already symmetric under the transposition (N_1+1, N_2+2). Also noting that each SYT of the sets, defining the two remaining operator combinations, has a fixed city block distance between the particles N_1+1 and N_2+2, we can verify that the combination of these two operator combinations is also right invariant under (N_1+1, N_1+2). A detailed verification of this result will be given later for the general IR of U(n).

Using arguments quite similar to the above we can also obtain an alternative to Eq. (14) as,

$$
\begin{aligned}
[(1^{m_{11}})(1^{m_{12}-m_{11}}2^{m_{22}})] &= \left\{\frac{(m_{12}-m_{11})(m_{12}+1)}{(m_{12}-m_{22}+1)(m_{12}+m_{22}-m_{11})}\right\}^{1/2}[(1^{m_{11}})(1^{m_{12}-m_{11}-1}2^{m_{22}})(1)] \\
&\quad+\left\{\frac{(m_{11}-m_{22}+1)m_{22}}{(m_{12}-m_{22}+1)(m_{12}+m_{22}-m_{11})}\right\}^{1/2}[(1^{m_{11}})(1^{m_{12}-m_{11}}2^{m_{22}-1})(2)],
\end{aligned}
\tag{16}
$$

which can also be verified to be right invariant under all $P\in S_{N_1}\otimes S_{N_2}$. In proving this we require further right [as against the left in Eq. (14)] decomposition of each term on the right of Eq. (16). It is evident that in this case we have to replace m_{12} by $m_{12}-1$ and m_{22} by $m_{22}-1$ respectively in applying Eq. (16) for the further decompositions of $[(1^{m_{11}})(1^{m_{12}-m_{11}-1}2^{m_{22}})(1)]$ and $[(1^{m_{11}})(1^{m_{12}-m_{11}}2^{m_{22}-1})(2)]$.

We note in Eqs. (14) and (16) that the $(1^{m_{11}})$ part of $[(1^{m_{11}})(1^{m_{12}-m_{11}}2^{m_{22}})]$ is not affected in the decompositions. This is a general result. Each of the parentheses in $[(m)]$ can be independently and successively decomposed. The coefficients of such decomposition depend only on the m_{ij}'s involved in the exponents of 1,2,...,etc. in the parenthesis. For generalizing the above considerations to the case of U(n) we identify the operator $[(m)]$ corresponding to the general U(n) Gel'fand state,

$$
\left|\begin{array}{ccccccc}
m_{1n} & & m_{2n} & & \cdots & m_{nn} \\
& m_{1n-1} & & m_{2n-1} & & m_{n-1n-1} \\
& & m_{12} & & m_{22} & & \\
& & & m_{11} & &
\end{array}\right\rangle
$$

as

$$
[(1^{m_{11}})(1^{m_{12}-m_{11}}2^{m_{22}})\cdots(1^{m_{1i}-m_{1i-1}}\cdots k^{m_{ki}-m_{ki-1}}\cdots i^{m_{ii}})\cdots(1^{m_{1n}-m_{1n-1}}\cdots n^{m_{nn}})]. \tag{17}
$$

The ith parenthesis in $[(m)]$ corresponds to the total symmetrization with respect to all particles occupying the orbital ϕ_i. Since the Gel'fand states of any IR of U(n) are sequence adapted to the chain of subgroups $U(n)\supset U(n-1)\supset\cdots\supset U(1)$, we will consider the decomposition of the nth parenthesis of Eq. (17). This requires the determination of $a_k{}^{(n)}$ in the decomposition

$$
(1^{m_{1n}-m_{1n-1}}\cdots k^{m_{kn}-m_{kn-1}}\cdots n^{m_{nn}})=\sum_{k=1}^{n}a_k{}^{(n)}(k)(1^{m_{1n}-m_{1n-1}}\cdots k^{m_{kn}-m_{kn-1}-1}\cdots n^{m_{nn}}). \tag{18}
$$

As for U(2) we first define $a_k{}^{(n)}$ as

$$
a_k{}^{(n)}=\left\{\frac{\Pi_{j=1}^{n}(m_{jn}-m_{kn-1}-j+k)}{\Pi_{j=1}^{n}{}'(m_{jn-1}-m_{kn-1}-j+k)}\right\}^{1/2}(N_n)^{-1/2} \tag{19}
$$

subject to $m_{nn-1}=0$. Substituting $n=2$ and $k=1$ in Eq. (19) reproduces the coefficient of $[(1^{m_{11}})(1)(1^{m_{12}-m_{11}-1}2^{m_{22}})]$ in the decomposition of $[(1^{m_{11}})(1^{m_{12}-m_{11}}2^{m_{22}})]$ in Eq. (14), just as substituting $n=2$, $k=2$ reproduces that of

$[(1^{m_{11}})(2)(1^{m_{12}-m_{11}}2^{m_{22}-1})]$. The verification that the right side of Eq. (18) is properly symmetrized is quite laborious. As an illustration we will show that the coefficients for $k = 1,2$ yield the correct symmetrization. The procedure is identical for all other pairs of k values. From Eq. (18) and Eq. (19) we have after omitting all the first $n-1$ parentheses for the U(n) state as in writing Eq. (18), the result

$$(1^{m_{1n}-m_{1n-1}}2^{m_{2n}-m_{2n-1}}\cdots n^{m_{nn}}) = (N_n)^{-1/2}\left\{\left[\frac{\Pi_{k=1}^{n}(m_{kn}-m_{1n-1}-k+1)}{\Pi_{k=1}^{\prime n}(m_{kn-1}-m_{1n-1}-k+1)}\right]^{1/2}(1)(1^{m_{1n}-m_{1n-1}-1}2^{m_{2n}-m_{2n-1}}\cdots n^{m_{nn}}) + \left[\frac{\Pi_{k=1}^{n}(m_{kn}-m_{2n-1}-k+2)}{\Pi_{k=1}^{\prime n}(m_{kn-1}-m_{2n-1}-k+2)}\right]^{1/2}(2)(1^{m_{1n}-m_{1n-1}}2^{m_{2n}-m_{2n-1}-1}\cdots n^{m_{nn}})+\cdots\right\}. \tag{20}$$

In the further decomposition of any of the operator combinations $(j)(\cdots j^{m_{jn}-m_{jn-1}-1}\cdots)$ on the right-hand side of Eq. (20), reasoning similar to that given in the case of U(2) shows that m_{jn-1} has to be replaced by $m_{jn-1}+1$ if $j\neq n$, and all m_{jn}'s and m_{jn-1}'s have to be decreased by 1 (keeping $m_{nn-1}=0$) if $j=n$. With these modifications we obtain

$$\begin{aligned}
&(1^{m_{1n}-m_{1n-1}}2^{m_{2n}-m_{2n-1}}\cdots n^{m_{nn}})\\
&= (N_n)^{-1/2}\left[\left\{\frac{\Pi_{k=1}^{n}(m_{kn}-m_{1n-1}-k+1)}{\Pi_{k=1}^{\prime n}(m_{kn-1}-m_{1n-1}-k+1)}\right\}^{1/2}(N_n-1)^{-1/2}\left[\left\{\frac{\Pi_{j=1}^{n}(m_{jn}-m_{1n-1}-j)}{\Pi_{j=1}^{\prime n}(m_{jn-1}-m_{1n-1}-j)}\right\}^{1/2}\right.\right.\\
&\times(1)(1)(1^{m_{1n}-m_{1n-1}-2}2^{m_{2n}-m_{2n-1}}\cdots n^{m_{nn}})+\left\{\frac{\Pi_{j=1}^{n}(m_{jn}-m_{2n-1}-j+2)}{\Pi_{j=1}^{\prime n}(m_{jn-1}-m_{2n-1}-j+2)}\right\}^{1/2}\left\{\frac{(m_{1n-1}-m_{2n-1}+1)}{(m_{1n-1}-m_{2n-1}+2)}\right\}^{1/2}\\
&\left.\times(1)(2)(1^{m_{1n}-m_{1n-1}-1}2^{m_{2n}-m_{2n-1}-1}\cdots n^{m_{nn}})\right]+\left\{\frac{\Pi_{k=1}^{n}(m_{kn}-m_{2n-1}-k+2)}{\Pi_{k=1}^{\prime n}(m_{kn-1}-m_{2n-1}-k+2)}\right\}^{1/2}\\
&\times(N_n-1)^{-1/2}\left[\left\{\frac{\Pi_{j=1}^{n}(m_{jn}-m_{1n-1}-j+1)}{\Pi_{j=1}^{\prime n}(m_{jn-1}-m_{1n-1}-j+1)}\right\}^{1/2}\right.\\
&\times\left\{\frac{(m_{2n-1}-m_{1n-1}-1)}{(m_{2n-1}-m_{1n-1})}\right\}^{1/2}(2)(1)(1^{m_{1n}-m_{1n-1}-1}2^{m_{2n}-m_{2n-1}-1}\cdots n^{m_{nn}})\\
&\left.\left.+\left\{\frac{\Pi_{j=1}^{n}(m_{jn}-m_{2n-1}-j+1)}{\Pi_{j=1}^{\prime n}(m_{jn-1}-m_{2n-1}-j+1)}\right\}^{1/2}(2)(2)(1^{m_{1n}-m_{1n-1}}2^{m_{2n}-m_{2n-2}}\cdots n^{m_{nn}})+\cdots\right]\right].
\end{aligned} \tag{21}$$

As for the case of U(2), the first and the last operator combinations are already symmetric under $(N_{n-1}+1, N_{n-1}+2)$. We have only to show that the combination of the middle two terms on the right of Eq. (21) is symmetric with respect to $(N_{n-1}+1, N_{n-1}+2)$. All the operators in Eq. (21) are assumed to be totally symmetric with respect to permutations of the last N_n-2 particles. [Strictly speaking the various parentheses in Eq. (21) are not operators by themselves. These parentheses taken along with the other $n-1$ parentheses corresponding to ϕ_i, $i = 1,2,\ldots,n-1$, are the operators we refer to.] We find that all the SYT's present in the second and the third operators on the right of Eq. (21) have the same "city-block distance"[15] of $(m_{1n-1}-m_{2n-1}+1)$ between the $(N_{n-1}+1)$ and $(N_{n-1}+2)$ particles. Applying on the right the transposition $(N_{n-1}+1, N_{n-1}+2)$ to either side of Eq. (21), and using Young orthogonal matrices[11,12] for this transposition we find that the coefficient of $(1)(2)(1^{m_{1n}-m_{1n-1}-1}2^{m_{2n}-m_{2n-1}-1}\cdots n^{m_{nn}})$ becomes the sum of two contributions. The diagonal contribution comes from the second operator in Eq. (21) and off-diagonal contribution from the third. A combination of these two yields

$$\begin{aligned}
&[N_n(N_n-1)]^{-1/2}\left\{\frac{\Pi_{k=1}^{n}(m_{kn}-m_{1n-1}-k+1)}{\Pi_{k=1}^{\prime n}(m_{kn-1}-m_{1n-1}-k+1)}\,\frac{\Pi_{k=1}^{n}(m_{kn}-m_{2n-1}-k+2)}{\Pi_{k=1}^{\prime n}(m_{kn-1}-m_{2n-1}-k+2)}\right\}^{1/2}\\
&\times(m_{1n-1}-m_{2n-1}+1)^{-1}\left[-\left\{\frac{(m_{1n-1}-m_{2n-1}+1)}{(m_{1n-1}-m_{2n-1}+2)}\right\}^{1/2}\right.\\
&\left.+\left\{\frac{(m_{1n-1}-m_{2n-1})(m_{1n-1}-m_{2n-1}+2)(m_{2n-1}-m_{1n-1}-1)}{(m_{2n-1}-m_{1n-1})}\right\}^{1/2}\right].
\end{aligned} \tag{22}$$

Simplification of the expression in (22) leads to the same expression as the coefficient of the second operator, i.e., $(1)(2)(1^{m_{1n}-m_{1n-1}-1}2^{m_{2n}-m_{2n-1}-1}\cdots n^{m_{nn}})$, on the right of Eq. (21). Similar verification can be carried out for the coefficient of the third operator in Eq. (21).

A right decomposition of the operator on the left-hand side of Eq. (18) is also possible, as in

$$(1^{m_{1i}-m_{1i-1}}2^{m_{2i}-m_{2i-1}}\cdots k^{m_{ki}-m_{ki-1}}\cdots i^{m_{ii}}) = \sum_{k=1}^{i} b_k{}^{(i)}(\cdots k^{m_{ki}-m_{ki-1}-1}\cdots)(k), \tag{23}$$

where

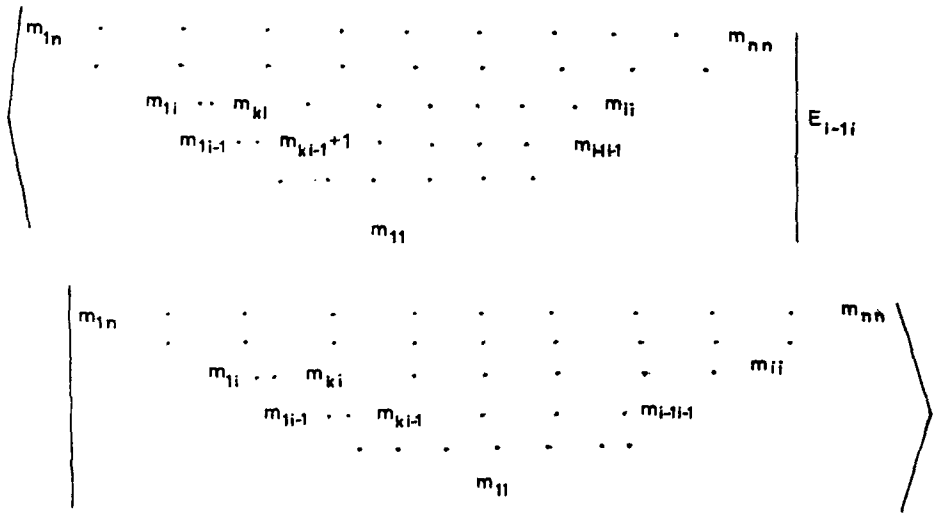
$$b_k^{(i)} = (-1)^{1/2}\left\{\frac{\Pi_{j=1}^{n}(m_{ji-1} - m_{ki} + k - j)}{\Pi'^{n}_{j=1}(m_{ji} - m_{ki} + k - j)}\right\}^{1/2}(N_i)^{-1/2}, \quad m_{ii-1} \equiv 0. \tag{24}$$

The verification of the right invariance of the operator of Eq. (23), with respect to $P \in \Pi_{k=1}^{n} \otimes S_{N_k}$, follows on the same lines as that for left decomposition given in Eqs. (18)–(22). For further decomposition [on the left as in Eq. (18) or on the right as in Eq. (23)] of the operator $(\cdots j^{m_{ji}-m_{ji-1}}\cdots i^{m_{ii}})(j)$, we replace m_{ji} by $m_{ji}-1$. Equation (23) is an alternative normalized operator combination right invariant under $P \in \Pi_{k=1}^{n} \otimes S_{N_k}$, the symmetrization affected being over N_i particles ($i = 1,2,\ldots,n$) in ϕ_i. It should be emphasized that Eq. (18) with n replaced by i and Eq. (23) are valid for decompositions of the ith parentheses, corresponding to symmetrization over the N_i particles in ϕ_i, $i = 1,2,\ldots,n$.

The operators $[(m)]$ to be written down in terms of single SYT's by using Eq. (18) or (23), or both, are complete because one, and only one, $[(m)]$ is associated with a Gel'fand pattern, and orthogonal because the symmetrization in each is done over distinct combinations of boxes in the 1st, 2nd, ...,etc. rows of the SYT's as dictated by the Weyl tableau corresponding to the Gel'fand pattern. Further $[(m)]$ are properly normalized on the assumption that the SYT's are. We have, therefore, obtained a complete set of operators which, acting on a subset of monomial tensor products of Eq. (1), in which no two elements are permutationally related, yield a complete orthonormal basis set for IR$[m]$ of U(n),

$$\left\{ |\,[(m)];(N_1,N_2,\ldots,N_n)\rangle \,\middle|\, \sum N_i = N, \quad \sum_{j=}^{i}(m_{ji} - m_{ji-1}) = N_i, \quad m_{ji} \geqslant m_{ji-1} \geqslant m_{j+1i} \right\}. \tag{25}$$

We now proceed to obtain the matrix elements of the elementary generators $E_{i-1,i}$ between the basis states of the set (25), making use of Eq. (9). The matrix element in Gel'fand and our notation is,



$$= M^{k}_{i-1,i} = \langle [(m')];(N_1,\ldots,N_{i-1}+1,N_i-1,\ldots,N_n)|E_{i-1i}|[(m)];(N_1,\ldots,N_{i-1},N_i,\ldots,N_n)\rangle, \tag{26}$$

where

$$[(m)] = [(1^{m_{11}})\cdots(1^{m_{1i}-m_{1i-1}}\cdots k^{m_{ki}-m_{ki-1}}\cdots)(1^{m_{1i}-m_{1i-1}}\cdots k^{m_{ki}-m_{ki-2}}\cdots)\cdots], \tag{27}$$

$$[(m')] = [(1^{m_{11}})\cdots(1^{m_{1i}-m_{1i-1}}\cdots k^{m_{ki}-m_{ki-2}+1}\cdots)(1^{m_{1i}-m_{1i-1}}\cdots k^{m_{ki}-m_{ki-1}}\cdots)\cdots]. \tag{28}$$

We have not written the parentheses which are identical for $[(m)]$ and $[(m')]$ in Eqs. (27), (28). From an inspection of Eqs. (27) and (28), we find that $[(m)]$ has an extra degree of symmetrization in the ith parenthesis in the k th row, whereas $[(m')]$ contains an extra degree of symmetrization in the $(i-1)$th parenthesis in the k th row. Apart from this the symmetrization is identical for $[(m)]$ and $[(m')]$ for other parentheses, i.e., for all $\phi_j, j \neq i, i-1$. This dictates that we use right decomposition [Eq. (23)] of the $(i-1)$th parenthesis of $[(m')]$ bringing out a (k) and matching with the $(i-1)$th parenthesis of $[(m)]$. Similarly, left decomposition [cf. Eq. (18)] of the ith parenthesis of $[(m)]$ again bringing out (k) matches the ith parentheses of $[(m)]$ and $[(m')]$. The operator combinations are now matched on both sides of $E_{i-1,i}$. Action of $E_{i-1,i}$ gives a factor of $[(N_{i-1}+1)N_i]^{1/2}$ according to Eq. (9), and matches the monomial products in front of the operators (i.e., identity is the matching permutation for monomial products). The overlap of matched operators is unity and the matrix element of Eq. (26) becomes

$$M^{k}_{i-1,i} = [(N_{i-1}+1)N_i]^{1/2} a_k^{(i)} b_k^{(i-1)}.$$

From Eq. (19) and (24) we obtain

$$M^{k}_{i-1,i} = (-1)^{1/2}\left\{\frac{\Pi_{j=1}^{i-1}(m_{ji-2} - m_{ki-1} - j + k - 1)\Pi_{j=1}^{i}(m_{ji} - m_{ki-1} - j + k)}{\Pi'^{i-1}_{j=1}(m_{ji-1} - m_{ki-1} - j + k - 1)\Pi'^{i}_{j=1}(m_{ji-1} - m_{ki-1} - j + k)}\right\}^{1/2}. \tag{29}$$

The matrix element of Eq. (29) is identical with the one obtained by Baird and Biedenharn[2] and Moshinsky and Nagel.[5] The matrix element of $E_{i,i-1}$ can be similarly obtained.

3. A PATTERN CALCULUS FOR THE DECOMPOSITION COEFFICIENTS: MATRIX ELEMENTS OF E_{ij}

The simplicity of the forms of left and right decompositions of Eqs. (18) and (23) enables us to determine the relevant coefficients by adapting the pattern calculus developed by Holman and Biedenharn[13] in their study of the tensor operators of U(n). The advantage of the method is that it also enables a ready determination of the matrix elements of general E_{ij} of U(n). We outline the method in what follows.

Consider the decomposition

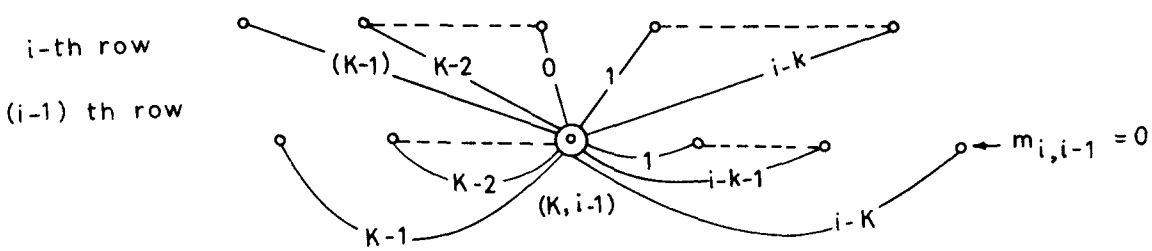
$$(1^{m_{1i}-m_{1i-1}}\ldots k^{m_{ki}-m_{ki-1}}\ldots i^{m_{ii}})\rightarrow(k)(1^{m_{1i}-m_{1i-1}}\ldots k^{m_{ki}-m_{ki-1}-1}\ldots i^{m_{ii}})$$

of the ith parentheses in $[(m)]$ of U(n) ($n\geqslant i$). The coefficient $a_k{}^{(i)}$ of Eq. (19) can be simply generated using the following rules:

(1) Consider the ith and $(i-1)$th rows of the Gel'fand tableau associated with $[(m)]$, adding an additional node of weight $m_{ii-1}=0$ to the $(i-1)$th row. Treat the $(k,i-1)$ node with weight m_{ki-1} as a reference node (a center).

(2) Join the center to all the nodes in the ith and $(i-1)$th rows by lines, assigning weights $k-j$ for $1\leqslant j\leqslant k$ and $j-k$ for $k<j\leqslant i$ to the line between the (k,i) or $(k,i-1)$ node and the center, as indicated in the pattern (A).

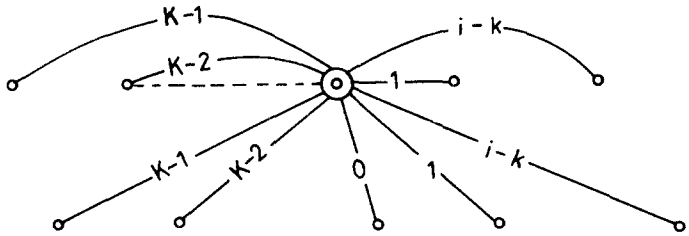
(A)



(3) Treat each of the lines connecting the center ot the ith row nodes as contributing a factor $(m_{ji}-m_{ki-1}-j+k)^{1/2}$ for $1\leqslant j\leqslant k$, and $(m_{ki-1}-m_{ji}+j-k)^{1/2}$ for $k<j\leqslant i$. Treat each of the lines connecting the center to the $(i-1)$ row nodes as contributing a factor $(m_{ji-1}-m_{ki-1}-j+k)^{-1/2}$ for $1\leqslant j<k$, and $(m_{ki-1}-m_{ji-1}+j-k)^{-1/2}$ for $k<j\leqslant 1$. The product of all the lines of the diagram is $a_k{}^{(i)}$ except for an additional overall factor, common to all $a_k{}^{(i)}$, of $(N_i)^{-1/2}$.

For the right decomposition of Eq. (23) we proceed similarly to a pattern for $b_k{}^{(i)}$, except that the center now is in the top row and lines joining the center to the ith row nodes now contribute inverse square root factors, and those joining the center to $(i-1)$ row nodes square root factors. This is indicated in the pattern (B).

(B)



Using such diagrammatic representations it is possible to considerably simplify the determination of the matrix elements of E_{ij} of U(n). As a means of obtaining general rules consider the matrix element,

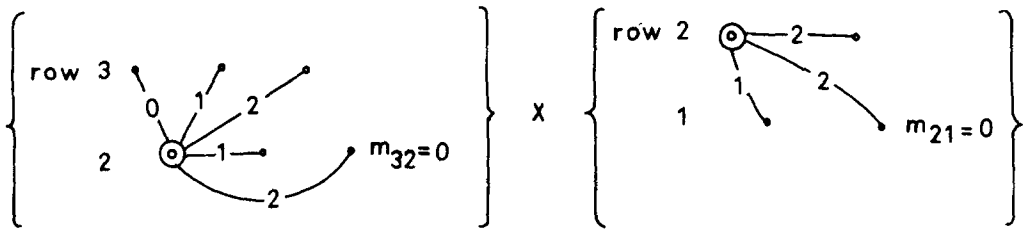
$$\langle[(1^{m_{11}+1})(1^{m_{12}-m_{11}}2^{m_{22}})(1^{m_{13}-m_{12}-1}2^{m_{23}-m_{22}}3^{m_{33}})]\phi_1^{N_1+1}\phi_2^{N_2}\phi_3^{N_3-1}|E_{13}|$$
$$\times[(1^{m_{11}})(1^{m_{12}-m_{11}}2^{m_{22}})(1^{m_{13}-m_{12}}2^{m_{23}-m_{22}}3^{m_{33}})]\phi_1^{N_1}\phi_2^{N_2}\phi_3^{N_3}\rangle$$

of U(3). Using $E_{13}=[E_{12},E_{23}]$, we find that the only intermediate state required for the nonzero matrix element of $E_{12}E_{23}$ is

$$|[(1^{m_{11}})(1^{m_{12}-m_{11}+1}2^{m_{22}})(1^{m_{13}-m_{12}-1}2^{m_{23}-m_{22}}3^{m_{33}})]\phi_1^{N_1}\phi_2^{N_2+1}\phi_3^{N_3-1}\rangle\ .$$

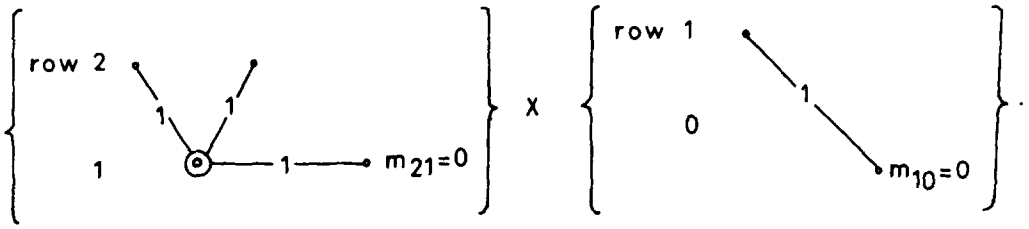
As in Eq. (29), the matrix element of E_{23} between this bra and the initial ket vector requires a right decomposition $(1^{m_{12}-m_{11}+1}2^{m_{22}})\rightarrow(1^{m_{12}-m_{11}}2^{m_{22}})(1)$, of the former and a left decomposition, $(1^{m_{13}-m_{12}}2^{m_{23}-m_{22}}3^{m_{33}})$ $\rightarrow(1)(1^{m_{13}-m_{12}-1}2^{m_{23}-m_{22}}3^{m_{33}})$, of the latter. Noting that for the intermediate state m_{12} is replaced by $m_{12}+1$ and hence

appropriately increasing/decreasing the line weights over those given by rule (2) above, we obtain the E_{23} part of the $E_{12}E_{23}$ matrix element as



(30)

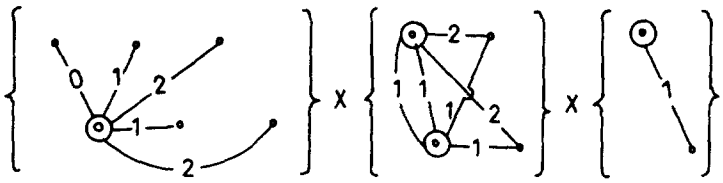
Similarly the E_{12} part is obtained as



(31)

The second pattern in (31) is introduced because the right decomposition factor for $(1^{m_{11}+1})\rightarrow(1_{m_{11}})(1)$ is $(m_{11}+1)^{1/2}$, and so it is necessary to artificially introduce a zeroth row to the Gel'fand pattern. Moreover we have omitted the factors $[N_i(N_{i-1}+1)]^{1/2}$, since they cancel with corresponding factors due to the action of $E_{i-1,i}$ at every stage. A combination of (30), (31) gives the $E_{12}E_{23}$ part of the matrix element of E_{13} as

$$\langle|E_{12}\ \ E_{23}|\rangle =$$



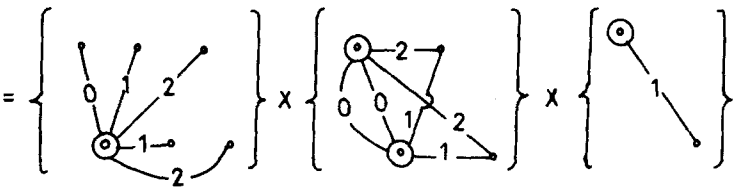
(32)

A similar procedure for $E_{23}E_{12}$ through the intermediate state

$$|(1^{m_{11}+1})(1^{m_{12}-m_{11}-1}2^{m_{22}})(1^{m_{13}-m_{12}}2^{m_{23}-m_{22}}3^{m_{33}});\phi_1^{N_1+1}\phi_2^{N_2-1}\phi_3^{N_3}\rangle\ ,$$

leads to

$$\langle|E_{23}\ \ E_{12}|\rangle =$$



(33)

By subtracting the RHS of Eq. (33) from that of Eq. (32) we find that all lines have the same weight except those defining a loop between the pair of centers in the "combined" second pattern. The difference between these loops is $(m_{12}-m_{11}+1)-(m_{12}-m_{11})=1$. Thus in the diagrammatic equivalent of the matrix element of E_{13}, we can omit the loop between the centers in the combined pattern and obtain,

$$\langle[(1^{m_{11}+1})(1^{m_{12}-m_{11}}2^{m_{22}})(1^{m_{13}-m_{12}-1}2^{m_{23}-m_{33}}3^{m_{33}})];\phi_1^{N_1+1}\phi_2^{N_2}\phi_3^{N_3-1}|E_{13}|$$
$$\times[(1^{m_{11}})(1^{m_{12}-m_{11}}2^{m_{22}})(1^{m_{13}-m_{12}}2^{m_{23}-m_{33}}3^{m_{33}})];\phi_1^{N_1}\phi_2^{N_2}\phi_3^{N_3}\rangle$$

$$= \left\{ \frac{(m_{13}-m_{12})(m_{12}-m_{23}+1)(m_{12}-m_{33}+2)(m_{11}-m_{22}+1)}{(m_{12}-m_{22}+1)(m_{12}-m_{22}+2)} \right\}^{1/2}$$

$$= \left\{ \begin{array}{c} 0\ 1\ 2 \\ -1- \\ \quad 2 \end{array} \right\} \times \left\{ \begin{array}{c} \quad 2 \\ 1\ 2 \\ -1 \end{array} \right\} \times \left\{ \begin{array}{c} \\ \quad 1 \end{array} \right\}. \tag{34}$$

Before attempting to obtain a generalized set of rules determining general E_{ij} matrix elements, let us consider one more matrix element,

$\langle [(1^{m_{11}+1})(1^{m_{12}-m_{11}-1}2^{m_{22}+1})(1^{m_{13}-m_{12}}2^{m_{23}-m_{22}-1}3^{m_{33}})]; \phi_1^{N_1+1}\phi_2^{N_2}\phi_3^{N_3-1} |E_{13}|$

$\times [(1^{m_{11}})(1^{m_{12}-m_{11}}2^{m_{22}})(1^{m_{13}-m_{12}}2^{m_{23}-m_{22}}3^{m_{33}})]; \phi_1^{N_1}\phi_2^{N_2}\phi_3^{N_3}\rangle$.

Proceeding exactly as above, with appropriate choice for the intermediate states, we obtain the $E_{12}E_{23}$ and $E_{23}E_{12}$ matrix elements as

$$\langle |E_{12}\ E_{23}| \rangle = \left\{ \begin{array}{c} 1\ 0\ 1 \\ -1- \\ \quad -1 \end{array} \right\} \times \left\{ \begin{array}{c} \quad 0 \\ 0\ 0\ 0\ 1 \\ \quad -1 \end{array} \right\} \times \left\{ \begin{array}{c} \\ \quad 1 \end{array} \right\}, \tag{35}$$

$$\langle |E_{23}\ E_{12}| \rangle = \left\{ \begin{array}{c} 1\ 0\ 1 \\ -1- \\ \quad -1 \end{array} \right\} \times \left\{ \begin{array}{c} \quad 0 \\ 0\ 1\ 1\ 1 \\ \quad -1 \end{array} \right\} \times \left\{ \begin{array}{c} \\ \quad 1 \end{array} \right\}. \tag{36}$$

Again, in subtracting the RHS of Eq. (36) from that of Eq. (35), we note that the only difference in the patterns is the loop connecting the centers (which this time are m_{12} and m_{11}). Subtraction of the loops gives the algebraic factor $(m_{12}-m_{11})-(m_{12}-m_{11}+1) = -1$. Hence, we have,

$\langle [(1^{m_{11}+1})(1^{m_{12}-m_{11}-1}2^{m_{22}+1})(1^{m_{13}-m_{12}}2^{m_{23}-m_{22}-1}3^{m_{33}})]; \phi_1^{N_1+1}\phi_2^{N_2}\phi_3^{N_3-1} |E_{13}|$

$\times [(1^{m_{11}})(1^{m_{12}-m_{11}}2^{m_{22}})(1^{m_{13}-m_{12}}2^{m_{23}-m_{22}}3^{m_{33}})]; \phi_1^{N_1}\phi_2^{N_2}\phi_3^{N_3}\rangle$

$$= (-1) \left\{ \begin{array}{c} 1\ 0\ 1 \\ -1- \\ \quad -1 \end{array} \right\} \times \left\{ \begin{array}{c} \quad 0 \\ 0\ \quad 1 \\ \quad -1 \end{array} \right\} \times \left\{ \begin{array}{c} \\ \quad 1 \end{array} \right\}$$

$$= (-1) \left[\frac{(m_{13}-m_{22}+1)(m_{23}-m_{22})(m_{22}-m_{33}+1)}{(m_{12}-m_{22}+1)(m_{22}+1)} \right]^{1/2} \left[\frac{(m_{12}-m_{11})(m_{22}+1)}{(m_{12}-m_{22})(m_{11}+1)} \right]^{1/2} (m_{11}+1)^{1/2}$$

$$= (-1) \left\{ \frac{(m_{13}-m_{22}+1)(m_{23}-m_{22})(m_{22}-m_{33}+1)(m_{12}-m_{11})}{(m_{12}-m_{22})(m_{12}-m_{22}+1)} \right\}^{1/2}. \tag{37}$$

We find that the results given in Eqs. (34), (37) coincide with those given by Moshinsky,[9] even including the phase factor. This permits us to propose a general set of rules for determining the matrix elements of E_{ij} $(i<j)$ of U(n). E_{ij} connects a Gel'fand state G with another Gel'fand state G' which has its k th $(1 \leqslant k < i$ or $n \geqslant k \geqslant j)$ rows identical with that of G. Each of the pth rows

of G' for $i \leqslant p \leqslant j$ differs from the corresponding row of G in that the k_p th entry in the pth row G' is the k_p th entry in the pth row of G increased by unity.

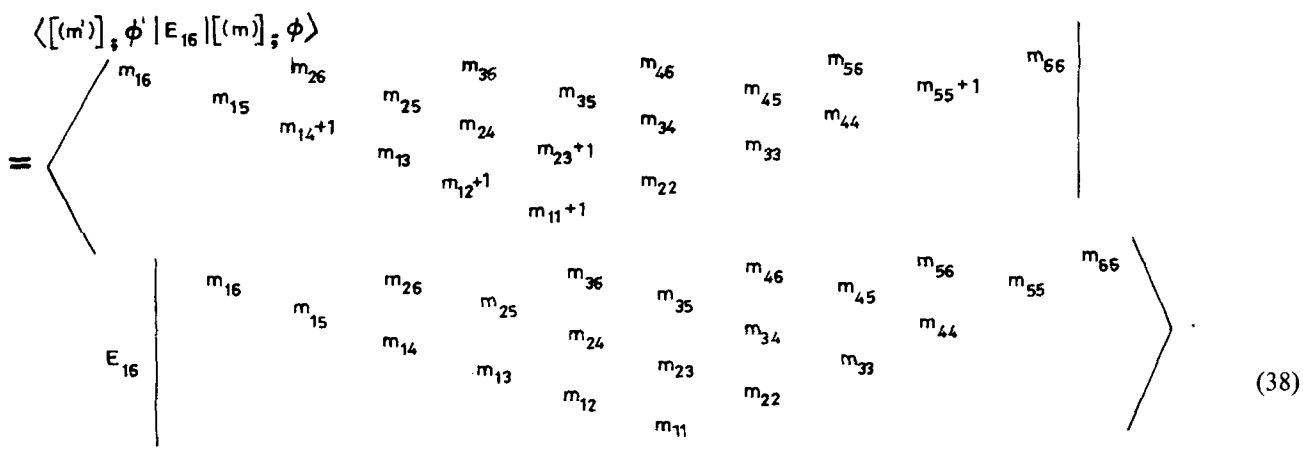
(1) For E_{ij} there are two simple patterns (SP's) and $(j-i-1)$ combined pattern (CP's).

(2) One of the SP's corresponds to left decomposition coefficient for taking out k_{j-1} on the left from ϕ_j parentheses in G. This is drawn as indicated in pattern (A) of Sec. 3 (rule 2). The second SP corresponds to the right decomposition coefficient for taking out k_i on the right from ϕ_i parenthesis of G'. This is again drawn as indicated in pattern (B) of Sec. 3. The line-weights are changed suitably to accomodate the fact that $m_{k_i,i}$ is replaced by $m_{k_i,i}+1$ for this decomposition, i.e., the lines joining the m_{pi} or m_{pi-1} nodes to the $m_{k_i,i}$ node have weights $k_i - p - 1$ for $1 \leqslant p < k_i$, and $p - k_i + 1$ for $k_i \leqslant p \leqslant i$.

(3) Each of the CP's has two centers which are not connected to each other. The first CP displays rows $j-1$ and $j-2$ explicitly, with an extra node $m_{j-1,j-1} = 0$ in the $(j-2)$ row. In evaluating the commutator $[E_{i,j-1}, E_{j-1,j}]$ we encounter two centers at $(k_{j-1}, j-1)$ and $(k_{j-2}, j-2)$ in the two rows. We first join the center $(k_{j-1}, j-1)$ to all the other nodes in the $j-1$ and $j-2$ rows except to the other center $(k_{j-2}, j-2)$. Noting that the center $(k_{j-1}, j-1)$ has a weight $m_{k_{j-1},j-1}+1$, due to the action of $E_{j-1,j}$, we assign weights $k_{j-1} - p - 1$ for $1 \leqslant p < k_{j-1}$ and $p - k_{j-1} + 1$ for $k_{j-1} \leqslant p \leqslant j-1$ and retain the weight $m_{k_{j-1},j-1}$ for the center in the CP. Similarly we draw lines connecting the center $(k_{j-2}, j-2)$ to all other nodes in the $j-1$ and $j-2$ rows except to the other center $(k_{j-1}, j-1)$. Since $(k_{j-2}, j-2)$ still has the weight $m_{k_{j-2},j-2}$, we assign weights $k_{j-2} - p$ for $1 \leqslant p \leqslant k_{j-2}$ and $p - k_{j-2}$ for $k_{j-2} \leqslant p \leqslant j-1$. In general each of the CP's displays two rows, pth and $(p-1)$th $(j-1 \geqslant p \geqslant i+1)$ with two centers at (k_p, p) and $(k_{p-1}, p-1)$. The centers are not connected to each other. We let m_{ij} with the same i constitute "planes." The lines joining the center (k_p, p) to other nodes have weights 0,1,2,...,etc., as the node joined lies in a plane successively to the left of the plane of the center, and 2,3,...,etc., as it lies in a plane successively to the right. The line joining the center $(k_{p-1}, p-1)$ has weight 1,2,...,etc., as the node lies in a plane successively to the left or right of the plane of the center. The weight of the line joining the center with a node in its own plane has weight 1 for center (k_p, p) and 0 for center $(k_{p-1}, p-1)$.

(4) Finally a CP with p and $p-1$ rows displayed, having centers at $(k_p, p-1)$, contributes an overall phase factor of (-1) if $k_p > k_{p-1}$. Otherwise the factor is $+1$.

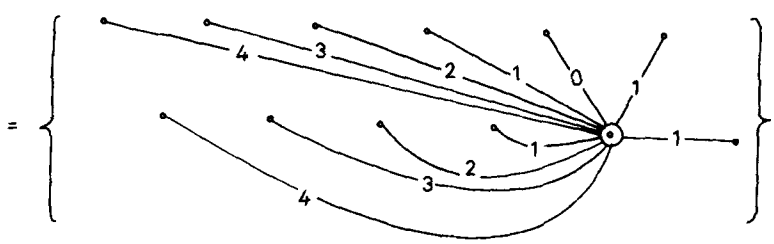
As a direct application of these rules we work out a reasonably complicated example of a matrix element of E_{16} of U(6). Consider the following matrix element

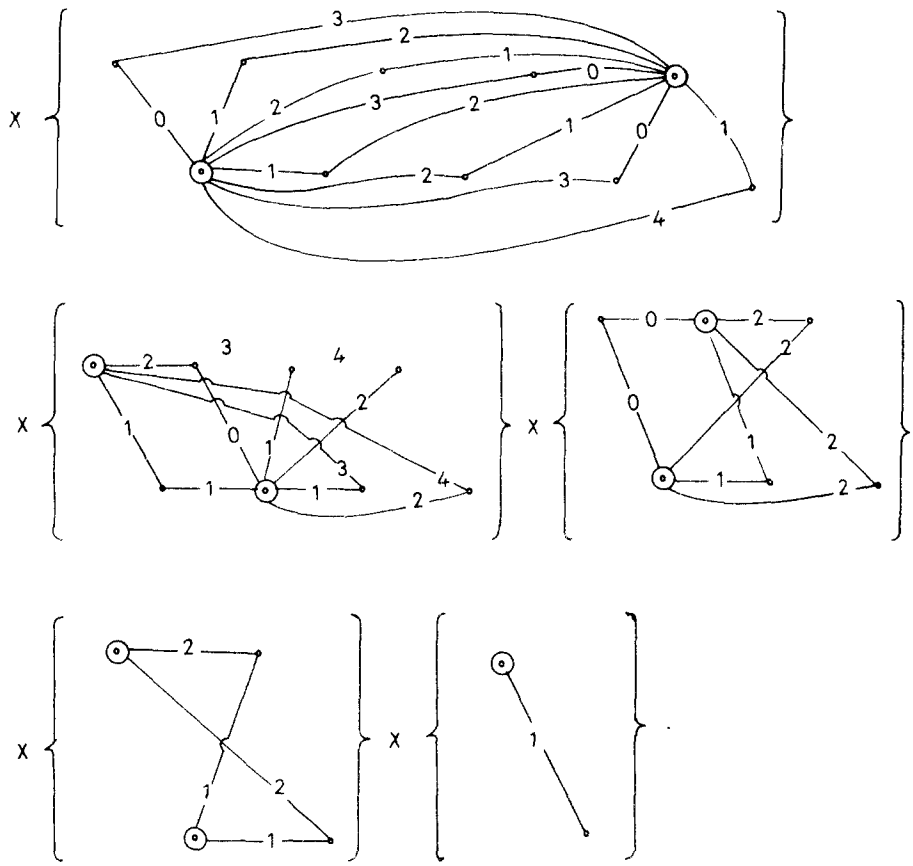


(38)

We apply the above rules (1)–(4) to evaluate the matrix element of Eq. (38).

$\langle [(m')]; \phi_1^{N_1+1} \phi_2^{N_2} \phi_3^{N_3} \phi_4^{N_4} \phi_5^{N_5} \phi_6^{N_6-1} | E_{16} | [(m)]; \phi_1^{N_1} \phi_2^{N_2} \phi_3^{N_3} \phi_4^{N_4} \phi_5^{N_5} \phi_6^{N_6} \rangle$





(39)

Using the rules for associating algebraic factors with the lines in the above patterns, we obtain after simplification, the result

$$\langle[(m')],\phi' \mid E_{16} \mid [(m)];\phi\rangle$$

$$= \left[\frac{(m_{16}-m_{55}+4)(m_{26}-m_{55}+3)(m_{36}-m_{55}+2)(m_{46}-m_{55}+1)(m_{56}-m_{55})(m_{55}-m_{66}+1)}{(m_{15}-m_{55}+4)(m_{25}-m_{55}+3)(m_{35}-m_{55}+2)(m_{45}-m_{55}+1)}\right]^{1/2}$$
$$\times\left[\frac{(m_{24}-m_{55}+2)(m_{34}-m_{55}+1)(m_{44}-m_{55})(m_{15}-m_{14})}{(m_{15}-m_{55}+3)(m_{25}-m_{55}+2)(m_{35}-m_{55}+1)(m_{45}-m_{55})}\right.$$
$$\left.\times\frac{(m_{14}-m_{25}+1)(m_{14}-m_{35}+2)(m_{14}-m_{45}+3)}{(m_{14}-m_{24}+1)(m_{14}-m_{34}+2)(m_{14}-m_{44}+3)}\right]^{1/2}$$
$$\times\left[\frac{(m_{14}-m_{13}+1)(m_{14}-m_{33}+3)(m_{24}-m_{23})(m_{23}-m_{34}+1)(m_{23}-m_{44}+2)}{(m_{14}-m_{24}+2)(m_{14}-m_{34}+3)(m_{14}-m_{44}+4)(m_{13}-m_{23}+1)(m_{23}-m_{33}+1)}\right]^{1/2}$$
$$\times\left[\frac{(m_{23}-m_{22}+1)(m_{13}-m_{12})(m_{12}-m_{33}+2)}{(m_{13}-m_{23})(m_{23}-m_{33}+2)(m_{12}-m_{22}+1)}\right]^{1/2}\left[\frac{(m_{11}-m_{22}+1)}{(m_{12}-m_{22}+2)}\right]^{1/2}. \tag{40}$$

The correctness of the above expression may be verified by either directly using the Baird–Biedenharn formula[2] for the matrix element of E_{ij}, or the full expansion of the commutator $[E_{12},[E_{23},[E_{34},[E_{45},E_{56}]]]]$, and the simple formula for $E_{i-1,i}$. Assuming the set of rules to be true for the evaluation of matrix elements of an arbitrary E_{ij} it is easy to show that the rules hold for matrix elements of $E_{i-1,j}$ (or $E_{i,j+1}$), i.e., the pattern calculus rules can be inductively shown to hold good.

4. DISCUSSION

The study presented in this note may be broadly categorized as consisting of two parts. In Sec. 2 we have carried the valuable suggestions of Moshinsky,[5] Biedenharn and Gifften,[3] Lezuo,[7] and Patterson and Harter[8,9] regarding the use of S_N in generating basis states of the IR's of U(n) to their logical conclusion. The similarity of the results obtained by us with those due to Biedenharn,[2] Moshinsky,[10] and others[1-5] implies the essential correctness of the procedure adapted. The only empiricism used was in weighting $E_{ij} \mid [(m)];\phi_1^{N_1}\phi_2^{N_2}\phi_3^{N_3}\cdots\phi_n^{N_n}\rangle$ by $[(N_i+1)N_j]^{1/2}$ for $i\neq j$. This was dictated by the desire to have as close a correspondence as possible with the results of earlier studies.[15] Equations (18), (19), (23), (24) are determined entirely from permutational symmetry of the monomial tensor products. The major difference between the present work, and that due to

Moshinsky and Nagel,[5] lies in replacing polynomials of lowering generators used in the latter work by the symmetrized combination of the matrix algebra associated with S_N.

The simplicity of form of the decomposition coefficients in Eqs. (19) and (23), and a very useful suggestion from the work of Holman and Biedenharn[13] enabled us to develop a pattern calculus approach in Sec. 3. This, as has been shown by the examples of that section, has led to a very straightforward procedure for determining the matrix elements of E_{ij} of U(n). The rules used are relatively simple and permit computer programs to be developed readily. The major simplification follows here in implicitly incorporating a large number of commutator brackets into the pattern calculus, simply by having "gaps" between two centers in a combined diagram. The phase factor can be readily determined as illustrated by the examples. The interesting feature of the results is that, including the correct phase factor, the reuslts of earlier formalisms have been reproduced using a relatively simple procedure. It should be pointed out that the use of commutator brackets is essential only to arrive at the pattern calculus. The matrix element $\langle [(m')];\phi' \mid E_{ij} \mid [(m)];\phi \rangle$ can be calculated directly by considering the action of the cyclic permutation P^{-1} on $[(m)]$ and comparing with $[(m')]$, where $\phi' = P(E_{ij}\phi)$. This procedure is, however, quite cumbersome and not suitable for generalization.

A computer program based on pattern calculus of Sec. 3 is being developed at present.

ACKNOWLEDGMENT

Sincere thanks are due to Mr. K.V. Dinesha for his critical comments and helpful suggestions.

[1]G.E. Baird and L.C. Biedenharn, J. Math. Phys. **4**, 463 (1963).
[2]G.E. Baird and L.C. Biedenharn, J. Math. Phys. **4**, 1449 (1963).
[3]M. Ciftan and L.C. Biedenharn, J. Math. Phys. **10**, 221 (1969).
[4]M. Ciftan, J. Math. Phys. **10**, 1635 (1969).
[5]J. Nagel and M. Moshinsky, J. Math. Phys. **6**, 682 (1965).
[6]I.M. Gel'fand and M.L. Tsetlin, Dok. Akad. Nauk. **71**, 825 (1950).
[7]K.J. Lezuo, J. Math. Phys. **13**, 1389 (1972).
[8]C.W. Patterson and W.G. Harter, J. Math. Phys. **17**, 1125 (1976).
[9]M. Moshinsky, *Group Theory and the Many-Body Problem*, Documents on Modern Physics (Gordon and Breach, New York, 1968), p. 26.
[10]C.W. Patterson and W.G. Harter, J. Math. Phys. **17**, (1976).
[11]I.G. Kaplan, *Symmetry of Many-Electron Systems* (Academic, New York, 1975).
[12]M. Hamermesh, *Group Theory and Its Application to Physical Problem* (Addison-Wesley, London, 1962).
[13]W.J. Holman, III and L.C. Biedenharn, in *Group Theory and Its Application*, edited by E.M. Loebl (Academic, New York, 1971), Vol. 2, p. 1.
[14]W.G. Harter and C.W. Patterson, Phys. Rev. A **13**, 1067 (1976).
[15]J.D. Louck, Am. J. Phys. **38**, 3 (1970).

Particle statistics from induced representations of a local current group [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 21 September 1979; accepted for publication 15 November 1979)

Representations of the nonrelativistic current group $\mathscr{S} \wedge \mathscr{K}$ are studied in the Gel'fand–Vilenkin formalism, where $\mathscr{S}$ is Schwartz' space of rapidly decreasing functions, and $\mathscr{K}$ is a group of diffeomorphisms of $\mathbb{R}^s$. For the case of N identical particles, information about particle statistics is contained in a representation of $\mathscr{K}_F$ (the stability group of a point $F \in \mathscr{S}'$) which factors through the permutation group S_N. Starting from a quasi-invariant measure μ concentrated on a $\mathscr{K}$ orbit Δ in $\mathscr{S}'$, together with a suitable representation of $\mathscr{K}_F$ for $F \in \Delta$, sufficient conditions are developed for inducing a representation of $\mathscr{S} \wedge \mathscr{K}$. The Hilbert space for the induced representation consists of square-integrable functions on a covering space of Δ, which transform in accordance with a representation of $\mathscr{K}_F$. The Bose and Fermi N-particle representations (on spaces of symmetric or antisymmetric wave functions) are recovered as induced representations. Under the conditions which are assumed, the following results hold: (1) A representation of $\mathscr{S} \wedge \mathscr{K}$ determines a well-defined representation of $\mathscr{K}_F$; (2) equivalent representations of $\mathscr{S} \wedge \mathscr{K}$ determine equivalent representations of $\mathscr{K}_F$; (3) a representation of $\mathscr{K}_F$ induces a representation of $\mathscr{S} \wedge \mathscr{K}$; and (4) equivalent representations of $\mathscr{K}_F$ determine equivalent induced representations.

I. INTRODUCTION

The nonrelativistic quantum mechanics of many identical particles may be described by means of a field $\psi(\mathbf{x})$ satisfying either canonical commutation relations (CCR) and describing bosons

$$[\psi(\mathbf{x}),\psi(\mathbf{y})]_- = [\psi^*(\mathbf{x}),\psi^*(\mathbf{y})]_- = 0\,,$$
$$[\psi(\mathbf{x}),\psi^*(\mathbf{y})]_- = \delta(\mathbf{x}-\mathbf{y})\,, \tag{1.1}$$

or satisfying canonical anticommutation relations (CAR) and describing fermions

$$[\psi(\mathbf{x}),\psi(\mathbf{y})]_+ = [\psi^*(\mathbf{x}),\psi^*(\mathbf{y})]_+ = 0\,,$$
$$[\psi(\mathbf{x}),\psi^*(\mathbf{y})]_+ = \delta(\mathbf{x}-\mathbf{y})\,. \tag{1.2}$$

The statistics of the system is thus determined by the algebra which is to be represented.

In the formulation of nonrelativistic quantum mechanics in terms of particle densities and currents, one defines

$$\rho(\mathbf{x}) = \psi^*(\mathbf{x})\psi(\mathbf{x}),$$
$$\mathbf{J}(\mathbf{x}) = (2i)^{-1}[\psi^*(\mathbf{x})\nabla\psi(\mathbf{x}) - (\nabla\psi^*(\mathbf{x}))\psi(\mathbf{x})]\,, \tag{1.3}$$

and obtains the commutation relations

$$[\rho(\mathbf{x}),\rho(\mathbf{y})] = 0,$$
$$[\rho(\mathbf{x}),J_k(\mathbf{y})] = -i\frac{\partial}{\partial x_k}[\rho(\mathbf{x})\delta(\mathbf{x}-\mathbf{y})]\,,$$
$$[J_j(\mathbf{x}),J_k(\mathbf{y})] = -i\frac{\partial}{\partial x_k}[J_j(\mathbf{x})\delta(\mathbf{x}-\mathbf{y})] + i\frac{\partial}{\partial y_j}[J_k(\mathbf{y})\delta(\mathbf{x}-\mathbf{y})]\,. \tag{1.4}$$

The calculations may be carried out formally, or performed explicitly in the Fock representation of the underlying field; but the same commutator algebra (1.4) is obtained whether one begins with the CCR or the CAR. Thus, in a nonrelativistic current theory, the particle statistics is not determined by the choice of an equal-time algebra, but instead may be determined by the choice of a *representation* of the algebra.[1-3]

The fundamental object of study in the present paper is the group obtained from (1.4) by exponentiating the local currents. This group is a semidirect product $\mathscr{S} \wedge \mathscr{K}$, where $\mathscr{S}$ is Schwartz' space of rapidly decreasing functions, and $\mathscr{K}$ is a certain group of diffeomorphisms of $\mathbb{R}^s$; representations of $\mathscr{K}$ and $\mathscr{S} \wedge \mathscr{K}$ have been studied by several authors.[3-8] In the Gel'fand–Vilenkin formalism for describing representations of nuclear Lie groups,[9] a continuous unitary representation of $\mathscr{S} \wedge \mathscr{K}$ may be characterized by means of (1) a quasi-invariant cylindrical measure μ on $\mathscr{S}'$, the space of tempered distributions, together with (2) a system of unitary mappings satisfying an algebraic compatibility condition, which in this paper we call a "system of multipliers."

To classify the representations of $\mathscr{S} \wedge \mathscr{K}$ with respect to statistics, we follow the approach introduced by Wigner for the Lorentz group and generalized by Mackey to semidirect products of second-countable locally compact groups.[10-12] For $F \in \mathscr{S}'$, we consider the "little group" $\mathscr{K}_F$, the subgroup of $\mathscr{K}$ which leaves F fixed. We shall establish a correspondence, under appropriate conditions, between the system of multipliers defined by a continuous unitary representation of $\mathscr{S} \wedge \mathscr{K}$, and a unitary representation of $\mathscr{K}_F$.

Our main result is that information about particle statistics is fully contained in the representation of the little

[a)]Work supported by the U.S. Department of Energy.

 0022-2488/80/040650-15$1.00

group; and that starting from the measure μ and a representation of $\mathscr{K}_F$, one can obtain a representation of the full group $\mathscr{S} \wedge \mathscr{K}$ by means of an inducing construction (in the sense of Mackey).

As part of this construction, we recover the N-particle representations of $\mathscr{S} \wedge \mathscr{K}$ in a Hilbert space of square-integrable functions. These functions must be chosen to be either symmetric or antisymmetric under exchange of particle coordinates, depending upon which of two inequivalent representations of the little group is selected. These representations of the little group correspond to Bose and Fermi statistics, respectively. Thus, starting from a representation of the little group, we recover the usual connection between particle statistics and the symmetry of wave functions.

The principal difficulty in obtaining our results arises from the fact that the diffeomorphism group $\mathscr{K}$ is not locally compact, and consequently does not possess Haar measure. (Girardello and Parravicini have indicated how the theory of induced representations can be extended to the case of a semidirect product of a nuclear space with a locally compact group.[13]) Thus, Mackey's theory of induced representations, which utilizes Haar measure heavily, cannot be carried over to our problem in a direct fashion. The main mathematical work of this paper will be to show how the topological properties of an orbit on which the measure is concentrated can be used to construct induced representations of $\mathscr{S} \wedge \mathscr{K}$. The Hilbert space for the induced representation consists of square-integrable functions on a covering space of the orbit which tranform in accordance with the little group representation.

Section II is devoted to the necessary background material. After giving a precise definition of the group $\mathscr{S} \wedge \mathscr{K}$, we review the needed concepts from the Gel'fand–Vilenkin and Mackey theories, emphasizing the parallels between them.

In Sec. III, we take a brief step away from the formal development in order to gain some physical intuition about what one should expect to be true in the abstract representation theory. We use the familiar N-particle representations of the current algebra to infer properties of the Gel'fand–Vilenkin system of multipliers. In particular, we note the connection between a system of multipliers and the (relative) phases of the ground state wave function at distinct points; and we examine the topological properties of the orbits and their covering spaces. The little group for an N-particle representation maps into S_N, the permutation group on N objects. We then associate systems of multipliers with representations of S_N, and hence (Bose or Fermi) representations of the little group.

Section IV contains the main mathematical results of the paper. First we review the concepts of a fundamental group, covering space and universal covering space for an orbit Δ in $\mathscr{S}'$. The following results (informally stated here) are then proved: (1) A system of multipliers satisfying appropriate conditions determines a well-defined representation of the little group; (2) systems of multipliers which correspond to unitarily equivalent representations of $\mathscr{S} \wedge \mathscr{K}$ determine unitarily equivalent representations of the little group; (3) under appropriate conditions, a representation of the little group determines an induced representation of $\mathscr{S} \wedge \mathscr{K}$; and (4) equivalent representations of the little group determine equivalent induced representations of $\mathscr{S} \wedge \mathscr{K}$, and hence equivalent systems of multipliers. Collectively, these theorems amount to a partial extension of Mackey's induced representation theory to the non-locally compact group $\mathscr{S} \wedge \mathscr{K}$. The theorems are proved for a class of orbits in $\mathscr{S}'$ which includes, but is not limited to, orbits defining the usual N-particle representations of the current algebra.

Section V summarizes the assumptions made on the orbit and the system of multipliers which are sufficient to permit the characterization of a representation by means of a little group representation. We will have occasion to refer to the interesting recent paper of Vershik, Gel'fand, and Graev, where an important class of representations of $\mathscr{S} \wedge \mathscr{K}$ is constructed.[8] Certain of these representations fit naturally into the framework we describe here. Some interesting open questions are also mentioned.

Appendix A contains the technical result used in Sec. IV, that in a space such as $\mathscr{S}'$ the weakly open sets are measurable. Appendix B reformulates some of our results using cohomology theory. For the class of representations considered here, the systems of multipliers can be identified with cocycles on the orbit.

II. DEFINITIONS AND BACKGROUND

A. The group $\mathscr{S} \wedge \mathscr{K}$

Representations of the current algebra (1.4) can be studied by averaging the densities with test functions from Schwartz' space $\mathscr{S}(\mathbb{R}^s)$ of C^∞ functions which, together with all their derivatives, decrease rapidly at infinity. Then, with

$$\rho(f) = \int d^s x\, \rho(\mathbf{x}) f(\mathbf{x}), \quad f \in \mathscr{S},$$

$$J(\mathbf{g}) = \sum_{i=1}^{s} \int d^s x J_i(\mathbf{x}) g_i(\mathbf{x}), \quad g_i \in \mathscr{S}, \tag{2.1}$$

the algebra becomes

$$\begin{aligned} &[\rho(f_1), \rho(f_2)] = 0, \\ &[\rho(f), J(\mathbf{g})] = i\rho(\mathbf{g}\cdot\nabla f), \\ &[J(\mathbf{g}), J(\mathbf{h})] = iJ(\mathbf{h}\cdot\nabla\mathbf{g} - \mathbf{g}\cdot\nabla\mathbf{h}). \end{aligned} \tag{2.2}$$

The argument of J on the right hand side of Eq. (2.2) is the Lie bracket of the vector fields $\mathbf{g}$ and $\mathbf{h}$. Thus, J is a representation of the Lie algebra of C^∞ vector fields on $\mathbb{R}^s$ which, together with all derivatives, decrease rapidly at infinity.

The Lie algebra (2.2) will in general be represented by self-adjoint unbounded operators. It is useful to study instead the corresponding infinite-dimensional Lie group and its continuous unitary representations. Each vector field $\mathbf{g}$ is the infinitesimal generator of a one-parameter group $\phi_t^{\mathbf{g}}$ of C^∞ diffeomorphisms or *flows* on $\mathbb{R}^s$, satisfying

$$\begin{aligned} &\frac{\partial \phi_t^{\mathbf{g}}(\mathbf{x})}{\partial t} = \mathbf{g}(\phi_t^{\mathbf{g}}(\mathbf{x})), \\ &\phi_{t=0}^{\mathbf{g}}(\mathbf{x}) = \mathbf{x}. \end{aligned} \tag{2.3}$$

A group containing these flows as continuous one-parameter subgroups is the following group $\mathscr{K}$ of C^∞ diffeomorphisms of $\mathbb{R}^s$. Let $\mathscr{K}_0$ be the group of all C^∞ diffeomorphisms from $\mathbb{R}^s$ onto $\mathbb{R}^s$ having compact support, with the operation of composition. $\mathscr{K}_0$ may be topologized by means of the countable family of metrics

$$\ll\phi,\psi\gg_n = \max_{0\le|m|\le n}\sup_{\mathbf{x}\in\mathbb{R}^s}\left|(1+|\mathbf{x}|^2)^n[\phi^{(m)}(\mathbf{x})-\psi^{(m)}(\mathbf{x})]\right|,$$

for $n=0,1,2,\ldots$; where $(m)=(m_1,\ldots m_s)$ is an s-tuple of non-negative integers; $|m|=\Sigma_{k=1}^s m_k$; and $\phi^{(m)}(\mathbf{x})$ $=\partial^{|m|}\phi(\mathbf{x})/(\partial x_1)^{m_1}\cdots(\partial x_s)^{m_s}$. $\mathscr{K}$ may be defined as the completion of $\mathscr{K}_0$ with respect to this topology. $\mathscr{K}$ is then a topological group; the topology is metrizable, and $\mathscr{K}$ has a countable basis of neighborhoods of each point. $\mathscr{K}$ contains diffeomorphisms which are not of compact support but which approximate the identity mapping as $|\mathbf{x}|\to\infty$. It has been conjectured, but not explicitly demonstrated, that if $\mathbf{g}$ has components in $\mathscr{S}$ then $\phi_t^{\mathbf{g}}$ is an element of $\mathscr{K}$ and the mapping $\mathbf{g}\to\phi_t^{\mathbf{g}}$ is continuous.[14]

We conjecture that $\mathscr{K}$ is pathwise connected and simply connected, although we are not aware of any easy proof. In any case, $\mathscr{K}$ is locally path-connected, and the path component of $\mathscr{K}$ which contains the identity also contains the one-parameter subgroups, which are the elements of $\mathscr{K}$ of physical interest.

Considering $\rho(f)$ and $J(\mathbf{g})$ to be self-adjoint operators, the corresponding one-parameter continuous unitary groups are

$$\begin{aligned}&\mathscr{U}(tf)=\exp[it\rho(f)],\\ &\mathscr{V}(\phi_t^{\mathbf{g}})=\exp[itJ(\mathbf{g})].\end{aligned}\tag{2.4}$$

From Eq. (2.2), these satisfy the group law

$$\begin{aligned}&\mathscr{U}(f_1)\mathscr{V}(\psi_1)\mathscr{U}(f_2)\mathscr{V}(\psi_2)\\ &\quad=\mathscr{U}(f_1+f_2\circ\psi_1)\mathscr{V}(\psi_2\circ\psi_1),\end{aligned}\tag{2.5}$$

for $f_1,f_2\in\mathscr{S}$ and $\psi_1,\psi_2\in\mathscr{K}$, where $(f_2\circ\psi_1)(\mathbf{x})$ denotes $f_2(\psi_1(\mathbf{x}))$. Thus, the appropriate object of study is the semidirect product $\mathscr{S}\wedge\mathscr{K}$, with the group law defined by $(f_1,\psi_1)(f_2,\psi_2)=(f_1+f_2\circ\psi_1,\psi_2\circ\psi_1)$. In the topology for $\mathscr{K}$ introduced above, the group operations in $\mathscr{S}\wedge\mathscr{K}$ are continuous (when $\mathscr{S}$ has its usual topology as a nuclear space).[5,14] Then $\mathscr{S}\wedge\mathscr{K}$ defines a topological group in which the one-parameter subgroups $(tf,\mathbf{e})$ and $(0,\phi_t^{\mathbf{g}})$ are continuous, where $\mathbf{e}(\mathbf{x})=\mathbf{x}$ is the identity in $\mathscr{K}$. A continuous unitary representation of $\mathscr{S}\wedge\mathscr{K}$ given by $W(f,\psi)=\mathscr{U}(f)\mathscr{V}(\psi)$ permits recovery of the self-adjoint operators $\rho(f)$ and $J(\mathbf{g})$ as the infinitesimal generators of $\mathscr{U}(tf)$ and $\mathscr{V}(\phi_t^{\mathbf{g}})$, respectively.

A topological group is said to be locally compact if every point has a neighborhood whose closure is compact. This is an important condition for utilization of Mackey's theory to classify its continuous unitary representations (see below). However, neither $\mathscr{S}$ nor $\mathscr{K}$ is locally compact in the above topologies.

B. Application of the Gel'fand–Vilenkin formalism (Refs. 3,5–9,15)

Let $\mathscr{S}$ be a nuclear space (such as Schwartz' space). Its continuous dual $\mathscr{S}'$ is equipped with a σ algebra of measurable sets generated by the cylinder sets with Borel base. A cylindrical measure μ is a countably additive normalized positive measure on this σ algebra. A functional $L(f)$ on $\mathscr{S}$ is the Fourier transform of a cylindrical measure μ, i.e.,

$$L(f)=\int_{\mathscr{S}'}\exp[i(F,f)]d\mu(F),\tag{2.6}$$

where $F\in\mathscr{S}'$, if and only if $L(f)$ satisfies the following conditions: (1) $L(f)$ is continuous with respect to the topology of $\mathscr{S}$; (2) $L(0)=1$; and (3) $L(f)$ is positive definite in the sense that $(\forall f_1,\ldots,f_n\in\mathscr{S})\,(\forall\lambda_1,\ldots,\lambda_n\in\mathbb{C})$

$$\sum_{j,k=1}^n\bar\lambda_k\lambda_j L(f_j-f_k)\ge 0.\tag{2.7}$$

Let $\mathscr{U}$ be a continuous unitary cyclic representation of $\mathscr{S}$ in a Hilbert space $\mathscr{H}$ with cyclic vector Ω. The functional $L(f)=(\Omega,U(f)\Omega)$ satisfies the above conditions, defining a cylindrical measure μ. Then $\mathscr{H}$ can be realized as $\mathscr{L}^2_\mu(\mathscr{S}')$, Ω as the μ-square-integrable function $\Omega(F)\equiv 1$, and $\mathscr{U}(f)$ as the multiplication operator $\exp[i(F,f)]$ in $\mathscr{L}^2_\mu(\mathscr{S}')$. Conversely, every cylindrical measure μ on $\mathscr{S}'$ defines a continuous representation of $\mathscr{S}$ in $\mathscr{L}^2_\mu(\mathscr{S}')$ with cyclic vector $\Omega(F)\equiv 1$. Two continuous unitary cyclic representations $\mathscr{U}_1$ and $\mathscr{U}_2$ of $\mathscr{S}$ are unitarily equivalent if and only if the corresponding cylindrical measures μ_1 and μ_2 are equivalent (i.e., have the same sets of measure zero). The second representation may then be thought of as differing from the first only in the choice of a different cyclic vector in $\mathscr{L}^2_\mu(\mathscr{S}')$.

With $\mathscr{U}(f)=\exp[i\rho(f)]$, the self-adjoint operator $\rho(f)$ is represented in $\mathscr{L}^2_\mu(\mathscr{S}')$ as multiplication by (F,f).

Suppose that $\mathscr{S}\wedge\mathscr{K}$ is the semidirect product of $\mathscr{S}$ with a topological group $\mathscr{K}$ (such as the diffeomorphism group above), so that the mapping $\mathscr{S}\times\mathscr{K}\to\mathscr{S}$ defining the group operation is jointly continuous in the topologies[5,14] of $\mathscr{S}$ and $\mathscr{K}$. For $\psi\in\mathscr{K}$, the action of ψ on $\mathscr{S}$ defines a dual action $\psi^*:\mathscr{S}'\to\mathscr{S}'$ by the equation $(\psi^*F,f)=(F,f\circ\psi)$. ψ^* is linear and continuous in the weak topology of $\mathscr{S}'$. In fact, ψ^*F is jointly continuous from $\mathscr{K}\times\mathscr{S}'$ into $\mathscr{S}'$, since (F,f) is jointly continuous and $f\circ\psi$ is continuous. If $X\subseteq\mathscr{S}'$ is a measurable set, then ψ^*X is measurable, since ψ^* maps cylinder sets onto cylinder sets. A cylindrical measure μ is called quasi-invariant for $\mathscr{K}$ if ψ^* preserves sets of μ-measure zero.

Let $\mathscr{U}(f)\mathscr{V}(\psi)$ be a continuous unitary representation of $\mathscr{S}\wedge\mathscr{K}$ in $\mathscr{H}$, with $\Omega\in\mathscr{H}$ cyclic for $\mathscr{U}$. Then the measure μ defined by Eq. (2.6) with $L(f)=(\Omega,\mathscr{U}(f)\Omega)$ is quasi-invariant for $\mathscr{K}$. In $\mathscr{L}^2_\mu(\mathscr{S}')$, the representation $\mathscr{U}(f)\mathscr{V}(\psi)$ becomes

$$\begin{aligned}&[\mathscr{U}(f)\Psi](F)=\exp[i(F,f)]\Psi(F),\\ &[\mathscr{V}(\psi)\Psi](F)=\chi_\psi(F)\Psi(\psi^*F)\left[\frac{d\mu(\psi^*F)}{d\mu(F)}\right]^{1/2},\end{aligned}\tag{2.8}$$

where $\Psi(F)$ is a function in $\mathscr{L}^2_\mu(\mathscr{S}')$; $[d\mu(\psi^*F)/d\mu(F)]$ is the Radon–Nikodym derivative of the transformed measure with respect to the original measure; and $\chi_\psi(F)$ is a complex-valued function of modulus one, depending on ψ, defined almost everywhere, and satisfying (for each pair ψ_1,ψ_2) the compatibility condition

$$\chi_{\psi_2}(F)\chi_{\psi_1}(\psi_2^* F) = \chi_{\psi_1\circ\psi_2}(F) \tag{2.9}$$

almost everywhere. That is to say, for each ψ, there is a set $\mathscr{Z}_\psi$ of measure zero outside of which $\chi_\psi(F)$ is defined. Likewise, for each pair ψ_1,ψ_2 there is a set $\mathscr{Z}_{\psi_1,\psi_2}$ of measure zero outside of which Eq. (2.9) holds. A set of functions $\chi_\psi(F)$ satisfying these conditions will be called a *system of multipliers*.

The choice $\chi_\psi(F)\equiv 1$, $\forall\psi$, satisfies Eq. (2.9), and defines a representation which may or may not be equivalent to the original.

The representation (2.8) of $\mathscr{S}\wedge\mathscr{K}$ is irreducible if and only if μ is ergodic for $\mathscr{K}$, i.e., if for any measurable invariant set X in $\mathscr{S}'$, either $\mu(X)=0$ or $\mu(\mathscr{S}'\setminus X)=0$. For $F\in\mathscr{S}'$, the set $\Delta=\{\psi^* F\,|\,\psi\in\mathscr{K}\}$ is called the *orbit* of F under the action of $\mathscr{K}$; an invariant set is the union of a (generally uncountable) family of mutually disjoint orbits. Supposing that the orbits comprising X are measurable, there are two ways in which a cylindrical measure μ can be ergodic: it can be concentrated on a single orbit, or else every orbit can be of μ-measure zero. Both situations occur when $\mathscr{K}$ is the diffeomorphism group discussed above.

C. Comparison with the Mackey theory

The Mackey theory[11,12] considers continuous unitary representations of topological groups satisfying the technical assumptions of being second countable (i.e., the open sets have a countable basis) and locally compact. Some minor changes in Mackey's notation emphasize the parallels with the preceding section.

Let $G=N\wedge H$, where N is an Abelian normal subgroup of G. $\hat{N}$ denotes the character group of N, which has an orbit structure under the action h^* of elements of H. For $F\in\hat{N}$, $(h^* F)(n)=F(hn)$. The closed subgroup H_F of H which leaves F fixed is called the stability group or little group associated with F. If F and F' are elements of the same orbit Δ of $\hat{N}$, H_F and $H_{F'}$ are isomorphic, and the points of Δ may be identified with the elements of the quotient space H/H_F (the space of right cosets).

Since H is locally compact, there exists a right-invariant measure (Haar measure) on the σ algebra of Borel sets in H. Letting $p: H\to H/H_F$ be the surjective mapping $p(h)=H_F h$, there exist quasi-invariant measures in H/H_F: letting ν be any finite measure in the measure class of Haar measure, define $\tilde{\nu}(X)=\nu(p^{-1}(X))$ for any set X in H/H_F such that $p^{-1}(X)$ is Borel in H. The quasi-invariant measures on H/H_F are equivalent.

Given an irreducible continuous unitary representation L of H_{F_0}, for some $F_0\in\hat{N}$, with values in a Hilbert space $\mathscr{M}$, one obtains the (irreducible) *induced representation* $\mathscr{U}(n)\mathscr{V}(h)$ of $N\wedge H$. Consider the Borel functions $\hat{\Psi}$ from H to $\mathscr{M}$ which satisfy $\hat{\Psi}(kh)=L_k(\hat{\Psi}(h))$ for every $k\in H_{F_0}$. If $\hat{\Psi}$ satisfies this equation, $\hat{\Psi}_g(h)=\hat{\Psi}(hg)$ also satisfies it. Then the function $(\hat{\Psi}(h),\hat{\Psi}(h))$ is a constant on every coset $H_{F_0}h$, and may be integrated with respect to the quasi-invariant measure $\tilde{\nu}$ in H/H_F. Define the Hilbert space $\mathscr{H}$ to be the set of all $\hat{\Psi}$ such that $\int_{H/H_F}(\hat{\Psi}(k),\hat{\Psi}(k))\,d\tilde{\nu}$ is finite. Then

$$\begin{aligned}(\mathscr{U}(n)\hat{\Psi})(h) &= (h^* F_0)(n)\hat{\Psi}(h),\\ (\mathscr{V}(g)\hat{\Psi})(h) &= \hat{\Psi}(hg)\left[\frac{d\tilde{\nu}(H_{F_0}hg)}{d\tilde{\nu}(H_{F_0}h)}\right]^{1/2},\end{aligned} \tag{2.10}$$

define the representation $\mathscr{U}(n)\mathscr{V}(g)$.

These results can be expressed in a way which brings out their relationship to the Gel'fand–Vilenkin theory more clearly. Consider a quasi-invariant measure μ on Borel sets in $\hat{N}$. Define the space $\mathscr{L}^2_\mu(\hat{N},\mathscr{M})$, where $\Psi\in\mathscr{L}^2_\mu(\hat{N},\mathscr{M})$ is a square-integrable Borel cross section with values in $\mathscr{M}$, i.e., to each point $F\in\hat{N}$, we associate a copy of $\mathscr{M}$ denoted $\mathscr{M}_F$ and for each F, $\Psi(F)\in\mathscr{M}_F$. To say that Ψ is Borel means that when $\hat{N}\times\mathscr{M}$ is given the product Borel structure it becomes a Hilbert bundle on which $(F,\Phi)\to(\Psi(F),\Phi)$ is a Borel function.

Let $(F,h)\to L_h(F)$ be a Borel mapping from $\hat{N}\times H$ into the unitary operators on $\mathscr{M}$, satisfying the equation

$$L_{h_1,h_2}(F)=L_{h_1}(F)L_{h_2}(h_1^* F). \tag{2.11}$$

$L_h(F)$ may be regarded as a mapping from $\mathscr{M}_{h^* F}$ to $\mathscr{M}_F$, and Eq. (2.11) as a compatibility condition on these mappings. Then in $\mathscr{L}^2_\mu(\hat{N},\mathscr{M})$ we have the representation

$$\begin{aligned}(\mathscr{U}(n)\Psi)(F) &= F(n)\Psi(F),\\ (\mathscr{V}(h)\Psi)(F) &= L_h(F)\Psi(h^* F)\left[\frac{d\mu(h^* F)}{d\mu(F)}\right]^{1/2}.\end{aligned} \tag{2.12}$$

For the representation (2.12) to be irreducible, μ must be ergodic with respect to the action of H on $\hat{N}$, and there are again two ways in which this can occur. Either μ is concentrated on a single orbit Δ, or else (the *strictly ergodic* case) the measure of every orbit is zero.

When μ is concentrated on a single orbit Δ, select $F_0\in\Delta$. Then Δ corresponds to H/H_{F_0} and $L_h(F_0)$ defines a representation of the little group H_{F_0}. The representation (2.12) is in this case equivalent to the induced representation (2.10) obtained from the representation of the little group. If it is possible to select a Borel subset of $\hat{N}$ which meets each orbit in just one point, then every irreducible representation of $N\wedge H$ can be obtained from an orbit Δ together with a representation of the little group H_{F_0} for $F_0\in\Delta$.

Thus, for locally compact groups $N\wedge H$ the Mackey theory can be expressed in a fashion which is analogous to the Gel'fand–Vilenkin theory: L_h plays the role of the Gel'fand–Vilenkin system of multipliers χ_ψ, $\hat{N}$ is the analog of $\mathscr{S}'$ and μ is in each case a suitable quasi-invariant measure. In Sec. IV, we will show how this parallelism can be extended to the non-locally compact group $\mathscr{S}\wedge\mathscr{K}$.

III. STATISTICS FROM N-PARTICLE REPRESENTATIONS OF THE CURRENT ALGEBRA

To understand how particle statistics is described in the local current formulation of quantum mechanics, we first discuss the N-particle representations of the algebra. This serves to identify the important properties of particle statistics which are to be carried over to the more general setting of the Gel'fand–Vilenkin representation theory.

A. N-particle representation of the current group $\mathscr{S} \wedge \mathscr{K}$ (Refs. 3,15,16)

We consider a system of N identical spinless particles for which there exists a time reversal-invariant ground state. As in Sec. II, we consider the unitary operators $\mathscr{U}(f) = \exp[i\rho(f)]$ and $\mathscr{V}(\phi_t^g) = \exp[itJ(g)]$. Let Ψ_N be an element of $\mathscr{L}^2_{\pm}(\mathbb{R}^{sN})$, the Hilbert space of symmetric (+) or antisymmetric (−) square integrable functions of N vector variables in s space dimensions. The equations

$$\mathscr{U}(f)\Psi_N = \exp\left[i\sum_{j=1}^{N} f(\mathbf{x}_j)\right]\Psi_N ,$$

$$\mathscr{V}(\psi)\Psi_N(\mathbf{x}_1,\cdots\mathbf{x}_N) = \Psi_N(\psi(\mathbf{x}_1),\cdots,\psi(\mathbf{x}_N))\prod_{j=1}^{N}\left[\det\left(\frac{\partial\psi^k}{\partial x^l}(\mathbf{x}_j)\right)\right]^{1/2}, \quad (3.1)$$

with $f\in\mathscr{S}$ and $\psi\in\mathscr{K}$, define irreducible representations, called the N-particle Bose (+) and Fermi (−) representations, of the group $\mathscr{S}\wedge\mathscr{K}$.

The N-particle representations can be described in the Gel'fand–Vilenkin formalism as follows: One first chooses a normalized vector Ω which is cyclic for the $\mathscr{U}(f)$'s, i.e., $\{\mathscr{U}(f)\Omega\}$ spans a dense subset of the Hilbert space. The cyclic vector can ordinarily be chosen to be the ground state of the system, but any other normalized wave function which is nonzero almost everywhere would suffice.

Next let $F_{\mathbf{x}_0}\in\mathscr{S}'$ be given by $(F_{\mathbf{x}_0}, f) = f(\mathbf{x}_0)$, i.e., $F_{\mathbf{x}_0} = \delta^{(s)}(\mathbf{x}-\mathbf{x}_0)$. Noting that $\psi^* F_{\mathbf{x}_0} = F_{\psi(\mathbf{x}_0)}$, we introduce the *single orbit*

$$\Delta_N^{(s)} = \left\{F = \sum_{j=1}^{N} F_{\mathbf{x}_j}\,\middle|\,\mathbf{x}_i\neq\mathbf{x}_j,\ \text{for } i\neq j\right\}\subset\mathscr{S}' . \quad (3.2)$$

The orbit $\Delta_N^{(s)}$ may be identified with the *configuration space* $\Gamma_N^{(s)}$ of N particles, which consists of the collection of (unordered) sets of N *distinct* points in $\mathbb{R}^s$. Thus, $F = \Sigma_{j=1}^{N} F_{\mathbf{x}_j} \in\Delta_N^{(s)}$ is uniquely identified with $\{\mathbf{x}_1,\cdots,\mathbf{x}_N\}\in\Gamma_N^{(s)}$. $\Delta_N^{(s)}$ is in the σ algebra generated by cylinder sets with Borel base in $\mathscr{S}'$. A quasi-invariant measure μ concentrated on $\Delta_N^{(s)}$ can be written as $\mu(X) = \int_X d\mu(F)$, where

$$d\mu(F_{\mathbf{x}_1} + \cdots + F_{\mathbf{x}_N}) = |\Omega(\mathbf{x}_1,\cdots,\mathbf{x}_N)|^2 d^s x_1\cdots d^s x_N . \quad (3.3)$$

Finally, we define a system of multipliers by

$$\chi_\psi(F) = \mathrm{Arg}[\Omega(\psi(\mathbf{x}_1),\cdots,\psi(\mathbf{x}_N))/\Omega(\mathbf{x}_1,\cdots,\mathbf{x}_N)] . \quad (3.4)$$

Both μ and χ are well defined as long as Ω is either symmetric or antisymmetric with respect to the exchange of any pair of coordinates.

The measure μ and the system of multipliers χ_ψ determine a continous unitary representation of the current group $\mathscr{S}\wedge\mathscr{K}$ on $\mathscr{L}^2_\mu(\mathscr{S}')$, according to Eq. (2.8) of Sec. II, equivalent to the representation given by Eq. (3.1). The self-adjoint operators $\rho(f)$ and $J(\mathbf{g})$ are recovered as infinitesimal generators of the corresponding one-parameter unitary groups:

$$\rho(f)\Psi_N = \sum_{j=1}^{N} f(\mathbf{x}_j)\Psi_N ,$$

$$J(\mathbf{g})\Psi_N = \sum_{j=1}^{N}\frac{1}{2i}[\mathbf{g}(\mathbf{x}_j)\cdot\nabla_j + \nabla_j\cdot\mathbf{g}(\mathbf{x}_j)]\Psi_N . \quad (3.5)$$

Since ρ and $\mathbf{J}$ are symmetric in the particle coordinates, and we are considering representations which are cyclic for the $\rho(f)$'s, the symmetry or antisymmetry of any wave function corresponds to that of the cyclic vector in the representation. Since the symmetry of a wave function $\Omega(\mathbf{x}_1,\cdots,\mathbf{x}_N)$ is determined by its relative phase at certain points, rather than by its magnitude, we find that when working with the local currents, information about particle statistics is contained in the properties of the system of multipliers, rather than in the measure μ.

From a physical point of view, it is to be expected that information about statistics is contained in the representation of the local current algebra (1.4). This is because $\mathbf{J}(\mathbf{x})$ can generate local rotations, and hence one can find an operator $\mathscr{V}(\psi)$ which exchanges a given pair of particle coordinates.[2] Information about the change in phase of the wave function under the action of $\mathscr{V}(\psi)$ is contained in χ_ψ. However, the case $s = 1$ is an exception, since there will be no local rotation which exchanges pairs of coordinates. In this case the representations (3.1) are unitarily equivalent for bosons and fermions and it is necessary to adjoin another operator such as the Hamiltonian in order to distinguish them.[3]

B. Statistics from the little group representation

Next we show how the information about statistics is contained in the representation of the little group determined by the system of multipliers.

Let $F\in\Delta_N^{(s)}$. The *little group* $\mathscr{K}_F$ is given by $\{\phi\in\mathscr{K}\,|\,\phi^* F = F\}$. It follows from Eq. (3.4) that $\chi_\phi(F)$ restricted to $\phi\in\mathscr{K}_F$ determines a continuous unitary representation of the little group by complex numbers.

Let $E\subset\mathscr{K}_F$ be the path component of the identity in $\mathscr{K}_F$, i.e., $\phi\in E$ if there is a continuous path ψ_t in $\mathscr{K}_F$ such that $\psi_{t=0} = \mathbf{e}$ and $\psi_{t=1} = \phi$. Then E is a normal subgroup of $\mathscr{K}_F$ and the path components of $\mathscr{K}_F$ correspond to the elements of the quotient group $\mathscr{K}_F/E$. In configuration space, with F identified with $\{\mathbf{x}_1,\cdots,\mathbf{x}_N\}$, we have $\phi\in\mathscr{K}_F$ if $\{\mathbf{x}_1,\cdots,\mathbf{x}_N\} = \{\phi(\mathbf{x}_1),\cdots,\phi(\mathbf{x}_N)\}$. We see that two elements of a path component of $\mathscr{K}_F$ must implement the same permutation of the N-tuple of points $\{\mathbf{x}_1,\cdots,\mathbf{x}_N\}$. Thus, there is a homomorphism from $\mathscr{K}_F/E$ to the group S_N of permutations on N objects. For the N-particle representations of the current group, the value of $\chi_\psi(F)$ is the same for all ψ in a component of $\mathscr{K}_F$, and the representation of $\mathscr{K}_F$ factors through S_N.

If $s\geqslant 2$, the N-particle multiplier $\chi_\psi(F)$ determines a representation of S_N, since the above homomorphism is surjective. This representation of S_N is just the representation of S_N determined by the symmetry or antisymmetry of the wave functions. The symmetric representation corresponds to bosons, the antisymmetric one to fermions. As a result we have the fact that the representations of S_N determined by $\chi_\psi(F)$ are equivalent for almost all $F\in\Delta_N^{(s)}$.

Only the one-dimensional representations of S_N occur here because we are considering spinless particles, having representations that are cyclic for the $\rho(f)$'s. It is because there are just two one-dimensional representations of S_N that only totally symmetric or totally antisymmetric cyclic

vectors generate well-defined systems of multipliers.

It has been shown that, for $s \geqslant 2$, the Bose and Fermi N-particle representations of the current algebra are unitarily inequivalent.[2,3] This may be regarded as a consequence of the inequivalence of the Bose and Fermi representations of the little group, as will be proved in Sec. IV.

The system of multipliers actually contains more information than is necessary for specifying the representation of the little group. For example, the system of multipliers $\chi_\psi(F) \equiv 1$ for all ψ and F defines a Bose representation, which results from choosing the cyclic vector to be real and strictly positive (e.g., the ground state of the system). Other Bose systems of multipliers correspond to other choices for the cyclic vector. Though the systems of multipliers thus obtained are different, they define unitarily equivalent representations of the group $\mathscr{S} \wedge \mathscr{K}$ and lead to equivalent representations of the the little group.

The system of multipliers corresponding to a given cyclic vector may be ill defined for some choices of ψ and F. For example, if the wave function for the cyclic vector vanishes at some points (a set of measure zero), its phase at such points is ill defined, and $\chi_\psi(F)$ is not well defined everywhere by Eq. (3.4). This is the case if the cyclic vector is the ground state of an N-particle Fermi system, since a continuous real-valued antisymmetric wave function $\Omega(\mathbf{x}_1,\cdots,\mathbf{x}_N)$ must have nodal surfaces in $\mathbb{R}^{sN}$. Moreover, the cyclic vector can be changed arbitrarily on a set of measure zero. For particle statistics to be well defined, we need to exclude the possibility that this can happen in such a way as to affect the representation of the little group.

Cyclic vectors describing physical systems are (almost everywhere) continuous nonzero wave functions, and lead to multipliers $\chi_\psi(F)$ which are jointly continuous in ψ and F except for $\psi^* F$ or $F \in \mathscr{Z}$, where $\mathscr{Z} \subset \Delta_N^{(s)}$ is a set of measure zero. This condition will be shown to be sufficient to uniquely determine a representation of the little group $\mathscr{K}_F$ for $F \notin \mathscr{Z}$. If the cyclic vector Ω is chosen to be the ground state of a Fermi system, the set $\mathscr{Z}$ becomes $\{F_{\mathbf{x}_1} + \cdots + F_{\mathbf{x}_N} | \Omega(\mathbf{x}_1,\cdots,\mathbf{x}_N) = 0\}$.

Although the phase of the cyclic vector may not have a natural definition everywhere, it can be defined everywhere (rather arbitrarily) in a manner consistent with the symmetry of the representation. Consequently, the system of multipliers defined by Eq. (3.4) can be chosen so that Eq. (2.9) actually holds everywhere.

C. The orbit $\Delta_N^{(s)}$ and its covering space $\mathbb{R}^{sN} \setminus D$

We have outlined how the usual formulation of the quantum mechanics of N identical particles in terms of wave functions can be expressed using a measure and a system of multipliers in the Gel'fand–Vilenkin formalism, leading to a representation of a "little group" which describes the particle statistics. We next discuss the converse: how one can start from a measure on a single orbit $\Delta_N^{(s)}$ and a representation of the little group, and recover the usual wave function formulation of quantum mechanics, in which the symmetry of the wave functions under interchange of coordinates characterizes the statistics.

In Sec. II we pointed out that in the Mackey theory one considers functions on the *group* $\mathscr{K}$ *itself*, which transform in accordance with the given representation of the little group. Since the group is locally compact, it possesses Haar measure, with respect to which the Hilbert space is constructed for a unitary representation of $\mathscr{K}$.

To generalize the inducing construction to the non-locally compact group of concern here, we will need to make use of the topological properties of the *orbit*, rather than those of the group. For the N-particle representations, the original wave functions can be defined on a *coordinate space*, which consists of ordered N-tuples of *distinct* coordinates in $\mathbb{R}^s$. We denote this space by $\mathbb{R}^{sN} \setminus D$, where $D = \{(\mathbf{x}_1,\cdots,\mathbf{x}_N) | \mathbf{x}_i = \mathbf{x}_j \text{ for some } i \neq j\}$. Since D is of measure 0 in $\mathbb{R}^{sN}$, omitting this set from the coordinate space does not affect the wave functions. This coordinate space is a *covering space* for the orbit $\Delta_N^{(s)}$. We denote the natural projection of the coordinate space onto the configuration space by $p: \mathbb{R}^{sN} \setminus D \to \Delta_N^{(s)}$ and use the notation $\tilde{F} \in \mathbb{R}^{sN} \setminus D$ if $p(\tilde{F}) = F \in \Delta_N^{(s)}$.

The action of the elements of $\mathscr{K}$ on the orbit $\Delta_N^{(s)}$ can be lifted to the covering space $\mathbb{R}^{sN} \setminus D$ so that if $p(\tilde{F}) = F$, $p(\psi^* \tilde{F}) = \psi^* F$; and so that $\psi_1^* \tilde{F} = \psi_2^* \tilde{F}$ if $\psi_1^{-1} \circ \psi_2$ is in the path component of $\mathscr{K}_F$ connected to the identity. For $\tilde{F} = (\mathbf{x}_1,\cdots,\mathbf{x}_N)$, this action is given by $\psi^* \tilde{F} = (\psi(\mathbf{x}_1),\cdots,\psi(\mathbf{x}_N))$. Due to the nontrivial connectivity of $\Delta_N^{(s)}$, the system of multipliers $\chi_\psi(F)$ cannot be expressed as a single-valued function of the two variables F and $\psi^* F$ in $\Delta_N^{(s)}$; however, $\chi_\psi(F)$ *can* be written as a function of $\tilde{F}$ and $\psi^* \tilde{F}$ by means of Eq. (3.4).[17]

The quasi-invariant measure μ on $\Delta_N^{(s)}$ can also be lifted to a measure $\tilde{\mu}$ on $\mathbb{R}^{sN} \setminus D$ which is equivalent to the usual Lebesgue measure. Thus, the usual wave function representation is recovered.

In Sec. IV we show how the above example can be generalized to carry out an inducing construction yielding a class of representations of $\mathscr{S} \wedge \mathscr{K}$.

Remarks: (1) In a recent paper, Bloore and Swarbrick also seek to describe quantum mechanical wave functions by means of functions on configuration space.[18] Defining the projection mapping from $\mathbb{R}^{3N} \setminus D$ onto $\Gamma_N^{(3)}$, they obtain a correspondence between rays of Bose or Fermi wave functions on $\mathbb{R}^{3N}$ and eqivalence classes of functions from $\mathbb{R}^{3N}/S_N$ to $\mathbb{C}/S_2$ which satisfy an appropriate (Bose or Fermi) homotopy condition. However, their classes of functions do not form a linear space.

In our opinion, the nonrelativistic current algebra in a Gel'fand–Vilenkin representation provides a natural context for the examination of wave functions describing identical particles as functions on configuration space.

(2) It is interesting that the topology of the N-particle orbit depends critically on the number of space dimensions s. The nontrivial connectedness properties of the orbit are a consequence of the fact that configurations in which two or more points coincide are not included.[19]

For $s \geqslant 3$, the coordinate space $\mathbb{R}^{sN} \setminus D$ is simply connected, and hence it may be identified with the *universal* covering space of $\Delta_N^{(s)}$. For $s = 2$, the coordinate space

$\mathbb{R}^{2N}\setminus D$ is multiply connected. In carrying out the inducing construction in Sec. IV, we are led to functions on the universal covering space of the orbit. Consequently, for $s = 2$ there may be representations other than the usual ones corresponding to wave functions on $\mathbb{R}^{2N}\setminus D$. These are not discussed further in this paper.

For $s = 1$, the orbit is simply connected. The configuration space $\mathbb{R}^N\setminus D$ decomposes into disconnected components, each of which is isomorphic to the orbit. The little group possesses only one component, the identity component. As a consequence, the representations of the current group corresponding to Bose and Fermi systems are unitarily equivalent.

IV. INDUCED REPRESENTATIONS OF THE CURRENT GROUP $\mathscr{S}\wedge\mathscr{K}$

This section is devoted to establishing the main mathematical results of the paper: sufficient conditions for the connection between systems of multipliers and representations of the little group; and the construction of induced representations of $\mathscr{S}\wedge\mathscr{K}$ starting from certain representations of the little group.

We assume a quasi-invariant cylindrical measure μ concentrated on a single measurable orbit Δ in $\mathscr{S}'$. In Appendix A it is noted that any weakly open set in $\mathscr{S}'$ is contained in the σ algebra generated by the cylinder sets with Borel base. Endowing Δ with the restricted weak topology, any open set in Δ is also measurable.

The main difficulty to be overcome is the absence of Haar measure on $\mathscr{K}$. Thus, there is no naturally defined measure on the product space $\Delta\times\mathscr{K}$ which would permit us to associate with a system of multipliers a well-defined representation of the little group. In fact, Eq. (2.9) holds only *almost* everywhere for each pair of elements of $\mathscr{K}$. If the elements of $\mathscr{K}$ are both chosen to belong to the little group $\mathscr{K}_{F_0}$, it is possible that F_0 happens to belong to the set of measure zero where Eq. (2.9) fails, and thus a representation of the little group is not even defined. This difficulty is met by requiring the system of multipliers to satisfy a certain continuity condition.

The main technique will be to construct the universal covering space $\tilde{\Delta}$ of Δ, and to consider representations of the little group which factor through a representation of the fundamental group of Δ. In this fashion we shall obtain induced representations of $\mathscr{S}\wedge\mathscr{K}$ which include the N-particle representations discussed in Sec. III, and which give us their classification as Bose or Fermi representations.

First we outline the necessary background concerning covering spaces.

A. Homotopy for orbits in $\mathscr{S}'$ under the action of $\mathscr{K}$ (Ref. 20)

Let Δ be a measurable orbit in $\mathscr{S}'$ under the action of $\mathscr{K}$, endowed with the restricted weak topology. The orbit Δ is *connected* because $\mathscr{K}$ is connected. A *path* in Δ is a continuous mapping from the unit interval I into Δ. We shall assume Δ to be *locally pathwise connected* (i.e., there exists a neighborhood basis of path-connected open sets), and *semilocally simply connected* (i.e., every point $F\in\Delta$ has a neighborhood U in which any loop can be shrunk continuously to a point).

Two paths having the same end points are *equivalent* if one can be continuously deformed into the other in Δ. Equivalence classes of closed paths (loops) based at a point $F\in\Delta$ form a group $\Pi(\Delta,F)$ called *fundamental group* (or first homotopy group) at the base point F, under the operation of composition of paths.

A *covering space* $(\tilde{\Delta}, p)$ of Δ is a topological space $\tilde{\Delta}$ and a continuous map $p:\tilde{\Delta}\to\Delta$ such that every point $F\in\Delta$ has a path-connected open neighborhood U with $p^{-1}(U)$ nonempty and each path component of $p^{-1}(U)$ mapped topologically onto U by p. Such an open neighborhood is called an *elementary* neighborhood.

If $\tilde{\gamma}:I\to\tilde{\Delta}$ is a path, then $p\tilde{\gamma}:I\to\Delta$ is a path. Conversely, if γ is a path in Δ with initial point F, and $\tilde{F}\in\tilde{\Delta}$ with $p(\tilde{F}) = F$, then there exists a unique path $\tilde{\gamma}$ in $\tilde{\Delta}$ with initial point $\tilde{F}$ such that $p\tilde{\gamma} = \gamma$.

There is a natural injection p_* from $\Pi(\tilde{\Delta},\tilde{F})$ into $\Pi(\Delta,F)$ called the *induced* homomorphism. A closed loop γ in Δ based at F lifts to a closed loop $\tilde{\gamma}$ in $\tilde{\Delta}$ based at $\tilde{F}$ if and only if the equivalence class $[\gamma]$ is an element of $p_*\,\Pi(\tilde{\Delta},\tilde{F})$.

If $\tilde{\Delta}$ is simply connected, it is called a *universal* covering space of Δ, and has the property that it can serve as a covering space for any other covering space of Δ. Then, $\Pi(\tilde{\Delta},\tilde{F}) = \{1\}$.

Under the condition that Δ be semilocally simply connected, for any given conjugacy class of subgroups of $\Pi(\Delta,F)$, there exists a covering space $(\tilde{\Delta}, p)$ such that $p_*\,\Pi(\tilde{\Delta},\tilde{F})$ belongs to the given conjugacy class. $(\tilde{\Delta}, p)$ is determined up to isomorphism by the conjugacy class. In particular, under this condition there always exists a universal covering space $(\tilde{\Delta}, p)$, unique up to isomorphism, corresponding to the conjugacy class of the trivial subgroup. All manifolds and manifolds with boundary are semilocally simply connected.

Let $\Delta_N^{(s)} = \{F_{\mathbf{x}_1}+\cdots+F_{\mathbf{x}_N} \mid \mathbf{x}_i\neq\mathbf{x}_j\in\mathbb{R}^s, \text{ for } i\neq j\}$ be an N-particle orbit in $\mathscr{S}'$. The weak topology of $\mathscr{S}'$, restricted to $\Delta_N^{(s)}$, yields the obvious topology where for $\mathscr{O}_i\subset\mathbb{R}^s$ open, $i = 1,\cdots,N$, and $\mathscr{O}_i\cap\mathscr{O}_j = \phi$ for $i\neq j$, $U_{\mathscr{O}_1,\cdots,\mathscr{O}_N}$ $= \{F = \Sigma_{i=1}^N F_{\mathbf{x}_i} \mid \mathbf{x}_i\in\mathscr{O}_i\}$ is a neighborhood basis for $\Delta_N^{(s)}$. In this topology $\Delta_N^{(s)}$ is connected, pathwise connected, and semilocally simply connected, and hence possesses a universal covering space $(\tilde{\Delta}_N, p)$.

In Sec. III we noted that for $s\geqslant 3$ the universal covering space of $\Delta_N^{(s)}$ is the space $\tilde{\Delta}_N^{(s)} = \mathbb{R}^{sN}\setminus D$, while special cases arise for $s = 1$ and $s = 2$. For $s\geqslant 3$, the fundamental group for $\Delta_N^{(s)}$ is the permutation group S_N.

B. Sufficient condition for recovering a representation of the little group

Next we establish a condition which allows the recovery of a well-defined representation of the little group from the system of multipliers for a representation of $\mathscr{S}\wedge\mathscr{K}$. In the framework of Sec. II.B, we assume that μ is concentrated on a single measurable orbit $\Delta\subset\mathscr{S}'$. This is not the case for every representation of $\mathscr{S}\wedge\mathscr{K}$. In particular, representations describing infinitely many particles do not have this

property.

Recall that $F \in \text{supp}\,\mu$ if and only if for any open neighborhood $\mathcal{O}$ containing F, $\mu(\mathcal{O}) > 0$. Supp μ is a closed set in $\mathcal{S}'$, and invariant under the action of $\mathcal{K}$. When μ is concentrated on the orbit Δ, supp μ is the weak closure of Δ.

Definition: A system of multipliers $\chi_\psi(F)$ is *continuous* (outside $\mathcal{Z}$) if $\mathcal{Z} \subset \text{supp}\,\mu$ is a closed set with $\mu(\mathcal{Z}) = 0$ such that for $F, \psi^* F \notin \mathcal{Z}$, $\chi_\psi(F)$ is defined and jointly continuous as a map from $(\text{supp}\,\mu) \times \mathcal{K} \to \mathcal{U}(1)$.

With μ concentrated on the orbit Δ, it is clear that $\chi_\psi(F)$ is continuous if $\mathcal{Z} \subset \Delta$ is closed with respect to the restriction of the weak topology to Δ, with $\mu(\mathcal{Z}) = 0$, and for $F, \psi^* F \in \Delta \setminus \mathcal{Z}$, $\chi_\psi(F)$ is jointly continuous as a mapping from $\Delta \times \mathcal{K}$ to $\mathcal{U}(1)$.

Theorem 1: Let the system of multipliers $\chi_\psi(F)$ be continuous (outside $\mathcal{Z}$). Then for all $F, \psi_2^* F$ and $\psi_1^* \psi_2^* F \in (\text{supp}\,\mu) \setminus \mathcal{Z}$, Eq. (2.9) holds.

Proof: Since $F \in \text{supp}\,\mu$, $\mu(\mathcal{O}) > 0$ for every neighborhood $\mathcal{O}$ of F. Since $\mathcal{Z}$ is closed we can select a neighborhood $\mathcal{O}$ of F such that the sets $\mathcal{O}$, $\psi_2^* \mathcal{O}$, and $\psi_1^* \psi_2^* \mathcal{O}$ all do not meet $\mathcal{Z}$. Since Eq. (2.9) holds almost everywhere, there is at least one point $G \in \mathcal{O}$ where it holds. Likewise, for any open subset of $\mathcal{O}$ containing F, Eq. (2.9) holds for at least one point. Then by continuity it holds for F. Q.E.D.

Corollary 1: If $F \in \text{supp}\,\mu \setminus \mathcal{Z}$, $\chi_\psi(F)$ defines a continuous representation of the little group $\mathcal{K}_F$.

Remark: The system of multipliers defined from a continuous wave function as in Eq. (3.4) is continuous outside the set $\mathcal{Z}$ of zeros of the wave function. We have $\mu(\mathcal{Z}) = 0$ since the wave function is taken to be a cyclic vector.

Corollary 2: If $F_1, F_2 \in \text{supp}\,\mu \setminus \mathcal{Z}$ are in the same orbit Δ, then the representations of the little groups $\mathcal{K}_{F_1}$ and $\mathcal{K}_{F_2}$ defined by $\chi_\psi(F)$ are unitarily equivalent.

Proof: Since F_1 and F_2 are in the same orbit, there exists $\psi \in \mathcal{K}$ with $\psi^* F_1 = F_2$. Let $\psi_1 \in \mathcal{K}_{F_1}$; then $\psi_2 = \psi \circ \psi_1 \circ \psi^{-1} \in \mathcal{K}_{F_2}$ defines an isomorphism $\psi_1 \to \psi_2$ between $\mathcal{K}_{F_1}$ and $\mathcal{K}_{F_2}$. The equivalence of the representations follows from Eq. (2.9) which holds for all points in question by Theorem 1. In particular

$$\chi_{\psi \circ \psi_1}(F_1) = \chi_{\psi_1}(F_1)\chi_\psi(\psi_1^* F_1) = \chi_{\psi_1}(F_1)\chi_\psi(F_1);$$

also

$$\chi_{\psi \circ \psi_1}(F_1) = \chi_{\psi_2 \circ \psi}(F_1) = \chi_\psi(F_1)\chi_{\psi_2}(\psi^* F_1) = \chi_\psi(F_1)\chi_{\psi_2}(F_2);$$

whence

$$\chi_\psi(F_1)^{-1}\chi_{\psi_1}(F_1)\chi_\psi(F_1) = \chi_{\psi_2}(F_2)\,. \qquad \text{Q.E.D.}$$

C. Sufficient conditions for the inducing construction

In what follows Δ is always a measurable orbit in $\mathcal{S}'$, locally path connected, and semilocally simply connected, with a quasi-invariant cylindrical measure μ concentrated on Δ. Our first objective will be to "lift" the action on Δ of $\psi \in \mathcal{K}$ to a covering space $(\hat{\Delta}, \hat{p})$. At times additional conditions on Δ will be required, which we now list:

(a) If ϕ_t for $t \in I$ is a closed loop in $\mathcal{K}$ based at $\mathbf{e}$ (i.e., $\phi_0 = \phi_1 = \mathbf{e}$), then for $F \in \Delta$, $\phi_t^* F$ is in the homotopy class of the identity in $\Pi(\Delta, F)$ i.e., the loop $\phi_t^* F$ shrinks to a point in Δ.

If $\mathcal{K}$ is simply connected, as conjectured in Sec. II, then this condition follows automatically from the continuity of $\phi \to \phi^* F$. We shall see below that condition (a) is not necessary to the development which follows, but results in some pedagogical simplification.

(b) If F_t is a closed loop in Δ based at F, there exists a path ψ_t in $\mathcal{K}$ with $\psi_{t=0}^* F = \psi_{t=1}^* F = F$, such that $\psi_t^* F$ is equivalent to F_t.

An equivalent, local version of condition (b) is the following:

(b') Given $F \in \Delta$, there exists an elementary neighborhood U_F of F, such that for any path F_t contained in U_F and originating at F, there exists a path ψ_t in $\mathcal{K}$ originating at $\mathbf{e}$ with $\psi_{t=1}^* F = F_{t=1}$ and with $\psi_t^* F$ equivalent to F_t. (Note that $\psi_t^* F$ need not be contained in U_F.)

Since Δ is an orbit under $\mathcal{K}$, it is sufficient to verify (b') for a single point $F \in \Delta$.

(c) Given $F \in \Delta$, there exists an elementary neighborhood U_F of F and an open neighborhood $\mathcal{O}$ of $\mathbf{e}$ in $\mathcal{K}$, such that $\mathcal{O}^* F \subseteq U_F$, and such that for any $\psi \in \mathcal{O}$ there exists a path ψ_t connecting $\mathbf{e}$ to ψ with $\psi_t^* F$ equivalent to a path in U_F connecting F to $\psi^* F$. (Note that ψ_t is not required to remain in $\mathcal{O}$, and $\psi_t^* F$ is not required to remain in U_F.)

We observe that if $\mathcal{K}$ is locally path connected (as is the diffeomorphism group), condition (c) follows from the continuity of $\psi \to \psi^* F$. Furthermore, if in condition (c) $\mathcal{O}^* F$ is itself assumed to be an elementary neighborhood of F, (b') follows as a consequence.

(d) For $F \in \Delta$, $\psi \to \psi^* F$ is an open mapping from $\mathcal{K} \to \Delta$, i.e., if $\mathcal{O}$ is open in $\mathcal{K}$, $\mathcal{O}^* F$ is open in Δ. Alternatively, if $F_i \to F$ in Δ, there exist $\psi_i \to \mathbf{e}$ in $\mathcal{K}$ with $\psi_i^* F = F_i$.

Again, if $\mathcal{K}$ is locally path connected, (d) implies the strengthened version of (c) above, and therefore (b) as well. It also implies that Δ is locally path connected, so that this does not have to be assumed separately. Note that $\psi \to \psi^* F$ is not an open mapping from $\mathcal{K} \to \mathcal{S}'$; for the orbits $\Delta_N^{(s)}$ are themselves not open in $\mathcal{S}'$.

Now we look at the consequences of these conditions, and demonstrate the above assertions as we go along.

Lemma 1: If Δ satisfies property (a) above, the action on Δ of $\psi \in \mathcal{K}$ lifts in a well-defined manner to the universal covering space $\tilde{\Delta}$.

Proof: For $\psi \in \mathcal{K}$, select a continuous path $\psi_t \in \mathcal{K}$ connecting $\mathbf{e}$ to ψ. Let $\tilde{F}_0 \in \tilde{\Delta}$ be any point such that $p(\tilde{F}_0) = F_0$. The path $\psi_t^* F_0$ is continuous in Δ, and therefore lifts to a unique path $\tilde{F}_t$ in $\tilde{\Delta}$ such that $\tilde{F}_{t=0} = \tilde{F}_0$ and $p(\tilde{F}_t) = \psi_t^* F_0$. Define $\psi^* \tilde{F}_0 = \tilde{F}_{t=1}$. This lifts the action of ψ to $\tilde{\Delta}$.

If ψ_t and ψ'_t are two paths in $\mathcal{K}$ connecting $\mathbf{e}$ to ψ, then $\psi_t^* F_0$ and $\psi_t'^* F_0$ are equivalent by property (a). Therefore, they lift to paths having the same terminal point, and $\psi^* \tilde{F}_0$ is well defined. Q.E.D.

At this point we shall consider what happens when property (a) does not hold. If $\mathcal{K}$ is not simply connected, then for $\psi \in \mathcal{K}$ and for distinct continuous paths ψ_t, ψ'_t connecting $\mathbf{e}$ to ψ, it is possible that the paths $\psi_t^* F_0$ and $\psi_t'^* F_0$ are not equivalent in Δ, and therefore do not lift to paths having

the same terminal point in $\tilde{\Delta}$. If this were to occur, one could consider the subgroup $\hat{\Pi}(\Delta,F_0)$ of $\Pi(\Delta,F_0)$ consisting of the homotopy classes of the loops $\phi_t^* F_0$ in Δ, where ϕ_t is a closed loop in $\mathscr{K}$ based at $\mathbf{e}$. Then the paths $\psi_t^* F_0$ and $\psi_t'^* F_0$ lift to paths in a covering space $(\hat{\Delta}, \hat{p})$ such that $\hat{p}_* \Pi(\hat{\Delta},\hat{F}_0)$ belongs to the conjugacy class of $\hat{\Pi}(\Delta,F_0)$ in $\Pi(\Delta,F_0)$, where $\hat{p}(\hat{F}_0) = F_0$. For appropriate choice of $\hat{F}_0$, $\hat{p}_* \Pi(\hat{\Delta},\hat{F}_0) = \hat{\Pi}(\Delta,F_0)$. In $\hat{\Delta}$, $\psi_t^* \hat{F}_0$ and $\psi_t'^* \hat{F}_0$ have the same terminal point. The covering space $\hat{\Delta}$ can be obtained from $\tilde{\Delta}$ by identifying all terminal points in $\tilde{\Delta}$ obtained by lifting the loops $\phi_t^* F_0$.

Two paths in Δ beginning at F_0, with the same end points, can now be defined as equivalent (mod $\hat{\Pi}$) if the homotopy class of the loop which they form is in the subgroup $\hat{\Pi}(\Delta,F_0)$. In conditions (b) and (c), the requirement of "equivalent" paths is replaced by the requirement of paths "equivalent (mod $\hat{\Pi}$)," and all of the subsequent development goes through.

For the orbits $\Delta_N^{(s)}$ considered in Sec. III, $\hat{\Pi}(\Delta_N^{(s)},F_0) = \{1\}$, i.e., property (a) holds and these procedures are not necessary. Henceforth, we shall assume this condition for simplicity; then we can work in the universal covering space $\tilde{\Delta}$.

We shall also need the following fact.

Lemma 2: For $\tilde{F}\in\tilde{\Delta}$, $(\psi_1\circ\psi_2)^*\tilde{F} = \psi_1^*(\psi_2^*\tilde{F})$.

Proof: Let ψ_{1t} and ψ_{2t} be paths in $\mathscr{K}$ from $\mathbf{e}$ to ψ_1 and ψ_2, respectively. Then $\psi_1^*(\psi_2^* F)$ is the terminal point of the path in $\tilde{\Delta}$ which begins at $\tilde{F}$ and which lifts the path $\psi_{2t}^* F$ followed by the path $\psi_{1t}^*(\psi_2^* F)$ in Δ, where $F = p(\tilde{F})$. However, this path in Δ is obtained from a path in $\mathscr{K}$ from $\mathbf{e}$ to $\psi_1\circ\psi_2$ which first traverses ψ_{2t} and then traverses $\psi_{1t}\circ\psi_2$, since in Δ, we have $(\psi_{1t}\circ\psi_2)^* F = \psi_{1t}^*(\psi_2^* F)$. Thus, $(\psi_1\circ\psi_2)^*\tilde{F} = \psi_1^*(\psi_2^*\tilde{F})$. Q.E.D.

We have noted that conditions (b) or (b′) and (c) above are ways of bypassing the assumption that $\mathscr{K}$ be locally path connected by looking at its action on Δ. The next lemma shows that $\mathscr{K}$ acts transitively on $\tilde{\Delta}$.

Lemma 3: If Δ satisfies condition (b) above, then it also satisfies (b′). Furthermore, the action of $\mathscr{K}$ on $\tilde{\Delta}$ defined in Lemma 1 is transitive.

Proof: Let F_t be any path in Δ commencing at F_0 and terminating at F_1. Pick $\phi\in\mathscr{K}$ with $\phi^* F_0 = F_1$, and let ϕ_t be any path from $\mathbf{e}$ to ϕ in $\mathscr{K}$. Then $\phi_t^* F_0 = F_t'$ also commences at F_0 and terminates at F_1. Let γ be the loop based at F_1 formed by traversing F_t' backwards to F_0 followed by F_t. By condition (b), there exists a path ψ_t of $\mathscr{K}$ with $\psi_t^* F_1$ equivalent to γ. We can select ψ_t to commence at $\mathbf{e}$ since if it does not, we may replace it by $(\psi_{t=0})^{-1}\circ\psi_t$ which also yields a path equivalent to γ. Define the path $(\psi\circ\phi)_t$ in $\mathscr{K}$ obtained by traversing ϕ_t from $\mathbf{e}$ to ϕ, followed by $\psi_t\circ\phi$. By construction, $(\psi\circ\phi)_t^* F_0$ is equivalent to F_t. Thus, for any path F_t in Δ we have obtained a path $(\psi\circ\phi)_t$ in $\mathscr{K}$ commencing at $\mathbf{e}$ with $(\psi\circ\phi)_t^* F_0$ having the same end points as F_t and equivalent to it. This implies that (b′) holds with any choice of elementary neighborhood. Pick $\tilde{F}_0$ in $\tilde{\Delta}$ with $p(\tilde{F}_0) = F_0$. Now any point $\tilde{F}_1$ in $\tilde{\Delta}$ with $p(\tilde{F}_1) = F_1$ corresponds to a certain equivalence class of paths from F_0 to F_1. Choosing F_t from this equivalence class, we find $(\psi\circ\phi)_t^*\tilde{F}_0$ connects $\tilde{F}_0$ to $\tilde{F}_1$. Thus, $\mathscr{K}$ acts transitively on $\tilde{\Delta}$. Q.E.D.

Lemma 4: If Δ satisfies condition (b′) above, then it also satisfies condition (b).

Proof: Let F_t be a closed loop in Δ, and for each point on F_t pick an elementary neighborhood U_{F_t} satisfying (b′). This gives an open covering of the loop, hence an open covering of the unit interval I, which is a compact set. Hence, we can decompose I into finitely many sufficiently small closed subintervals I_i such that the path F_t for $t\in I_i$ lies in a single elementary neighborhood U_{F_i}.

From (b′) it is easy to show that for *any* path F_t^i contained in U_{F_i} (not necessarily originating at F_i), with end points F_0^i and F_1^i, there is a path ψ_t^i in $\mathscr{K}$ originating at $\mathbf{e}$ with $\psi_t^{i*} F_0^i$ equivalent to F_t^i. Thus, for each of the finitely many path segments F_t^i, we obtain a path ψ_t^i in $\mathscr{K}$ originating at $\mathbf{e}$ with $\psi_t^{i*} F_0^i$ equivalent to F_t^i. Finally, we compose these paths together in $\mathscr{K}$ by traversing ψ_t^1 followed by $(\psi_t^2\circ\psi_{t=1}^1)$ followed by $(\psi_t^3\circ\psi_{t=1}^2\circ\psi_{t=1}^1)$ and so forth, to satisfy condition (b). Q.E.D.

Next we show that $\mathscr{K}$ acts continuously on $\tilde{\Delta}$.

Lemma 5: If Δ satisfies condition (c) above, then the mapping $\psi\to\psi^*\tilde{F}$ defined in Lemma 1 gives a continuous mapping from $\mathscr{K}\to\tilde{\Delta}$.

Proof: If $\psi^i\to\mathbf{e}$ in $\mathscr{K}$, and $F = p(\tilde{F})$, let $\mathscr{O}$ be the neighborhood of $\mathbf{e}$ described in (c) with $\mathscr{O}^* F\subseteq U_F$. Then for $\psi^i\in\mathscr{O}$, there is a path ψ_t^i connecting $\mathbf{e}$ to ψ^i with $\psi_t^{i*} F$ equivalent to a path in U_F. Therefore, $\psi^{i*}\tilde{F}$ is in the same path component of $p^{-1}(U_F)$ as $\tilde{F}$. Since p maps this path component topologically onto U_F, and since $p(\psi^{i*}\tilde{F}) = \psi^{i*} F\to F$, we have $\psi^{i*}\tilde{F}\to\tilde{F}$. Since the group operations in $\mathscr{K}$ are continuous, continuity at $\mathbf{e}$ implies continuity elsewhere in $\mathscr{K}$. Q.E.D.

Finally, we look at consequences of property (d).

Lemma 6: If Δ satisfies condition (d) above, and if $\mathscr{K}$ is locally path connected, then Δ also satisfies (b), (b′), and (c).

Proof: If $\mathscr{K}$ is locally path connected, (c) is automatic: since $\psi\to\psi^* F$ is continuous, the inverse image of any elementary neighborhood U_F of F is open, and contains a path-connected open neighborhood $\mathscr{O}$ of $\mathbf{e}$. For $\psi\in\mathscr{O}$, ψ_t connecting $\mathbf{e}$ to ψ may be chosen to lie wholly in $\mathscr{O}$, whence $\psi_t^* F$ lies wholly in U_F and (c) is satisfied.

$\mathscr{O}^* F$ is path connected and contained in U_F; by condition (d) it is also open. Thus, $\mathscr{O}^* F$ is itself an elementary neighborhood of F. It qualifies as the elementary neighborhood of condition (b′).

We have already seen in Lemma 4 that (b′) implies (b). Q.E.D.

Lemma 7: Let Δ satisfy conditions (a)–(d). Then the mapping $\psi\to\psi^*\tilde{F}$ defined in Lemma 1 is an open mapping from $\mathscr{K}\to\tilde{\Delta}$.

Proof: Suppose $\tilde{F}_i\to\tilde{F}$ in $\tilde{\Delta}$. Then $p(\tilde{F}_i)\to p(\tilde{F})$ in Δ. Let $F = p(\tilde{F})$ and let U_F be an elementary neighborhood of F in accordance with (c). Since $p(\tilde{F}_i)\to F$, we may eventually take $p(\tilde{F}_i)\in U_F$. By condition (d), there exists $\psi_i\to\mathbf{e}$ in $\mathscr{K}$ with $\psi_i^* F = p(\tilde{F}_i)$. It follows from (c) that in some neighborhood $\mathscr{O}$ of $\mathbf{e}$, $\psi_i^*\tilde{F}$ is in the same path component of

$p^{-1}(U_F)$ as $\tilde{F}$. Then there is an open neighborhood of $\tilde{F}$ in which $\psi_i^*\tilde{F} = \tilde{F}_i$, which is sufficient for the result.

Q.E.D.

Lemma 8: Under conditions (a)–(d), the mapping $\mathscr{K}\times\tilde{\Delta} \to \tilde{\Delta}$ given by $(\psi,\tilde{F}) \to \psi^*\tilde{F}$ is jointly continuous.

Proof: Let $\psi_i \to \mathbf{e}$ in $\mathscr{K}$, and $\tilde{F}_j \to \tilde{F}$ in $\tilde{\Delta}$. By Lemma 7, pick $\phi_j \to \mathbf{e}$ in $\mathscr{K}$ such that $\phi_j^*\tilde{F} = \tilde{F}_j$. Then $\psi_i^*\tilde{F}_j = \psi_i^*(\phi_j^*\tilde{F}) = (\psi_i\circ\phi_j)^*\tilde{F}$ by Lemma 2. However, $\psi_i\circ\phi_j \to \mathbf{e}$ by continuity of the group operation; and by Lemma 5 we have the result.

Q.E.D.

Remark: We have seen how conditions (a)–(c) take the place of connectedness assumptions on $\mathscr{K}$. While the latter may be difficult to prove, it is straightforward to verify (a)–(c), as well as (d), for the orbits $\Delta\binom{s}{N}$. The effect of these conditions is to permit all of the local properties of the action of $\mathscr{K}$ on Δ to be lifted to $\tilde{\Delta}$.

With the above results we can proceed with the development. Under property (a), there is a natural homomorphism from $\mathscr{K}_{F_0}$ to $\Pi(\Delta,F_0)$ as follows: for $\psi\in\mathscr{K}_{F_0}$, let ψ_t be a path in $\mathscr{K}$ from $\mathbf{e}$ to ψ. The action $\psi_t^*F_0$ then yields a closed loop in Δ, which defines an element of $\Pi(\Delta,F_0)$. Again, if ψ'_t is a second path in $\mathscr{K}$ from $\mathbf{e}$ to ψ, $\psi_t'^*F_0$ and $\psi_t^*F_0$ are equivalent and thus define the same element of $\Pi(\Delta,F_0)$. Calling this mapping $h_{F_0}: \mathscr{K}_{F_0} \to \Pi(\Delta,F_0)$, it is evident that h_{F_0} is a homomorphism. Where there is no possible ambiguity we suppress the subscript F_0. We note that for $\psi_1, \psi_2\in\mathscr{K}_{F_0}$ and for $p(\tilde{F}_0) = F_0$, we have $\psi_1^*F_0 = \psi_2^*F_0$ if and only if $h(\psi_1) = h(\psi_2)$. If $h(\psi_1)\neq h(\psi_2)$ in $\Pi(\Delta,F_0)$, then ψ_1 and ψ_2 are in different path components of $\mathscr{K}_{F_0}$. Property (b) above implies that the homomorphism h is surjective.

For $\tilde{F}_0\in\tilde{\Delta}$ with $p(\tilde{F}_0) = F_0$, and for $g\in\Pi(\Delta,F_0)$ represented by the closed loop F_t based at F_0, define $g\tilde{F}_0$ to be the terminal point of the unique path $\tilde{F}_t$ with initial point $\tilde{F}_0$ such that $p(\tilde{F}_t) = F_t$. If $g = h(\phi)$ for $\phi\in\mathscr{K}_{F_0}$, then $g\tilde{F}_0 = \phi^*\tilde{F}_0$. (Note that if $\phi_1, \phi_2\in h^{-1}(g)$, $\phi_1^*F_0 = \phi_2^*F_0$.)

Next we construct an isomorphism from $\Pi(\Delta,F)$ to $\Pi(\Delta,F')$ as follows: If $F' = \psi^*F$ and F_t is a loop in Δ based at F defining an element of $\Pi(\Delta,F)$, then ψ^*F_t is a loop in Δ based at F' defining an element of $\Pi(\Delta,F')$. Let α denote this map from loops based at F to those based at F'. It is obvious that α maps equivalent loops to equivalent loops, and defines an isomorphism between $\Pi(\Delta,F)$ and $\Pi(\Delta,F')$.

Lemma 9: The isomorphism α is independent of the choice of ψ.

Proof: Suppose $\psi'^*F = F'$. Then we can write $\psi' = \phi\circ\psi$, where $\phi\in\mathscr{K}_{F'}$. Define $\phi_t\in\mathscr{K}$ to be a path from $\mathbf{e}$ to ϕ. Let F_t be a loop based at F. Then $\phi_t^*(\psi^*F_t)$ defines a continuous deformation of the loop ψ^*F_t into the loop ψ'^*F_t. Thus, these two loops are equivalent and ψ, ψ' define the same isomorphism between $\Pi(\Delta,F)$ and $\Pi(\Delta,F')$.

Q.E.D.

It follows from the definitions of h and α that if $h_F(\phi) = g\in\Pi(\Delta,F)$ for $\phi\in\mathscr{K}_F$, then $h_{\psi^*F}(\psi\circ\phi\circ\psi^{-1}) = \alpha(g)\in\Pi(\Delta,\psi^*F)$.

Lemma 10: Given $g\in\Pi(\Delta,F)$, $F' = \psi^*F$, and $\tilde{F}\in\tilde{\Delta}$ with $p(\tilde{F}) = F$, then $\alpha(g)\psi^*\tilde{F} = \psi^*g\tilde{F}$.

Proof: Suppose g corresponds to the loop F_t based at F, and the path ψ_t connects $\mathbf{e}$ to ψ in $\mathscr{K}$. Then $\alpha(g)$ corresponds to the loop $F'_t = \psi^*F_t$ based at F'. $\psi^*g\tilde{F}$ is the terminal point of the lifting of the path F_t, followed by the path ψ_t^*F, while $\alpha(g)\psi^*\tilde{F}$ is the terminal point of the lifting of the path ψ_t^*F followed by ψ^*F_t. Both liftings have the same initial point $\tilde{F}$. However, the two paths are equivalent in Δ: $\psi_s^*F_t$ continuously deforms F_t into ψ^*F_t as s moves from 0 to 1. Thus, both liftings have the same terminal point in $\tilde{\Delta}$.

Q.E.D.

Now we are ready to proceed with the inducing construction.

Let $g \to T_g^{F_0}$ be a representation of $\Pi(\Delta,F_0)$ in $\mathscr{U}(\mathbb{1})$. For any other point $F\in\Delta$, $T^F_{\alpha(g)} = T^{F_0}_g$ defines a representation $g \to T^F_g$ of $\Pi(\Delta,F)$, where $\alpha:\Pi(\Delta,F_0) \to \Pi(\Delta,F)$ is the isomorphism defined above. If $\hat{\Psi}:\tilde{\Delta} \to \mathbb{C}$ satisfies

$$\hat{\Psi}(g\tilde{F}) = T^F_g\hat{\Psi}(\tilde{F}), \tag{4.1}$$

then the function $\overline{\hat{\Psi}(\tilde{F})}\,\hat{\Psi}(\tilde{F})$ depends only on F and not on the particular choice of $\tilde{F}$; thus, it is a function on Δ.

We may endow $\tilde{\Delta}$ with the σ algebra of measurable sets generated by the path components of $p^{-1}(U)$ such that U is an elementary neighborhood in Δ. Now let $\mathscr{H}$ be the Hilbert space of all measurable functions $\hat{\Psi}$ on $\tilde{\Delta}$, satisfying Eq. (4.1) everywhere, such that $\hat{\Psi}(\tilde{F})$ is square integrable on Δ, i.e.,

$$\int_\Delta \overline{\hat{\Psi}(\tilde{F})}\,\hat{\Psi}(\tilde{F})d\mu(F) < \infty\,. \tag{4.2}$$

The Hilbert space $\mathscr{H}$ is defined up to equivalence classes of functions which differ on sets of measure 0 in Δ, and an inner product on $\mathscr{H}$ is defined from Eq. (4.2) in the usual way.

Finally, defining

$$(\mathscr{V}(\psi)\hat{\Psi})(\tilde{F}) = \hat{\Psi}(\psi^*\tilde{F})\left[\frac{d\mu(\psi^*F)}{d\mu(F)}\right]^{1/2}, \tag{4.3}$$

where $F = p(\tilde{F})$, we obtain a representation of $\mathscr{K}$ in $\mathscr{H}$. Moreover, with $(\mathscr{U}(f)\hat{\Psi})(\tilde{F}) = [\exp i(F,f)]\hat{\Psi}(\tilde{F})$, we recover a representation of $\mathscr{S}\wedge\mathscr{K}$. This is called the representation of $\mathscr{S}\wedge\mathscr{K}$ *induced* by the representation $T_g^{F_0}$ of the fundamental group.

The induced representation is cyclic for $\{\mathscr{U}(f)\}$. A cyclic vector $\hat{\Phi}$ is any function in $\mathscr{H}$ such that $\mu(p\{\tilde{F}\in\tilde{\Delta}\mid\hat{\Phi}(\tilde{F}) = 0\}) = 0$.

Remarks: (1) If for any $\psi_i \to \mathbf{e}$ in $\mathscr{K}$, μ has the properties that (a) $\mu(M\triangle\psi_i^*M) \to 0$ for any measurable set M in Δ where $\triangle$ is the symmetric difference, and (b) $[d\mu(\psi_i^*F)/d\mu(F)]^{1/2} \to \Omega(F)\equiv 1$ in $L^2_\mu(\Delta)$; then the induced representation is a continuous representation of $\mathscr{K}$. These correspond to standard assumptions in the definition of a dynamical system.[4]

(2) A representation of $\Pi(\Delta,F_0)$ defines a continuous representation of the little group $\mathscr{K}_{F_0}$ utilizing h: $\mathscr{K}_{F_0} \to \Pi(\Delta,F_0)$. For the N-particle representations of $\mathscr{S}\wedge\mathscr{K}$, with space dimension $s\geqslant 3$, the fundamental group of $\Delta\binom{s}{N}$ is S_N, and the representations induced by the even and odd representations of S_N are, respectively, the Bose and Fermi representations of the current group. $\tilde{\Delta}$ is an sN dimensional space, and $\hat{\Psi}$ is simply a wavefunction.

(3) Comparing Eq. (4.3) to Eq. (2.8), we see that by using a larger space ($\tilde{\Delta}$ instead of Δ) the system of multipliers has disappeared from the equation.

In the next theorem, we show that the induced representation just defined is equivalent to a representation determined by the Gel'fand–Vilenkin theory.

Theorem 2: Given a representation T^{F_0} of $\Pi(\Delta,F_0)$ in $\mathbb{C}$, the induced representation is equivalent to the representation defined by Eq. (2.8) in the Gel'fand–Vilenkin formalism; with a system of multipliers given by $\chi_\psi(F) = [\hat{\Omega}(\psi^*\tilde{F})/\hat{\Omega}(\tilde{F})]$, where $\tilde{F}$ is any point with $p(\tilde{F}) = F$, and where $\hat{\Omega}:\tilde{\Delta} \to \mathbb{C}$ is a cyclic vector for $\{\mathscr{U}(f)\}$ in the induced representation satisfying $|\hat{\Omega}(\tilde{F})| = 1$ almost everywhere.

Proof: Let $\hat{\Phi}$ be any cyclic vector for $\{\mathscr{U}(f)\}$ in the induced representation, and set $\hat{\Omega}(\tilde{F}) = \hat{\Phi}(\tilde{F})/|\hat{\Phi}(\tilde{F})|$. $\hat{\Omega}$ is defined wherever $\hat{\Phi}$ is nonzero, i.e., outside a set $\tilde{\mathscr{Z}}_0$ with $\mu(p(\tilde{\mathscr{Z}}_0)) = 0$. Now $\chi_\psi(F)$ is defined as long as $\hat{\Omega}(\tilde{F})$ and $\hat{\Omega}(\psi^*\tilde{F})$ are defined. It is independent of the choice of $\tilde{F}$, since

$$\begin{aligned}\hat{\Omega}(\psi^*g\tilde{F})/\hat{\Omega}(g\tilde{F}) &= \hat{\Omega}(\alpha(g)\psi^*\tilde{F})/\hat{\Omega}(g\tilde{F}) \\ &= T^{\psi^*F}_{\alpha(g)}\hat{\Omega}(\psi^*\tilde{F})/T^F_g\hat{\Omega}(\tilde{F}) \\ &= \hat{\Omega}(\psi^*\tilde{F})/\hat{\Omega}(\tilde{F}).\end{aligned}$$

Moreover, $\chi_\psi(F)$ satisfies Eq. (2.9) wherever it is defined, since

$$\begin{aligned}&\chi_{\psi_1}(F)\chi_{\psi_2}(\psi_2^*F) \\ &\quad= [\hat{\Omega}(\psi_2^*\tilde{F})/\hat{\Omega}(\tilde{F})]\cdot[\hat{\Omega}(\psi_1^*\psi_2^*\tilde{F})/\hat{\Omega}(\psi_2^*\tilde{F})] \\ &\quad= [\hat{\Omega}(\psi_1^*\psi_2^*\tilde{F})/\hat{\Omega}(\tilde{F})] = [\hat{\Omega}(\psi_1\circ\psi_2)^*\tilde{F})/\hat{\Omega}(\tilde{F})] \\ &\quad= \chi_{\psi_1\circ\psi_2}(F),\end{aligned}$$

making use of Lemma 2.

We shall show that the induced representation is equivalent to the representation defined by Eq. (2.8) by showing that they have the same generating functionals $L(f,\psi) = (\Omega, \mathscr{U}(f)\mathscr{V}(\psi)\Omega)$, where Ω is the cyclic vector.

For Eq. (2.8), with $\Omega(F) \equiv 1$, the generating functional is

$$L_{\text{G-V}}(f,\psi) = \int d\mu(F) e^{i(F,f)}\chi_\psi(F)\left[\frac{d\mu(\psi^*F)}{d\mu(F)}\right]^{1/2}.$$

For the induced representation, on the other hand, the cyclic vector is $\hat{\Omega}(\tilde{F})$, so

$$\begin{aligned}&L_{\text{induced}}(f,\psi) \\ &\quad= \int d\mu(F)\,\overline{\hat{\Omega}(\tilde{F})}\,e^{i(F,f)}\left[\frac{d\mu(\psi^*F)}{d\mu(F)}\right]^{1/2}\hat{\Omega}(\psi^*\tilde{F}).\end{aligned}$$

However, $|\hat{\Omega}(\tilde{F})| = 1$ almost everywhere in Δ; hence, almost everywhere

$$\begin{aligned}\overline{\hat{\Omega}(\tilde{F})}\,\hat{\Omega}(\psi^*\tilde{F}) &= \overline{\hat{\Omega}(\tilde{F})}\,\hat{\Omega}(\psi^*\tilde{F})/\overline{\hat{\Omega}(\tilde{F})}\,\hat{\Omega}(\tilde{F}) \\ &= \hat{\Omega}(\psi^*\tilde{F})/\hat{\Omega}(\tilde{F}) = \chi_\psi(F).\end{aligned}$$

Thus, the generating functionals are equal.

Q.E.D.

Remarks:(1) If the cyclic vector $\hat{\Phi}$ in the induced representation is continuous almost everywhere in Δ [i.e., outside a set $\tilde{\mathscr{Z}}$ with $\mu(p(\tilde{\mathscr{Z}})) = 0$)], then $\chi_\psi(F)$ satisfies the continuity condition of Theorem 1 (outside $p(\tilde{\mathscr{Z}})$).

(2) The system of multipliers $\chi_\psi(F)$ defined from an induced representation as in Theorem 2 can now be modified so that it is defined and satisfies Eq. (2.9) *everywhere* (though of course the continuity condition will not hold everywhere). In fact, if $\tilde{F}$ is a particular point at which $\hat{\Omega}$ is undefined [i.e., at which $\hat{\Phi}(\tilde{F}) = 0$], then $\hat{\Omega}(\tilde{F})$ may be defined to be an arbitrary complex number of modulus 1, and $\hat{\Omega}(g\tilde{F})$ defined to be $T^F_g\hat{\Omega}(\tilde{F})$ for $g\in\Pi(\Delta,F)$. Now $\chi_\psi(F) = [\hat{\Omega}(\psi^*\tilde{F})/\hat{\Omega}(\tilde{F})]$ is well defined and satisfies Eq. (2.9) everywhere. This remark is utilized in Appendix B.

Given a system of multipliers $\chi_\psi(F)$, the continuity condition implies that a well-defined representation of $\mathscr{K}_F$ is determined for $F\in\Delta \setminus \mathscr{Z}$. We next establish sufficient conditions for knowing that equivalent representations of the little group correspond to equivalent representations of $\mathscr{S}\wedge\mathscr{K}$.

Let h: $\mathscr{K}_F \to \Pi(\Delta,F)$ be the homomorphism defined previously. If for all $\psi_1,\psi_2\in\mathscr{K}_F$ such that $h(\psi_1) = h(\psi_2)$, $\chi_{\psi_1}(F) = \chi_{\psi_2}(F)$, we shall say that the representation of the little group obtained from the multiplier system *factors through* the fundamental group. Assuming the path connectedness of $h^{-1}(g)$ for $g\in\Pi(\Delta,F)$, this occurs under the physical condition that the cyclic vector be time-reversal invariant [implying that $\chi_\psi(F) = \pm 1$].

Theorem 3: Given a representation $\mathscr{U}(f)\mathscr{V}(\psi)$ of $\mathscr{S}\wedge\mathscr{K}$ in the Gel'fand–Vilenkin formalism with μ concentrated on a measurable orbit Δ having properties (a)–(d). Suppose that $\chi_\psi(F)$ is continuous and defines, in accordance with Corollary 1, a representation of $\mathscr{K}_{F_0}$ (for $F_0\in\Delta\setminus\mathscr{Z}$) which factors through $\Pi(\Delta,F_0)$. Then $\mathscr{U}(f)\mathscr{V}(\psi)$ is unitarily equivalent to the representation induced from a representation T^{F_0} of $\Pi(\Delta,F_0)$ satisfying $T^{F_0}_{h(\psi)} = \chi_\psi(F_0)$ for $\psi\in\mathscr{K}_{F_0}$, with h: $\mathscr{K}_{F_0} \to \Pi(\Delta,F_0)$ defined as above.

Proof: Select $\tilde{F}_0\in\tilde{\Delta}$ with $p(\tilde{F}_0) = F_0$. Define $\hat{\Omega}(\psi^*\tilde{F}_0) = \chi_\psi(F_0)$ for $\psi^*F_0\notin\mathscr{Z}$. Then $\hat{\Omega}$ is well defined, since if $\psi_1^*\tilde{F}_0 = \psi_2^*\tilde{F}_0$, $\psi_1^{-1}\circ\psi_2$ is an element of $\mathscr{K}_{F_0}$ with $h(\psi_1^{-1}\circ\psi_2)$ equal to the identity in $\Pi(\Delta,F_0)$. Thus, $\chi_{\psi_1^{-1}\circ\psi_2}(F_0) = 1$ and it follows from Eq. (2.9) that $\chi_{\psi_1}(F_0) = \chi_{\psi_2}(F_0)$.

Since ψ acts transitively on $\tilde{\Delta}$, we have $\hat{\Omega}$ defined everywhere except on $p^{-1}(\mathscr{Z})$. Furthermore, $\hat{\Omega}$ is continuous wherever it is defined. If $\tilde{F}_i \to \tilde{F}$ in $\tilde{\Delta}$, Lemma 7 permits us to obtain $\psi_i \to \mathbf{e}$ in $\mathscr{K}$ with $\psi_i^*\tilde{F} = \tilde{F}_i$. With $\tilde{F} = \phi^*\tilde{F}_0$, we have $\hat{\Omega}(\tilde{F}_i) = \hat{\Omega}(\psi_i^*\phi^*\tilde{F}_0) = \hat{\Omega}((\psi_i\circ\phi)^*\tilde{F}_0) = \chi_{\psi_i\circ\phi}(F_0)$ $\to \chi_\phi(F_0) = \hat{\Omega}(\phi^*\tilde{F}_0) = \hat{\Omega}(\tilde{F})$, whence $\hat{\Omega}$ is continuous. Then by Appendix A, $\hat{\Omega}$ is also a measurable function.

We next show that $\hat{\Omega}$ satisfies Eq. (4.1). Let $\psi\in\mathscr{K}_{F_0}$ with $h(\psi) = g$ and let $\tilde{F} = \phi^*\tilde{F}_0$. Then $\alpha(g)\tilde{F} = \phi^*\psi^*\phi^{-1*}\tilde{F}$, where $\alpha:\Pi(\Delta,F_0) \to \Pi(\Delta,F)$, and $T^F_{\alpha(g)} = T^{F_0}_g = \chi_\psi(F_0)$. We have $\hat{\Omega}(\alpha(g)\tilde{F}) = \hat{\Omega}(\phi^*\psi^*\tilde{F}_0) = \chi_{\phi\circ\psi}(F_0)$ $= \chi_\psi(F_0)\chi_\phi(\psi^*F_0) = \chi_\psi(F_0)\chi_\phi(F_0) = T^F_{\alpha(g)}\hat{\Omega}(\tilde{F})$.

Finally, we show that $\chi_\psi(F) = \hat{\Omega}(\psi^*\tilde{F})/\hat{\Omega}(\tilde{F})$ for any $\psi\in\mathscr{K}$ and $F\in\Delta$ such that F, $\psi^*F\notin\mathscr{Z}$. Let $\tilde{F} = \phi^*\tilde{F}_0$. Then,

$$\begin{aligned}\hat{\Omega}(\psi^*\tilde{F}) = \hat{\Omega}(\psi^*\phi^*\tilde{F}_0) &= \chi_{\psi\circ\phi}(F_0) \\ &= \chi_\phi(F_0)\chi_\psi(\phi^*F_0) \\ &= \hat{\Omega}(\tilde{F})\chi_\psi(F).\end{aligned}$$

As in the previous theorem, the unitary equivalence of the representations follows from the equality of the generating functionals.

Q.E.D.

Corollary 3: Two representations of $\mathscr{S} \wedge \mathscr{K}$ in the Gel'fand–Vilenkin formalism, with equivalent measures, satisfying the conditions of Theorem 3 and having multiplier systems $\chi^1_\psi(F)$ and $\chi^2_\psi(F)$, are equivalent if the representations of the little group defined by the multiplier systems are equivalent.

Proof: If the representations of the little group are equivalent, then these representations are unitarily equivalent to the same induced representation.

Q.E.D.

Remark: If the representations are unitarily equivalent, there is a measurable function $Q:\Delta \to \mathbb{C}$ with $|Q(F)| = 1$ almost everywhere, such that

$$\chi^1_\psi(F) = Q^{-1}(F)\chi^2_\psi(F)Q(\psi^* F) \tag{4.4}$$

almost everywhere for each $\psi \in \mathscr{K}$. From Theorem 3, each $\chi^{1,2}_\psi(F)$ corresponds to a cyclic vector $\hat{\Omega}_{1,2}(\tilde{F})$ in the induced representation. Then $Q(F) = \mathrm{Arg}[\hat{\Omega}_2(\tilde{F})/\hat{\Omega}_1(\tilde{F})]$, which is independent of the choice of $\tilde{F}$.

Furthermore, since the phases of $\hat{\Omega}_1$ and $\hat{\Omega}_2$ can be defined everywhere (albeit in a nonunique manner), the systems of multipliers $\chi^1_\psi(F)$ and $\chi^2_\psi(F)$ can also be chosen to satisfy Eq. (2.9) everywhere, and $Q(F)$ can be defined everywhere so that Eq. (4.4) actually holds everywhere. This remark is utilized in Appendix B.

We have stated sufficient conditions on the orbit Δ and the multiplier system $\chi_\psi(F)$ to ensure that a representation of the little group $\mathscr{K}_F$ is well defined, and that equivalent representations of the little group which factor through the fundamental group result in equivalent representations of $\mathscr{S} \wedge \mathscr{K}$. In the next theorem, we complete the program of this section by stating conditions which are sufficient to ensure that *inequivalent* representations of the fundamental group induce *inequivalent* representations of $\mathscr{S} \wedge \mathscr{K}$. An additional condition on the orbit Δ appears to be needed.

Definition: An elementary neighborhood $\mathscr{O}$ of F_0 in Δ will be called a *splitting set* for Δ at the point F_0, associated with the element $g \in \Pi(\Delta, F_0)$, if there exists a family $\mathscr{F} = \{\mathscr{O}_i\}$ of open subsets of $\mathscr{O}$ such that (i) every open subset of $\mathscr{O}$ is the union of countably many mutually disjoint members of $\mathscr{F}$, together with a set of measure zero; and (ii) for each $\mathscr{O}_i \in \mathscr{F}$, there exists $\psi_i \in \mathscr{K}$ satisfying, for $\tilde{\mathscr{O}}_i$ contained in a fixed path component of $p^{-1}(\mathscr{O})$ with $p(\tilde{\mathscr{O}}_i) = \mathscr{O}_i$, (a) $\psi_i^* \tilde{\mathscr{O}}_i = g\tilde{\mathscr{O}}_i$ and (b) ψ_i is *weakly measure preserving* on $\tilde{\mathscr{O}}_i$ in the sense that there exist constants $c_1, c_2 > 0$ (independent of i) such that for any measurable subset M of $\tilde{\mathscr{O}}_i$, $c_2\mu(p(M)) \geqslant \mu(p(\psi_i^* M)) \geqslant c_1\mu(p(M))$.

We will also use the following lemma.

Lemma 11: For any measurable set $X \subset \mathscr{S}'$ with $\mu(X) > 0$, and for all $\delta > 0$, there exists an open set $\mathscr{O}$ such that $\mu(\mathscr{O} \cap X) > (1 - \delta)\mu(\mathscr{O})$. (In other words, "most" of $\mathscr{O}$ is contained in X.)

Proof: For any measurable set X, $\mu(X) = \inf_Z \mu(Z)$, where Z runs through all countable unions of open cylinder sets such that $X \subset Z$. (This is the property of regularity in the sense of Carotheodory.) Z may easily be selected to satisfy the condition of the lemma.

Q.E.D.

Theorem 4: Let μ be a quasi-invariant cylindrical measure concentrated on the measurable orbit $\Delta \subset \mathscr{S}'$ satisfying conditions (a)–(d). Let $\chi^1_\psi(F)$ and $\chi^2_\psi(F)$ be continuous systems of multipliers on Δ, and $\mathscr{Z}^1$ and $\mathscr{Z}^2$ be the associated sets of measure zero outside of which the continuity condition holds. Let $\chi^1_\psi(F)$ and $\chi^2_\psi(F)$ define, for each $F_0 \in \Delta \setminus (\mathscr{Z}^1 \cup \mathscr{Z}^2)$, inequivalent representations of $\mathscr{K}_{F_0}$ which factor through $\Pi(\Delta, F_0)$. Assume that for $F_0 \in \Delta$ and for some $g \in \Pi(\Delta, F_0)$ such that $T^1_g \neq T^2_g$ there exists a splitting set $\mathscr{O}$ for Δ. Then the representations of $\mathscr{K}$ defined by $\chi^1_\psi(F)$ and $\chi^2_\psi(F)$ are unitarily inequivalent.

Proof: The representations defined by $\chi^{1,2}_\psi(F)$ are equivalent, respectively, to induced representations on

$$\mathscr{H}_{1,2} = \{\hat{\Psi}:\tilde{\Delta} \to \mathbb{C} \mid \hat{\Psi}(g\tilde{F}) = T^{1,2}_g \hat{\Psi}(\tilde{F}); \hat{\Psi} \text{ measurable and square integrable}\}.$$

Suppose the representations are unitarily equivalent by virtue of the unitary mapping $Q:\mathscr{H}_1 \to \mathscr{H}_2$. Since $\mathscr{U}_1(f)$ is equivalent to $\mathscr{U}_2(f)$, Q must be the operation of multiplication by a measurable function $Q(\tilde{F})$.

Let $\mathscr{O}$ be the splitting set for Δ at F_0, let $\tilde{\mathscr{O}}$ be a path component of $p^{-1}(\mathscr{O})$, and choose $\tilde{F}_0 \in \tilde{\mathscr{O}}$. Let $T^2_g(T^1_g)^{-1} = e^{i\beta} \neq 1$ for $\beta \neq 0$ between $-\pi$ and π. Let I denote the interval of the unit circle between $e^{-i\beta/3}$ and $e^{i\beta/3}$. Then $e^{i\beta} I \cap I = \phi$ and for $\tilde{F} \in \tilde{\mathscr{O}}$,

$$Q(g\tilde{F}) = T^2_g(T^1_g)^{-1} Q(\tilde{F}) = e^{i\beta} Q(\tilde{F}) \text{ almost everywhere}.$$

Define $X_I = \{\tilde{F} \in \tilde{\mathscr{O}} \mid Q(\tilde{F}) \in I\}$. There exists some rotation I' of I so that $\mu(p(X_{I'})) > 0$, and $e^{i\beta} I' \cap I' = \phi$. The set $p(X_{I'})$ is measurable; hence, by Lemma 11 there is for any $\delta > 0$ an open set $\mathscr{O}' \subset \mathscr{O}$ such that $\mu(\mathscr{O}' \cap p(X_{I'})) > (1 - \delta)\mu(\mathscr{O}')$. Write $\mathscr{O}'$ as the countable union of a mutually disjoint subfamily $\{\mathscr{O}'_j\}$ of the family $\mathscr{F}$ of open subsets of $\mathscr{O}$, together with a set of measure zero. Then there is at least one such set $\mathscr{O}'_0$ such that $\mu(\mathscr{O}'_0 \cap p(X_{I'})) > (1 - \delta)\mu(\mathscr{O}'_0)$.

Let $\tilde{\mathscr{O}}'_0$ be $p^{-1}(\mathscr{O}'_0) \cap \tilde{\mathscr{O}}$. Then by hypothesis there exists ψ such that $\psi^* \tilde{\mathscr{O}}'_0 = g\tilde{\mathscr{O}}'_0$. However, since $Q\mathscr{V}_1(\psi) = \mathscr{V}_2(\psi)Q$ for all ψ, $Q(\psi^* \tilde{F}) = Q(\tilde{F})$ almost everywhere.

Thus we have (a) Since $Q(\psi^* \tilde{F}) = Q(\tilde{F})$ almost everywhere, and since ψ^* is weakly measure preserving,

$$\mu(p(\{\tilde{F} \in \psi^* \tilde{\mathscr{O}}'_0 \mid Q(\tilde{F}) \in I'\})) \geqslant c_1\mu(p(\{\tilde{F} \in \tilde{\mathscr{O}}'_0 \mid Q(\tilde{F}) \in I'\})) > c_1(1 - \delta)\mu(\mathscr{O}'_0).$$

However, we also have (b) since $\psi^* \tilde{\mathscr{O}}'_0 = g\tilde{\mathscr{O}}'_0$, and $Q(g\tilde{F}) = e^{i\beta} Q(\tilde{F})$ with $e^{i\beta} I' \cap I' = \phi$,

$$\mu(p(\{\tilde{F} \in \psi^* \tilde{\mathscr{O}}'_0 \mid Q(\tilde{F}) \in I'\})) \leqslant c_2\mu(p(\{\tilde{F} \in \tilde{\mathscr{O}}'_0 \mid Q(\tilde{F}) \notin I'\})) < c_2\delta\mu(\mathscr{O}'_0).$$

Thus, by choosing δ sufficiently small in relation to c_1 and c_2 we have a contradiction, and the postulated unitary equivalence Q cannot exist.

Q.E.D.

Remark: Theorem 4 is a generalization of a theorem proved in Ref. 3 for the N-particle orbits. The N-particle case provides an example of a splitting set as follows:

Let $g \in S_N$ and $F_0 = F_{x_1^0} + \cdots + F_{x_N^0} \in \Delta_N^{(s)}$, $s \geqslant 2$. Define $R = \min_{i \neq j} |\mathbf{x}_i^0 - \mathbf{x}_j^0|$ and $B_r(\mathbf{x}_1, \cdots, \mathbf{x}_N) = \{(\mathbf{y}_1, \ldots, \mathbf{y}_N) \in \mathbb{R}^{sN} \setminus D; |\mathbf{y}_i - \mathbf{x}_i| < r\}$. Now let the splitting set be $\mathcal{O} = p[B_{R/3}(\mathbf{x}_1^0, \cdots, \mathbf{x}_N^0)]$.

Next consider the family of sets $\{p[B_r(\mathbf{x}_1, \cdots \mathbf{x}_N)] \subset \mathcal{O}\}$, with r and $\mathbf{x}_j$ rational numbers. This is a countable family of sets, and they may be labeled $\mathcal{O}_i$. Any open subset of $\mathcal{O}$ may be written as the union of a mutually disjoint subfamily of $\{\mathcal{O}_i\}$, together with a set of measure zero.

Furthermore, given the N disjoint balls $U_j = \{\mathbf{x} \in \mathbb{R}^s, |\mathbf{x} - \mathbf{x}_j| < r\}$, there is a diffeomorphism ψ, Lebesgue measure preserving on the U_j, such that $\psi(U_j) = U_{g_j}$, where g is a permutation of the N indices. Finally, for $\mathcal{O}_i = p[B_r(\mathbf{x}_1, \cdots, \mathbf{x}_N)]$, let $\tilde{\mathcal{O}}_i = B_r(\mathbf{x}_1, \cdots, \mathbf{x}_N)$ be in one of the connected components of $p^{-1}(\mathcal{O})$. Then $\psi^* \tilde{\mathcal{O}}_i = B_r(\mathbf{x}_{g_1}, \cdots, \mathbf{x}_{g_N}) = g\tilde{\mathcal{O}}_i$. The N-particle measure may be written $d\mu(F_{\mathbf{x}_1} + \cdots + F_{\mathbf{x}_N}) = |\hat{\Omega}(\mathbf{x}_1, \cdots, \mathbf{x}_N)|^2 d^s x_1 \cdots d^s x_N$, where the cyclic vector $\hat{\Omega}$ is nonzero on $p^{-1}(\mathcal{O})$. Then ψ is weakly measure preserving with respect to μ, with $c_1 = c_2^{-1} = \inf_{p^{-1}(\mathcal{O})} |\hat{\Omega}|^2 / \sup_{p^{-1}(\mathcal{O})} |\hat{\Omega}|^2$. Thus, $\mathcal{O}$ has all properties of a splitting set.

V. CONCLUSIONS

Let us summarize the assumptions which permitted us to prove Theorems 1–4, and then note in what sense we have a partial extension of the Mackey theory.

We have assumed a quasi-invariant measure μ in $\mathcal{S}'$ which is concentrated on a single measurable orbit Δ. In the restricted weak topology, Δ is locally path connected and semilocally simply connected. The action of $\mathcal{K}$ on Δ (given by $\psi \to \psi^* F$) is such that (a) closed loops in $\mathcal{K}$ map into loops which shrink to a point in Δ; (b) any loop in Δ is equivalent to one obtained as the image of a path in $\mathcal{K}$; (c) there is an open neighborhood $\mathcal{O}$ of $\mathbf{e}$ in $\mathcal{K}$ that maps into an elementary neighborhood U in Δ, such that a path exists in $\mathcal{K}$ connecting $\mathbf{e}$ to each $\psi \in \mathcal{O}$ which maps into a path equivalent to one in U; (d) the map $\psi \to \psi^* F$ is open from $\mathcal{K} \to \Delta$; and (e) for some element of the fundamental group, there is a splitting set for Δ under the action of $\mathcal{K}$.

We have noted that (a) is not necessary but simplifies the presentation; and that if $\mathcal{K}$ is locally path connected, (c) is automatic, while (b) follows from (d). With these conditions, a continuous system of multipliers determines a representation of the little group $\mathcal{K}_F$. Conversely, a representation of $\mathcal{K}_F$ which factors through the fundamental group of Δ induces a representation of $\mathcal{S} \wedge \mathcal{K}$ which has a continuous system of multipliers. Two induced representations are unitarily equivalent if and only if the corresponding representations of $\mathcal{K}_F$ are equivalent. This paper has focused on the sufficiency but not the necessity of the assumptions made. We especially do not exclude the likelihood of weakening or eliminating (e) above.

The development has been motivated by the N-particle representations in the Gel'fand–Vilenkin formalism, and their connection with ordinary wave function representations. All of the assumptions hold for the N-particle orbits $\Delta_N^{(s)}$ in $\mathcal{S}'$.

In Ref. 13, it was shown how the Mackey theory could be extended naturally to the semidirect product of a nuclear space with a locally compact group. In the case of a discrete little group, induced representations constructed using the techniques of the present paper would fall within this extension of Mackey theory. At the other extreme, if the little group is locally compact and path connected, the techniques of the present paper are inapplicable to recovery of the representations obtained by the Mackey theory. In this respect only a very partial extension of the theory has been achieved. However, a covering space is of course a special case of a fiber space, in which the fiber is discrete. It is possible that under appropriate conditions, the consideration of covering spaces of orbits discussed in this paper could be replaced by consideration of fiber spaces over orbits. The representations of the little group could be required to factor through the structural group of a fiber bundle. Possibly, a wider extension of Mackey theory could then be achieved.

Important work on representations of $\mathcal{S} \wedge \mathcal{K}$ related to this paper has been carried out by Vershik, Gel'fand, and Graev, who construct a representation of $\mathcal{K}$ from a measure on configuration space and a representation of S_N.[8] They consider both the N-particle case and the case of infinitely many particles with Poisson measure; they also treat representations of S_N with dimension greater than one. These representations form a ring (under the tensor product operation), and one can extend a representation of $\mathcal{K}$ from this ring to a representation of $\mathcal{S} \wedge \mathcal{K}$ in a natural way, representing elements of $\mathcal{S}$ by multiplication operators.

The representations of Ref. 8 do not exhaust those of physical interest. For example, let μ be concentrated on the orbit $\Delta = \{F = \Sigma_{i=1}^{M} \lambda_i \Sigma_{j=1}^{n_i} F_{\mathbf{x}_j^i}$; with the $\mathbf{x}_j^i$ all distinct, and $\lambda_i \in \mathbb{R}$ fixed$\}$. We then obtain a representation describing M distinct species of particles.[16] Since μ is concentrated on a single orbit, the representation is irreducible. If $\rho(f)$ describes the total mass density (the charge density), then the λ_i are the masses (the charges) of each species. As representations of $\mathcal{K}$, these representations are in the Vershik–Gel'fand–Graev ring; but not as representations of $\mathcal{S} \wedge \mathcal{K}$. From the physicist's standpoint, it is perhaps more natural to begin with a representation of $\mathcal{S}$ (describing local particle density), and induce representations of $\mathcal{S} \wedge \mathcal{K}$.

We conclude with mention of two possible next steps: (1) to establish conditions which permit extension of the results of this paper to the infinite-particle, multiorbit case; and (2) to examine the case of many-dimensional representations of the fundamental group, which induce representations of $\mathcal{S} \wedge \mathcal{K}$ that are cyclic for $\{\mathcal{U}(f)\, \mathcal{V}(\psi)\}$ but not cyclic for $\{\mathcal{U}(f)\}$. For the N-particle orbits, we believe this would correspond to parastatistics.

In addition, enlarging the current algebra to include particle spin densities is of physical interest.

ACKNOWLEDGMENTS

The authors are indebted to Arnold A. Dicke for stimulating discussions and active collaboration during initial

phases of this research. We are also grateful to Professor Robert F. Wheeler for outlining the results in Appendix A. One of the authors (GG) would like to thank the Los Alamos Scientific Laboratory for hospitality during the course of this research.

APPENDIX A

In this Appendix $\mathscr{S}$ is any nuclear Fréchet space, i.e., a complete metrizable nuclear space (Schwartz' space falls in this category). We summarize some results about $\mathscr{S}'$, culminating in the fact that every weakly closed (or open) set in $\mathscr{S}'$ is contained in the σ algebra generated by the cylinder sets with Borel base. Therefore, every continuous complex-valued function on $\mathscr{S}'$ is measurable.

$\mathscr{S}$ is separable (Gel'fand–Vilenkin, p. 73).[9] Let $U\subset\mathscr{S}$. $U^0=\{F\in\mathscr{S}'\,|\,|(F,f)|\leqslant 1\ (\forall f\in U)\}$ is called the *polar* of U. Since $\mathscr{S}$ is separable and metric, any subset $U\subset\mathscr{S}$ contains a countable everywhere dense set $\{f_n\}$. Since $F\in\mathscr{S}'$ is continuous, we have $F\in U^0$ if and only if $|(F,f_n)|\leqslant 1\ (\forall f_n)$. Thus, U^0 is the countable intersection of the (closed) cylinder sets $\{F\,|\,|(F,f_n)|\leqslant 1\}$ and is measurable (Gel'fand–Vilenkin, p. 313).

If $\mathscr{E}$ and $\mathscr{F}$ are Hausdorff topological vector spaces, and $\mathscr{L}(\mathscr{E},\mathscr{F})$ denotes the space of continuous linear maps from $\mathscr{E}$ into $\mathscr{F}$, a subset H of $\mathscr{L}(\mathscr{E},\mathscr{F})$ is *equicontinuous* if and only if for every neighborhood V of 0 in $\mathscr{F}$ there is a neighborhood $\mathscr{O}$ of 0 in $\mathscr{E}$ such that $\cup_{F\in H}F(\mathscr{O})\subset V$ (Schaefer, p. 83).[21] Suppose that U is a neighborhood of zero in $\mathscr{S}$. Then U^0 is equicontinuous in $\mathscr{L}(\mathscr{S},\mathbb{R})=\mathscr{S}'$; for if the neighborhood V of 0 in $\mathbb{R}$ contains $\{\lambda\,|\,|\lambda|<r\}$, pick $\mathscr{O}=(r/2)U$; then $(\forall F\in U^0)\ (\forall f\in\mathscr{O}),\ |(F,f)|\leqslant(r/2)$, and $(F,f)\in V$. Similarly, if U is a neighborhood of any other point in $\mathscr{S}$, U^0 is equicontinuous in $\mathscr{S}'$. For a topological vector space $\mathscr{E}$, every equicontinuous set in $\mathscr{E}'$ is relatively compact for the weak topology (Schaefer, p. 84). Thus, U^0 is relatively compact; since it is closed, it is compact.

If $H\subset\mathscr{L}(\mathscr{E},\mathscr{F})$ is equicontinuous where $\mathscr{E}$ is separable and $\mathscr{F}$ is metrizable, then the restriction to H of the topology of simple convergence is metrizable. If $\mathscr{F}$ is also separable, then H is separable for this topology (Schaefer, p. 87). Thus, the restriction to U^0 of the weak topology in $\mathscr{S}'$ is metrizable and U^0 is separable for this topology (where U is a neighborhood in $\mathscr{S}$).

Let U_n be a decreasing sequence of neighborhoods of 0 in $\mathscr{S}$, constituting a base for the topology at 0; then $\cap_{n=1}^{\infty}U_n=\{0\}$. If $F\in\mathscr{S}'$, there is a neighborhood U_n with $F\in U_n^0$ (since F is continuous and equal to zero at $f=0$). Thus, $\mathscr{S}'=\cup_{n=1}^{\infty}U_n^0$, and we have expressed $\mathscr{S}'$ as the union of a countable increasing family of compact metric spaces.

Finally, let C be a weakly closed subset of $\mathscr{S}'$. To show that C is measurable it is sufficient to show $C_{\cap}U_n^0$ to be measurable. Since U_n^0 is compact, $C_{\cap}U_n^0$ is compact. Since U_n^0 is metric, $C_{\cap}U_n^0$ is the countable intersection of a decreasing sequence of open sets (in U_n^0): $C_{\cap}U_n^0=\cap_{m=1}^{\infty}\mathscr{O}_m$, where if d is the metric in U_n^0, $\mathscr{O}_m=\cup_{G\in C\cdot U^0}\{F\in U_n^0\,|\,d(F,G)<\delta_m\}$, and where δ_m is a sequence of positive real numbers tending towards zero. However, the open cylinder sets (intersected with U_n^0) form a base for the topology in U_n^0. Thus, $\mathscr{O}_m=\cup_{\gamma\in\Gamma_m}\mathscr{O}_\gamma^m$, where Γ_m is an index set and $\mathscr{O}_\gamma^m$ is a cylinder set intersected with U_n^0. Each family $\{\mathscr{O}_\gamma^m\,|\,\gamma\in\Gamma_m\}$ is an open covering for $C_{\cap}U_n^0$, and thus has a finite subcovering $\{\mathscr{O}_1^m,\cdots\mathscr{O}_{J_m}^m\}$. Now with $\mathscr{O}'_m=\cup_{\gamma=1}^{J_m}\mathscr{O}_\gamma^m$, we have $C_{\cap}U_n^0\subset\mathscr{O}'_m\subset\mathscr{O}_m$ with $\mathscr{O}'_m$ measurable. Thus, $C_{\cap}U_n^0=\cap_{m=1}^{\infty}\mathscr{O}'_m$ is measurable, which is the desired result.

It may be further remarked that for a nuclear Fréchet space $\mathscr{S}$, the strong topology in $\mathscr{S}'$ coincides with the weak topology on each U_n^0; thus, any strongly closed set in $\mathscr{S}'$ is also measurable, and strongly continuous complex-valued functions on $\mathscr{S}'$ are measurable.

APPENDIX B

For the representations of $\mathscr{S}\wedge\mathscr{K}$ considered here, the systems of multipliers define cocycles in Δ, as outlined below. Standard definitions of n-simplex, n-chain, n-cycle, and the boundary mapping are assumed.[22]

A continuous mapping $\sigma:I\to\Delta$ (i.e., a path in Δ) is a singular 1-simplex in Δ. The p-chain group $C^p(\Delta)$ contains the formal finite linear combinations $\alpha=\Sigma_{i=1}^{n}a_i\sigma_i$ of singular p simplexes on Δ, with integer coefficients a_i. A 1-cycle on Δ is a 1-chain with zero boundary. A 1-cochain is a homomorphism from $C^1(\Delta)$ into the multiplicative group $\mathscr{U}(1)$ of complex numbers of modulus one, and a 1-cocycle on Δ is a 1-cochain which has the value $+1$ on every boundary.

For $\psi\in\mathscr{K}$ and $F\in\Delta$, let $\psi_t\in\mathscr{K}$ be a path from $\mathbf{e}$ to ψ in $\mathscr{K}$, and consider the singular 1-simplex given by $\psi_t^*F\in\Delta$. Two singular 1-simplexes obtained in this fashion from paths in $\mathscr{K}$ differ by a boundary because of condition (a) in Sec. IV. Thus, a 1-cocycle takes on the same value on both singular 1-simplexes, and this value may be written $\chi_\psi(F)$. Then Eq. (2.9) states the fact that a 1-cocycle is a homomorphism from $C^1(\Delta)$ into $\mathscr{U}(1)$ (where we note that the sum of the singular 1-simplexes ψ_{2t}^*F and $\psi_{1t}^*\psi_2^*F$ differs from the singular 1-simplex $(\psi_1\circ\psi_2)_t^*F$ by a boundary).

A 1-cochain is a coboundary if it is a boundary mapping followed by a 0-cochain. A 0-cochain Q maps points in Δ into $\mathscr{U}(1)$. Thus, a 1-coboundary when evaluated on a path connecting F to ψ^*F in Δ may be written $Q(\psi^*F)\,Q(F)^{-1}$. Two 1-cocycles are cohomologous if they differ by a coboundary. Thus, $\chi_\psi^1(F)$ and $\chi_\psi^2(F)$ are cohomologous if they satisfy

$$\chi_\psi^1(F)=Q(F)^{-1}\chi_\psi^2(F)Q(\psi^*F),$$

which we recognize as Eq. (4.4) relating systems of multipliers obtained from unitarily equivalent representations of $\mathscr{S}\wedge\mathscr{K}$.

What we have called a system of multipliers differs from a cocycle in that it is known to be measurable (for each ψ), but on the other hand is defined and satisfies Eq. (2.9) only up to sets of measure zero. In Sec. IV we observe that a system of multipliers arising from an induced representation of $\mathscr{S}\wedge\mathscr{K}$ can be chosen so that Eq. (2.9) holds everywhere and likewise unitarily equivalent systems of multipliers can be chosen so that Eq. (4.4) holds everywhere. For these representations, it is therefore possible to identify Gel'fand–

Vilenkin systems of multipliers with cocycles. However, if the little group representation determined by the system of multipliers does not factor through the fundamental group of Δ, this correspondence appears to break down.

[1]R.F. Dashen and D.H. Sharp, Phys. Rev. **165**, 1857 (1968).
[2]J. Grodnik and D.H. Sharp, Phys. Rev. D **1**, 1546 (1970).
[3]G.A. Goldin, J. Math. Phys. **12**, 462 (1971).
[4]R.S. Ismagilov, Izv. Akad. Nauk SSSR Ser. Mat. **36**, (1972) [English translation: Math. USSR Izv. **6**, 181 (1972)].
[5]G.A. Goldin, J. Grodnik, R.T. Powers, and D.H. Sharp, J. Math. Phys. **15**, 88 (1974).
[6]R. Menikoff, J. Math. Phys. **15**, 1138 (1974).
[7]R. Menikoff, J. Math. Phys. **15**, 1394 (1974).
[8]A.M. Vershik, I.M. Gel'fand, and M.I. Graev, Usp. Mat. Nauk **30**, 3 (1975) [English translation: Russ. Math. Surveys **30**, 1 (1975)].
[9]I. Gel'fand and N. Vilenkin, *Generalized Functions* (Academic, New York, 1964), Vol. 4.
[10]E.P. Wigner, Ann. Math. **40**, 149 (1939).
[11]G.W. Mackey, Ann. Math. **55**, 101 (1952).
[12]G.W. Mackey, *Induced Representations of Groups and Quantum Mechanics* (Benjamin, New York, 1968).
[13]L. Girardello and G. Parravicini, "Some Remarks on the Representations of Infinite Parameter Groups and of the BMS Group," in *Group Theoretical Methods in Physics* (Proc. 3rd International Colloquium), Vol. I, pp. 217–26 (Centre Nat. Recherche Sci., Centre Phys. Theor., Marseilles, 1974).
[14]G.A. Goldin, Ph. D. thesis, Princeton University, 1968 (unpublished).
[15]G.A. Goldin and D.H. Sharp, "Lie Algebras of Local Currents and Their Representations," in *Group Representations in Mathematics and Physics, Battelle Seattle* 1969 *Rencontres*, edited by V. Bargmann (Springer, New York, 1970), pp. 300–311.
[16]J. Grodnik and D.H. Sharp, Phys. Rev. D **1**, 1531 (1970).
[17]This statement corrects an error of notation in Ref. 3, p. 474.
[18]F.J. Bloore and S.J. Swarbrick, J. Math. Phys. **19**, 878 (1978).
[19]For references on the topology of configuration space see D. McDuff, Topology **14**, 91 (1975).
[20]The homotopy theory used here is adapted from W.S. Massey, *Algebraic Topology: An Introduction* (Harcourt, Brace and World, New York, 1967).
[21]H.H. Schaefer, *Topological Vector Spaces* (MacMillan, New York, 1966).
[22]A.H. Wallace, *Algebraic Topology: Homology and Cohomology* (Benjamin, New York, 1970).

On the dynamical symmetries of the Kepler problem[a)]

Yilmaz Akyildiz
Department of Mathematics, Middle East Technical University, Ankara, Turkey

(Received 20 October 1978; accepted for publication 4 May 1979)

We try to understand the geometry of the SO(n + 1,2) action on the Kepler Manifold of the n-dimensional hydrogen atom. We show that the SO(n + 1,2) symmetry of the Kepler Problem is closely related to the fact that the geodesic flow on T^*S^n is periodic. We also exhibit the orbit picture analog of the peculiar property of the corresponding SO(n + 1,2) representation; that is, it stays irreducible when restricted to SO(n + 1,1) subgroups.

INTRODUCTION

For negative values of the energy, the phase space of the n-dimensional two-body problem (also known as the hydrogen atom or Kepler Problem), after "regularization" of the flow, is symplectomorphic to the geodesic flow on the cotangent bundle of S^n minus the image of the zero section, $T^*S^n - \{0\}$.[2,3] Hence the group Diff(S^n) operates symplectomorphically on the Kepler manifold $T^*S^n - \{0\}$.

The transformation of the Kepler Problem to the geodesic flow on the sphere is ultimately related to V. Fock's treatment (1935)[4] of the quantum theoretical problem for the hydrogen atom which has an "accidental degeneracy". To exhibit the maximum invariance group of the problem, Fock applies the stereographic projection to the momentum variables and shows that the Kepler Problem in 3-dimensional space is reduced to the 4-dimensional Laplace equation (a problem of free motion on the surface of a 4-dimensional sphere). This conclusion beautifully explains the accidental degeneracy in the energy with respect to the total angular momentum, and has been generalized to the n-dimensional Kepler Problem by Alliluev.[5]

Hence the eigenfunctions of the discrete spectrum of the n-dimensional hydrogen atom, in the momentum representation, are homogeneous harmonic polynomials in (n +1)-variables and the wavefunctions belonging to one level realize a finite dimensional irreducible representation of SO(n +1) which has been regarded as the maximum invariance group for this problem. In this way one is directly able to read off the degeneracy of an energy level as the number of independent homogeneous harmonic polynomials of the corresponding degree.

For the 3-dimensional problem, Barut *et al.*[6] have shown that one can operate in the Hilbert space of these wavefunctions with a still larger group, isomorphic to the de Sitter group SO(4,1), in such a way that one gets an irreducible unitary (therefore infinite dimensional) representation of this noncompact group, hence introducing the concept of the dynamical or spectrum-generating symmetry in physics. They use Majorana's description of this representation in terms of creation and annihilation operators and observe that it can be unitarily extended to a representation of SO(4,2) on the same Hilbert space. This means that SO(4,2) is also a dynamical group for the 3-dimensional hydrogen atom.

Returning to the geometry of the Kepler manifold $T^*S^n - \{0\}$ we realize that the SO(n + 1) symmetry of the hydrogen atom problem corresponds to the SO(n + 1)-invariance of the geodesic flow on T^*S^n. Also, we notice that the isometries are a particular class of conformal transformations, consequently the action of SO(n + 1,1) $\subset$ Diff(S^n) on S^n extends the action of SO(n + 1). The natural realization of the carrier space of the representation of SO(n + 1,1), corresponding to its action on S^n, is given by 1/2-densities $D_{1/2}(S^n)$. After identifying $D_{1/2}(S^n)$ with $C^\infty(S^n)$ via a Riemannian metric on S^n, as a basis of the Hilbert space we can take the spherical harmonics which are at the same time the wavefunctions of the corresponding quantum mechanical problem.

Furthermore, the action of SO(n +1,1) lifted to the cotangent bundle of S^n is transitive on $T^*S^n - \{0\}$, therefore its action on the Kepler manifold is symplectomorphic to the action of SO(n +1,1) on an orbit of the (co-) adjoint representation.[3]

$T^*S^n - \{0\}$ is also a symplectic homogeneous space for SO(n +1,2) which is not a subgroup of Diff(S^n).[3] Therefore one expects that the new generators of SO(n +1,2) have quite a different nature than the generators of SO(n +1,1)-action.

We tried to understand the geometry of the SO(n + 1,2) action. First we review the meaning of invariance and dynamical groups of an equation, then we continue with Alliluev's generalization to arbitrary dimensions of V. Fock's method to explain the degeneracy of the energy levels in terms of the symmetry group SO(n + 1). We make a Fourier transformation and write the Schrödinger equation in momentum space. Then we perform a stereographic projection and arrive at the Laplace equation on an n-dimensional sphere as the equivalent equation. This shows that the problem is rotationally invariant in an (n + 1)-dimensional space. In this space we exhibit the so(n + 1), so(n + 2), so(n + 3), so(n + 1,1), and so(n + 1,2) symmetries of the problem in terms of first order differential operators acting on the same space. Hence we obtain the Lie algebras of their representations. At this point we notice that in these representations of the generators of the algebras there are two ways of extending so(n + 1) to so(n + 2), or to so(n + 1,1), but there is only one way to extend the representation of

[a)]These results are taken from the author's dissertation under Professor Alan Weinstein.

 0022-2488/80/040665-06$1.00

so($n+2$) to so($n+3$), and the same is true for the extension so($n+1,1$) to so($n+1,2$). Furthermore, in the case of so($n+1,2$) we show that the new generator of the maximum compact subalgebra is unique and

$$= \left[\Delta_{S^n} + \left(\frac{n-1}{2}\right)^2\right]^{1/2}.$$

We notice that this generator has the same spectrum and multiplicity as the operator $(-2H)^{-1/2}$ where H is the usual quantum mechanical Hamiltonian for the n-dimensional hydrogen atom (negative energies only). Hence we show that the inclusion of the inverse of the generator of the Hamiltonian flow leads to SO($n+1,2$) as one of the dynamical groups of the Kepler Problem.

We also notice that the Laplacian Δ_{S^n} acting on the basis vectors chosen for the representation space is diagonal with eigenvalues $\mu(\mu+n-1)$, $\mu\in\mathbb{Z}^+$. Then $[(n-1)/2]^2$ is the only quantity that completes $\mu(\mu+n-1)$ into a perfect square ($\forall\mu$) so that the operator $(\Delta_{S^n}+?)^{1/2}$ has integer eigenvalues (up to an additive constant 1/2 in case n is even. This is exactly the Maslov (Morse) index mod 4 of a simple closed geodesic on an even-dimensional sphere. It is equal to 0 mod 4 for n odd.) We also find a uniqueness property for the same operator in the representation theory context. And, the group theoretical quantization picks up this very specific quantity. This fact is, of course, connected with the fact that the geodesic flow on T^*S^n is periodic. Professor Guillemin told us that similar additional terms must be added to the Laplacian in other contexts such as scattering theory and Hadamard's conjecture on Huygens' Principle.[7] See also p. 6 of Ref. 8, p. 114 of Ref. 9, and Ref. 10.

In this work, we will usually work in the framework of Lie algebras rather than the groups. We can always go over from the groups to the algebras by differentiation. Conversely, by using a theorem of Harish–Chandra we show that the representations of the algebras so($n+1,1$) and so($n+1,2$) constructed here are integrable to the group levels. We also show the irreducibility of these representations, hence prove that the set of wavefunctions of the hydrogen atom can be put into a single irreducible representation of SO($n+1,1$) which can be uniquely extended to a representation of SO($n+1,2$). In this way we identify SO($n+1,1$) and SO($n+1,2$) as the dynamical groups of the Kepler Problem.

In the last section we go to the Kirillov–Kostant–Souriau picture[11] of the same problem and there find the orbit theory analogue of the peculiar fact that this representation of SO($n+1,2$) stays irreducible when it is restricted to SO($n+1,1$).

1. SYMMETRIES OF EQUATIONS

We first define the meaning of symmetry of an equation. Suppose we have a function $\varphi(x_1,x_2,\ldots,x_n)$. We consider the relation

$$A\varphi = 0, \tag{1}$$

where A will be regarded, for simplicity, as a linear differential operator, athough this is immaterial in the reasoning that follows. The symmetries of Eq. (1) may be defined as the set of coordinate transformations that do not change the form of Eq. (1). Clearly they form a group. Or the symmetry of Eq. (1) may be taken to mean the set of operators $\{B_i\}$, forming a Lie algebra and satisfying.

$$[A,B_i] = 0, \quad \forall i, \tag{2}$$

In physics literature this kind of symmetry is called the *invariance algebra* of the equation. Clearly, this definition of the symmetry is broader than the previous one. In fact, if there is a Lie group of transformations that do not change the form of Eq. (1), one can go over to its infinitesimal transformations and obtain a Lie algebra of operators satisfying Eq. (2). We shall understand the symmetry of an equation in an even broader sense: Whenever we speak of the symmetry of Eq. (1) we mean that if some solution is obtained, then the symmetry transformations can be used to obtain also other solutions; i.e., the solution space of Eq. (1) is a basis of a representation of that Lie group (or Lie algebra) which is regarded to be the symmetry of the problem. When this representation is irreducible we call it the *dynamical symmetry* of the system. It is more convenient to talk about the Lie algebra rather than the group because one is then working with linear equations and we can always go over from the group level to the algebra by differentiation, so any representation of the group automatically gives a representation of its Lie algebra (but not vice versa). We can find such a symmetry if we find operators C_i forming a Lie algebra and satisfying

$$[A,C_i]\varphi = 0, \tag{3}$$

i.e., $\{C_i\}$ commute with the operator A not identically, but only on solution space of Eq. (1). Obviously, if φ is a solution of Eq. (1) so is also $C_i\varphi$, $\forall i$.

One can ask the following question: what irreducible representations of the symmetry algebra of Eq. (1) are realized as solutions. It is known from some examples such as the hydrogen atom and harmonic oscillator that not all representations possessed by the algebra are realized on solutions. As in the hydrogen atom case, it may happen that only one representation of the algebra is realized.

If the obtained symmetry algebra of the equation is so broad that all the solutions of the equation can be connected by operators from this algebra, then, in a sense, the algebra and its representation realized on the solution space replace the equation itself. This enables us to formulate the problem in pure algebraic form rather than in the language of equations.

2. SYMMETRIES OF THE KEPLER PROBLEM

In 1926 W. Pauli[12] found the spectrum of the Kepler Problem in a very elegant way by the use of the conservation of a second vector $\mathbf{R}$ beside the angular momentum:

$$\mathbf{R} = \tfrac{1}{2}[(\mathbf{p}\times\mathbf{L}) - (\mathbf{L}\times\mathbf{p})] - \frac{\mathbf{x}}{r} \text{ (Runge–Lenz Vector)}$$

is a second order differential operator in $\mathbf{x}$-space. However, it is a linear polynomial in $\mathbf{x}$. This suggests that we make a Fourier transformation from $\mathbf{x}$-space to $\mathbf{p}$-space and define a representation of the generators of the invariance algebra by differential operators in momentum space for we know that properly parametrized infinitesimal generators are linear

differential operators. This explains the second approach to the symmetry of the Kepler Problem due to V. Fock.

Following Fock and Alliluev we write the Schrödinger equation in momentum representation:

$$\left(\frac{p^2}{2} - E\right)\Psi(p) = \frac{\Gamma[(n-1)/2]}{2\pi^{(n+1)/2}} \int \frac{\Psi(p')\, d^n p'}{|p - p'|^{n-1}}.$$

This has those square integrable solutions which, in the new variables $(\xi_1,\xi_2,...,\xi_{n+1})$ defined by

$$p_i = \frac{p_0 \xi_i}{\xi_{n+1} + \Sigma_{j=1}^{n+1} \xi_j^2}, \quad i = 1,2,...,n,$$

are harmonic polynomials of ξ on the n-dimensional sphere S^n, i.e., they satisfy the equation

$$\Delta_{n+1}\, \Phi(\xi) = 0, \tag{4}$$

where

$$\Delta_{n+1} = \frac{\partial^2}{\partial \xi_1^2} + \cdots + \frac{\partial^2}{\partial \xi_{n+1}^2}.$$

Hence we arrive at the Laplace equation on S^n as the equivalent problem. (We notice that the mapping $(p_1,p_2,...,p_n) \to (\xi_1,\xi_2,...,\xi_n,\xi_{n+1})$ which is just the stereographic projection of the n-dimensional linear space to the unit sphere in $(n+1)$-dimensions, is such that the components of angular momentum

$$\mathscr{L}_{ij} = -i\left(\xi_i \frac{\partial}{\partial \xi_j} - \xi_j \frac{\partial}{\partial \xi_i}\right), \quad i,j = 1,2,...,n+1,$$

are constants of the motion.

Now we want to apply the philosophy of Part A to Eq. (4): A direct verification shows that the operators

$$M_{jk} = -i(\xi_j \partial_k - \xi_k \partial_j)\left(\partial_j \equiv \frac{\partial}{\partial \xi_j}\right), \quad j,k = 1,2,...(n+1),$$

commute with Δ_{n+1}. This is nothing else than the rotational invariance of the Laplace equation in $(n+1)$-dimensions. Now we search for more operators, preferably first order differential, which commute with Δ_{n+1} on the solution set of Eq. (4). We also want to close in a finite dimensional Lie algebra. Clearly

$$P_j = -i\partial_j, \quad j = 1,2,...,(n+1),$$

have the required properties:

$$[M_{jk},P_l] = i(\delta_{lj}P_k - \delta_{lk}P_j),$$

and

$$[P_i,P_j] = 0.$$

An other set of operators having the same properties is

$$Q_i = \xi^2 \partial_i - 2\xi_i \xi_l \partial_l + (1-n)\xi_i, \quad i = 1,2,...,(n+1),$$

$$\partial_k Q_i \Phi = [2\xi_k \partial_i + \xi^2 \partial_{ik}^2 - 2\delta_{ik}\xi_l \partial_l - 2\xi_i \partial_k - 2\xi_i \xi_l \partial_{kl}^2 + (1-n)\delta_{ki} + (1-n)\xi_l \partial_k]\Phi.$$

We are always assuming $\Delta\Phi = \partial_k \partial_k \Phi = \partial_{kk}^2 \Phi = 0$. Similarly we show that

$$\partial_k^2 Q_l \Phi = 0.$$

δ_{ij} denotes the Kronecker δ function. We are also employing the Einstein's summation convention on repeated indices. After a tedious computation we arrive at

$$[Q_i,Q_j] = 0.$$

Finally we compute

$$[P_k,Q_i] = 2\left[-i(\xi_k \partial_i - \xi_i \partial_k) + i\partial_{ki}\left(\xi_l \partial_l + \frac{n-1}{2}\right)\right]$$
$$= 2\left[M_{ki} + i\delta_{ki}\left(\xi_l \partial_l + \frac{n-1}{2}\right)\right].$$

This formula immediately forces us to take

$N \equiv \xi_l \partial_l + (n-1)/2$ as the new operator in order to close $\{M_{ij},P_i,Q_j\}$ into a finite dimensional Lie algebra. Furthermore we notice that N is the only operator with this property.

Summary: The $(1/2)(n+2)(n+3)$ operators

$$M_{jk} = -i(\xi_j \partial_k - \xi_k \partial_j),$$
$$P_i = -i\partial_i,$$
$$Q_j = \xi^2 \partial_j - 2\xi_j \xi_l \partial_l + (1-n)\xi_i,$$
$$N = \xi_l \partial_l + \frac{n-1}{2}, \quad i,j,k,l = 1,2,...,(n+1),$$

commute with Δ_{n+1} on the solution space of Eq. (4) and we have

$$[M_{jk},M_{ik}] = iM_{ji}, \quad [M_{ij},M_{kl}] = 0, \quad \text{for } ijkl \neq,$$
$$[M_{jk},P_i] = i(\delta_{ij}P_k - \delta_{ik}P_j),$$
$$[M_{jk},Q_i] = i(\delta_{ij}Q_k - \delta_{ik}Q_j),$$
$$[P_i,Q_j] = 2\{M_{ij} + i\delta_{ij}N\},$$
$$[N,P_i] = -P_i,$$
$$[N,Q_i] = Q_i,$$

i.e., $\{M_{ij},P_i,Q_j,N\}$ generate a finite dimensional Lie algebra. From these operators we can form the generators of all the compact and noncompact forms of complex Lie algebra so$(n+3)$ each of which containing the compact form of so$(n+1)$. We shall always define $L_{ij} = -L_{ji}$.

For example

$$L_{ij} = M_{ij},$$
$$L_{i\,n+2} = \tfrac{1}{2}(Q_i + iP_i),$$

generate the Lie algebra of SO$(n+2)$. The same is true for the set

$$L_{ij} = M_{ij},$$
$$L_{i\,n+3} = \tfrac{1}{2}(iQ_i + P_i).$$

We check that

$$[L_{i\,n+2},L_{j\,n+2}] = iM_{ij} = iL_{ij},$$
$$[L_{i\,n+3},L_{j\,n+3}] = iM_{ij} = iL_{ij}.$$

The equality

$$[L_{i\,n+2},L_{j\,n+3}] = i\delta_{ij}N,$$

implies that the operators

$$L_{ij} = M_{ij},$$
$$L_{i\,n+2} = \tfrac{1}{2}(Q_i + iP_i),$$
$$L_{j\,n+3} = \tfrac{1}{2}(iQ_j + P_j),$$
$$L_{n+2,n+3} = -N,$$

generate the Lie algebra of SO$(n+3)$. Now we define

$$\mathscr{L}_{in+2} = \tfrac{1}{2}(iQ_i - P_i),$$

and compute

$$[\mathscr{L}_{in+2}, \mathscr{L}_{jn+2}] = -iM_{ij}.$$

Hence the operators

$$\mathscr{L}_{ij} = M_{ij},$$
$$\mathscr{L}_{in+2} = \tfrac{1}{2}(iQ_i - P_i),$$

generate the Lie algebra of SO($n+1,1$). Also, the same is true for the set

$$\mathscr{L}_{ij} = M_{ij},$$
$$\mathscr{L}_{in+3} = \tfrac{1}{2}(iP_i - Q_i).$$

Similarly, the equality

$$[\mathscr{L}_{in+2}, \mathscr{L}_{jn+3}] = i\delta_{ij}N,$$

implies that the operators

$$L_{ij} = M_{ij},$$
$$\mathscr{L}_{in+2} = \tfrac{1}{2}(iQ_i - P_i),$$
$$\mathscr{L}_{jn+3} = \tfrac{1}{2}(iP_j - Q_j),$$
$$\mathscr{L}_{n+2,n+3} = N,$$

generate the Lie algebra of SO($n+1,2$).

Group	Generators of its Lie algebra
SO($n+1$)	M_{jk}
SO($n+2$)	$M_{ij}, \tfrac{1}{2}(Q_i + iP_i)$
SO($n+2$)	$M_{ij}, \tfrac{1}{2}(iQ_i + P_i)$
SO($n+1,1$)	$M_{ij}, \tfrac{1}{2}(iQ_i - P_i)$
SO($n+1,1$)	$M_{ij}, \tfrac{1}{2}(iP_i - Q_i)$
SO($n+3$)	$M_{ij}, \tfrac{1}{2}(Q_i + iP_i), \tfrac{1}{2}(iQ_i + P_i), -N$
SO($n+1,2$)	$M_{ij}, \tfrac{1}{2}(iQ_i - P_i), \tfrac{1}{2}(iP_i - Q_i), N.$

At this point we notice that in these differential operator representations of the generators of the Lie algebras, there are two ways of extending so($n+1$) to so($n+2$), or to so($n+1,1$), but there is only one way to extend the representation of so($n+2$) to so($n+3$) and the same is true for the extension of so($n+1,1$) to so($n+1,2$). Furthermore, in the case of so($n+1,2$) the new generator of the maximum compact subalgebra is unique and equals $\xi_l\partial_l + (n-1)/2$.

Now we consider $\{M_{ij}, P_i, Q_j, N\}$ acting on the solution space of Eq. (4), i.e., on the space of harmonic functions in ($n+1$)-variables. A natural basis for this space is the set of homogeneous polynomials $T^{\nu}_{i_1,\ldots,i_\nu}$ ($\nu =$ the degree of homogeneity) in the variables ξ_i that are totally symmetric in all indices and with all pair traces zero. Since $\partial_k\partial_k T^{\nu}_{i_1 i_2 \cdots i_\nu} = 0$ they, at the same time, represent harmonic polynomials of degree ν in an ($n+1$)-dimensional space Ref. 13, p. 400. The number of linearly independent harmonic polynomials of a given degree is equal to $M_\nu = (2\nu + n - 1)(\nu + n - 2)!/\nu!(n-1)!$, and they make up a basis of a finite dimensional irreducible representation of so($n+1$) algebra, and no subspace of our solution space is invariant under the action of the algebras containing linear combinations of P_k and Q_k, i.e., from any state we can obtain the entire set of states of the hydrogen atom. Therefore, for these algebras (compact and noncompact) we obtain irreducible representations. The representation space is infinite dimensional, therefore the representations of the compact algebras so($n+2$) and so($n+3$) cannot be Hermitian. (That is why we do not find these algebras in physics literature.) This fact also tells us that these representations are not coming from geometric actions on the Kepler manifold.

We go from a representation of the Lie algebra of a group to a representation of the group by the process of exponentiation, and the question arises when is this exponentiation possible. To prove the integrability of our representations for the algebras so($n+1,1$) and so($n+1,2$), we use a theorem of Harish–Chandra.[14] Thus we show that the ($n+1$)-dimensional spherical harmonics form the basis for a single irreducible representation of the groups SO($n+1,1$) and SO($n+1,2$); i.e., the n-dimensional Kepler Problem has SO($n+1,1$) and SO($n+1,2$) as its dynamical groups.

We easily observe that the representation of SO($n+1,2$) remains irreducible when we restrict ourselves to any one of its two SO($n+1,1$)-subgroups, if we realize that the generators of these subgroups leave no subspace of our original space invariant. In the next section we shall find the corresponding phenomenon in the orbit picture corresponding to this peculiar fact of this SO($n+1,2$)-representation. Professor I.T. Todorov told us that there is an infinite family of representations of SO(4,2) with this property.

Finally we want to relate the new generator of the maximum compact subalgebra of so($n+1,2$), $N = \xi_l\partial_l + [(n-1)/2]$, to more familiar objects: ($n+1$)-dimensional Laplacian in spherical coordinates Ref. 15, p. 507, reads

$$\Delta_{n+1} = \frac{1}{r^n}\partial_r r^n \partial_r + \frac{1}{r^2}\Omega,$$

where

$$r^2 = \xi_1^2 + \cdots + \xi_{n+1}^2,$$

and

$$\Omega = -\sum_{i<j}^{n+1} M_{ij}^2 = \text{the angular part of the Laplacian,}$$
$$= \text{Laplacian on } S^n = \Delta_{S^n},$$
$$= \text{Casimir operator of SO}(n+1).$$

In the same coordinates we have

$$N = r\partial_r + \frac{n-1}{2},$$

and

$$N^2 = \left(r\partial_r + \frac{n-1}{2}\right)\left(r\partial_r + \frac{n-1}{2}\right)$$
$$= r^2\left(\frac{1}{r^n}\partial_r r^n \partial_r\right) + \left(\frac{n-1}{2}\right)^2$$
$$= r^2\left\{\Delta_{n+1} - \frac{1}{r^2}\Omega\right\} + \left(\frac{n-1}{2}\right)^2$$
$$= r^2\Delta_{n+1} - \Omega + \left(\frac{n-1}{2}\right)^2.$$

Therefore, on the solution space of the Eq. (4); i.e., for $\Delta_{n+1}\Phi = 0$ we have

$$N^2\Phi = \left[-\Omega + \left(\frac{n-1}{2}\right)^2\right]\Phi,$$

or

$$N\Phi = \pm\left[\Delta_{S^n} + \left(\frac{n-1}{2}\right)^2\right]^{1/2}\Phi.$$

We realize that the Laplacian Δ_{S^n} acting on the basis vectors chosen for our representation space is diagonal with eigenvalues $\nu(\nu+n-1)$, $\nu\in\mathbb{Z}^+$. Then $[(n-1)/2]^2$ is the only quantity that completes $\nu(\nu+n-1)$ into a perfect square $(\forall\nu)$ so that the operator $(\Delta_{S^n}+?)^{1/2}$ has integer eigenvalues (up to an additive constant 1/2 in case n is even). (This quantity is exactly the Maslov index (which is the Morse index for geodesics) mod 4 of a simple closed geodesic on an even-dimensional sphere. Its actual value for S^n is $2(n-1)$, which is equal to 0 mod 4 when n is odd.) We recall that we have also found a uniqueness property for the operator N in the representation theory context. And the group theoretical quantization of the Kepler Problem picks up this very specific quantity. This fact is, of course, connected with the fact that the geodesic flow on T^*S^n is periodic. We also notice that N has the same spectrum: $\nu+(n-1)/2$, $\nu=0,1,2,\cdots$, and multiplicity M_ν as the operator $(-2H)^{-1/2}$ where H is the usual quantum mechanical Hamiltonian for the n-dimensional hydrogen atom (negative energies only). Hence we show that the inclusion of the inverse of the generator of the Hamiltonian flow on the Kepler Manifold leads to so$(n+1,2)$ as the other dynamical algebra of the Kepler Problem.

3. ON THE ORBIT PICTURE OF THE KEPLER MANIFOLD

As we mentioned it before, the Kepler manifold $T^*S^n-\{0\}$ is an SO$(n+1,1)$, and SO$(n+1,2)$, homogeneous symplectic space, corresponding to singular orbits in the (co-) adjoint representations. (It was observed that these orbits are nilpotent. Conversely, Wolf proved that in the semisimple case the nilpotent orbits are open subsets of cotengent bundles with zero sections deleted.[16]

Elhadad (1973) constructed the dynamical quantization of the SO(4,1)-orbit and obtained the energy levels and the wavefunctions of the Schrödinger equation in this picture.[17] But, he could not apply the same method, due to Kostant–Souriau, to SO(4,2)-orbits because he showed that this orbit does not carry an SO(4,2)-invariant polarization.

Later, Onofri (1975) came up with an alternate procedure and obtained a quantization of the SO$(n+1,2)$-orbit through a limit of quantizable manifolds, admitting invariant complex polarizations.[18] In this way he also recovered the usual quantum mechanical hydrogen atom model.

In this section we take Onofri's description of this singular SO$(n+1,2)$-orbit and show that it is transitive under the two SO$(n+1,1)$ subgroups. Comparing with the Theorem 6.1 of Ref. 11 we understand that this is the orbit picture analogue of the peculiar fact about the SO$(n+1,2)$-representation we have constructed in the previous section; that is, it stays irreducible when restricted to SO$(n+1,1)$ subgroups.

We also show that a family of regular SO$(n+1,2)$-orbits do not have this property; furthermore, we establish the fact that the SO$(n+1,1)$ actions foliate these nonsingular orbits with Lagrangian submanifolds.

For convenience we use new names for the generators of so$(n+1,2)$ algebra:

$$\left.\begin{aligned} M_{ij} &= \mathcal{L}_{ij} \\ N &= \mathcal{L}_{n+2,n+3} \end{aligned}\right\} \text{maximal compact subalgebra so}(n+1)\oplus\text{so}(2),$$

$$Z_i = \mathcal{L}_{i\,n+2},$$

$$W_j = \mathcal{L}_{j\,n+3}, \quad i,j = 1,2,\ldots,(n+1).$$

Since the orthogonal groups are semisimple their Killing forms

$$B(X,Y) = \frac{1}{2(n+1)}\,\text{Trace}(\text{ad}X\,\text{ad}Y),$$

are nonsingular. Explicitly, on the basis vectors of the so$(n+1,2)$ algebra we have

$$B(\mathcal{L}_{ij},\mathcal{L}_{hk}) = g_{ik}\,g_{jh} - g_{ih}\,g_{jk}, \tag{5}$$

where

$$g_{11} = g_{22} = \cdots = -g_{n+2,n+2} = -g_{n+3,n+3} = -1,$$
$$g_{ij} = 0 \text{ if } i\neq j.$$

Therefore through the Killing form we can identify the coadjoint representation with the adjoint representation.

A general point $w\in$so$(n+1,2)$ is given by

$$w = lN + \underset{i<j}{m_{ij}M_{ij}} + z_iZ_i + w_iW_i.$$

We shall denote by O_w the orbit of SO$(n+1,2)$ through w. Onofri shows that the orbit through

$$w_0 = N + M_{12} + Z_1 + W_2,$$

corresponds to the Kepler Manifold. To find the Lie algebra of the stability group of w_0 we solve the following equation for n, m_{ij}, z_i, and w_j:

$$[lN + m_{ij}M_{ij} + z_iZ_i + w_jW_j, N + M_{12} + Z_1 + W_2] = 0,$$

we obtain

$$\begin{aligned} &z_2 = 0,\\ &m_{12} + l = 0,\\ &m_{13} = 0,\\ &\vdots\\ &m_{1\,n+1} = 0,\\ &z_3 + m_{23} = 0,\\ &z_4 + m_{24} = 0,\\ &\vdots\\ &z_{n+1} + m_{2\,n+1} = 0. \end{aligned} \tag{6}$$

Here we have $2n$ independent equations and this tells us that the dimension of $O_{w_0} = 2n$, as it has to be. We notice that these equations do not involve w_i. This suggests that we look at O_{w_0} under the action of the SO$(n+1,1)$ subgroup whose Lie algebra is generated by $\{M_{ij}, Z_i\}$. Clearly, the stability subalgebra of this action is given by the same number of linearly independent equations. (We just have to remove l from the second equation in Eq. (6) and this does not change the linear independence.) This proves that the SO$(n+1)$ subgroup, corresponding to $\{M_{ij},Z_i\}$, acts transitively on O_{w_0}. It is also a routine computation to show that the stabil-

ity algebra of $\{M_{ij},Z_i\}$-action on w_0 is given by the same number of equations, so the same transitivity property holds for the corresponding SO($n+1,1$) subgroup.

Finally we look at some regular orbits of the form O_{w_l} with $w_l = lN$. (These are the orbits Onofri approximates O_{w_0} with.) The stability subgroup of w_l is given by the maximal compact subgroup SO($n+1$)×SO(2), therefore $\dim O_{w_l} = 2(n+1)$, and

$$TO_{w_l} \approx \mathrm{so}(n+1,2)/\mathrm{so}(n+1)\oplus\mathrm{so}(2)$$
$$\approx \{Z_i,\ W_j\} \equiv \text{the linear span of } Z_i, W_j.$$

The stabilty subalgebra of $\{M_{ij},Z_i\}$ at w_l is given by $\{M_{ij}\}$, therefore the tangent space at w_l of the SO($n+1,1$)-orbit through w_l is spanned by Z_i. (We are dealing with the SO($n+1,1$) subgroup corresponding to the Lie algebra generated by $\{M_{ij},Z_i\}$; similar results, of course, hold for the other SO($n+1,1$) subgroup when Z_i and W_i are interchanged).

Claim: $\{Z_i\}\subset\{Z_i,W_j\}$ is a Lagrangian subspace.

Proof: Clearly, $\dim\{Z_i\} = \frac{1}{2}\dim\{Z_i,W_j\}$.

The symplectic form on O_{w_l} at w_l is given by

$$(v_1,v_2)\rightarrow\langle w_l,[v_1,v_2]\rangle$$
$$= B(w_l,[v_1,v_2]),\quad v_i\in\{Z_i,W_j\}.$$

On the basis vectors of $\{Z_i\}$ we get

$$(Z_i,Z_j)\rightarrow \quad B(lN,[Z_i,Z_j])$$
$$= -ilB(N,M_{ij})$$
$$= -ilB(\mathscr{L}_{n+2,n+3},\mathscr{L}_{ij}).$$

Equation

$$(5)\Rightarrow \quad = -il\{g_{n+2\,j}\,g_{n+3\,i} - g_{n+2\,i}\,g_{n+3\,j}\}$$
$$= 0, \quad \text{because } i,j = 1,2,\ldots,(n+1).$$

Hence we have shown that SO($n+1,1$) actions foliate O_{w_l} with Lagrangian submanifolds. At the moment we do not know if this fact can be of any use.

ACKNOWLEDGMENTS

I wish to thank my supervisor, Professor Alan Weinstein for his assistance during this research. I would like to acknowledge the hospitality of I.H.E.S., and I am very much indebted to all the mathematicians and physicists there during the 1975–76 academic year.

[1] Y. Akyildiz, Ph.D. thesis, Univ. of Calif., Berkeley, 1976.
[2] J. Moser, Commun. Pure Appl. Math. **23**, 609 (1970).
[3] J.M. Souriau, Convegno di Geometria Simplettica (Rome, 1973).
[4] V. Fock, Z. Phys. **98**, 145 (1935).
[5] Alliluev, Zh. Eksp. Teor. Fiz. **33**, 200 (1957) [Sov. Phys. JETP **6**, 156 (1958)].
[6] A.O. Barut and G.L. Bornzin, J. Math. Phys. **15**, 1000 (1974).
[7] S. Helgason, Acta. Math. **102**, 239–99 (1959); Am. J. Math. **86**, 565–601 (1964).
[8] B. Kostant, Adv. Math. **20**, 179–212 (1976).
[9] V. Guillemin, Adv. Math. **22**, 85–119 (1976).
[10] A. Weinstein, Convegno di Geometria Simplettica (Rome, 1973); Lectures on Symplectic Manifolds, CBM S/29 Regional Conference Series in Mathematics, 1977.
[11] A.A. Kirillov, Russ. Math. Surv. **17**, 53–104 (1962).
[12] W. Pauli, *Die Allgemeinen Prinzipien der Wellenmechanick*, Handbuk der Physik (Springer–Verlag, Berlin, 1958).
[13] M. Hamermesh, *Group Theory and its Application to Physical Problems* (Addison–Wesley, Reading, Massachusetts, 1962).
[14] Harish-Chandra, Trans. Am. Math. Soc. **76**, 26 (1954).
[15] N. Vilenkin, *Fonctions Spéciales et Theories de la Représentation des Groupes* (Dunod, Paris, 1969).
[16] J.A. Wolf, Proc. Am. Math. Soc., **51**, 213–6 (1975).
[17] J. Elhadad, Convegno di Geometria Simplettica (Rome, 1973).
[18] E. Onofri, Lecture Notes in Math. (Springer-Verlag, Berlin, 1977), Vol. 570.

Missing label operators in the reduction Sp(2n)↓Sp(2n −2)

Adam M. Bincer
Physics Department, University of Wisconsin-Madison, Madison, WI 53706

(Received 28 September 1979; accepted for publication 10 December 1979)

I consider the "missing label" problem for basis vectors of an Sp($2n$) representation corresponding to a group reduction chain with links Sp(2ν)↓Sp($2\nu-2$)×U(1), $2\leqslant\nu\leqslant n$. I obtain two different sets of $\nu-1$ missing label operators. These operators, together with the ν Casimir operators of the Sp(2ν) group, the $\nu-1$ Casimir operators of the Sp($2\nu-2$) subgroup and the one generator of the U(1) subgroup, form a complete set of labeling operators whose eigenfunctions provide a canonical basis in the representation space of Sp(2ν). When the number of missing labels exceeds one the most general solution to the labeling problem is not known. The two particular solutions presented here have certain appealing aspects of symmetry and simplicity.

1. INTRODUCTION

An important problem in the applications of group theory is the explicit construction of a "canonical basis" for the irreducible representations of the group. A canonical basis consists of a complete nondegenerate set of orthonormal basis functions. Clearly this can be achieved by choosing as basis functions the (normalized) eigenfunctions of a complete set of commuting Hermitian operators. In two remarkable but obscure papers, Gel'fand and Zetlin[1] completely solved this problem for all finite-dimensional irreducible representations of the unitary and orthogonal groups. Their results have been rederived since in various ways, most recently by me using shift operators in the form of tensor operators.[2] The Gel'fand–Zetlin patterns, that label a basis vector, consist of an array of integers or semi-integers which can be put into a one-to-one correspondence with the eigenvalues of the Casimir operators of the group and its subgroups in a canonical chain. Thus, e.g., for the orthogonal group in n dimensions, O(n), the canonical chain is

$$\mathrm{O}(n)\downarrow\mathrm{O}(n-1)\downarrow\cdots\mathrm{O}(3)\downarrow\mathrm{O}(2), \tag{1.1}$$

and the complete set of commuting operators consists of the Casimir operators of O(ν), $2\leqslant\nu\leqslant n$. Because the reduction O(n)↓O($n-1$) is multiplicity-free (i.e., a given irrep of O($n-1$) occurs in an irrep of O(n) at most once] it is appropriate to apply to this procedure the term "canonical" as defined by Baird and Biedenharn.[3] Similarly, because the reduction U(n)↓U($n-1$)×U(1) is multiplicity-free, the Gel'fand–Zetlin procedure also works for the unitary groups.

In the case of the remaining series of the classical groups, the symplectic group in $2n$ dimensions Sp($2n$), no such canonical basis appears in the literature. Here the links in the natural reduction chain are of the form

$$\mathrm{Sp}(2\nu)\downarrow\mathrm{Sp}(2\nu-2)\times\mathrm{U}(1),\quad 2\leqslant\nu\leqslant n. \tag{1.2}$$

This reduction is *not* multiplicity-free, the resultant chain of subgroups is *not* canonical and one is faced with the so-called "missing label problem." The branching law for the splitting of an irrep of Sp($2n$) under restriction to Sp($2n-2$) is known;[4] implicit in it is the solution to the state-labeling problem and the corresponding Gel'fand–Zetlin pattern has been used by Mickelsson[5] to construct a basis. It is clear from his work, however, that the basis functions are neither mutually orthogonal nor eigenfunctions of a complete set of commuting operators.

This difficulty is common to many other situations and arises whenever a noncanonical chain of subgroups is involved. The reduction SU(3)↓O(3) is perhaps the most famous such case;[6] it involves one missing label. The solution to the problem is, of course, to find a sufficient number of additional operators that would, together with the Casimir operators of group and subgroups, form a complete set. The form of the general missing label operator for seven cases with *one* missing label has been given by Sharp.[7] The problem becomes more complicated when two or more labels are missing because of the requirement that the missing label operators must commute with each other. A particular solution for the case SU(4)↓SU(2)×SU(2), where two labels are missing, was given by Moshinsky and Nagel.[8]

In the case of Sp($2n$)↓Sp($2n-2$)×U(1) the number of missing labels is $n-1$. In what follows I find two different solutions for the $n-1$ missing label operators. In Sec. 2 the notation is explained and the symmetric and antisymmetric tensors, as well as the invariants, are introduced. In Sec. 3 the missing label operators are explicitly constructed with the proof of their crucial property of commutativity relegated to the Appendix. I conclude with a discussion of the results in Sec. 4.

2. GENERATORS AND TENSORS

I denote the generators of Sp($2n$) by G^a_b with the indices ranging from $-n$ to n, zero excluded; the Sp($2n-2$) subgroup is obtained by omitting $-n$ and n from the range of the indices. In the Racah[9] basis the commutation relations are

$$[G^a_b,G^c_d]=\delta^c_b G^a_d-\delta^a_d G^c_b+\epsilon^a\epsilon^b(\delta^{\bar b}_d G^c_{\bar a}-\delta^c_{\bar a}G^{\bar b}_d), \tag{2.1}$$

where

$$\bar a\equiv -a,\quad \epsilon^a\equiv a/|a|. \tag{2.2}$$

These G's obey

$$G^a_b=-\epsilon^a\epsilon^b G^{\bar b}_{\bar a}, \tag{2.3}$$

 0022-2488/80/040671-04$1.00

so that the number of independent generators (order of the group) is $n(2n+1)$. Moreover, in unitary representations one has

$$G^{a\dagger}_{b} = G^{b}_{a}. \tag{2.4}$$

The Cartan subalgebra is generated by the n generators

$$G^{a}_{a}, \quad 1 \leqslant a \leqslant n, \tag{2.5}$$

so that the rank of Sp($2n$) is n. Thus the number of labels needed to specify an irrep of Sp($2n$) is n. The additional number of labels needed to uniquely specify a vector within an irrep is given by half the difference between order and rank, i.e., n^2 for Sp($2n$).

Consider the reduction Sp($2n$)$\downarrow$Sp($2n-2$)$\times$U(1) where the U(1) is generated by G^{n}_{n}. Assuming that the labeling has been solved for Sp($2n-2$) this subgroup will provide a number of labels equal to

$$n-1+\tfrac{1}{2}[(n-1)(2n-1)-(n-1)] = n^2 - n, \tag{2.6}$$

and since the U(1) subgroup provides one more label this reduction results in $n-1$ missing labels.

I define an Sp($2n$) tensor operator T^{a}_{b} by the commutation relations

$$[G^{a}_{b}, T^{c}_{d}] = \delta^{c}_{b} T^{a}_{d} - \delta^{a}_{d} T^{c}_{b} + \epsilon^{a}\epsilon^{b}(\delta^{\bar{b}}_{d} T^{c}_{\bar{a}} - \delta^{c}_{\bar{a}} T^{\bar{b}}_{d}). \tag{2.7}$$

Such a tensor is reducible into a symmetric and antisymmetric part:

$$S^{a}_{b} = \tfrac{1}{2}(T^{a}_{b} + \epsilon^{a}\epsilon^{b} T^{\bar{b}}_{\bar{a}}) = \epsilon^{a}\epsilon^{b} S^{\bar{b}}_{\bar{a}}, \tag{2.8}$$

$$A^{a}_{b} = \tfrac{1}{2}(T^{a}_{b} - \epsilon^{a}\epsilon^{b} T^{\bar{b}}_{\bar{a}}) = -\epsilon^{a}\epsilon^{b} A^{\bar{b}}_{\bar{a}}, \tag{2.9}$$

and the symmetric tensor may be reduced further into a traceless part

$$S^{a}_{b} - \delta^{a}_{b} C, \tag{2.10}$$

and the invariant

$$C = (2n)^{-1} \sum_{a=\bar{n}}^{n} S^{a}_{a}. \tag{2.11}$$

The most general Sp($2n$) tensor formed out of the generators may be expressed as some linear combination of the $2n-1$ linearly independent basic tensors

$$(G^{k})^{a}_{b}, \quad 1 \leqslant k \leqslant 2n-1, \tag{2.12}$$

where I define

$$(G^{k})^{a}_{b} = \sum_{c=\bar{n}}^{n} G^{a}_{c} (G^{k-1})^{c}_{b}, \quad (G^{0})^{a}_{b} = \delta^{a}_{b}. \tag{2.13}$$

An equivalent basic set is provided by the n antisymmetric tensors $A(k)$ together with the $n-1$ symmetric tensors $S(k)$ defined by

$$A(k)^{a}_{b} = \tfrac{1}{2}\{(G^{2k+1})^{a}_{b} - \epsilon^{a}\epsilon^{b}(G^{2k+1})^{\bar{b}}_{\bar{a}}\}, \quad 0 \leqslant k \leqslant n-1, \tag{2.14}$$

$$S(k)^{a}_{b} = \tfrac{1}{2}\{(G^{2k})^{a}_{b} + \epsilon^{a}\epsilon^{b}(G^{2k})^{\bar{b}}_{\bar{a}}\}, \quad 1 \leqslant k \leqslant n-1. \tag{2.15}$$

Similarly a complete set of Sp($2n$) scalars is given by the n invariants $C_n(k)$ defined by

$$C_n(k) = (2n)^{-1} \sum_{a=\bar{n}}^{n} (G^{2k})^{a}_{a}, \quad 1 \leqslant k \leqslant n; \tag{2.16}$$

these are the well-known n Casimir operators of Sp($2n$) whose eigenvalues may be used to provide the n labels needed to specify an irrep of Sp($2n$).

The validity of all the statements in the preceding paragraph may be deduced from the work of Green[10] and Nwachuku and Rashid[11] on characteristic polynomial identities satisfied by Sp($2n$) generators. Perhaps two aspects require comment. First, the fact that G^{2n}, or equivalently $S(n)$, is *not* an independent tensor—this is a consequence of the characteristic polynomial identity which allows one to reexpress this tensor in terms of those given by Eq. (2.12), or equivalently Eq. (2.15), and the invariant unit tensor I (here $I^{a}_{b} = \delta^{a}_{b}$). Second, the fact that antisymmetric tensors formed from even powers of the generators and symmetric tensors formed from odd powers are *not* independent—this is further elucidated in the Appendix and is ultimately due to the antisymmetry of the generators themselves [Eq. (2.3)].

3. THE MISSING LABEL OPERATORS

It is clear that in the reduction Sp($2n$)$\downarrow$Sp($2n-2$)$\times$U(1) the basis functions may be taken as simultaneous eigenfunctions of G^{n}_{n}, of the n Casimir operators $C_n(k)$ of Sp($2n$), and of the $n-1$ Casimir operators $C_{n-1}(k)$ of Sp($2n-2$). Here

$$C_{n-1}(k) = (2n-2)^{-1} \sum_{a=\bar{n}+1}^{n-1} (G^{2k})^{a}_{a}, \quad 1 \leqslant k \leqslant n-1, \tag{3.1}$$

where also the indices suppressed by the notation omit the values n and $\bar{n}$.

If this reduction is to be the desired first link in the formation of a chain of subgroups leading to a canonical basis I need $n-1$ missing label operators that (a) are Hermitian, (b) are polynomially independent of each other, and of G^{n}_{n}, (c) are polynomially independent of all the $C_n(k)$ and $C_{n-1}(k)$, (d) are scalars under the Sp($2n-2$)$\times$U(1) subgroup, (e) commute with all the $C_n(k)$ and (f) commute with each other. Here the statement "X is polynomially independent of all X_i, $i = 1,2,3,\cdots$" means that X cannot be expressed as a polynomial in the X_i.

I now show that the following components of the $n-1$ Sp($2n$) tensors from Eq. (2.14)

$$A(k)^{n}_{n}, \quad 1 \leqslant k \leqslant n-1, \tag{3.2}$$

satisfy all the above requirements. They are manifestly Hermitian [see Eq. (2.4)]. They are polynomially independent of each other and of $G^{n}_{n} = A(0)^{n}_{n}$ since they and G^{n}_{n} are particular components of linearly independent Sp($2n$) tensors. They are polynomially independent of all the $C_n(k)$ being components of Sp($2n$) tensors while the $C_n(k)$ are Sp($2n$) scalars, and they are polynomially (even functionally) independent of all the $C_{n-1}(k)$ since they contain generators with indices n and/or $\bar{n}$, which are absent in the $C_{n-1}(k)$. By Eq. (2.7) they commute with G^{n}_{n} and with all the G^{a}_{b}, $|a| \neq n$, $|b| \neq n$, hence are Sp($2n-2$)$\times$U(1) scalars. They commute with all the $C_n(k)$ since they are polynomials formed out of Sp($2n$) generators and the $C_n(k)$ are Sp($2n$) scalars. Lastly, as is shown in the Appendix.

$$[A(k)^{n}_{n}, A(k')^{n}_{n}] = 0, \quad 1 \leqslant k, k' \leqslant n-1. \tag{3.3}$$

Thus the $n-1$ operators, Eq. (3.2), or any $n-1$ linearly independent combinations of them, are satisfactory choices for the $n-1$ missing label operators.

Consider next the following components of the $n-1$ Sp($2n$) tensors from Eq. (2.15)

$$S(k)^n_n, \quad 1 \leqslant k \leqslant n-1. \tag{3.4}$$

The proof that these $n-1$ operators satisfy requirements (a) through (e) follows word for word the proof given for the $A(k)^n_n$ above. Requirement (f) is also satisfied because, as is shown in the Appendix

$$[S(k)^n_n, S(k')^n_n] = 0, \quad 1 \leqslant k,k' \leqslant n-1. \tag{3.5}$$

Thus I have *two* sets of missing label operators and it is worth noting that members of one set fail to commute with at least one member from the other set. Moreover, for the set involving symmetric tensors the links in the chain are [compare with Eq. (1.2)]

$$\mathrm{Sp}(2\nu) \downarrow \mathrm{Sp}(2\nu-2) \times \mathrm{Sp}_\nu(2), \quad 2 \leqslant \nu \leqslant n, \tag{3.6}$$

where the $\mathrm{Sp}_\nu(2)$ is the one generated by $G^\nu_\nu, G^\nu_{\bar\nu}, G^{\bar\nu}_{\bar\nu}$. To see this note the easily proved identity involving quadratic Casimir operators:

$$\nu C_\nu(1) = (\nu-1)\, C_{\nu-1}(1) - C_1(1) + 2S(1)^\nu_\nu, \tag{3.7}$$

where

$$C_1(1) = G^\nu_\nu(G^\nu_\nu + 2) + G^{\bar\nu}_\nu G^\nu_{\bar\nu}, \tag{3.8}$$

is the quadratic Casimir of $\mathrm{Sp}_\nu(2)$.

4. DISCUSSION

The approach of Judd, Miller, Patera, and Winternitz[6] and of Sharp[7] to the internal labeling problem consists essentially of two steps. In the first step all possible independent subgroup scalars, treated as c numbers, are determined. For this step they provide a complete algorithm. In the second step the noncommutativity of these operators is tackled and for this step no general algorithm is given. Of course, in the case of *one* missing label the second step is trivial.

My work in this paper may be viewed as complementary to theirs in the sense that I provide an algorithm for obtaining the appropriate number of *commuting* operators. The set of $2n-2$ operators

$$A(k)^n_n, \; S(k)^n_n, \quad 1 \leqslant k \leqslant n-1, \tag{4.1}$$

that I start from, even when augmented by all the Casimir operators, is *not* a complete set of all possible independent $\mathrm{Sp}(2n-2) \times \mathrm{U}(1)$ scalars; e.g., the operators

$$A(k)^n_{\bar n} A(k')^{\bar n}_n, \quad 0 \leqslant k,k' \leqslant n-1, \tag{4.2}$$

are again $\mathrm{Sp}(2n-2) \times \mathrm{U}(1)$ scalars (which can be easily formed into Hermitian combinations) and which, for at least some values of k, k', will be independent of the operators in Eq. (4.1) and the Casimir operators.

Thus, the two solutions to the internal labeling problem for $\mathrm{Sp}(2n) \downarrow \mathrm{Sp}(2n-2) \times \mathrm{U}(1)$ that I have given leave unanswered the question of what the most general solution might be. Because of their symmetry and simplicity the two solutions do have certain appeal and so, perhaps, deserve to be called canonical.

Another problem, left unsolved here, is the determination of the eigenvalue spectrum of these missing label operators.

For low values of n my results may be compared with the literature. For $n = 2$ there is only one missing label. If I choose the symmetric operator $S(1)^2_2$ the reduction is in fact $\mathrm{Sp}(4) \downarrow \mathrm{Sp}(2) \times \mathrm{Sp}(2)$ with no missing labels. Because of the isomorphisms between the corresponding algebras this reduction can be compared with $\mathrm{O}(5) \downarrow \mathrm{O}(4)$, which is well-known to be canonical. If I choose the antisymmetric operator $A(1)^2_2$ the reduction may be compared with $\mathrm{O}(5) \downarrow \mathrm{SU}(2) \times \mathrm{U}(1)$. This is one of the cases discussed by Sharp,[7] who states that the most general missing label operator is a linear combination of a certain cubic and quartic operator. His cubic may be related to $A(1)^2_2$; his quartic does not occur on my list, Eq. (4.1), but it can be related to those operators from the supplementary list, Eq. (4.2), for which $k + k' = 1$. Lastly for $n = 3$, the reduction $\mathrm{Sp}(6) \downarrow \mathrm{Sp}(4) \times \mathrm{Sp}(2)$ has in my formulation the symmetric $S(2)^3_3$ as the missing label operator. This can be compared with $\mathrm{Sp}(6) \downarrow \mathrm{Sp}(4) \times \mathrm{SU}(2)$ for which Sharp[7] gives a linear combination of a certain quartic and hexic operator. Again his quartic may be related to $S(2)^3_3$, his hexic to those operators from the supplementary list for which $k + k' = 2$.

APPENDIX

In this Appendix I prove Eqs. (3.3) and (3.5) of the text. Define the antisymmetric Sp($2n$) tensors

$$N(k)^a_b \equiv \tfrac{1}{2}\{(G^k)^a_b - \epsilon^a\epsilon^b(G^k)^{\bar b}_{\bar a}\}, \tag{A1}$$

the symmetric Sp($2n$) tensors

$$M(k)^a_b \equiv \tfrac{1}{2}\{(G^k)^a_b + \epsilon^a\epsilon^b(G^k)^{\bar b}_{\bar a}\}, \tag{A2}$$

and the Sp($2n$) invariants

$$D(k) \equiv (2n)^{-1} \sum_{a=\bar n}^{n} (G^k)^a_a. \tag{A3}$$

Making use of Eqs. (2.13), (2.7), and (2.3), I get

$$\begin{aligned}(G^k)^{\bar b}_{\bar a} &= \sum_{c=\bar n}^{n} G^{\bar b}_c (G^{k-1})^c_{\bar a} \\ &= \sum_{c=\bar n}^{n} \{[G^{\bar b}_c, (G^{k-1})^c_{\bar a}] + (G^{k-1})^c_{\bar a} G^{\bar b}_c\} \\ &= (2n+1)(G^{k-1})^{\bar b}_{\bar a} - \epsilon^a\epsilon^b(G^{k-1})^a_b - 2nD(k-1) \\ &\quad - \sum_{c=\bar n}^{n} \epsilon^b\epsilon^c(G^{k-1})^{\bar c}_{\bar a} G^c_b,\end{aligned} \tag{A4}$$

from which it follows by induction on k that

$$\epsilon^a\epsilon^b(G^k)^{\bar b}_{\bar a} = (-)^k(G^k)^a_b + \sum_{j=0}^{k-1} g_j(k)(G^j)^a_b, \tag{A5}$$

where the $g_j(k)$ are some functions of the Sp($2n$) invariants. Therefore,

$$N(k) = \sum_{j=0}^{k} n_j(k) G^j, \quad M(k) = \sum_{j=0}^{k} m_j(k) G^j, \tag{A6}$$

with $n_j(k)$, $m_j(k)$ some functions of the Sp($2n$) invariants. In particular

$$n_k(k) = \tfrac{1}{2}[1 - (-)^k], \quad m_k(k) = \tfrac{1}{2}[1 + (-)^k]. \tag{A7}$$

Hence given that the G^k, $1 \leqslant k \leqslant 2n-1$, are independent it follows that the $N(k)$ for even k, the $M(k)$ and $D(k)$ for odd k are not independent but can be related to the independent tensors and invariants given in the text, which in the notation of this Appendix become

$$A(k) = N(2k+1), \quad S(k) = M(2k),$$
$$C_n(k) = D(2k). \tag{A8}$$

Next I observe that if X and Y are any two tensors from the enveloping algebra of Sp($2n$), then

$$(XY)^a_b = (YX)^a_b, \tag{A9}$$

where the product is defined as in Eq. (2.13). This is obvious if X and Y are linear combinations of tensors of the type G^k since $G^k G^{k'} = G^{k+k'} = G^{k'} G^k$. The important point is that because of Eq. (A6) the Abelian nature of this product also holds if X and Y are tensors of given symmetry.

Now let A be an arbitrary antisymmetric Sp($2n$) tensor. Then

$$\begin{aligned}
2[N(k)^n_n, A^{\,n}_n] &= \sum_{r=1}^{k} \sum_{a=\bar{n}}^{n} \sum_{k=\bar{n}}^{n} \{(G^{k-r})^n_a [G^a_b, A^{\,n}_n](G^{r-1})^b_n \\
&\quad - (G^{k-r})^{\bar{n}}_a [G^a_b, A^{\,n}_n](G^{r-1})^b_{\bar{n}}\} \\
&= \sum_{r=1}^{k} \{(G^{k-r}A)^n_n (G^{r-1})^n_n - (G^{k-r})^n_n (AG^{r-1})^n_n \\
&\quad - (G^{k-r}A)^n_{\bar{n}} (G^{r-1})^{\bar{n}}_n + (G^{k-r})^n_{\bar{n}} (AG^{r-1})^{\bar{n}}_n \\
&\quad - (G^{k-r}A)^{\bar{n}}_n (G^{r-1})^n_{\bar{n}} + (G^{k-r})^{\bar{n}}_n (AG^{r-1})^n_{\bar{n}} \\
&\quad + (G^{k-r}A)^{\bar{n}}_{\bar{n}} (G^{r-1})^{\bar{n}}_{\bar{n}} - (G^{k-r})^{\bar{n}}_{\bar{n}} (AG^{r-1})^{\bar{n}}_{\bar{n}}\} \\
&= \sum_{r=1}^{k} \{[(G^{k-r}A)^n_n, (G^{r-1})^n_n] \\
&\quad + [(G^{k-r}A)^{\bar{n}}_{\bar{n}}, (G^{r-1})^{\bar{n}}_{\bar{n}}] \\
&\quad + [(G^{r-1})^n_n, (G^{k-r}A)^n_{\bar{n}}] \\
&\quad + [(G^{r-1})^n_{\bar{n}}, (G^{k-r}A)^{\bar{n}}_n]\},
\end{aligned} \tag{A10}$$

where the last step involves the use of Eq. (A9) and the observation that $k-r$ and $r-1$ may be interchanged whenever convenient. If I now denote by $\widetilde{A}$ the antisymmetric, by $\widetilde{S}$ the symmetric, part of the tensor $G^{k-r}A$, I arrive at the final result

$$\begin{aligned}
[N(k)^n_n, A^{\,n}_n] = \sum_{r=1}^{k} \{ &[\widetilde{A}^{\,n}_n, N(r-1)^{\,n}_n] + [\widetilde{S}^n_n, M(r-1)^{\,n}_n] \\
&+ \tfrac{1}{2}[N(r-1)^{\,n}_{\bar{n}}, \widetilde{A}^{\,\bar{n}}_n] \\
&+ \tfrac{1}{2}[N(r-1)^{\,\bar{n}}_n, \widetilde{A}^{\,n}_{\bar{n}}]\}.
\end{aligned} \tag{A11}$$

By the same procedure I get

$$\begin{aligned}
&[N(k)^n_{\bar{n}}, A^{\,\bar{n}}_n] + [N(k)^{\bar{n}}_n, A^{\,n}_{\bar{n}}] \\
&\quad = 4 \sum_{r=1}^{k} \{[\widetilde{S}^n_n, M(r-1)^{\,n}_n] + [N(r-1)^{\,n}_n, \widetilde{A}^{\,n}_n]\},
\end{aligned} \tag{A12}$$

and lastly,

$$\begin{aligned}
[M(k)^n_n, S^{\,n}_n] = \sum_{r=1}^{k} \{ &[\widehat{S}^{\,n}_n, M(r-1)^{\,n}_n] + [\widehat{A}^{\,n}_n, N(r-1)^{\,n}_n] \\
&+ \tfrac{1}{2}[\widehat{A}^{\,n}_{\bar{n}}, N(r-1)^{\,\bar{n}}_n] \\
&+ \tfrac{1}{2}[\widehat{A}^{\,\bar{n}}_n, N(r-1)^{\,n}_{\bar{n}}]\},
\end{aligned} \tag{A13}$$

where S is an arbitrary symmetric Sp($2n$) tensor and $\widehat{S}$ is the symmetric, $\widehat{A}$ the antisymmetric, part of the tensor $G^{k-r}S$.

Since $N(0) = 0$, $M(0) = I$, the sums over r in Eqs. (A11)–(A13) actually begin with $r = 2$ and therefore

$$[N(1)^n_n, A^{\,n}_n] = 0, \tag{A14}$$
$$[N(1)^n_{\bar{n}}, A^{\,\bar{n}}_n] + [N(1)^{\bar{n}}_n, A^{\,n}_{\bar{n}}] = 0, \tag{A15}$$
$$[M(1)^n_n, S^{\,n}_n] = 0. \tag{A16}$$

Combining Eqs. (A11)–(A16) I obtain by induction on k that for any $k \geqslant 0$

$$[N(k)^n_n, A^{\,n}_n] = 0, \tag{A17}$$
$$[N(k)^n_{\bar{n}}, A^{\,\bar{n}}_n] + [N(k)^{\bar{n}}_n, A^{\,n}_{\bar{n}}] = 0, \tag{A18}$$
$$[M(k)^n_n, S^{\,n}_n] = 0, \tag{A19}$$

where A is an arbitrary antisymmetric, S an arbitrary symmetric, Sp($2n$) tensor. Eq. (3.3) of the text follows from Eq. (A17) and Eq. (3.5) from Eq. (A19).

ACKNOWLEDGMENTS

This research was supported in part by the University of Wisconsin Research Committee with funds granted by the Wisconsin Alumni Research Foundation, and in part by the Department of Energy under contract EY-76-C-02-0881, number COO-881-108.

By acceptance of this article, the publisher and/or recipient acknowledges the U.S. Government's right to retain a nonexclusive, royalty-free license in and to any copyright covering this paper.

[1] I.M. Gel'fand and M.L. Zetlin, Dokl. Akad. Nauk. SSSR **71**, 825 (1950); **71**, 1017 (1950).
[2] A.M. Bincer, J. Math. Phys. **18**, 1870 (1977); **19**, 1173 (1978); **19**, 1179 (1978).
[3] G.E. Baird and L.C. Biedenharn, J. Math. Phys. **4**, 1449 (1963).
[4] D.P. Zhelobenko, Usp. Mat. Nauk. **17**, 27 (1962) [Russ. Math. Surveys **17**, 1 (1962)]; G.C. Hegerfeldt, J. Math. Phys. **8**, 1195 (1967).
[5] J. Mickelsson, Rep. on Math. Phys. **3**, 193 (1972).
[6] B.R. Judd, W. Miller, Jr., J. Patera, and P. Winternitz, J. Math. Phys. **15**, 1787 (1974), and references cited therein.
[7] R.T. Sharp, J. Math. Phys. **16**, 2050 (1975).
[8] M. Moshinsky and J.G. Nagel, Phys. Lett. **5**, 173 (1963).
[9] G. Racah, Springer Tracts Mod. Phys. **37**, 28 (1965), Eq. (76). The relation between Racah's and my notation is $X_{ik} = \epsilon^i G^i_k$.
[10] H.S. Green, J. Math. Phys. **12**, 2106 (1971).
[11] C.O. Nwachuku and M.A. Rashid, J. Phys. A (GB) **11**, 457 (1978).

On the invariant scalar products and the UIR of SO (n,1)

Takayoshi Maekawa
Department of Physics, Kumamoto University, Kumamoto, Japan 860

(Received 14 August 1979; accepted for publication 30 November 1979)

A quantity, which is invariant under the transformation of $h\in$SO (n,1), is found in a form of the series expansion in terms of the D matrix elements of SO (n) and an invariant scalar product, which is needed for the complementary series of the unitary irreducible representation of SO (n,1), is introduced in the Hermitian form by using the invariant quantity. The action of the intertwining operator defined through the Hermitian scalar product to the bases of H_c, on which the complementary series of the UIR of SO (n,1) are realized, is found.

The linear representations of the group SO(n,1) are studied by many authors.[1,2] The principal series of the unitary irreducible representation (UIR) are completed in the sense that the invariant scalar product together with their classifications are known explicitly and the representation matrix elements are given with the computation formula. Even for the nonunitary representation, the computation formula for the representation matrix elements is known.[2] For the complementary series of the UIR, the intertwining operator is used through the existence of the Hermitian scalar product in Refs. 1 but in Refs. 2, the computation formula of the representation matrix elements is determined provided that the Hermitian scalar product exists. The purpose of this article is to give an invariant quantity under the transformation of $h\in$SO(n,1) and then to introduce the invariant scalar product in the Hermitian form by using the invariant quantity. As a by-product of our results, the action of the intertwining operator to the bases of the space H_c, on which the complementary series of the UIR of SO(n,1) are realized, is found and it follows that our procedure is equivalent to those in Refs. 1.

We first summarize briefly the results needed for our discussion.[2] In what follows, the same notations as in Ref. 2. will be used unless otherwise stated. Let H be a linear space of functions on SO(n) such that

$$\int_{\mathrm{SO}(n)} |\Phi(g^{(n)})|^2\, dV_n < \infty\,,$$

holds for $\Phi\in H$. Then any $\Phi(g^{(n)})\in H$ can be expanded in terms of the D matrix elements of SO(n) as follows[3]:

$$\Phi(g^{(n)}) = \sum a^{(\lambda_n)}_{\{\lambda'_n\}\{\lambda_{n-1}\}} \sqrt{\frac{N(\lambda_n)}{V_n}}\, D^{(\lambda_n)}_{\{\lambda'_n\}\{\lambda_{n-1}\}}(g^{(n)})\,, \tag{1}$$

where the D matrix elements of SO(n) satisfy the orthonormal relation

$$\frac{\sqrt{N(\lambda'_n)(\lambda_n)}}{V_n} \int_{\mathrm{SO}(n)} dV_n\, \overline{D^{(\lambda_n)}_{\{\lambda''_n\}\{\lambda_{n-1}\}}(g^{(n)})} \times D^{(\lambda'_n)}_{\{\lambda'''_n\}\{\lambda'_{n-1}\}}(g^{(n)}) = \delta_{\{\lambda_n\}\{\lambda'_n\}}\delta_{\{\lambda''_n\}\{\lambda'''_n\}}\,. \tag{2}$$

The coefficients $a^{(\lambda)}$ are given by

$$a^{(\lambda_n)}_{\{\lambda'_n\}\{\lambda_{n-1}\}} = \int_{\mathrm{SO}(n)} dV_n \sqrt{\frac{N(\lambda_n)}{V_n}} \times \overline{D^{(\lambda_n)}_{\{\lambda'_n\}\{\lambda_{n-1}\}}(g^{(n)})}\Phi(g^{(n)}) \tag{3}$$

and satisfy the Parseval relation

$$\int_{\mathrm{SO}(n)} |\Phi(g^{(n)})|^2\, dV_n = \sum \left|a^{(\lambda_n)}_{\{\lambda'_{n-1}\}\{\lambda_{n-1}\}}\right|^2. \tag{4}$$

We can consider the representation of SO(n,1) in the subspace of H. The action of the representation operator corresponding to $h\in$SO(n,1) on the function $\Phi\in H$ is defined through the multiplier representation as follows[2]:

$$R(h)\Phi(g^{(n)}) = (h_{n+1n+1} + (g^{(n)}h)_{nn+1})^{\rho_{n+1}}\Phi(\bar{g}_{(n)})\,, \tag{5}$$

where ρ_{n+1} is a complex number and the quantity $\bar{g}^{(n)}$ transformed by h on the right-hand side is given by

$$\bar{g}^{(n)}_{nj} = \frac{h_{n+1j} + (g^{(n)}h)_{nj}}{h_{n+1n+1} + (g^{(n)}h)_{nn+1}} \quad (j = 1,2,\ldots,n)\,, \tag{6}$$

$$\bar{g}^{(n)}_{ij} = \frac{1}{h_{n+1n+1} + (g^{(n)}h)_{nn+1}}\left[(h_{n+1n+1} + (g^{(n)})_{nn+1}) \times (g^{(n)}h)_{ij} - (g^{(n)}h)_{in+1}(h_{n+1j} + (g^{(n)}h)_{nj})\right] \quad (i = 1,2,\ldots,n-1; j = 1,2,\ldots,n)\,, \tag{7}$$

with $(g^{(n)})_{n+1j} = (g^{(n)})_{jn+1} = \delta_{jn+1}$. The replacements $g^{(n)}\rightleftarrows\bar{g}^{(n)}$ and $h\to h^{-1}$ in Eqs. (6) and (7) lead to the expressions of $g^{(n)}$ in terms of $\bar{g}^{(n)}$. The invariant measure (dV_n) of SO(n) transforms under $h\in$SO(n,1) as follows:

$$d\bar{V}_n = \frac{dV_n}{(h_{n+1n+1} + (g^{(n)}h)_{nn+1})^{n-1}}\,. \tag{8}$$

It follows from Eqs. (6) and (7) that the following relation holds:

$$1 - (\bar{g}^{(n)'}\bar{g}^{(n)-1})_{nn} = \frac{1 - (g^{(n)'}g^{(n)-1})_{nn}}{\{h_{n+1n+1} + (g^{(n)'}h)_{nn+1}\}\{h_{n+1n+1} + (g^{(n)}h)_{nn+1}\}}\,, \tag{9}$$

where $\bar{g}^{(n)'}$ is the transformed quantity of $g^{(n)'}$ given by Eqs. (6) and (7).

It is straightforward to confirm the representation condition, i.e.,

$$R(h')R(h)\Phi(g^{(n)}) = R(h'h)\Phi(g^{(n)})\,. \tag{10}$$

The infinitesimal operators of the representations are easily obtained from Eq. (5) and their explicit expressions are given

 0022-2488/80/040675-05$1.00

in Ref. 2 with the Euler angle parametrization. An important operator is one corresponding to the boost in the nth direction and given as follows:

$$J_{nn+1} = i \sin^{\rho_{n+1}}\theta \sin\theta \frac{\partial}{\partial\theta} \sin^{-\rho_{n+1}}\theta$$
$$= i(\sin\theta \frac{\partial}{\partial\theta} - \rho_{n+1}\cos\theta), \tag{11}$$

where $(g^{(n)})_{nn} = \cos\theta\ (0 \leqslant \theta \leqslant \pi)$.

The irreducible representations of SO(n,1) are characterized by the numbers Λ_{n-1}, which are determined by the irreducible representation of the centralizer [SO($n-1$)] of the Abelian subgroup (boost) in the maximal compact subgroup [SO(n)] in the Iwasawa decomposition of SO(n,1),[1,2] and ρ_{n+1}, which is related to a linear form on the Abelian subgroup in the decomposition. Explicitly, the representations can be treated with the operators of the first and second parameter groups of SO(j) ($j = 2,3,...,n$) and those of the multiplier representation such as (11).[2] In order to discuss the irreducible representation of SO(n,1) in H, it is sufficient for us to consider the subspace H_1 of H because the infinitesimal operators of the first and second parameter groups commute with each other. Thus, we may take the bases of H_1 as follows[2]:

$$\Phi^{(\rho_{n+1},\Lambda_{n-1})}_{\{\lambda_n\}}(g^{(n)}) = N(\rho_{n+1};\lambda_n)\sqrt{\frac{N(\lambda_n)}{V_n}}\, D^{(\lambda_n)}_{\{\Lambda_{n-1}\}\{\lambda_{n-1}\}}(g^{(n)}), \tag{12}$$

where

$$\frac{N(\rho_{n+1};\lambda_n)}{N(\rho_{n+1};\lambda'_n)} = \left[\prod_{j=1}^{[n/2]} \frac{\Gamma(m_{nj} - \rho_{n+1} - j + 1)\Gamma(m'_{nj} + \rho_{n+1} - j + n)}{\Gamma(m_{nj} + \rho_{n+1} - j + n)\Gamma(m_{nj} - \rho_{n+1} - j + 1)}\right]^{1/2} \tag{13}$$

As noted in Ref. 2, we may fix $N(\rho_{n+1};\lambda_n)$ without a factor.

The action of J_{nn+1} on the bases (12) is known to give the matrix elements,[2] i.e.,

$$J_{nn+1}\Phi^{(\rho_{n+1},\Lambda_{n-1})}_{\{\lambda_n\}}(g^{(n)}) = \sum_{j=1}^{n/2} A(m_{nj})\Phi^{(\rho_{n+1},\Lambda_{n-1})}_{(\{\lambda_{n-1}\}\{\lambda^+_{nj}\})}(g^{(n)}) - \sum_{j=1}^{n/2} A(m_{nj}-1)\,\Phi^{(\rho_{n+1},\Lambda_{n-1})}_{(\{\lambda_{n-1}\}\{\lambda^-_{nj}\})}(g^{(n)}) \tag{14}$$

for $n+1$ odd and

$$J_{nn+1}\Phi^{(\rho_{n+1},\Lambda_{n-1})}_{\{\lambda_n\}}(g^{(n)}) = \sum_{j=1}^{(n-1)/2} B(m_{nj})\Phi^{(\rho_{n+1},\Lambda_{n-1})}_{(\{\lambda_{n-1}\}\{\lambda^+_{nj}\})}(g^{(n)}) - \sum_{j=1}^{(n-1)/2} B(m_{nj}-1)\,\Phi^{(\rho_{n+1},\Lambda_{n-1})}_{(\{\lambda_{n-1}\}\{\lambda^-_{nj}\})}(g^{(n)})$$
$$+ C_{n+1}\Phi^{(\rho_{n+1},\Lambda_{n-1})}_{\{\lambda_n\}}(g^{(n)}), \tag{15}$$

for $n+1$ even. The explicit expressions for the matrix elements A, B, and C are omitted here but it should be noted that A and B contain $\rho_{n+1} + (n-1)/2$ as squared in the form $[\rho_{n+1} + (n-1)/2]^2$ and C as linear in the form $\rho_{n+1} + (n-1)/2$. The classification of the irreducible representations of SO(n,1) is made on the basis of the expressions (14) and (15) for the matrix elements.[4] It is noted that for the complementary series of the UIR with $n+1$ even the number $m_{n+1(n+1)/2}$, which is contained in Λ_{n-1}, is at least zero, i.e., $C_{n+1} = 0$.

The representation matrix elements (D) of the operator $R(h)$ [$h \in$ SO(n,1)] may be defined by[2]

$$R(h)\Phi^{(\rho_{n+1},\Lambda_{n-1})}_{\{\lambda_n\}}(g^{(n)}) = \sum_{\{\lambda'_n\}} D^{(\rho_{n+1},\Lambda_{n-1})}_{\{\lambda'_n\}\{\lambda_n\}}(h)\Phi^{(\rho_{n+1},\Lambda_{n-1})}_{\{\lambda'_n\}}(g^{(n)}). \tag{16}$$

Equation (16) gives due to Eqs. (2) and (5)

$$D^{(\rho_{n+1},\Lambda_{n-1})}_{\{\lambda'_n\}\{\lambda_n\}}(h)$$
$$= \frac{N(\rho_{n+1};\lambda_n)}{N(\rho_{n+1};\lambda'_n)} \frac{\sqrt{N(\lambda'_n)N(\lambda_n)}}{V_n} \int_{SO(n)} dV_n\, \overline{D^{(\lambda'_n)}_{\{\Lambda_{n-1}\}\{\lambda'_{n-1}\}}(g^{(n)})}\,(h_{n+1n+1} + (g^{(n)}h)_{nn+1})^{\rho_{n+1}} D^{(\lambda_n)}_{\{\Lambda_{n-1}\}\{\lambda_{n-1}\}}(\bar{g}^{(n)}). \tag{17}$$

Equation (17) may be regarded as the computation formula for the representation D matrix elements of SO(n,1).[2] It is noted that Eq. (17) is obtained independently of the existence of the scalar product in H_1.

For $h = g \in$ SO(n), Eq. (17) gives the desired D matrix elements of SO(n), i.e.,

$$D^{(\rho_{n+1},\Lambda_{n-1})}_{\{\lambda'_n\}\{\lambda_n\}}(g) = \delta_{\lambda'_n\lambda_n} D^{(\lambda_n)}_{\{\lambda'_{n-1}\}\{\lambda_{n-1}\}}(g). \tag{18}$$

The representation condition (10) becomes

$$D^{(\rho_{n+1},\Lambda_{n-1})}_{\{\lambda'_n\}\{\lambda_n\}}(h'h) = \sum_{\{\lambda''_n\}} D^{(\rho_{n+1},\Lambda_{n-1})}_{\{\lambda'_n\}\{\lambda''_n\}}(h')D^{(\rho_{n+1},\Lambda_{n-1})}_{\{\lambda''_n\}\{\lambda_n\}}(h). \tag{19}$$

Taking the conjugate complex of Eq. (17) and changing the integration variables $g^{(n)}$ into $\bar{g}^{(n)}$, we get the relation

$$\overline{D^{(\rho_{n+1},\Lambda_{n-1})}_{\{\lambda'_n\}\{\lambda_n\}}(h)} = \overline{\left(\frac{N(\rho_{n+1};\lambda_n)}{N(\rho_{n+1};\lambda'_n)}\right)} \frac{N(1-n-\overline{\rho}_{n+1};\lambda_n)}{N(1-n-\overline{\rho}_{n+1};\lambda'_n)} D^{(1-n-\overline{\rho}_{n+1};\Lambda_{n-1})}_{\{\lambda_n\}\{\lambda'_n\}}(h^{-1}), \tag{20}$$

which means that in general the representation is not unitary. Therefore, we examine the matrix elements and determine the scalar product under which the irreducible representation becomes unitary.

(i) For $\rho_{n+1} = (1-n)/2 + i\nu_{n+1}$, ν_{n+1} real: The relation (20) becomes due to Eq. (13)

$$\overline{D^{(\rho_{n+1},\Lambda_{n-1})}_{\{\lambda'_n\}\{\lambda_n\}}(h)} = D^{(\rho_{n+1},\Lambda_{n-1})}_{\{\lambda_n\}\{\lambda'_n\}}(h^{-1}) . \tag{21}$$

It is, therefore, expected that the invariant scalar product exists. We consider the subspace (H_p) of H_1 consisting of the bases (12) with $\rho_{n+1} = (1-n)/2 + i\nu_{n+1}$ and define the scalar product of $\Phi_1, \Phi_2 \in H_p$ as follows:

$$\langle \Phi_1, \Phi_2 \rangle = \int_{SO(n)} dV_n \, \overline{\Phi_1(g^{(n)})} \, \Phi_2(g^{(n)}) . \tag{22}$$

Then, the space H_p becomes the Hilbert space with the scalar product (22). It is straightforward to show the invariance of Eq. (22) under the transformation of $h \in SO(n,1)$, to derive the expression (17) for the D matrix elements and to confirm Eq. (21). Detailed discussions will be seen in Refs. 2.

(ii) For $\rho_{n+1} = (1-n)/2 + \sigma_{n+1}$, σ_{n+1} real: The relation (20) becomes due to Eq. (13)

$$\overline{D^{(\rho_{n+1},\Lambda_{n-1})}_{\{\lambda'_n\}\{\lambda_n\}}(h)} = D^{((1-n)/2-\sigma_{n+1},\Lambda_{n-1})}_{\{\lambda_n\}\{\lambda'_n\}}(h^{-1}) . \tag{23}$$

It is, therefore, seen that the existence of the unitary representation is intimately connected with the dependence of the D matrix elements on σ_{n+1}. In order to examine the dependence, we rewrite $h \in SO(n,1)$ as follows:

$$h = g_1 h^0 g_2 , \tag{24}$$

where $g_1, g_2 \in SO(n)$ and $h^0 \in SO(n,1)$ is the boost to the nth direction with a parametrization, e.g., $h^0_{ij} = \delta_{ij}$ (i or $j < n$), $h^0_{nn} = h^0_{n+1\,n+1} = \cosh\zeta$, $h^0_{nn+1} = h^0_{n+1\,n} = -\sinh\zeta$. Then we obtain from Eqs. (18) and (19)

$$D^{(\rho_{n+1},\Lambda_{n-1})}_{\{\lambda'_n\}\{\lambda_n\}}(h) = \sum_{\{\lambda''_{n-1}\}} D^{(\lambda'_n)}_{\{\lambda'_{n-1}\}\{\lambda''_{n-1}\}}(g_1) \, {}^b d^{(\rho_{n+1},\Lambda_{n-1})}_{\lambda'_n(\lambda''_{n-1})\lambda_n}(h^0) D^{(\lambda_n)}_{\{\lambda''_{n-1}\}\{\lambda_{n-1}\}}(g_2) . \tag{25}$$

Because the D matrix elements of $SO(n)$ satisfy the unitarity condition, it is sufficient for us to examine the dependence of the ${}^b d$ (boost) matrix elements on σ_{n+1}. The explicit expression for the ${}^b d$ is given by

$$\begin{aligned} &{}^b d^{(\rho_{n+1},\Lambda_{n-1})}_{\lambda'_n(\lambda_{n-1})\lambda_n}(\zeta) \\ &= \frac{N(\rho_{n+1};\lambda_n)}{N(\rho_{n+1};\lambda'_n)} \frac{\sqrt{N(\lambda'_n)N(\lambda_n)}}{V_n} \int_{SO(n)} dV_n \, D^{(\lambda'_n)}_{\{\Lambda_{n-1}\}\{\lambda_{n-1}\}}(g^{(n)}) (\cosh\zeta - \cos\theta_{n1}\sinh\zeta)^{\rho_{n+1}} D^{(\lambda_n)}_{\{\Lambda_{n-1}\}\{\lambda_{n-1}\}}(\bar{g}^{(n)}) . \end{aligned} \tag{26}$$

The quantity which changes under h^0 is only θ_{n1} given by

$$g^{(n)}_{nn} = \cos\theta_{n1}, \quad \cos\bar{\theta}_{n1} = \frac{\cos\theta_{n1}\cosh\zeta - \sinh\zeta}{\cosh\zeta - \cos\theta_{n1}\sinh\zeta} . \tag{27}$$

It is easily seen that the following relation holds for any nonnegative integers k:

$$\begin{aligned} &\frac{\partial^k}{\partial\zeta^k}(\cosh\zeta - \cos\theta_{n1}\sinh\zeta)^{\rho_{n+1}} D^{(\lambda_n)}_{\{\Lambda_{n-1}\}\{\lambda_{n-1}\}}(\bar{g}^{(n)}) \Big|_{\zeta=0} \\ &= \sin^{\rho_{n+1}}\theta_{n1} \left(\sin\theta_{n1}\frac{\partial}{\partial\theta_{n1}}\right)^k \sin^{-\rho_{n+1}}\theta_{n1} D^{(\lambda_n)}_{\{\Lambda_{n-1}\}\{\lambda_{n-1}\}}(g^{(n)}) = (iJ_{n+1\,n})^k D^{(\lambda_n)}_{\{\Lambda_{n-1}\}\{\lambda_{n-1}\}}(g^{(n)}) , \end{aligned} \tag{28}$$

because

$$\frac{\partial\bar{\theta}_{n1}}{\partial\zeta} = \sin\bar{\theta}_{n1} .$$

As is expected, Eq. (26) is equivalent to the exponentiation of the matrix elements of $J_{n+1\,n}$ and thus it follows from Eqs. (14) and (15) together with the remarks given below Eq. (15) that the ${}^b d$ matrix elements (26) become even functions of σ_{n+1} in general for $n+1$ odd but $m_{n+1(n+1)/2} = 0$ ($C_{n+1} = 0$) for $n+1$ even. In these cases, the relation (23) becomes

$$\overline{D^{(\rho_{n+1},\Lambda_{n-1})}_{\{\lambda'_n\}\{\lambda_n\}}(h)} = D^{(\rho_{n+1},\Lambda_{n-1})}_{\{\lambda_n\}\{\lambda'_n\}}(h^{-1}) . \tag{29}$$

In what follows, we restrict our discussion to the case when Eq. (29) holds, i.e., with no restriction for $n+1$ odd but with $m_{n+1(n+1)/2} = 0$ for $n+1$ even. These cases correspond to the complementary series of the UIR of $SO(n,1)$ and σ_{n+1} may be fixed to the negative value.[2]

We consider the subspace H_c of H_1 consisting of the bases (12) with $\rho_{n+1} = (1-n)/2 + \sigma_{n+1}$ and $m_{n+1(n+1)/2} = 0$ for $n+1$ even. Before introducing the invariant scalar product in H_c, we derive an invariant quantity from Eq. (29). The following relation is obvious from Eqs. (19) and (29):

$$\delta_{\{\lambda'_n\}\{\lambda_n\}} = \sum_{\{\lambda''_n\}} \overline{D^{(\rho_{n+1},\Lambda_{n-1})}_{\{\lambda''_n\}\{\lambda'_n\}}(h)} \, D^{(\rho_{n+1},\Lambda_{n-1})}_{\{\lambda''_n\}\{\lambda_n\}}(h) . \tag{30}$$

Substitution of the expression from Eq. (17) into Eq. (30) gives

$$\delta_{\{\lambda_n''\}\{\lambda_n\}} = \int_{SO(n)}\int_{SO(n)} dV_n' dV_n \left[\sum_{\lambda_n''} \frac{N(\lambda_n'')}{V_n} \left|N(\rho_{n+1};\lambda_n'')\right|^{-2} D^{(\lambda_n'')}_{\{\Lambda_{n-1}\}\{\Lambda_{n-2}\}}(g^{(n)'}g^{(n)-1})\right]$$
$$\times \{h_{n+1n+1} + (g^{(n)'}h)_{nn+1}\}^{\rho_{n+1}} \{h_{n+1n+1} + (g^{(n)}h)_{nn+1}\}^{\rho_{n+1}} \overline{\Phi^{(\rho_{n+1},\Lambda_{n-1})}_{\{\lambda_n'\}}(g^{(n)'})}\, \Phi^{(\rho_{n+1},\Lambda_{n-1})}_{\{\lambda_n\}}(g^{(n)}) . \quad (31)$$

Equation (31) holds for all h and we obtain for h = unit element

$$\delta_{\{\lambda_n'\}\{\lambda_n\}} = \int_{SO(n)}\int_{SO(n)} dV_n' dV_n \left[\sum_{\lambda_n''} \frac{N(\lambda_n'')}{V_n} \left|N(\rho_{n+1};\lambda_n'')\right|^{-2} D^{(\lambda_n'')}_{\{\Lambda_{n-1}\}\{\Lambda_{n-2}\}}(g^{(n)'}g^{(n)-1})\right] \overline{\Phi^{(\rho_{n+1},\Lambda_{n-1})}_{\{\lambda_n'\}}(g^{(n)'})}\, \Phi^{(\rho_{n+1},\Lambda_{n-1})}_{\{\lambda_n\}}(g^{(n)}) , \quad (32)$$

whose validity is also confirmed directly. The right-hand sides of Eqs. (31) and (32) are equal to each other for all bases in the space H_c and thus we deduce on making use of Eqs. (8) and (9)

$$[1-(g^{(n)'}g^{(n)-1})_{nn}]^{[(n-1)/2]+\sigma_{n+1}} \sum_{\lambda_n} \frac{N(\lambda_n)}{V_n} \left|N(\rho_{n+1},\lambda_n)\right|^{-2} D^{(\lambda_n)}_{\{\Lambda_{n-1}\}\{\Lambda_{n-2}\}}(g^{(n)'}g^{(n)-1})$$
$$= [1-(\bar{g}^{(n)'}\bar{g}^{(n)-1})_{nn}]^{[(n-1)/2]+\sigma_{n+1}} \sum_{\lambda_n} \frac{N(\lambda_n)}{V_n} \left|N(\rho_{n+1};\lambda_n)\right|^{-2} D^{(\lambda_n)}_{\{\Lambda_{n-1}\}\{\Lambda_{n-2}\}}(\bar{g}^{(n)'}\bar{g}^{(n)-1})$$
$$= K(g^{(n)'}g^{(n)-1}) = inv . \quad (33)$$

In this way, we have obtained the invariant quantity (33) under the transformation of $h\in SO(n,1)$.

The invariant quantity K is expressed in terms of a definite function for the special case of $\Lambda_{n-1}=0$, i.e., then K becomes

$$K(g^{(n)'}g^{(n)-1}) = \frac{\sqrt{\pi}\, 2^{\sigma-[(n-3)/2]}}{V_n} \frac{\Gamma(\sigma+\frac{n-1}{2})}{\Gamma(n/2)\Gamma(-\sigma)} . \quad (34)$$

where the value of σ is restricted to $0<-\sigma<(n-1)/2$ according to the classification of the complementary series.[2,4] It follows from Eq. (34) that K is positive. In the general cases, it is important to examine whether the series (33) converge or not. If we multiply the series of Eq. (33) by $\overline{\Phi(g^{(n)'})}$ and $\Phi(g^{(n)})(\Phi\in H_c)$ and integrate over dV_n' and dV_n, we obtain

$$\sum_{\{\lambda_n\}} \left|N(\rho_{n+1};\lambda_n)\right|^{-2} \left|a^{(\lambda_n)}_{\{\Lambda_{n-1}\}\{\lambda_{n-1}\}}\right|^2$$
$$< \sum_{\{\lambda_n\}} \left|a^{(\lambda_n)}_{\{\Lambda_{n-1}\}\{\lambda_{n-1}\}}\right|^2 ,$$

where $a^{(\lambda)}$'s are defined by Eq. (3) and the value of σ_{n+1} is negative for the complementary series. Thus, we assume that the series in Eq. (33) converge in each classes of the complementary series of the UIR of $SO(n,1)$ and give a definite K. Then it is easy to find the invariant scalar product in the Hermitian form.

For $\Phi_1,\Phi_2\in H_c$, we define the scalar product as follows:

$$\langle\Phi_1,\Phi_2\rangle_c = \int_{SO(n)}\int_{SO(n)} dV_n' dV_n K(g^{(n)'}g^{(n)-1})$$
$$\times [1-(g^{(n)'}g^{(n)-1})_{nn}]^{[(1-n)/2]-\sigma_{n+1}} \overline{\Phi_1(g^{(n)'})}\,\Phi_2(g^{(n)}) . \quad (35)$$

The scalar product is the same as in Ref. 2. It is easily seen from Eqs. (8) and (9) and the invariance of K that the positive definite scalar product (35) is invariant under the transformation of $h\in SO(n,1)$, i.e.,

$$\langle R(h)\Phi_1,R(h)\Phi_2\rangle_c = \langle\Phi_1,\Phi_2\rangle_c . \quad (36)$$

Thus, the H_c becomes the Hilbert space with the scalar product (35).

In this way, the complementary series of the UIR of $SO(n,1)$ can be treated in H_c and their representation matrix elements are easily found.[2] It is noted that the invariant K is not a constant in general because Eq. (35) becomes zero for all bases except for $\Lambda_{n-1}=0$ in the case of K = const. The scalar product (35) for the constant K is given in Ref. 5.

The intertwining operator (T) may be defined for Φ_1, $\Phi_2\in H_c$ by

$$\langle\Phi_1,\Phi_2\rangle_c = \langle\Phi_1,T\Phi_2\rangle , \quad (37)$$

where the scalar product on the right-hand side is given by (22). Thus it is sufficient for us to find the action of T on the bases of H_c. Taking the bases for Φ_1 and Φ_2 and making use of Eqs. (35), (33), and (2), we easily get the relation

$$\left\langle \Phi^{(\rho_{n+1},\Lambda_{n-1})}_{\{\lambda_n'\}}, \Phi^{(\rho_{n+1},\Lambda_{n-1})}_{\{\lambda_n\}} \right\rangle_c$$
$$= \left|N(\rho_{n+1};\lambda_n)\right|^{-2} \left\langle \Phi^{(\rho_{n+1},\Lambda_{n-1})}_{\{\lambda_n'\}}, \Phi^{(\rho_{n+1},\Lambda_{n-1})}_{\{\lambda_n\}} \right\rangle . \quad (38)$$

Therefore, the action of T on the bases is given as follows:

$$T\Phi^{(\rho_{n+1},\Lambda_{n-1})}_{\{\lambda_n\}}(g^{(n)}) = \left|N(\rho_{n+1};\lambda_n)\right|^{-2} \Phi^{(\rho_{n+1},\Lambda_{n-1})}_{\{\lambda_n\}}(g^{(n)}) . \quad (39)$$

Thus, it follows that the bases of H_c are the eigenvectors of T with the eigenvalue $\left|N(\rho_{n+1};\lambda_n)\right|^{-2}$. It follows from (37) that the following relations hold:

$$\langle R(h)\Phi_1,R(h)\Phi_2\rangle_c = \langle R(h)\Phi_1,TR(h)\Phi_2\rangle$$
$$= \langle\Phi_1,T\Phi_2\rangle , \quad (40)$$

due to Eq. (36). It is obvious that the following relations hold

$$\langle\Phi_1,T\Phi_2\rangle = \langle T\Phi_1,\Phi_2\rangle, \langle\Phi,T\Phi\rangle \leqslant \langle\Phi,\Phi\rangle ,$$

because we may take $\left|N(\rho_{n+1};\lambda_n)\right|^{-2}\leqslant 1$ for the complementary series.

Conversely, if we define the scalar product by the right-hand side of (37) together with the action (39) of T on the bases of the space H_c, it is easily seen that the condition of the invariant scalar product given by the right-hand sides of (40) is equivalent to the following

$$D^{(\rho_{n+1},\Lambda_{n-1})}_{\{\lambda'_n\}\{\lambda_n\}}(h) = D^{(1-n-\rho_{n+1},\Lambda_{n-1})}_{\{\lambda'_n\}\{\lambda_n\}}(h)\,, \tag{41}$$

where the matrix elements $D(h)$ are defined by

$$D^{(\rho_{n+1},\Lambda_{n-1})}_{\{\lambda'_n\}\{\lambda_n\}}(h) = \left\langle \Phi^{(\rho_{n+1},\Lambda_{n-1})}_{\{\lambda'_n\}}, TR(h)\Phi^{(\rho_{n+1},\Lambda_{n-1})}_{\{\lambda_n\}} \right\rangle,$$

and satisfy the relation

$$\overline{D^{(\rho_{n+1},\Lambda_{n-1})}_{\{\lambda'_n\}\{\lambda_n\}}(h)} = D^{(1-n-\rho_{n+1},\Lambda_{n-1})}_{\{\lambda_n\}\{\lambda'_n\}}(h^{-1})\,.$$

The definition of the matrix elements $D(h)$ gives the same expression as (17). It follows that the condition (41) is the same as that given by (23) and (29) and the representation becomes unitary.

[1] A.W. Knapp and E.M. Stein, Ann. Math **93**, 489 (1971); E. Thieleker, Trans. Am. Math. Soc. **199**, 327 (1974); G. Schiffmann, Bull. Soc. Math. Fr. **99**, 3 (1971); A.M. Gavrilik and A.U. Klimyk, Preprint of Academy of Sciences of the Ukrainian SSR, Kiev 1975.

[2] T. Maekawa, J. Math. Phys. **19**, 2028 (1978); **20**, 691 (1979); **20**, 1460 (1979).

[3] N. Ja. Vilenkin, *Special Functions and the Theory of Group Representations*, Mathematical Monographs Vol. 22, translated by V.N. Singh (Am. Math. Soc. Providence, 1968).

[4] For example, F. Schwartz, J. Math. Phys. **12**, 131 (1971).

[5] R. Takahashi, Bull. Soc. Math. Fr. **91**, 289 (1963).

Canonical transforms. IV. Hyperbolic transforms: Continuous series of SL(2,*R*) representations

Kurt Bernardo Wolf[a)]
Centre de Recherches Mathématiques, Université de Montréal, Montréal, Québec, Canada

(Received 5 January 1979; accepted for publication 23 April 1979)

We consider the sl(2,*R*) Lie algebra of second-order differential operators given by the Schrödinger Hamiltonians of the harmonic, repulsive, and free particle, all with a strong centripedal core placing them in the C^{∞}_{q} continuous series of representations. The corresponding SL(2,*R*) Lie group is shown to be a group of integral transforms acting on a (two-component) space of square-integrable functions, with an integral (matrix) kernel involving Hankel and Macdonald functions. The subgroup bases for irreducible representations consist of Whittaker, power, Hankel, and Macdonald functions. We construct the operator which intertwines this realization of SL(2,*R*) with the more familiar Bargmann realization on functions on the unit circle. This operator implements the canonical transformation of the above Schrödinger systems to action and angle variables.

1. INTRODUCTION

The program to explore the role of canonical transformations in quantum mechanics followed by Moshinsky and collaborators[1,2] has lead to advances and applications in three related fields: (a) It has given a better understanding of the dynamical *groups* (as opposed to dynamical or similarity algebras) for quantum systems and partial differential equations,[3,4] (b) it has brought a significant unification into the theory of integral transforms,[5-7] and (c) it has complemented the study of the three-dimensional Lorentz group generated by algebras of second-order differential operators.[8-10] In this article, the fourth of a series,[5,6,11] we would like to explore the following territory: Consider the three operators

$$\mathbb{J}_1 = \frac{1}{4}\left(-\frac{d^2}{d\rho^2} - \frac{\mu}{\rho^2} - \rho^2\right), \tag{1.1a}$$

$$\mathbb{J}_2 = -\frac{i}{2}\left(\rho\frac{d}{d\rho} + \frac{1}{2}\right), \tag{1.1b}$$

$$\mathbb{J}_3 = \frac{1}{4}\left(-\frac{d^2}{d\rho^2} - \frac{\mu}{\rho^2} + \rho^2\right), \quad \mu > \frac{1}{4}, \tag{1.1c}$$

which form an sl(2,*R*)$\simeq$sp(2,*R*)$\simeq$so(2,1) Lie algebra, with the well-known commutation relations

$$[\mathbb{J}_1,\mathbb{J}_2] = -i\mathbb{J}_3, \quad [\mathbb{J}_2,\mathbb{J}_3] = i\mathbb{J}_1, \quad [\mathbb{J}_3,\mathbb{J}_1] = i\mathbb{J}_2. \tag{1.2}$$

Among the algebra elements we have the Schrödinger Hamiltonians corresponding to a strongly attractive centripedal well ($\mathbb{J}_1 + \mathbb{J}_3$), and similarly welled harmonic ($2\mathbb{J}_3$) and repulsive ($2\mathbb{J}_1$) oscillators. The algebra (1.1) constitutes the dynamical algebra for these systems. On calculating the value of the Casimir invariant of Eqs. (1.1), we find

$$\mathbb{Q} = \mathbb{J}_1^2 + \mathbb{J}_2^2 - \mathbb{J}_3^2 = q\mathbb{1}, \tag{1.3a}$$

$$q = \tfrac{1}{4}\mu + \tfrac{3}{16} = k(1-k) = \tfrac{1}{4}(1+\lambda^2) > \tfrac{1}{4}, \tag{1.3b}$$

$$k = \tfrac{1}{2}(1 + i\lambda), \quad \lambda^2 = \mu - \tfrac{1}{4} > 0, \tag{1.3c}$$

i.e., this set of operators belongs to the continuous or principal series of representations C_q as defined by Bargmann.[12] In the proper function domain—so that Eqs. (1.1) be self-adjoint—their spectra will have no lower bound.[13] The potential singularity at the origin is indicative of the rather delicate domain problems we would find should we meet the problem starting from the algebra. This has been emphasized by Mukunda and Radhakrishnan,[10] who also considered this realization.

In Sec. 2, we shall embed the sl(2,*R*) algebra (1.1) as a subalgebra of sp(4,*R*), reduced with respect to a "hyperbolic" subalgebra so(1,1) $\oplus$ sl(2,*R*). This chain is distinct from the "radial" so(2) $\oplus$ sl(2,*R*) chain considered in Refs. 3, 6 (Appendix B), and 14. The parameterization of the plane in hyperbolic coordinate will lead to a two-component space $\mathscr{L}^2_{\rm II}(\mathscr{R}^+) = \mathscr{L}^2(\mathscr{R}^+) + \mathscr{L}^2(\mathscr{R}^+)$ of square-integrable functions on the half-line, as the appropriate domain for Eqs. (1.1), carrying both the C^0_q and $C^{1/2}_q$ representations.

In Sec. 3 we consider the Lie group SL(2,*R*)$\simeq$Sp(2,*R*) generated by Eqs. (1.1), associated with the corresponding group of matrices through

$$\exp(i\alpha\mathbb{J}_1) : \begin{pmatrix} \cosh(\alpha/2) & -\sinh(\alpha/2) \\ -\sinh(\alpha/2) & \cosh(\alpha/2) \end{pmatrix}, \tag{1.4a}$$

$$\exp(i\beta\mathbb{J}_2) : \begin{pmatrix} \exp(-\beta/2) & 0 \\ 0 & \exp(\beta/2) \end{pmatrix}, \tag{1.4b}$$

$$\exp(i\gamma\mathbb{J}_3) : \begin{pmatrix} \cos(\gamma/2) & -\sin(\gamma/2) \\ \sin(\gamma/2) & \cos(\gamma/2) \end{pmatrix}, \tag{1.4c}$$

whose adjoint action of the algebra—which is independent of the realization—is given by

$$\begin{pmatrix}\mathbb{J}_1\\ \mathbb{J}_2\\ \mathbb{J}_3\end{pmatrix} \overset{\begin{pmatrix}a & b\\ c & d\end{pmatrix}}{\longrightarrow} \begin{pmatrix} \tfrac{1}{2}(a^2-b^2-c^2+d^2) & bd-ac & \tfrac{1}{2}(a^2-b^2+c^2-d^2) \\ cd-ab & ad+bc & -cd-ab \\ \tfrac{1}{2}(a^2+b^2-c^2-d^2) & -bd-ac & \tfrac{1}{2}(a^2+b^2+c^2+d^2) \end{pmatrix} \begin{pmatrix}\mathbb{J}_1\\ \mathbb{J}_2\\ \mathbb{J}_3\end{pmatrix}. \tag{1.5}$$

[a)]On sabbatical leave from IIMAS, Universidad Nacional Autónoma de México. Apdo. Postal 20-726, México 20, D.F.

 0022-2488/80/040680-09$1.00

This group of automorphisms of the algebra will induce a corresponding group SL(2,R) of integral transforms of $\mathscr{L}^2_{\rm II}(\mathscr{R}^+)$. In the first paper of this series,[5] the algebra whose group of automorphisms was studied was the Heisenberg–Weyl algebra of quantum mechanics. The group turned out to be, as here, SL(2,R), but the integral transform carried the oscillator (or *metaplectic*) representation $D^+_{1/4} + D^+_{3/4}$. In the second paper[6] it was the sl(2,R) algebra—as here—which provided the "quantum mechanics" out of which we built the group of automorphisms (1.5) carrying the discrete D^+_k series of representations. The integral transform kernel consisted of a Gaussian times a Bessel function. Here, it will involve Gaussian functions times Hankel and Macdonald functions of imaginary index. In contradistinction with the previous cases,[5,6] this integral transform group does not allow a complex extension in the group parameters to a unitary semigroup of transforms.

In Sec. 4 we build the intertwining operator (i.e., the quantum mechanical canonical transform to action-and-angle variables) between the realization (1.1) of sl(2,R) and the well-known Bargmann realization[12] of the algebra in terms of first-order differential operators on the circle S_1:

$$\mathbb{J}^{\rm I}_1 = i\,e^{-i\epsilon\phi}(\cos\phi\,\frac{d}{d\phi} - k\,\sin\phi\,)e^{i\epsilon\phi}, \tag{1.6a}$$

$$\mathbb{J}^{\rm I}_2 = i\,e^{-i\epsilon\phi}(\sin\phi\,\frac{d}{d\phi} + k\,\cos\phi\,)e^{i\epsilon\phi}, \tag{1.6b}$$

$$\mathbb{J}^{\rm I}_3 = -i\,e^{-i\epsilon\phi}\frac{d}{d\phi}e^{i\epsilon\phi},\ k=\tfrac{1}{2}(1+i\lambda),\ \lambda\in\mathscr{R},\ \epsilon=0,\tfrac{1}{2}, \tag{1.6c}$$

which also carry the C^ϵ_q representations of the continuous series. In the third paper of this series,[11] we solved the same problem for the D^+_k case, being faced with the construction of an appropiate inner product to define a Hilbert space[15] where the spectrum of Eq. (1.6c) has a lower bound,[16] leading to the definition of a nonlocal measure on S_1. Here the problem is simpler as the appropriate Hilbert space in plainly $\mathscr{L}^2(S_1)$.

From the point of view of the program on nonlinear canonical transformations outlined Ref. 17, our case presents a challenge which merits deeper study, since the classical canonical transformation to action-and-angle variables

$$p_\phi = J^c_3,\quad \phi = \arctan(J^c_2/J^c_1), \tag{1.7}$$

[where J^c_k is the "classical counterpart" ($-id/d\rho \mapsto p_\rho$) of Eqs. (1.1) and Poisson brackets replace commutators] has the same "ambiguity group"[16] for all $\mu > 0$. Moreover, the interval $\frac{1}{4} \geqslant \mu > -\frac{3}{4}$, $(0 < k < 1)$ is particularly troublesome, since various choices of boundary conditions[18] lead to representations which may belong to the lower-bounded discrete series[19] or to the unbounded supplementary series—a problem still to be solved for the algebra (1.1)—which are not quite apparent in the formal expressions in Eqs. (1.1), and invisible in the classical Poisson-bracket construct. In establishing our results from the point of view of groups of integral transforms, we hope to settle some of the uncertainties which may arise in the algebraic approach to canonical transformations in quantum mechanics. Finally in Sec. 5 we outline some applications and offer some concluding remarks.

2. THE CHAIN sp(4,R)⊃so(1,1)⊕sl(2,R) AND HYPERBOLIC COORDINATES

We consider the usual quantum mechanical operators of position and momentum in two dimensions [$\mathbb{Q}_m f(\mathbf{q}) = q_m f(\mathbf{q})$ and $\mathbb{P}_m f(\mathbf{q}) = -i\partial f(\mathbf{q})/\partial q_m$, $m = 1,2$] and out of these we build the symmetrized quadratic expressions $\mathbb{Q}_m\mathbb{Q}_n$, $\mathbb{P}_m\mathbb{P}_n$, $\frac{1}{2}\{\mathbb{Q}_m,\mathbb{P}_n\}_+$. These ten operators span under Lie commutation the four-dimensional real symplectic algebra sp(4,R), isomorphic to the pseudo-orthogonal algebra so(3,2). Let us denote the latter's generators in the Cartesian basis by

$$\begin{aligned}
&\mathbb{M}_{12} = \tfrac{1}{2}(\mathbb{Q}_1\mathbb{P}_2 - \mathbb{Q}_2\mathbb{P}_1), \mathbb{M}_{13} = -\tfrac{1}{2}(\mathbb{P}_1\mathbb{P}_2 + \mathbb{Q}_1\mathbb{Q}_2),\\
&\mathbb{M}_{14} = -\tfrac{1}{2}(\mathbb{Q}_1\mathbb{P}_2 + \mathbb{Q}_2\mathbb{P}_1),\ \mathbb{M}_{15} = -\tfrac{1}{2}(\mathbb{P}_1\mathbb{P}_2 - \mathbb{Q}_1\mathbb{Q}_2),\\
&\mathbb{M}_{23} = \tfrac{1}{4}(\mathbb{P}_1^2 - \mathbb{P}_2^2 + \mathbb{Q}_1^2 - \mathbb{Q}_2^2),\ \mathbb{M}_{24} = \tfrac{1}{2}(\mathbb{Q}_1\mathbb{P}_1 - \mathbb{Q}_2\mathbb{P}_2),
\end{aligned} \tag{2.1}$$

$$\begin{aligned}
&\mathbb{M}_{25} = \tfrac{1}{4}(\mathbb{P}_1^2 - \mathbb{P}_2^2 - \mathbb{Q}_1^2 + \mathbb{Q}_2^2),\\
&\mathbb{M}_{34} = -\tfrac{1}{4}(\mathbb{P}_1^2 + \mathbb{P}_2^2 - \mathbb{Q}_1^2 - \mathbb{Q}_2^2),\\
&\mathbb{M}_{35} = \tfrac{1}{4}(\{\mathbb{Q}_1,\mathbb{P}_1\}_+ + \{\mathbb{Q}_2,\mathbb{P}_2\}),\\
&\mathbb{M}_{45} = \tfrac{1}{4}(\mathbb{P}_1^2 + \mathbb{P}_2^2 + \mathbb{Q}_1^2 + \mathbb{Q}_2^2),
\end{aligned}$$

where the metric is $(+\ +\ +\ -\ -)$. The set of operators generating the compact subgroup SO(2)⊗SO(3)⊂SO(3,2) [i.e., those which have a discrete spectrum in $\mathscr{L}^2(\mathscr{R}^2)$] is $\{\mathbb{M}_{45};\mathbb{M}_{12},\mathbb{M}_{13},\mathbb{M}_{23}\}$. The set generating the "radial" subgroup SO(2)⊗SL(2,R) of Refs. 3 and 6 is $\{\mathbb{M}_{12};\mathbb{M}_{34},\mathbb{M}_{35},\mathbb{M}_{45}\}$. Here, we shall consider the set $\{\mathbb{M}_{14};\mathbb{M}_{23},\mathbb{M}_{25},\mathbb{M}_{35}\}$ generating the "hyperbolic" subgroup SO(1,1)⊗SL(2,R)⊂Sp(4,R). The so(1,1) element is the Lorentz boost generator in the plane, while the sl(2,R) elements are built out of the harmonic (h) and repulsive (r) one-dimensional Schrödinger Hamiltonians $\mathbb{H}^{(\cdot)}_k, k = 1,2$ as $\mathbb{M}_{23} = \frac{1}{2}(\mathbb{H}^{(h)}_1 - \mathbb{H}^{(h)}_2)$ and $\mathbb{M}_{25} = \frac{1}{2}(\mathbb{H}^{(r)}_1 - \mathbb{H}^{(r)}_2)$, [rather than $\mathbb{M}_{45} = \frac{1}{2}(\mathbb{H}^{(h)}_1 + \mathbb{H}^{(h)}_2)$ and $\mathbb{M}_{34} = -\frac{1}{2}(\mathbb{H}^{(r)}_1 + \mathbb{H}^{(r)}_2)$ as in the radial case]. The generator $\mathbb{M}_{35}$ is common to the hyperbolic and radial subgroups. In $\mathscr{L}^2(\mathscr{R}^2)$, thus, the eigenfunctions of $\mathbb{M}_{23}$ will be $\Psi_{n_1n_2}(\mathbf{q}) = \Psi^h_{n_1}(q_1)\Psi^h_{n_2}(q_2)$ [where $\Psi^h_n(q) = (-1)^n\Psi^h_n(-q)$ are the simple harmonic oscillator wavefunctions], and its spectrum will be given by $m = \frac{1}{2}(n_1 - n_2)$, $n_1,n_2 = 0,1,2,\cdots$. This set of functions will thus constitute a basis for the two continuous series representations of sl(2,R): C^0_q spanned by the subset with $n_1 + n_2$

even [so that m is integer $\Psi_{n_1n_2}(-\mathbf{q})=\Psi_{n_1n_2}(\mathbf{q})$], and $C_q^{1/2}$ by the subset with n_1+n_2 odd [m in half-integer and $\Psi_{n_1n_2}(-\mathbf{q})=-\Psi_{n_1n_2}(\mathbf{q})$].

We shall now parametrize the plane in hyperbolic coordinates (ρ,ϕ,σ), dividing it into two regions labeled by σ as

$$\text{for } q_1^2-q_2^2>0:\quad \sigma=+1,\quad q_1=\rho\cosh\phi,\quad q_2=\rho\sinh\phi,\quad \rho,\phi\in R;\tag{2.2a}$$

$$\text{for } q_1^2-q_2^2<0:\quad \sigma=-1,\quad q_1=\rho\sinh\phi,\quad q_2=\rho\cosh\phi,\tag{2.2b}$$

and disregard the cone $q_1^2-q_2^2=0$, as this is a submanifold of lower dimension. The elements $f(\mathbf{q})$ of the space of functions $\mathscr{L}^2(\mathscr{R}^2)$ on the plane will be correspondingly represented by *pairs* of functions $f_\sigma(\rho,\phi)$, $\sigma=\pm1$, elements of a space $\mathscr{L}^2_1(\mathscr{R})\dotplus\mathscr{L}^2_{-1}(\mathscr{R})$ which can be arranged as a two-component vector column

$$\mathbf{f}(\rho,\phi)=\begin{pmatrix} f_1(\rho,\phi)\\ f_{-1}(\rho,\phi)\end{pmatrix},\quad f_\sigma(\rho,\phi)=f(\mathbf{q}(\rho,\phi,\sigma)).\tag{2.3}$$

The inner product in $\mathscr{L}^2(\mathscr{R}^2)$ becomes

$$\begin{aligned}(\mathbf{f},\mathbf{g})_2&=\int_{-\infty}^{\infty}dq_1\int_{-\infty}^{\infty}dq_2 f(q_1,q_2)^*g(q_1,q_2)\\&=\sum_{\sigma=\pm1}\int_{-\infty}^{\infty}|\rho|d\rho\int_{-\infty}^{\infty}d\phi f_\sigma(\rho,\phi)^*g_\sigma(\rho,\phi),\end{aligned}\tag{2.4}$$

in terms of the hyperbolic coordinates. Finally, the generators of $SO(1,1)\otimes SL(2,R)$ can be written as

$$\mathbb{K}_0=-\mathbb{M}_{14}=-i\frac{1}{2}\frac{\partial}{\partial\phi},\tag{2.5}$$

$$\mathbb{K}_1=\mathbb{M}_{25}=\sigma\tfrac{1}{4}\rho^{-1/2}\Big[-\frac{\partial^2}{\partial\rho^2}-\rho^{-2}\times\Big(\frac{1}{4}-\frac{\partial^2}{\partial\phi^2}\Big)-\rho^2\Big]\rho^{1/2},\tag{2.6a}$$

$$\mathbb{K}_2=\mathbb{M}_{35}=-i\tfrac{1}{2}\rho^{-1/2}\Big[\rho\frac{\partial}{\partial\rho}+\frac{1}{2}\Big]\rho^{1/2},\tag{2.6b}$$

$$\mathbb{K}_3=\mathbb{M}_{23}=\sigma\tfrac{1}{4}\rho^{-1/2}\Big[-\frac{\partial^2}{\partial\rho^2}-\rho^{-2}\times\Big(\frac{1}{4}-\frac{\partial^2}{\partial\phi^2}\Big)+\rho^2\Big]\rho^{1/2}.\tag{2.6c}$$

The operators (2.6) exhibit commutation relations analogous to Eq. (1.2). Acting on the column-vector function (2.3), the generators above will be represented by 2×2 diagonal matrices with operator elements, which for Eqs. (2.6a) and (2.6c) have opposite signs. The adjoint action of the group generated by Eqs. (2.6) on themselves can be verified to be formally identical to Eq. (1.5), as it should be, since the latter is a relation independent of the particular operator realization. For the $\sigma=-1$ components, we have a reversal of the signs of α and γ in Eqs. (1.4), i.e., of b and c in the elements of the 2×2 matrix realization in Eq. (1.5). This leaves the 3×3 matrix in Eq. (1.5) invariant.

The subalgebras so(1,1) and sl(2,R) generated by Eqs.(2.5) and (2.6) are *conjugate* in sp(4,R); the reduction to an irreducible subspace (irrep) of the former leads to a corresponding irrep of the latter. Since for sp(4,R) itself we do not have a single irrep space but a direct sum of two—those with a basis with integer and with half-integer eigenvalues m under $\mathbb{M}_{45}$ or $\mathbb{M}_{23}$—the corresponding reduction of the sl(2,R) generators will be the direct sum of two irrep's C_q^0 and $C_q^{1/2}$, respectively. An irrep space for $\mathbb{K}_0$ within Eqs. (2.1) is provided by functions $f_\sigma^\lambda(\rho,\phi)=f_\sigma^\lambda(\rho)\exp(i\lambda\phi)$, $\lambda\in\mathscr{R}$. This will replace the operator $-\partial^2/\partial\phi^2$ in Eqs. (2.6) by λ^2 and bring the $\mathbb{K}_k$ to within a similarity transformation (by $\rho^{-1/2}$) of the forms (1.1).

In the following sections we shall be interested in certain discrete operations on the plane in Cartesian and hyperbolic coordinates which are, nevertheless, elements of the parent Sp(4,R) group and which can be connected to the identity. These will be identified using the notation of Mukunda and Radhakrishnan[10]. First, we have the full space inversion

$$\mathbb{P}:(q_1,q_2)\to(-q_1,-q_2),\text{i.e.,}\mathbb{P}:(\rho,\phi,\sigma)\to(-\rho,\phi,\sigma),\tag{2.7a}$$

$$\mathbb{P}\mathbb{K}_\nu=\mathbb{K}_\nu\mathbb{P},\quad \nu=0,1,2,3,\tag{2.7b}$$

$$\mathbb{P}=\exp(2\pi i\mathbb{M}_{23})=\exp(2\pi i\mathbb{M}_{45}),\tag{2.7c}$$

i.e., it is the rotation-by-2π element of SL(2,R) which commutes with the algebra $so(2,1)\simeq sl(2,R)$ and which can be used to distinguish the vector and spinor constituent irrep's C_q^0 and $C_q^{1/2}$ by demanding that $\mathbb{P}$ be diagonal. We use its eigenvalues $p=\pm1$ to distinguish the irrep spaces for C_q^ϵ through

$$\epsilon=\tfrac{1}{4}(1-p),\text{i.e., }\epsilon=0(1/2)\text{ for }p=+1(-1).\tag{2.7d}$$

Second, we have the inversion of the second Cartesian coordinate

$$\mathbb{B}:(q_1,q_2)\to(q_1,-q_2),\text{i.e., }\mathbb{B}(\rho,\phi,\sigma)\to(\sigma\rho,-\phi,\sigma),\tag{2.8a}$$

$$\mathbb{B}\mathbb{K}_0=-\mathbb{K}_0\mathbb{B};\ \mathbb{B}\mathbb{K}_k\mathbb{B},\quad k=1,2,3,\tag{2.8b}$$

$$\mathbb{B}=\exp(i\pi[\mathbb{M}_{45}-\mathbb{M}_{23}]).\tag{2.8c}$$

This element commutes with the sl(2,R) algebra and with $\mathbb{P}$, but will intertwine the λ and $-\lambda$ representations of so(1,1), and hence those of sl(2,R). Its effect on the properly reduced irrep space C_q^ϵ will be to change the sign of the lower component of the $\epsilon=\frac{1}{2}$ function pair.

Third, we have the element $\mathbb{B}\mathbb{P}$, which will not interest us separately, and fourth, the operator

$$\mathbb{A}:(q_1,q_2)\to(q_2,q_1),\text{i.e., }\mathbb{A}:(\rho,\phi,\sigma)\to(\rho,\phi,-\sigma),\tag{2.9a}$$

$$\mathbb{A}\mathbb{K}_j=\mathbb{K}_j\mathbb{A},\quad j=0,2;\ \mathbb{A}\mathbb{K}_k=-\mathbb{K}_k\mathbb{A},\quad k=1,3,\tag{2.9b}$$

$$\mathbb{A}=\mathbb{B}\exp(i\pi\mathbb{M}_{12}).$$

This element does not commute with $\mathbb{B}$ (instead, $\mathbb{A}\mathbb{B}=\mathbb{B}\mathbb{P}\mathbb{A}$), but it commutes with $\mathbb{P}$ and $\mathbb{K}_0$ and is thus representable as a unitary transformation in each C_q^ϵ irrep which reverses the sign of the $\mathbb{K}_3$ eigenvalues. Its own eigenvalues ($a=\pm1$) will be used to classify the double-multiplicity $\mathbb{K}_2$ eigenfunctions. It is representable as a $\boldsymbol{\sigma}_1$ Pauli matrix in the two-component function space (2.3). The $\mathbb{A}$ and $\mathbb{B}$ automorphisms are outer to SL(2,R), while $\mathbb{P}$ is inner.

3. THE INTEGRAL TRANSFORM GROUP

The integral transform action of the Sp(4,R) group generated by (2.1) on $\mathscr{L}^2(\mathscr{R}^2)$ is known[20,21]. In particular, for the SL(2,R) subgroup generated by Eqs. (2.6), represented by the matrices[22,23]

$$\mathbf{M}=\begin{pmatrix} a\mathbf{1} & b\boldsymbol{\sigma}_3 \\ c\boldsymbol{\sigma}_3^{-1} & d\,\mathbf{1}\end{pmatrix},\quad ad-bc=1,$$
$$\mathbf{1}=\begin{pmatrix}1&0\\0&1\end{pmatrix},\quad \boldsymbol{\sigma}_3=\begin{pmatrix}1&0\\0&e^{-i\pi}\end{pmatrix}, \tag{3.1}$$

it is

$$f(\mathbf{q})\overset{\mathbf{M}}{\to}[\mathbb{C}_{\mathbf{M}}\mathbf{f}](\mathbf{q})=\int_{R^2}d^2\mathbf{q}'C_{\mathbf{M}}(\mathbf{q},\mathbf{q}')f(\mathbf{q}'), \tag{3.2a}$$

where the integral kernel is, for $b\neq 0$,

$$C_{\mathbf{M}}(\mathbf{q},\mathbf{q}')=(2\pi|b|)^{-1}\exp[i(a\{q_1'^2-q_2'^2\}-2\{q_1'q_1-q_2'q_2\}+d\{q_1^2-q_2^2\})/2b], \tag{3.2b}$$

while for $b=0$, it is

$$C_{\mathbf{M}(b=0)}(\mathbf{q},\mathbf{q}')=a^{-1}\exp[ic(q_1^2-q_2^2)/2a]\times\delta^2(\mathbf{q}'-a^{-1}\mathbf{q}). \tag{3.2c}$$

For $\mathbf{M}=\mathbf{1}$ we have thus the reproducing kernel under Eq. (3.2a). This integral transform group provides a vector representation[24] of SL(2,R):

$$\int_{\mathscr{R}^2}d^2\mathbf{q}'C_{\mathbf{M}_1}(\mathbf{q},\mathbf{q}')C_{\mathbf{M}_2}(\mathbf{q}',\mathbf{q}'')=C_{\mathbf{M}_1\mathbf{M}_2}(\mathbf{q},\mathbf{q}''), \tag{3.3}$$

and the transforms are unitary in $\mathscr{L}^2(\mathscr{R}^2)$.

We now introduce hyperbolic coordinates (ρ,ϕ,σ) as given by Eqs. (2.2). The kernel (3.2b) and (3.2c) then appears as

$$\begin{aligned}C_{\mathbf{M}}(\rho,\phi,\sigma;\rho',\phi',\sigma')&=C_{\mathbf{M}}(-\rho,\phi,\sigma;-\rho',\phi',\sigma')\\&=C_{\mathbf{M}}(\rho,\phi-\phi',\sigma;\rho',0,\sigma')\\&=(2\pi|b|)^{-1}\exp[i(a\sigma'\rho'^2-2\rho\rho'\mathrm{hyp}_{\sigma,\sigma'}(\phi'-\phi)\\&\quad+d\sigma\rho^2)/2b],\end{aligned} \tag{3.4a}$$

$$\begin{aligned}\mathrm{hyp}_{1,1}(z)&=\cosh(z)=-\mathrm{hyp}_{-1,-1}(z),\\ \mathrm{hyp}_{1,-1}(z)&=\sinh(z)=-\mathrm{hyp}_{-1,1}(z),\end{aligned} \tag{3.4b}$$

and can be arranged into a 2×2 matrix with rows and columns as the functions are represented by Eq. (2.3). We can display the eigenspaces of $\mathbb{P}$ and $\mathbb{K}_0$ through the operator

$$\begin{aligned}f_\sigma^{p,\lambda}(\rho)&=(\mathbb{T}^{p,\lambda}\mathbf{f}_\sigma)(\rho)=pf_\sigma^{p,\lambda}(-\rho)\\&=|\rho|^{1/2}(\mathbb{1}+p\mathbb{P})\int_{-\infty}^{\infty}d\phi f_\sigma(\rho,\phi)\exp(-i\lambda\phi),\end{aligned} \tag{3.5a}$$

thus allowing us to reduce the domain of the functions to the interval $\rho\geqslant 0$. Conversely,

$$\begin{aligned}f_\sigma(\rho,\phi)&=(\mathbb{T}^{\phi}\mathbf{f}_\sigma^{\sim})(\rho,\phi)\\&=\frac{1}{4\pi}|\rho|^{-1/2}\sum_{p=\pm 1}\int_{-\infty}^{\infty}d\lambda f_\sigma^{p\lambda}(\rho)\exp(i\lambda\phi).\end{aligned} \tag{3.5b}$$

We define an inner product in the (p,λ) subspace $\mathscr{L}^2_{\mathrm{II}}(\mathscr{R}^+)_{p,\lambda}=\mathscr{L}^2_1(\mathscr{R}^+)\dot{+}\mathscr{L}^2_{-1}(\mathscr{R}^+)$ as

$$(\mathbf{f},\mathbf{g})_{p,\lambda}=\sum_{\sigma=\pm 1}\int_0^\infty d\rho f_\sigma^{p,\lambda}(\rho)^*g_\sigma^{p,\lambda}(\rho), \tag{3.6}$$

and note that it will relate to Eq. (2.4) through

$$(\mathbf{f},\mathbf{g})_2=\frac{1}{4\pi}\sum_{p=\pm 1}\int_{-\infty}^{\infty}d\lambda\,(\mathbf{f},\mathbf{g})_{p,\lambda}. \tag{3.7}$$

The properties of $\mathbb{T}^{p,\lambda}$ are such that

$$\mathbb{T}^{p,\lambda}\mathbb{P}=p\mathbb{T}^{p,\lambda},\quad \mathbb{T}^{p,\lambda}\mathbb{K}_0=\tfrac{1}{2}\lambda\mathbb{T}^{p,\lambda}, \tag{3.8a}$$

$$(\mathbb{J}_1^{\mathrm{II}},\mathbb{J}_2^{\mathrm{II}},\mathbb{J}_3^{\mathrm{II}})=\mathbb{T}^{p,\lambda}(\mathbb{K}_1,\mathbb{K}_2,\mathbb{K}_3)=(\sigma\mathbb{J}_1,\mathbb{J}_2,\sigma\mathbb{J}_3)\mathbb{T}^{p,\lambda}, \tag{3.8b}$$

$$\mathbb{T}^{p,\lambda}\mathbb{A}=\mathbb{A}\mathbb{T}^{p,\lambda},\quad \mathbb{T}^{p,\lambda}\mathbb{B}=\mathbb{T}^{p,-\lambda}. \tag{3.8c}$$

Equations (3.8a) only state that $\mathbb{T}^{p,\lambda}$ indeed projects out eigenspaces of $\mathbb{P}$ and $\mathbb{K}_0$, while Eqs. (3.8c) give relations which will be used later on. Equations (3.8b), finally, bring the three algebra generators (1.1) into the picture and, besides telling us that the special Sp(4,R) transform (3.2) leaves the (p,λ) subspace invariant, allows us to calculate the integral transform representing the operator $\mathbb{C}_{\mathbf{M}}^{p,\lambda}=\mathbb{T}^{p,\lambda}\mathbb{C}_{\mathbf{M}}$ which maps $\mathscr{L}^2_{\mathrm{II}}(\mathscr{R}^+)_{p,\lambda}$ onto itself unitarily. Since the inner product (3.6) does not explicitly contain the labels p,λ, we shall henceforth drop them from specifying the space $\mathscr{L}^2_{\mathrm{II}}(\mathscr{R}^+)$.

For functions $f_\sigma(\rho)\in\mathscr{L}^2_{\mathrm{II}}(\mathscr{R}^+)$, thus, the SL(2,$R$) group generated by the operators $\mathbb{J}_k^{\mathrm{II}}$, $k=1,2,3$, acts as

$$\begin{aligned}f_\sigma(\rho)&\overset{\mathbf{M}}{\to}[\mathbb{C}_{\mathbf{M}}^{p,\lambda}\mathbf{f}]_\sigma(\rho)\\&=\sum_{\sigma'=\pm 1}\int_0^\infty d\rho' C_{\mathbf{M},\sigma,\sigma'}^{p,\lambda}(\rho,\rho')f_{\sigma'}(\rho'),\end{aligned} \tag{3.9a}$$

with the integral kernel

$$\begin{aligned}C_{\mathbf{M},\sigma,\sigma'}^{p,\lambda}(\rho,\rho')&=\rho^{1/2}(\mathbb{T}^{p,\lambda}\mathbf{C}_{\mathbf{M},\sigma,\sigma'})(\rho,\rho')\\&=(\rho\rho')^{1/2}\int_{-\infty}^{\infty}d\psi[C_{\mathbf{M},\sigma,\sigma'}(\rho,\psi;\rho',0)\\&\quad+pC_{\mathbf{M},\sigma,\sigma'}(\rho,\psi;-\rho',0)]\exp(-i\lambda\psi)\\&=G_{\mathbf{M},\sigma,\sigma'}(\rho,\rho')H_{\sigma,\sigma'}^{p,\lambda}(\rho\rho'/b),\end{aligned} \tag{3.9b}$$

where, on evaluating this expression from Eqs. (3.4) for $b\neq 0$, we find it to be a product of a Gaussian factor

$$G_{\mathbf{M},\sigma,\sigma'}(\rho,\rho')=(2\pi|b|)^{-1}(\rho\rho')^{1/2}\times\exp[i(d\sigma\rho^2+a\sigma'\rho'^2)/2b], \tag{3.10}$$

and a factor $H_{\sigma,\sigma'}^{p,\lambda}(z)$ which contains the integration over ψ and which can be performed in terms of Hankel and Macdonald[25] functions, yielding[26]

$$\begin{aligned}H_{1,1}^{p,\lambda}(z)&=pH_{-1,-1}^{p,\lambda}(z)\\&=4p\int_0^\infty d\psi\mathrm{trig}_p(z\cosh\psi)\cos(\lambda\psi)\\&=i\pi[pe^{-\lambda\pi/2}H_{i\lambda}^{(1)}(z)-e^{\lambda\pi/2}H_{i\lambda}^{(2)}(z)]\\&=pH_{1,1}^{p,\lambda}(-z)=H_{1,1}^{p,-\lambda}(z),\end{aligned} \tag{3.11a}$$

$$\begin{aligned}H_{1,-1}^{p,\lambda}(z)&=pH_{-1,1}^{p,\lambda}(z)\\&=4p\int_0^\infty d\psi\mathrm{trig}_p(z\sinh\psi)\mathrm{trig}_p(\lambda\psi)\\&=4(\mathrm{sign}z)^{2\epsilon}\mathrm{hyp}_{1,p}(\lambda\pi/2)K_{i\lambda}(|z|)\\&=pH_{1,-1}^{p,\lambda}(-z)=pH_{1,-1}^{p,-\lambda}(z),\end{aligned} \tag{3.11b}$$

$$\mathrm{trig}_{+1}(z)=\cos(z),\quad \mathrm{trig}_{-1}(z)=i\sin(z). \tag{3.11c}$$

The case $b = 0$ may be obtained either from Eq. (3.11) for $b \to 0$ and the use of the asymptotic properties of the cylinder functions,[27] or directly from Eqs. (3.2c) and (3.9), as

$$C^{p,\lambda}_{\mathbf{M}(b=0),\sigma,\sigma'}(\rho,\rho') = |a|^{-1/2}(\mathrm{sign}a)^{2\epsilon}\exp(i\sigma c\rho^2/2a)\,\delta_{\sigma,\sigma'}\delta(\rho'-\rho/|a|). \tag{3.12}$$

This integral transform is unitary[28] on $\mathscr{L}^2_{\mathrm{II}}(\mathscr{R}^+)$ with the inner product (3.6).

The group properties of this matrix kernel are directly inherited from Eq. (3.3) via Eqs. (3.5a), namely,

$$\sum_{\sigma'=\pm 1}\int_0^\infty d\rho' C^{p,\lambda}_{\mathbf{M}_1,\sigma,\sigma'}(\rho,\rho')C^{p,\lambda}_{\mathbf{M}_2,\sigma',\sigma''}(\rho',\rho'') = C^{p,\lambda}_{\mathbf{M}_1\mathbf{M}_2,\sigma,\sigma''}(\rho,\rho''). \tag{3.13}$$

We should point out that the property which distinguishes the C^0_q and $C^{1/2}_q$ representations is clearly displayed:

$$C^{p,\lambda}_{-\mathbf{M},\sigma,\sigma'}(\rho,\rho') = (-1)^{2\epsilon}C^{p,\lambda}_{\mathbf{M},\sigma,\sigma'}(\rho,\rho'). \tag{3.14}$$

This is a consequence of Eqs. (3.12) and (3.13) which can also be seen from the explicit expressions (3.9)–(3.11), noting that the Gaussian factor is the same for $\mathbf{M}$ and $-\mathbf{M}$, while $H^{p,\lambda}_{\sigma,\sigma'}(-z) = pH^{p,\lambda}_{\sigma,\sigma'}(z)$.

Regarding the operator $\mathbb{B}$ defined in Eqs. (2.8), the last equality in Eqs. (3.12a) and (3.12b) shows that the $\sigma = -1$ components of the $\epsilon = 1/2$ irrep functions indeed consistently invert their signs and that this inversion is thus representable by a σ_3 Pauli matrix in the two-component $\mathscr{L}^2_{\mathrm{II}}(\mathscr{R}^+)$ space, intertwining the $k = \frac{1}{2}(1+i\lambda)$ and $k^* = \frac{1}{2}(1-i\lambda)$ representations $C^{1/2}_q$. Eigenfunctions of $\mathbb{B}$ with eigenvalue $b = \pm 1$ in $C^{1/2}_q$ can be built as functions with only a nonzero upper ($b = +1$) or a lower ($b = -1$) component. In Sec. 4, however, we shall gloss over this classification scheme in favor of others. In C^0_q, $\mathbb{B}$ is equivalent to the identity transformation.

Having given the kernel for the general hyperbolic canonical transforms, we would like to present a peculiarity of the "hyperbolic Fourier transform," i.e., the transform $\mathbb{C}^{p,\lambda}_{\mathbf{F}}$ corresponding to the matrix $\mathbf{M}$ given by $\mathbf{F} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ which from Eq. (1.4c) is $\mathbb{C}^{p,\lambda}_{\mathbf{F}} = \exp(-i\pi\mathbb{J}^{\mathrm{II}}_3)$. In the case of linear canonical transforms[5,7] this is $e^{-i\pi/4}$ times the ordinary Fourier transform. For the radial case[6], it is the Hankel transform. Here, as $a = 0 = d$, the Gaussian factor (3.11) is simply $(\rho\rho')^{1/2}/2\pi$ and hence

$$\left[\mathbb{C}^{p,\lambda}_{\mathbf{F}}\begin{pmatrix} \mathbf{f}_1 \\ \mathbf{f}_{-1} \end{pmatrix}\right](\rho) = p\left[\mathbb{C}^{p,\lambda}_{\mathbf{F}}\begin{pmatrix} \mathbf{f}_1 \\ \mathbf{f}_{-1} \end{pmatrix}\right](\rho)$$
$$= \frac{1}{2\pi}\int_0^\infty d\rho'(\rho\rho')^{1/2}\begin{pmatrix} H^{p,\lambda}_{1,1}(\rho\rho') & H^{p,\lambda}_{1,-1}(\rho\rho') \\ pH^{p,\lambda}_{1,-1}(\rho\rho') & pH^{p,\lambda}_{1,1}(\rho\rho') \end{pmatrix} \times \begin{pmatrix} f_1(\rho') \\ f_{-1}(\rho') \end{pmatrix}. \tag{3.15}$$

The inverse hyperbolic Fourier transform is thus identical to the direct one for the C^0_q irrep, while it differs by a minus sign for the $C^{1/2}_q$ irrep[29]. The origin of this property is the behavior of the sl(2,R) algebra under the $\mathbb{A}$ operator in Eq. (2.9b): $\mathbb{C}^{p,\lambda}_{\mathbf{F}}\mathbb{A} = \mathbb{A}\mathbb{C}^{p,\lambda}_{\mathbf{F}} = p\mathbb{C}^{p,\lambda}_{\mathbf{F}}$. For $p = +1$ ($\epsilon = 0$) we may thus construct eigenspaces of $\mathbb{A}$ consisting of functions of even and odd parity under this operator: It suffices to apply the unimodular matrix $2^{-1/2}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$ to Eq. (3.15) in order to obtain through similarity a diagonal transformation matrix kernel with elements $(H^{1,\lambda}_{1,1} + H^{1,\lambda}_{1,-1})(\rho\rho')$ and $(H^{1,\lambda}_{1,1} - H^{1,\lambda}_{1,-1})(\rho\rho')$, corresponding to eigenvalues $a = +1$ and $a = -1$, respectively, under $\mathbb{A}$. Each $\mathbb{A}$-classified subspace is then transformed into itself under $\mathbb{C}^{1,\lambda}_{\mathbf{F}}$ and consists of functions such that $f_\sigma(\rho) = af_{-\sigma}(\rho)$. Besides providing a distinguishing label for the two $\mathbb{J}^{\mathrm{II}}_2$ eigenfunctions (Sec. 4), the operator $\mathbb{A}$ may be thus used to construct and distinguish between these two $p = +1$ Fourier transforms in the space where—as for the Hankel transform—the square of the transform is the identity. A similar construction for the $C^{1/2}_q$ irrep yields an antidiagonal matrix kernel.

In closing this section, it should be noted that the analytic continuation in the group parameters of Eqs. (3.1)—so fruitfully exploited in Ref. 14—turns out to be impossible here: If one applies the criterion of Ref. 20 to this matrix, one sees that for no complex values of the parameters does one have a Hilbert–Schmidt operator. This becomes intuitively clear as the analog of the heat diffusion transform[4] ($a = 1 = d, c = 0, b = -2it$) is forward in time t for the first Cartesian coordinate, but backward in the second one. Examination of the kernel in Eqs. (3.9) or introduction of complex hyperbolic coordinates in Ref. 5 (Appendix B) corroborates this conclusion. This seems to be thus a major and inescapable distinction between the discrete and continuous SL(2,R) representation series.

4. THE INTERTWINING OPERATOR

In this section we shall build the operator which intertwines the two algebra realizations (1.1) and (1.6) or, more precisely, the unitary transform kernel mapping the space $\mathscr{L}^2_{\mathrm{II}}(\mathscr{R}^+)$ described in Sec. 3 onto the more usual $\mathscr{L}^2(S_1)$ space, in such a way that the second-order differential operators $\mathbb{J}^{\mathrm{II}}_k$ defined in Eqs. (3.8b) map onto the first-order ones $\mathbb{J}^{\mathrm{I}}_k$ given in Eqs. (1.6). This is the proper quantum analog of the canonical transformation to action-and-angle variables (1.7)[30].

Let $\Psi^{p,\lambda}_{m,\sigma}(\rho)$ and $X^{p,\lambda}_{\nu,\sigma}(\rho)$ be the (proper or generalized) eigenfunctions of $\mathbb{P}$ and two operators in the set $\mathbb{J}^{\mathrm{II}}_k$, and $\psi^{p,\lambda}_m(\phi)$ and $\chi^{p,\lambda}_\nu(\phi)$ for the corresponding operators in the set $\mathbb{J}^{\mathrm{I}}_k$. We can choose the first operator to be elliptic, specifically $\mathbb{J}^{\bullet}_3$, and the second to be either hyperbolic[31] ($\mathbb{J}^{\bullet}_1$ or $\mathbb{J}^{\bullet}_2$), or parabolic ($\mathbb{J}^{\bullet}_3 \pm \mathbb{J}^{\bullet}_1$)—specifically, we shall employ $\mathbb{J}^{\bullet}_3 - \mathbb{J}^{\bullet}_1$. The last choice will be followed, as it is the simplest: The generalized eigenfunctions are Dirac δ's while we are assured that the spectrum of this operator covers the real line *once*[32]. The intertwining integral kernel will then be computable as the generating function

$$K^{p,\lambda}_\sigma(\phi,\rho) = \sum_m \Psi^{p,\lambda}_{m,\sigma}(\rho)^*\psi^{p,\lambda}_m(\phi)\exp[i\Phi^{\psi}(p,\lambda,m,\sigma)] = \int_{-\infty}^{\infty} d\nu X^{p,\lambda}_{\nu,\sigma}(\rho)^*\chi^{p,\lambda}_\nu(\phi)\exp[i\Phi^{\chi}(p,\lambda,\nu,\sigma)]. \tag{4.1}$$

The correct choice of phase[33] for Φ^{ψ} and Φ^{χ} is nontrivial for two reasons. First, it actually may change the generating function: Assume we apply $\mathbb{C}^{p,\lambda}_{\mathbf{M}(\gamma)} = \exp[i\gamma(\mathbb{J}^{\mathrm{II}}_3 - \mathbb{J}^{\mathrm{II}}_1)]$ to Eq. (4.1)[34], multiplying the integrand by $e^{i\gamma\nu}$ and thus producing a new generating function which, as a sum, will consist of eigenfunctions of[35] $\mathbb{C}^{p,\lambda}_{\mathbf{M}(\gamma)}\mathbb{J}^{\mathrm{II}}_3\mathbb{C}^{p,\lambda}_{\mathbf{M}(-\gamma)} \neq \mathbb{J}^{\mathrm{II}}_3$. Second, certain phase requirements exist, notably Bargmann's convention[36] for the $\mathbb{J}^{\mathrm{I}}_3$ eigenbasis, which involves definite transformation phases under the operator $\mathbb{A}$ in Eqs. (2.9). However, once we have used two generators [algebraic basis for sl(2,R)] to determine the phases for the intertwining kernel, no further requirement is imposed by the third (vector basis) generator, as its matrix elements are fixed by the first two.

It should be clear, however, that independent of the appropriate choice of phases, the kernel (4.1) will intertwine $\mathscr{L}^2_{\mathrm{II}}(\mathscr{R}^+)$ and $\mathscr{L}^2(S_1)$ as

$$f^{\mathrm{I}}(\phi) = \sum_{\sigma=\pm 1}\int_0^{\infty} d\rho K^{p,\lambda}_{\sigma}(\phi,\rho) f^{\mathrm{II}}_{\sigma}(\rho), \tag{4.2a}$$

$$f^{\mathrm{II}}_{\sigma}(\rho) = \int_{-\pi}^{\pi} d\phi f^{\mathrm{I}}(\phi) K^{p,\lambda}_{\sigma}(\phi,\rho)^*, \tag{4.2b}$$

for $f^{\mathrm{II}}_{\sigma}(\rho)$ and $f^{\mathrm{I}}(\phi)$ in the two spaces, respectively. The unitarity of the transformation is guaranteed by the assumed Dirac orthonormality and completeness of the two eigenbases—including any similarity transformation as mentioned above—which, from Eq. (4.1) alone, implies

$$\int_{-\pi}^{\pi} d\phi K^{p,\lambda}_{\sigma}(\phi,\rho) K^{p,\lambda}_{\sigma'}(\phi,\rho')^* = \delta_{\sigma\sigma'}\delta(\rho-\rho'), \tag{4.3a}$$

$$\sum_{\sigma=\pm 1}\int_0^{\infty} d\rho K^{p,\lambda}_{\sigma}(\phi,\rho) K^{p,\lambda}_{\sigma}(\phi',\rho)^* = \delta(\phi-\phi'). \tag{4.3b}$$

The phase definition we shall impose will stem from the requirement that if $f^{\mathrm{II}}_{\sigma}(\rho)$ is the $K^{p\lambda}$ transform of $f^{\mathrm{I}}(\phi)$ then the $K^{p,\lambda}$ transform of $(\mathbb{J}^{\mathrm{I}}_k f^{\mathrm{I}})(\phi)$ should be $(\mathbb{J}^{\mathrm{II}}_k f^{\mathrm{II}})_{\sigma}(\rho)$, with $\mathbb{J}^{\mathrm{II}}_k$ and $\mathbb{J}^{\mathrm{I}}_k$ given precisely by Eqs. (3.8b) and (1.6b), respectively, supplemented by the discrete transformation $\mathbb{A}$, as imposed by Bargmann's convention[36].

The $\mathbb{J}^{\mathrm{I}}_k$-basis eigenvectors are easy to obtain as they are solutions of first-order differential equations, and Fourier anaysis techniques allow us to find the correct constants for ordinary or Dirac orthonormality in $\mathscr{L}^2(S_1)$. This is more difficult for those of $\mathbb{J}^{\mathrm{II}}_k$ since, as will be borne out below, these are two-component, in general, Whittaker functions whose orthonormality and completeness relations certainly imply a careful analysis. For the parabolic operator in sl(2,R), the simplest one we can choose is

$$\mathbb{J}^{\mathrm{II}}_3 - \mathbb{J}^{\mathrm{II}}_1 = \tfrac{1}{2}\sigma\rho^2, \tag{4.4}$$

since the set of generalized eigenfunctions is readily found as

$$X^{p,\lambda}_{\nu,\sigma}(\rho) = (2|\nu|)^{-1/4}\delta(\rho - [2|\nu|]^{1/2})\delta_{\sigma,\mathrm{sign}\nu}\exp(i\Phi^{\chi}) = \rho^{1/2}\delta(|\nu| - \tfrac{1}{2}\rho^2)\delta_{\sigma,\mathrm{sign}\nu}\exp(i\Phi^{\chi}), \quad \nu\in\mathscr{R}, \tag{4.5}$$

where we have left a phase factor to be determined later on.

Note that $X^{p\lambda}_{\nu}(\rho)$ is a two-component function which has only an upper component for $\nu > 0$ and only a lower one for $\nu < 0$. As they stand, these functions may only involve the representation indices (p,λ), if at all[37], in the phase factor $\Phi^{\chi}(p,\lambda,\nu,\sigma)$.

Now, in $\mathscr{L}^2(S_1)$ the operator corresponding to Eqs. (4.2) for C^{ϵ}_q is

$$\mathbb{J}^{\mathrm{I}}_3 - \mathbb{J}^{\mathrm{I}}_1 = -ie^{-i\epsilon\phi}\left[(1+\cos\phi)\frac{d}{d\phi} - k\sin\phi\right]e^{i\epsilon\phi}, \tag{4.6}$$

with q, k, and λ related as in Eqs. (1.3), and p and ϵ as in Eq. (2.7d). Through the change of variables $\xi = \tan(\phi/2)$ we can find the generalized Dirac-normalized eigenfunctions to be[38]

$$\chi^{p,\lambda}_{\nu}(\phi) = (4\pi)^{-1/2}(\cos(\phi/2))^{-2k}e^{-i\epsilon\phi}\exp[i\nu\tan(\phi/2)]. \tag{4.7}$$

The generating function (4.1) is thus readily calculated from the integral as

$$K^{p,\lambda}_{\sigma}(\phi,\rho) = \rho^{1/2}\chi^{p,\lambda}_{\sigma\rho^2/2}(\phi)\exp[i\Phi^{\chi}(p,\lambda,\sigma\rho^2/2,\sigma)]. \tag{4.8}$$

In order to determine the phase function, consider the orthonormal $\mathscr{L}^2(S_1)$ eigenbasis for $\mathbb{J}^{\mathrm{I}}_3$ in C^{ϵ}_q:

$$\psi^{p,\lambda}_m(\phi) = [\eta^{p,\lambda}_m]^{-1}(2\pi)^{-1/2}\exp[i(m-\epsilon)\phi], \tag{4.9}$$

where m is the integer for $\epsilon = 0$ ($p = +1$) and half-integer for $\epsilon = 1/2$ ($p = -1$). The phase factors $\eta^{p,\lambda}_m$ will be those of Bargmann[36]:

$$\eta^{1,\lambda}_0 = 1 = \eta^{-1,\lambda}_{1/2}, \tag{4.10a}$$

$$\eta^{p,\lambda}_m = (-1)^{m-\epsilon} \times \prod_{l=\epsilon+1/2}^{m-1/2}[(l - i\lambda/2)/(l + i\lambda/2)]^{1/2}, \quad m \geqslant 1, \tag{4.10b}$$

$$\eta^{p,\lambda}_{-m} = (-1)^{m+\epsilon}(i\lambda/|\lambda|)^{2\epsilon}\eta^{p,\lambda}_m, \quad m \geqslant 1/2, \tag{4.10c}$$

where the running index in Eq. (4.10b) takes the $m-\epsilon$ values $l = \epsilon + 1/2, \epsilon + 3/2, \ldots, m - 1/2$. The basis vectors (4.9) of $\mathscr{L}^2(S_1)$ should, upon their transformation to $\mathscr{L}^2_{\mathrm{II}}(\mathscr{R}^+)$, provide the properly normalized eigenbasis for $\mathbb{J}^{\mathrm{II}}_3$. Thus, introducing Eq. (4.9) in (4.2b) with the intertwining (4.8) (with the as yet undetermined phase), we find, under a division of the integration range in two, trigonometric identities and an integration[39], that

$$\Psi^{p,\lambda}_{m,\sigma}(\rho) = \int_{-\pi}^{\pi} d\phi \psi^{p,\lambda}_m(\phi) K^{p,\lambda}_{\sigma}(\phi,\rho)^* = [\eta^{p,\lambda}_m]^{-1}2^{1/2-i\lambda}[\Gamma(k+\sigma m)]^{-1}\rho^{-1/2+i\lambda} \times W_{\sigma m,-i\lambda/2}(\rho^2)\exp[-i\Phi^{\chi}(p,\lambda,\sigma\rho^2/2,\sigma)], \tag{4.11}$$

which is valid for integer as well as half-integer values of m.

Notice that the phase factor cannot depend on m. Now, the (unnormalized) solutions of $\mathbb{J}_3\Psi(\rho) = m\Psi(\rho)$ which are bounded at infinity are of the form[40] $\rho^{-1/2}W_{m,\pm i\lambda/2}(\rho^2)$; the phase factor is thus constrained to be $\rho^{-i\lambda}$ times any other ρ-independent phase. We can set

$$\Phi^{\chi}(p,\lambda,\sigma\rho^2/2,\sigma) = \lambda\ln(\rho/2) \tag{4.12}$$

and declare the proper eigenfunctions of $\mathbb{J}^{\mathrm{II}}_3$ corresponding to the eigenvalue m (integer or half-integer) to be

$$\Psi^{p\lambda}_{m,\sigma}(\rho) = [\eta^{p,\lambda}_m\Gamma(k+\sigma m)]^{-1} \times (\rho/2)^{-1/2}W_{\sigma m,-i\lambda/2}(\rho^2) \tag{4.13}$$

spanning the C^{ϵ}_q irrep for sl(2,R). On Eq. (4.13) we can ver-

ify immediately that we have an eigenfunction of $\mathbb{J}_3^{II}$, as this operator acts as $\sigma\mathbb{J}_3$ [see Eq. (1.1c)] on the two $\sigma = +1$ and $\sigma = -1$ components. Hance the eigenvalue is indeed m. Normalization under the inner product (3.6) can be checked straightforwardly[41]. In order to support our claim that Eq. (4.12) is indeed an appropriate phase, we may verify that the action of $\mathbb{J}_{\pm}^{I} = \mathbb{J}_1^I \pm i\mathbb{J}_2^I$ on the simple functions $\psi_m^{p,\lambda}(\phi)$, namely, $(\eta_m^{p\lambda}/\eta_{m\pm1}^{p\lambda})(k \pm m)\psi_{m\pm1}^{p,\lambda}(\phi)$ is the same as that of $\mathbb{J}_{\pm}^{II}$ on $\Psi_{m,\sigma}^{p,\lambda}(\rho)$. This has to be done separately on the upper and lower components as $\mathbb{J}_{\pm}^{II} = \sigma\mathbb{J}_1 \pm i\mathbb{J}_2$ and yields the same result through the recurrence realtions for Whittaker functions[42]. Finally, the transformation properties under $\mathbb{A}$ in Eqs. (2.9) can be defined explicitly, as their action on $\mathscr{L}_{II}^2(\mathscr{R}^+)$ is to exchange the two component functions

$$\mathbb{A}:\Psi_{m,\sigma}^{p,\lambda}(\rho) = \Psi_{m,-\sigma}^{p,\lambda}(\rho) = (\eta_{-m}^{p,\lambda}/\eta_m^{p,\lambda})\Psi_{-m,\sigma}^{p,\lambda}(\rho), \quad (4.14)$$

and it can be readily seen that $\mathbb{A}^2 = \mathbb{1}$. From Eqs. (4.10), the factor in Eq. (4.14) is $(-1)^m$ for $\epsilon = 0$ and [43] $i(-1)^{m+\epsilon} \times \text{sign}\lambda\,\text{sign}m$ for $\epsilon = 1/2$. The intertwining kernel can thus be written as

$$K_{\sigma}^{p,\lambda}(\phi,\rho) = (2\pi)^{-1/2}e^{-i\epsilon\phi}(\rho/2)^{1/2+i\lambda}(\cos(\phi/2))^{-1-i\lambda} \times \exp(i\tfrac{1}{2}\sigma\rho^2\tan(\phi/2)). \quad (4.15)$$

The generalized eigenfunctions of the parabolic generator $\mathbb{J}_3^{II} + \mathbb{J}_1^{II}$ can be found from those of $\mathbb{J}_3^{II} - \mathbb{J}_1^{II}$ in Eq. (4.5) through the Fourier transformation (3.15) representing a rotation by $-\pi$ around the 3-axis. Since Eqs. (4.5) are essentially Dirac δ's in ρ, the $\mathbb{J}_3^{II} + \mathbb{J}_1^{II}$ eigenfunctions will include $H_{\sigma,\sigma'}^{p,\lambda}([2|\nu|]^{1/2}\rho)$—Hankel and Macdonald functions of imaginary index—times $\rho^{1/2}$. The corresponding $\mathbb{J}_3^I + \mathbb{J}_1^I$ generalized eigenfunctions are obtained from Eq. (4.7) simply by a rotation of π in the argument. These basis functions and their transformation properties are particularly interesting, since from Eqs. (2.1) it can be seen that $\mathbb{M}_{23} + \mathbb{M}_{25}$ is the *Klein-Gordon* operator in a two-dimensional space–time. We reserve some observations pertaining to this subject and the Kontorovich–Lebedev transform for future development.

As a final calculation, let us use the preceding information in order to find the generalized eigenfunctions of the hyperbolic operators $\mathbb{J}_2^{\bullet}$ and $\mathbb{J}_1^{\bullet}$ in $\mathscr{L}_{II}^2(\mathscr{R}^+)$ and $\mathscr{L}^2(S_1)$. The four functions are related by pairs by a rotation by $\pi/2$ over the 3-axis (i.e., the square root of the hyperbolic Fourier transform) and by the intertwining operator. The simplest of the four are the $\mathscr{L}_{II}^2(\mathscr{R}^+)$ generalized eigenfunctions of $\mathbb{J}_2^{II}$ and $\mathbb{A}$:

$$\Upsilon_{\tau,a,\sigma}^{p,\lambda}(\rho) = (2\pi)^{-1/2}(\delta_{\sigma,1} + a\delta_{\sigma,-1})\rho^{-1/2+2i\tau}, \quad (4.16)$$

with eigenvalues $\tau \in \mathscr{R}$ and $a = \pm 1$, respectively. The spectrum of this hyperbolic operator thus covers the real line twice[44]. The functions (4.16) are Dirac orthonormal and complete with respect to Eq. (3.6) as can be ascertained through bilateral Mellin transformation[45]. The corresponding $\mathbb{J}_2^I$ eigenfunctions can be found through Eqs. (4.2a) and (4.15) using the Fourier transform of the complex power functions[46]. Defining the "cut" functions

$$x_+ = \begin{cases} x, & x \geqslant 0 \\ 0 & x < 0 \end{cases}, \quad x_- = \begin{cases} 0, & x \geqslant 0, \\ -x, & x < 0, \end{cases} \quad (4.17a)$$

we find

$$\begin{aligned} \upsilon_{\tau,a}^{p,\lambda}(\phi) &= (4\pi)^{-1}2^{i(\tau-\lambda/2)}\Gamma(\eta)e^{-i\epsilon\phi}(\cos(\phi/2))^{-1-i\lambda} \\ &\quad \times (e^{i\pi\eta/2} + ae^{-\pi\eta/2}) \\ &\quad \times [(\tan(\phi/2))_+^{-\eta} + a(\tan(\phi/2))_-^{-\eta}] \\ &= \pi^{-1}2^{-1/2+i\tau}\Gamma(\eta)e^{-i\epsilon\phi}\,\text{trig}_a(\pi\eta/2) \\ &\quad \times (\delta_{1,\text{sign}\phi} + a\delta_{-1,\text{sign}\phi})|\sin\phi|^{-k}|\tan(\phi/2)|^{-i\tau}, \end{aligned} \quad (4.17b)$$

where $\eta = k + i\tau = \frac{1}{2} + i(\tau + \lambda/2)$. The eigenfunctions of $\mathbb{J}_1^{\bullet}$ can be now found from Eqs. (4.16) and (4.17), as $\mathbb{J}_1^{\bullet} = \exp(i\tfrac{1}{2}\pi\mathbb{J}_3^{\bullet})\mathbb{J}_2^{\bullet}\exp(-i\tfrac{1}{2}\pi\mathbb{J}_3^{\bullet})$. This amounts to a rotation by $\pi/2$ in S_1:

$$\omega_{\tau,a}^{p,\lambda}(\phi) = \upsilon_{\tau,a}^{p,\lambda}(\phi + \pi/2). \quad (4.18)$$

In $\mathscr{L}_{II}^2(\mathscr{R}^+)$ this is the $2^{-1/2}\begin{pmatrix}1 & -1\\ 1 & 1\end{pmatrix}$ transform (3.9) of the chosen $(\mathbb{J}_2^{II},\mathbb{A})$ eigenfunctions (4.16). The hyperbolic canonical transform involves three integrals for each component[47]. After several cancellations and factorizations, we obtain

$$\Omega_{\tau,a,\sigma}^{p,\lambda}(\rho) = C_{\tau}^{\lambda}V_{a,\sigma}^{p,\lambda}e^{-\sigma\pi\tau/2}\rho^{-1/2}W_{-i\tau,i\lambda/2}(i\sigma\rho^2), \quad (4.19a)$$

$$C_{\tau}^{\lambda} = 2^{1/2+i\tau}(2\pi)^{-3/2}\Gamma(\tfrac{1}{2} + i\tau + i\tfrac{1}{2}\lambda)\Gamma(\tfrac{1}{2} + i\tau - i\tfrac{1}{2}\lambda), \quad (4.19b)$$

$$V_{a,\sigma}^{p,\lambda} = (\delta_{\sigma,1} + ap\delta_{\sigma,-1})[2a\,\text{hyp}_{1,p}(\lambda\pi/2) - i\tfrac{1}{2}(p-1)], \quad (4.19c)$$

which are indeed eigenfunctions of $\mathbb{J}_1^{II}$ with eigenvalue τ. They are not eigenfunctions of $\mathbb{A}$, or course; rather, $\mathbb{A}$ can be seen to map Eq. (4.19a) through $\sigma \to -\sigma$ into an eigenfunction of $\mathbb{J}_1^{II}$ with eigenvalue $-\tau$. As $W_{\mu,\nu}(z)$ and $W_{-\mu,\nu}(-z)$ are independent solutions to the Whittaker equation, their relation is not simple. In fact,

$$\begin{aligned} \mathbb{A}\Omega_{\tau,a}^{p,\lambda} &= \mathbb{A}\exp(i\tfrac{1}{2}\pi\mathbb{J}_3^{II})\Upsilon_{\tau,a}^{p,\lambda} \\ &= \exp(-i\tfrac{1}{2}\pi\mathbb{J}_3^{II})\mathbb{A}\Upsilon_{\tau,a}^{p,\lambda} = a\mathbb{C}_F^{p,\lambda}\Omega_{\tau,a}^{p,\lambda}, \end{aligned} \quad (4.20)$$

where $\mathbb{C}_F^{p,\lambda}$ is the hyperbolic Fourier transform as given by Eq. (3.15).

5. APPLICATIONS AND CONCLUSION

The analytic properties of the basis functions and transformations belonging to the continuous series of the SL(2,R) group generated by Eqs. (1.1) have been seen to be rather arduous. Their group-theoretic properties are, however, as simple as that of any other realization, and herein lies the advantage of using the latter to derive relations for the former. These relations take the form of integral identities involving Hankel, Macdonald, Whittaker, power, and exponential functions, some with imaginary indices and parameters, which are now endowed with a group-theoretic interpretation. In what follows, we outline five examples of applications of these concepts.

First, of course, we have the Hankel and Macdonald function integral relations implicit in the kernel composition (3.13). Second, Whittaker functions of the kind (4.13) and (4.19) are displayed as being *self-reciprocating*[48] under hyperbolic canonical transforms. This can be seen in the following way: Consider the matrix identity

$$\mathbf{M}=\begin{pmatrix} a & b \\ c & d \end{pmatrix}=\begin{pmatrix} \alpha & 0 \\ \gamma & \alpha^{-1} \end{pmatrix}\begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}, \quad a\neq 0, \tag{5.1a}$$

$$\alpha=(a^2+b^2)^{1/2}, \quad \gamma=(ac+bd)/\alpha, \quad \tan t=-b/a. \tag{5.1b}$$

The integral transforms associated to these matrices will follow suit through Eq. (3.3). Now, apply these transforms to the $\mathbb{J}_3^{\mathrm{II}}$ eigenfunctions $\Psi_m^{p,\lambda}(\rho)$ in Eq. (4.13), noting that the rightmost transform will multiply the functions by $\exp(2imt)$, while the second transform is purely geometric and given by Eq. (3.12). Their composition thus leads to the integral relation

$$\sum_{\sigma'=\pm 1}\int_0^\infty d\rho' C^{p,\lambda}_{\mathbf{M},\sigma,\sigma'}(\rho,\rho')\Psi^{p,\lambda}_{m,\sigma'}(\rho')=|\alpha|^{-1/2}(\mathrm{sign}\alpha)^{2\epsilon} \times e^{2imt}\exp(i\sigma\gamma\rho^2/2\alpha)\Psi^{p,\lambda}_{m,\sigma}(\rho/|\alpha|), \tag{5.2}$$

which, if written out explicitly [Eqs. (2.7d), (3.9)–(3.11), (4.13), and (5.1)], is rather difficult to solve by elementary methods. Decompositions analogous to Eqs. (5.1) can be made for the parabolic- and hyperbolic-operator eigenfunctions seen in Sec. 4.

Third, the intertwining operator (4.2) can be used to "close the fourth side of a rectangle" in applying a hyperbolic canonical transform to a given function in $\mathscr{L}^2_{\mathrm{II}}(\mathscr{R}^+)$: we pass to $\mathscr{L}^2(S_1)$, transform the function there [this is an easy task since the group SL(2,R) in that space acts geometrically as its generators are of first order], and transform back to $\mathscr{L}^2_{\mathrm{II}}(\mathscr{R}^+)$. Fourth, the intertwining integral may be solved if the functions involved are recognized to be canonical transforms of eigenfunctions of SL(2,R) generators. We use formulas such as Eq. (5.2) in order to transform them to the simplest eigenfunction of the orbit such as Eq. (4.5) for the parabolic and Eq. (4.16) for the hyperbolic cases, intertwine the resulting simpler function with the aid of the results of Sec. 4, and transform back in $\mathscr{L}^2(S_1)$. Fifth, $\mathscr{L}^2_{\mathrm{II}}(\mathscr{R}^+)$ inner products between basis functions such as the right-hand side of Eq. (5.2) may be intertwined to their $\mathscr{L}^2(S_1)$ counterparts and the simpler ϕ — integral solved. The latter is nothing more than a SL(2,R) representation matrix element (same or mixed basis) and thus expressible in terms of ${}_2F_1$ hypergeometric functions[49].

From the point of view of canonical transformations in quantum mechanics, we have been occupied with potentials which are not realistic. Our approach, however, suggests that any other classical-quantum correspondence method of solution[17] tackling Eq. (1.7) should, when extended to strongly centripetal potentials, lead to the results in this article.

As regards SL(2,R) representation theory, only the supplementary series ($0<q<\frac{1}{4}$) remains to be worked out, in particular, the peculiar properties of the representations at the values $q=0$ and $\frac{1}{4}$ of the Casimir operator.

Finally, on the terrain of the integral transform theory, we have previously shown that[5–7] Fourier and Hankel transforms are particular cases of real linear and radial canonical transforms and that, through complex extension, one can reach the bilateral Laplace, Gauss–Weierstrass, Bargmann, and Barut–Girardello transforms. Hyperbolic canonical transforms do not seem to include any well-known particular cases, yet they come within close range: The Meijer-K, Kontorovich–Lebedev, and Neumann transforms[50]. The first ones, involving kernels with Macdonald functions of real index and related to the Laplace transform, may be reached if a valid analytic continuation of the kernel can be implemented. This may require nonunitary SO(1,1) representations in Eq. (2.5). The second transform involves Hankel functions of imaginary index, where the integrations take place on the argument and on the index. This seems to require either a different subgroup reduction of Sp(4,R) or operators other than Eq. (3.5) in the representation decomposition. As both of these cases involve single-component functions, we surmise that they correspond to the $\mathbb{A}$-diagonal Fourier transform (3.15). Lastly, Neumann transforms—and, indeed, Hankel transforms as well—are suggested by the analytic continuation in λ of the kernel elements (3.11a), as even the Struve function contained in the inverse Neumann transform appears to be closely related[51] to the use of the representations of a compact subgroup. It is our intention to address these extensions and further the study of the Klein–Gordon operator elsewhere.

ACKNOWLEDGMENTS

I would like to thank Professor James D. Finley (University of New Mexico) for a conversation on the Klein–Gordon operator Green function which brought to my attention the subject of this article. It is a pleasure to acknowledge the hospitality of the Centre de Recherches Mathématiques and the Université de Montréal, where this work was developed, and Dr. Marcel Perroud for his interest in this subject.

[1]M. Moshinsky and C. Quesne, J. Math. Phys. **12**, 1772 (1971).
[2]M. Moshinsky and C. Quesne, "Oscillator Systems," in *Proceedings of the* 15th *Solvay Conference in Physics* (1974) (Gordon and Breach, New York, 1974); M. Moshinsky, SIAM J. Appl. Math. **25**, 193 (1973).
[3]M. Moshinsky, T.H. Seligman, and K.B. Wolf, J. Math. Phys. **13**, 901 (1972).
[4]K.B. Wolf, J. Math. Phys. **17**, 602 (1976); S. Steinberg and K.B. Wolf, Technical Report No. 351, Dept. of Mathematics and Statistics, University of New Mexico (1978), to be published in Nuovo Cimento.
[5]K.B. Wolf, J. Math. Phys. **15**, 1295 (1974).
[6]K.B. Wolf, J. Math. Phys. **15**, 2102 (1974).
[7]K.B. Wolf, *Integral Transforms in Science and Engineering* (Plenum, New York, 1979).
[8]In this regard, it should be noted that this problem has been tackled independently by other authors, notably L. Infeld and J. Plebański, Acta Phys. Polon. **14**, 41 (1955); A. Weil, Acta Math. **111**, 143 (1963) calls this the *metaplectic* representation of the symplectic group; C. Itzykson, J. Math. Phys. **10**, 1109 (1969); G. Burdet, M. Perrin, and M. Perroud, Commun. Math. Phys. **58**, 241 (1978).
[9]D. Basu, J. Math. Phys. **19**, 1667 (1978).
[10]N. Mukunda and B. Radhakrishnan, J. Math. Phys. **15**, 1320, 1332, 1643, 1656 (1974).
[11]C.P. Boyer and K.B. Wolf, J. Math. Phys. **16**, 1493 (1975).
[12]V. Bargmann, Ann. Math. **48**, 568 (1947); see also L. Pukański, Math. Ann. **156**, 96 (1964) and P.J. Sally Jr., Proc. R. Soc. (London) Ser. A **287**, 532 (1967).
[13]Reference 12, Sec. 5g.1 and Appendix.

[14]P. Kramer, M. Moshinsky, and T.H. Seligman, "Complex Extensions of Canonical Transformations in Quantum Mechanics," in *Group Theory and its Applications*. III, edited by E.M. Loebl (Academic, New York, 1975).

[15]C.P. Boyer, Comunicaciones Técnicas IIMAS, UNAM, México D.F., (1976, unpublished).

[16]A different approach to the intertwining operators between operators whose $\mathscr{L}^2$ spectra are different has been proposed by M. Moshinsky and T.H. Seligman, Ann. Phys. **114**, 243 (1978), and hinges on the examination of the "ambiguity group" of the classical canonical transformation. See also M. Moshinsky, Proceedings of the VII international Colloquium on Group Theory and its Applications, Austin Texas, September 1978), Lecture Notes in Physics, Vol. 94, p. 164 (Springer, Berlin, 1979).

[17]P.A. Mello and M. Moshinsky, J. Math. Phys. **16**, 2017 (1975).

[18]E.C. Titchmarsh, *Eigenfunction Expansions* (Oxford U.P., London, 1962).

[19]C.P. Boyer and K.B. Wolf, Rev. Mex. Fís. **25**, 31 (1976).

[20]M. Brunet and P. Kramer, *Proceedings of the* IV *International Colloquium on Group Theory and its Applications (Nijmegen*, 1975) (Springer, Berlin, 1976).

[21]Reference 4, Eqs. (2.5); Ref. 14, Eq. (3.24); the specific one-parameter subgroups can be related to their generators as in S. Steinberg and K.B. Wolf, Ref. 4.

[22]These results can be obtained from the analysis of $SL(2,R)_1 \otimes SL(2,R)_2$ on the two Cartesian coordinates. Our introduction of $Sp(4,R)$ is primarily for the purpose of exhibiting its various subgroup reductions, related integral transforms and automorphisms of $SL(2,R)$ which are inner in $Sp(4,R)$.

[23]The specification of the phase of the 2–2 element of the Pauli matrix insures that the q_2 factor in the transform has the appropriate phase. It is otherwise imposed by the inversion of (1.4a) and (1.4b) subgroups and the phase consistency of this operation. See Ref. 7, p. 389, where negative-b elements of a real 2×2 matrix must be set to have phase $-\pi$.

[24]The two ray representation signs in $SL(2,R)_1 \otimes SL(2,R)_2$ cancel to unity. This property is also present in the case of radial canonical transforms which include a single D_k^+ irrep.

[25]I.S. Gradshteyn and I.M. Ryzhik, *Tables of Integrals, Sums and Products* (Academic, New York, 1965).

[26]Reference 25, Eqs. 3.996,1,2,4 and 5.

[27]Reference 25, Eqs. 8.451,3,4 and 6, and the limit of oscillating Gaussians, Ref. 7, Eq. (9.20).

[28]When we consider the three one-parameter subgroups (1.4), the integral kernel (3.11) and (3.12) is the Green's function of N. Mukunda and B. Radhakrishnan, J. Math. Phys. **13**, 254 (1973).

[29]This assertion may be verified independently using Eqs. 8.407.1,2 and 8.476.8,9 of Ref. 25 to convert Hankel to Macdonald functions of imaginary argument, and for the integral (Eq. 6.521.3) of two Macdonald functions. Upon a limiting argument, one finds the $\mathbf{M} = \mathbf{1}, -\mathbf{1}$ transform reproducing kernel.

[30]Since no continuous Heisenberg–Weyl group can be constructed on $\mathscr{L}^2(S_1)$. See K.B. Wolf, "The Heisenberg–Weyl ring in Quantum Mechanics," in *Group Theory and its Applications*. III, edited by F.M. Loebl (Academic, New York, 1975), Sec. VI, and also Ref. 11.

[31]The choice of Basu, Mukunda, and Radhakrishnan (Refs. 9 and 10) was $\mathbb{J}_2^{\text{II}}$, leading to transforms of the Mellin type.

[32]V. Bargmann, Ref. 12, Appendix.

[33]Mello and Moshinsky (Ref. 17) also comment on the definition of a phase for their canonical transformation kernel.

[34]This corresponds to the subgroup of lower-triangular matrices in Eq. (3.12) with $c = \gamma$ and $a = 1 = d, b = 0$.

[35]The new diagonal operator can be found from Eq. (1.5) with the parameters of Ref. 34.

[36]V. Bargmann, Ref. 12, Eqs. (6.23) and (7.10). The phase factors are chosen so that the raising operator, his Eq. (6.26), has only positive matrix elements.

[37]This is the case for the $\mathbb{J}_2^{\text{II}}$ eigenfunctions. Mukunda and Radhakrishnan devote a lengthier commentary on this fact; see Ref. 10, p. 1323.

[38]In writing Eq. (4.7) we are proposing a definite choice of phase. As far as the construction of the intertwining kernel is concerned, all phases may be ascribed to eigenfunctions (4.5).

[39]Reference 25, Eq. 3.718.6.

[40]Reference 25, Secs. 9.22 and 9.23.

[41]Reference 25, Eqs. 7.6114, 8.334.2, and 8.365.9.

[42]*Handbook of Mathematical Functions*, edited by M. Abramowitz and I. Stegun (National Bureau of Standards, Washington, D.C., 1964), Eqs. 13.4.31 and 13.4.33.

[43]It should be noted that the phase appearing in Ref. 10, I, Eq. (1.5) seems to conflict with the fact that $\eta^{\rho,\lambda}{}_m/\eta^{\rho,\lambda}_m$ is an imaginary quantity for $C_q^{1/2}$ in Bargmann's convention; see Ref. 36.

[44]They are eigenfunctions selected in Ref. 10, I, Eq. (1.20).

[45]See, for example, Ref. 7, Sec. 8.2.2.

[46]I.M. Gel'fand *et al.*, *Generalized Functions* (Academic, New York, 1964), Vol. I, Sec. 4.4; see also Ref. 7, Sec. 7.5.13.

[47]Hankel functions are converted to Macdonald ones of imaginary argument through Eqs. 8.407 and 8.476.8 of Ref. 25, and the integrals performed through Eq. 6.631.3 for the principal sheet of α. For $p = +1$ only the off-diagonal kernel elements contribute.

[48]On self-reciprocating functions under linear and radial canonical transforms, see K.B. Wolf, J. Math. Phys. **18**, 1046 (1977).

[49]In fact, for the $D_k^{\pm}$ series, the $\mathscr{L}^2(\mathscr{R}^+)$ integrals are easier to perform and have been used to find all same and mixed-basis representation matrix elements in Ref. 19.

[50]I.N. Sneddon, *The Use of Integral Transforms* (McGraw-Hill, New York, 1972).

[51]Reference 42, Eq. 12.1.7.

Classification of real simple Lie superalgebras of classical type

Monique Parker
Department of Mathematics, Université Libre de Bruxelles, Campus Plaine C.P. 218, B-1050 Brussels, Belgium

(Received 18 July 1979; accepted for publication 15 November 1979)

Finite-dimensional simple Lie superalgebras (also called $\mathbb{Z}_2$-graded Lie algebras) over an algebraically closed field of characteristic zero were classified in 1976. All simple Lie superalgebras over the reals, whose Lie subalgebra is reductive, are determined here up to isomorphism. As is the theory of simple Lie algebras, this is done by classifying the involutive semimorphisms of the complex Lie superalgebras. One sees in particular that the real form of the Lie subalgebra completely determines the real form of the Lie superalgebra.

INTRODUCTION

In the last 15 years, algebraic structures called Lie superalgebras or graded Lie algebras have appeared in various contexts, e.g., in particle physics and in deformation theory. For a review of the subject we refer to Ref. 1.

Soon there were attempts to develop a structure theory and special interest was devoted to the properties and classification of the simple Lie superalgebras over an algebraically closed field of characteristic zero. These appear to be of two types according to whether the Lie subalgebra is reductive or not. The Lie superalgebras for which this is the case are called classical and it turns out that all those encountered in physics are, after complexification, of that type (eventually after a Wigner contraction). In the present paper, we give a classification of the real classical Lie superalgebras (theorem 2.5); this inserts the examples known from the physical literature or from Ref. 2 in a general setting.

1. REAL AND COMPLEX LIE SUPERALGEBRAS

Definition 1.1: A *Lie superalgebra* G is an algebra graded over $\mathbb{Z}_2$, i.e., G is a direct sum of vector spaces $G = G_0 \oplus G_1$, and such that the bracket satisfies

(i) $[G_i, G_j] \subset G_{i+j(\mathrm{mod}2)}$,

(ii) $[x_i, y_j] = -(-1)^{ij}[y_j, x_i]$, for all $x_i \in G_i$, $y_j \in G_j$,

(iii) $[x_i, [y_j, z]] = [[x_i, y_j], z] + (-1)^{ij}[y_j, [x_i, z]]$.

In what follows, G is an algebra over a field K of characteristic zero and is finite dimensional. Moreover, we will always assume $G_0 \neq 0 \neq G_1$.

Definition 1.2: A Lie superalgebra G is said to be *classical if* G is simple, i.e., does not contain any nontrivial graded ideal and if the representation of G_0 on G_1 is completely reducible.

Definition 1.3: If G is a complex Lie superalgebra, a *semimorphism* C of G is a semilinear transformation of G such that $[Cx, Cy] = C[x, y]$ for all $x, y \in G$. All homomorphisms and semimorphisms of Lie superalgebras will be assumed to preserve G_0 and G_1.

Proposition 1.4: Let G be a complex classical Lie superalgebra and let C be an involutive semimorphism of G. Then $G_C = \{x + Cx \mid x \in G\}$ is a real classical Lie superalgebra.

Proof: One sees immediately that G_C is a real simple Lie superalgebra, the complexification of which is G. We need to show that the representation of $G_{0C} = \{x + Cx \mid x \in G_0\}$ on $G_{1C} = \{x + Cx \mid x \in G_1\}$ is completely reducible. Let V be a subspace of G_{1C} which is invariant by G_{0C}. Then $\tilde{V} = V \otimes \mathbb{C}$ is invariant by G_0; hence, there exists a subspace W' supplementary to $\tilde{V}$ in G_1 and invariant by G_0. The subspace $W = (I + C)\{w' \in W' \mid (I - C)\, W' \in iV\}$ of G_{1C} is supplementary to V and invariant by G_{0C}.

To see this, notice first that $C(I + C) = C + I$ implies W included in G_{1C}. Second, if $g \in G_{0C}$, $[g, w] = [g, (I + C)w']$ where $(I - C)w' \in iV$, hence $[g, w] = (I + C)[g, w']$ where $(I - C)[g, w'] = [g, (I - C)w']$ belongs to iV and thus W is invariant by G_{0C}. Finally, W is supplementary to V for if $w = (I + C)w' \in V$, one has $2w' = (I + C)w' + (I - C)w' \in V + iV = \tilde{V}$; hence, $w' = 0$ and thus $W \cap V = 0$. On the other hand, if $x \in G_{1C}$, one has $x = w' + v'$ where $w' \in W'$ and $v' \in \tilde{V}$. However, $(I - C)x = 0$ implies $(I - C)w' = (I - C)(-v') \in iV$ and thus $(I + C)w' \in W$; hence, $2x = (I + C)x = (I + X)w' + (I + C)v' \in W + V$.

Proposition 1.5: If G is a real classical Lie superalgebra, its complexification $\tilde{G} = G \otimes \mathbb{C}$ is a Lie superalgebra which is either classical or the direct sum of two isomorphic ideals which are classical.

Proof: Let C be the conjugation in $\tilde{G}$ with respect to G. We show first that the representation of $\tilde{G}_0 = G_0 \otimes \mathbb{C}$ on $\tilde{G}_1 = G_1 \otimes \mathbb{C}$ is completely reducible. Let indeed V be a complex subspace of $\tilde{G}_1$ which is invariant by $\tilde{G}_0$. Hence, V is invariant by G_0 and so $V' = (I + C)V$ is real and invariant by G_0. Thus, there exists a subspace W' supplementary to V' in G_1 and invariant by G_0. Then $\tilde{G}_1 = \tilde{V}' \oplus \tilde{W}'$ and $\tilde{W}'$ is invariant by $\tilde{G}_0$, which proves the first point. If $\tilde{G}$ is not simple, it contains a simple graded ideal S. Then $(I + C)S$ is a graded ideal of G so either $(I + C)S = 0$ or $(I + C)S = G$. However, $(I + C)S = 0$ is impossible since $s + Cs = 0$ implies $is + C(is) \neq 0$. Hence, $(I + C)S = G$ and $\tilde{G} = G + iG = (I + C)S + (I - C)S = S + CS$. Since $S \cap CS$ is an ideal of S, we have $S \cap CS = 0$, which shows that $\tilde{G}$ is the direct sum of the two ideals S and CS.

Proposition 1.6: Let G be a complex Lie superalgebra and let C and C' be two involutive semimorphisms of G. The real forms G_C and $G_{C'}$ are isomorphic if and only if there exists an automorphism φ of G such that $C' = \varphi C \varphi^{-1}$.

 0022-2488/80/040689-09$1.00

Proof: If $C' = \varphi C \varphi^{-1}$, it is clear that $G_C = \varphi G_{C'}$. Conversely, assume there exists an isomorphism θ from G_C onto $G_{C'}$. The linear extension φ of θ to $G = G_C + iG_C$ defined by $\varphi(g + ig) = \theta g + i\theta g$ is an automorphism of G. Moreover, if $g' = \theta g \in G_{C'}$ we have $C'g' = g' = \theta g = \varphi g = \varphi C g = \varphi C \varphi^{-1} g'$; hence, $C' \varphi C \varphi^{-1}$ is the identity on $G_{C'}$ and thus also on G.

The above propositions show that any real classical Lie superalgebra G is obtained by one of the following two procedures: (a) If the complexification of G is not simple, G is a complex classical Lie superalgebra considered as real algebra. (b) If the complexification $\tilde{G}$ of G is simple, G is the subalgebra of fixed points of an involutive semimorphism of $\tilde{G}$. The complex classical Lie superalgebras are known.[3,4] Classifying the real classical Lie superalgebras is thus equivalent to classifying the involutive semimorphisms of each complex Lie superalgebra G in the group of automorphisms of G.

Some properties of the complex classical Lie superalgebras: Proofs of the following results are given in Refs. 3–5.

(a) There exist six families of complex classical Lie superalgebras with unbounded dimension, denoted by $A(m,n)$, $d(n)$, $b(n)$, $B(m,n)$, $C(m)$, and $D(m,n)$, and three isolated algebras denoted by $D(1,2,\alpha)$, $G(3)$, and $F(4)$ of respective dimension 17, 31, and 40.

(b) Let $G = G_0 + G_1$ be a complex classical Lie superalgebra, and let ρ denote the representation of G_0 on G_1. Then $[G_0,G_1] = G_1$, $[G_1,G_1] = G_0$, G_0 is reductive, and ρ is faithful.

Either ρ is irreducible or G_1 is the direct sum of two irreducible subspaces Y' and Y'' such that $[Y',Y'] = [Y'',Y''] = 0$ and $[Y',Y''] = G_0$; this decomposition $G_1 = Y' \oplus Y''$ is unique.

If G_0 is not semisimple, ρ is reducible and the center of G_0 is one dimensional; it contains an element k_0 such that $\rho(k_0)\, y' = y'$ and $\rho(k_0)\, y'' = -y''$ for all $y' \in Y', y'' \in Y''$.

(c) A bilinear form B on G is said to be *invariant if* $B([x,y],z) = B(x,[y,z])$ for all $x,y,z \in G$.

If G is simple, any invariant bilinear form is nondegenerate or zero. Moreover, we will always assume $B(G_0,G_1) = 0$.

All complex classical Lie superalgebras admit a nondegenerate invariant bilinear form such that $B(G_0,G_1) = 0$, except $d(n)$ and $b(n)$.

2. METHOD TO CLASSIFY THE INVOLUTIVE SEMIMORPHISMS

(A) If C (or φ) is a semimorphism (homomorphism) of G, we will denote by C_0 and C_1 (φ_0 and φ_1, respectively) its restriction to G_0 and G_1. If C is an involutive semimorphism of G, then C_0 is an involutive semimorphism of G_0; hence G_{0C} is a real form of G_0. The following proposition shows that G_{0C} may be chosen up to isomorphism:

Proposition 2.1: Let G be a complex classical Lie superalgebra and let $C = C_0 + C_1$ be an involutive semimorphism of G. Assume C_0' is an involutive semimorphism of G_0 conjugate to C_0 in $\mathrm{Aut}G_0$. Then there exists an involutive semimorphism $C' = C_0' + C_1'$ of G which is conjugate to C in $\mathrm{Aut}G$.

Proof: Assume $C_0' = \varphi_0 C_0 \varphi_0^{-1}$, where $\varphi_0 \in \mathrm{Aut}G_0$. If there exists $\varphi = \varphi_0 + \varphi_1 \in \mathrm{Aut}G$, then $C' = \varphi C \varphi^{-1}$ is an involutive semimorphism of G and the proposition is proved. The existence of φ will follow from Lemma 2.2 when φ_0 is inner and G is not $d(n)$ or $b(n)$. In the other cases, φ will be explicitly constructed.

Lemma 2.2: If G is not $d(n)$ or $b(n)$ and φ_0 is an inner automorphism of G_0, there exists an automorphism $\varphi = \varphi_0 + \varphi_1$ of G.

Proof: Let ρ denote the representation of G_0 on G_1 and let B be any nondegenerate invariant bilinear form on G. If $\varphi_0 = e^{ad(n)}, \varphi_1 = e^{\rho(n)}$ satisfies $\rho(\varphi_0 g)\varphi_1 = \varphi_1 \rho(g)$ for all $g \in G_0$. On the center of G_0, φ_0 is the identity. On each of the simple ideals of G_0, the bilinear form B is a multiple of the Killing form; hence, φ_0 is an isometry for B. On the other hand, $B(\rho(n)\, x,y) + B(x,\rho(n)\, y) = 0$ for all $x,y \in G_1$ implies that φ_1 is also an isometry for B. We then have for all $g \in G_0$

$$\begin{aligned} B([\varphi_1 x,\varphi_1 y],\varphi_0 g) &= B(\varphi_1 x,[\varphi_1 y,\varphi_0 g]) \\ &= B(\varphi_1 x,\varphi_1[y,g]) \\ &= B(x,[y,g]) = B([x,y],g) \\ &= B(\varphi_0[x,y],\varphi_0 g), \end{aligned}$$

and thus $[\varphi_1 x,\varphi_1 y] = \varphi_0[x,y]$, which completes the proof.

(B) The next lemmas show that, for a given real form G_{0C} of the Lie subalgebra G_0, there exists, up to isomorphism, at most two real forms G_C which contain G_{0C}. In fact, we will show case by case that these two real forms are isomorphic. Let $C = C_0 + C_1$ and $C' = C_0 + C_1'$ be two involutive semimorphisms of G having the same restriction to G_0.

Lemma 2.3: If the representation ρ of G_0 on G_1 is irreducible, then $C_1' = \pm C_1$.

Proof. The linear transformation $C_1 C_1'$ of G_1 commutes with $\rho(G_0)$; hence, $C_1' = \lambda C_1$ and $C_1^2 = C_1'^2 = I$ implies $\lambda\bar{\lambda} = 1$. If $x,y \in G_1$, we have $C_0[x,y] = [C_1' x, C_1' y] = \lambda^2 [C_1 x, C_1 y] = \lambda^2 C_0[x,y]$ and thus $\lambda^2 = 1$.

If the representation ρ of G_0 on G_1 is reducible, we write $G_1 = Y' \oplus Y''$ for the sum of the invariant subspaces and if G_0 is not semisimple, we denote by k_0 the element of the center of G_0 such that $\rho(k_0)|_{Y'} = I$ and $\rho(k_0)|_{Y''} = -I$.

Lemma 2.4: (i) If C_1 preserves Y' and Y'', $C_0 k_0 = k_0$ and if C_1 permutes Y' and Y'', $C_0 k_0 = -k_0$. (ii) If $C = C_0 + C_1$ and $C' = C_0 + C_1'$ preserve Y' and Y'', they are conjugate in $\mathrm{Aut}G$. (iii) If they permute Y' and Y'', then C' is conjugate in $\mathrm{Aut}G$ to $C_0 + C_1$ or to $C_0 - C_1$.

Proof: (i) We first notice that since the decomposition $G_1 = Y' \oplus Y''$ is unique, any semimorphism of G either preserves or permutes Y' and Y''. If G_0 is not semisimple, we set $C_0 k_0 = a k_0$, where $a\bar{a} = 1$. Then $[C_0 k_0, C_1 y] = C_1[k_0, y] = C_1 y$ if $y \in Y'$ implies $a = 1$ if $C_1 Y' = Y'$ and $a = -1$ if $C_1 Y' = Y''$. (ii) We set $C_1 = C' + C''$, where $C' = C_1|_{Y'}$ and $C'' = C_1|_{Y''}$. By the same argument as in Lemma 2.3, we know that C' and C'' are unique up to a factor of modulus 1, so we may write $C_1' = \lambda C' + \mu C''$. If $x \in Y'$ and $y \in Y''$, we have $C_0[x,y] = [\lambda C' x, \mu C'' y] = \lambda\mu [C' x, C'' y] = \lambda\mu C_0[x,y]$ and thus $\lambda\mu = 1$. However, the linear transformation ψ defined by $\psi y' = \lambda^{1/2} y'$, $\psi y'' = \lambda^{-1/2} y''$, and $\psi g_0 = g_0$ is an automorphism of G and $C_0 + C_1' = \psi(C_0 + C_1)\psi^{-1}$. (iii) Since

$C_1 C_1'$ preserves Y' and Y'' and commutes with $\rho(G_0)$, we may write $C_1' x = \lambda C_1 x$ and $C_1' y = \mu C_1 y$ for all $x \in Y'$ and $y \in Y''$. From $C_1^2 = C_1'^2 = I$ we deduce $\bar{\lambda}\mu = 1$ and $C_0[x,y] = [C_1' x, C_1' y]$ implies as above $\lambda\mu = 1$. We define an automorphism ψ of G by $\psi y' = \lambda^{-1/2} y'$, $\psi y'' = \lambda^{1/2} y''$, and $\psi g_0 = g_0$. If the real number λ is positive, we have $C_0 + C_1' = \psi(C_0 + C_1)\psi^{-1}$ and if λ is negative, we have $C_0 + C_1' = \psi(C_0 - C_1)\psi^{-1}$.

(C) We recall some notions related to representations of complex semisimple Lie algebras.

Let L be a complex semisimple Lie algebra and let ρ be a representation of L in a vector space V. If L_C is a real form of L defined by the involutive semimorphism C, the representation ρ is said to be *real* for L_C if there exists a semilinear transformation C_1 of V such that

$$\rho(Ca)C_1 = C_1\rho(a), \quad \text{for all } a \in L \tag{1}$$

and $C_1^2 = I$.

If there exists C_1 satisfying Eq. (1) and such that $C_1^2 = -I$, ρ is said to be *antireal*. If any C_1 satisfying Eq. (1) is singular, ρ is said to be *areal*. If $L = \oplus_i L_i$ is a direct sum of simple ideals, any irreducible representation ρ of L is a tensor product $\rho = \otimes_i \rho_i$, where ρ_i is an irreducible representation of L_i. If $CL_i = L_i$ for all i, then $\rho = \otimes_i \rho_i$ is real for L_C if and only if an even number of ρ_i are antireal for the real forms L_{iC} and the remaining ρ_j are real for the real forms L_{jC}.

The isomorphism type of the simple complex Lie algebras will be denoted by upper case script letters: $\mathscr{A}_l$, $\mathscr{B}_l$, $\mathscr{C}_l$, $\mathscr{D}_l$, $\mathscr{G}_2$. The real forms of these Lie algebras are denoted as in Ref. 6.

For each of the complex classical Lie superalgebras we will give a definition; the possible real forms of G_0, using the condition that the representation ρ must be real for G_{0C}; one involutive semimorphism C extending the chosen C_0; in case Lemma 2.2 does not apply, an argument to prove that C_0 may be chosen up to conjugacy by $\mathrm{Aut}G_0$; and an automorphism $\varphi = \varphi_0 + \varphi_1$ of G such that $\varphi_0 C_0 = C_0 \varphi_0$ and $\varphi_1 C_1 = -C_1 \varphi_1$ in order to prove that the real forms defined by $C_0 + C_1$ and $C_0 - C_1$ are isomorphic. This will lead to the following conclusion:

Theorem 2.5: Up to isomorphism, the real forms of the classical Lie superalgebras are uniquely determined by the real form G_{0C} of the Lie subalgebra G_0, and G_{0C} is one of the following:

$A(m,n)$ $\quad G_{0C} = \mathrm{Sl}(m,\mathbb{R}) \oplus \mathrm{Sl}(n,\mathbb{R}) \oplus \mathbb{R}$,
$G_{0C} = \mathrm{Su}^*(m) \oplus \mathrm{Su}^*(n) \oplus \mathbb{R}$, if m and n are even,
$G_{0C} = \mathrm{Su}(p,m-p) \oplus \mathrm{Su}(r,n-r) \oplus i\mathbb{R}$,
$G_{0C} = \mathrm{Sl}(n,\mathbb{C})$, if $m = n$;

$d(n)$: $\quad G_{0C} = \mathrm{Su}(p,n-p)$,
$G_{0C} = \mathrm{Su}^*(n)$, if n is even
$G_{0C} = \mathrm{Sl}(n,\mathbb{R})$;

$b(n)$: $\quad G_{0C} = \mathrm{Su}(n)$, if n is even,
$G_{0C} = \mathrm{Sl}(n,\mathbb{R})$;

$B(m,0)$: $\quad G_{0C} = \mathrm{Sp}(m,\mathbb{R})$;

$B(m,n)$ with $n > 0$, or $D(m,n)$:
$G_{0C} = \mathrm{Sp}(m,\mathbb{R}) \oplus \mathrm{So}(p,q)$,
$G_{0C} = \mathrm{Sp}(r,s) \oplus \mathrm{So}^*(2p)$;

$C(m)$: $\quad G_{0C} = \mathrm{Sp}(m,\mathbb{R}) \oplus \mathrm{So}(2)$,
$G_{0C} = \mathrm{Sp}(r,s) \oplus \mathrm{So}(2)$;

$D(1,2,\alpha)$: $\quad G_{0C} = \mathrm{Sl}(2,\mathbb{R}) \oplus \mathrm{Sl}(2,\mathbb{R}) \oplus \mathrm{Sl}(2,\mathbb{R})$,
$G_{0C} = \mathrm{Su}(2) \oplus \mathrm{Su}(2) \oplus \mathrm{Sl}(2,\mathbb{R})$,
$G_{0C} = \mathrm{Sl}(2,\mathbb{C}) \oplus \mathrm{Sl}(2,\mathbb{R})$.

The first two real forms G_{0C} only occur when the parameter α is real; the third one occurs when $\alpha + \bar{\alpha} = -1$:

$G(3)$: $\quad G_{0C} = \mathrm{Sl}(2,\mathbb{R}) \oplus \mathscr{G}_{2,0}$,
$G_{0C} = \mathrm{Sl}(2,\mathbb{R}) \oplus \mathscr{G}_{2,2}$;

$F(4)$: $\quad G_{0C} = \mathrm{Sl}(2,\mathbb{R}) \oplus \mathrm{So}(7)$,
$G_{0C} = \mathrm{Sl}(2,\mathbb{R}) \oplus \mathrm{So}(3,4)$,
$G_{0C} = \mathrm{Su}(2) \oplus \mathrm{So}(2,5)$,
$G_{0C} = \mathrm{Su}(2) \oplus \mathrm{So}(1,6)$.

The following paragraphs are devoted to the proof of theorem 2.5.

3. THE SPECIAL LINEAR LIE SUPERALGEBRAS

Let $V = V_0 \oplus V_1$ be a complex graded vector space. The set L of linear transformations of V admits a natural structure of Lie superalgebra by setting $L = L_0 + L_1$, where L_0 is the set of elements of L which preserve V_0 and V_1 and L_1 is the set of elements of L which permute them. The bracket $[f, g]$ is the usual commutator $fg - gf$ if f or g belongs to L_0 and it is the anticommutator $fg + gf$ if both f and g belong to L_1.

If $\dim V_0 = m$ and $\dim V_1 = n$, the Lie superalgebra L will be denoted by $L(m,n)$. We will always write the matrices of the elements of $L(m,n)$ with respect to a basis of V adapted to the decomposition $V_0 \oplus V_1$; thus, $X = \begin{pmatrix} X_1 & X_3 \\ X_4 & X_2 \end{pmatrix}$ belongs to L_0 if $X_3 = X_4 = 0$ and belongs to L_1 if $X_1 = X_2 = 0$.

The subalgebra consisting of the elements X such that $\mathrm{tr}X_1 = \mathrm{tr}X_2$ is denoted by $\mathrm{SL}(m,n)$.

The natural representation of the simple Lie algebra $\mathscr{A}_n$ will be denoted by π_1 and its contragredient by π_n.

We will use the following notation for matrices in block form: $\mathrm{diag}(A,B,C)$ stands for

$$\begin{pmatrix} A & & \\ & B & \\ & & C \end{pmatrix}$$

while

$$\begin{pmatrix} & & A \\ & B & \\ C & & \end{pmatrix}$$

is denoted by $\mathrm{antidiag}(A,B,C)$.

A. Real forms of $A(m,n)$ where $m \neq n$

For $m \neq n$, the Lie superalgebra $A(m,n)$ is just $\mathrm{SL}(m,n)$; since $A(m,n)$ and $A(n,m)$ are isomorphic, we may assume $m > n \geqslant 1$. Besides, $A(2,1)$ is isomorphic to $\mathrm{OSp}(2,2)$ which will be handled in Sec. 4, so we assume here $(m,n) \neq (2,1)$. The Lie subalgebra G_0 is the direct sum of its center K_0 and of

the two simple ideals K_1 and K_2 of respective type $\mathscr{A}_{m-1}$ and $\mathscr{A}_{n-1}$ where

$$K_0 = \{X \in A(m,n) | X_3 = X_4 = 0, X_1 = naI_m, \quad X_2 = maI_n, a \in \mathbb{C}_0\},$$

$$K_1 = \{X \in A(m,n) | X_2 = X_3 = X_4 = 0\},$$

$$K_2 = \{X \in A(m,n) | X_1 = X_3 = X_4 = 0\}.$$

The subspace G_1 is the direct sum of the two invariant subspaces

$$Y' = \{X \in A(m,n) | X_1 = X_2 = X_4 = 0\},$$

$$Y'' = \{X \in A(m,n) | X_1 = X_2 = X_3 = 0\}.$$

The representation ρ' of $K_1 \oplus K_2$ on Y' is the tensor product of the natural representation π_1 of K_1 and the contragredient representation π_{n-1} of K_2, which we will abbreviate $\rho' = \pi_1(K_1) \otimes \pi_{n-1}(K_2)$. Similarly, the representation ρ'' of $K_1 \oplus K_2$ on Y'' is $\rho'' = \pi_{m-1}(K_1) \otimes \pi_1(K_2)$.

Both the natural representation π_1 of $\mathscr{A}_l$ and its contragredient π_l are real for the real form $\mathrm{Sl}(l+1,\mathbb{R})$, they are antireal for $\mathrm{Su}^*(l+1)$, and areal for $\mathrm{Su}(p, l+1-p)$. Hence, the only real forms of $K_1 \oplus K_2$ for which the irreducible representations ρ' and ρ'' are real are $\mathrm{Sl}(m,\mathbb{R}) \oplus \mathrm{Sl}(n,\mathbb{R})$ and $\mathrm{Su}^*(m) \oplus \mathrm{Su}^*(n)$ if m and n are even.

For the real form $\mathrm{Su}(p, m-p) \oplus \mathrm{Su}(q, n-q)$ of $K_1 \oplus K_2$, we will see that there exists an extension to G of the semimorphism C_0 which permutes Y' and Y''. We now consider the real forms of $K_1 \oplus K_2$ in which K_{1C} and K_{2C} are of different types. Since the representations ρ' and ρ'' are then never real, any extension of C_0 to G must permute Y' and Y''. However, the existence of a semilinear involution C_1 permuting Y' and Y'' such that $\rho'(C_0X)C_1 = C_1\rho''(X)$ implies that the weights of ρ'' are conjugate to those of $\rho' \circ C$. We check that this is never the case:

If $K_{1C} \oplus K_{2C}$ is of the form	the weights of $\rho' \circ C$ are conjugate to those of
$\mathrm{Su}(p, m-p) \oplus \mathrm{Sl}(n,\mathbb{R})$ $\mathrm{Su}(p, m-p) \oplus \mathrm{Su}^*(n)$	$\pi_{m-1}(K_1) \otimes \pi_{n-1}(K_2)$,
$\mathrm{Sl}(m,\mathbb{R}) \oplus \mathrm{Su}^*(n)$	$\pi_1(K_1) \otimes \pi_{n-1}(K_2)$,
$\mathrm{Sl}(m,\mathbb{R}) \oplus \mathrm{Su}(q, n-q)$	$\pi_1(K_1) \otimes \pi_1(K_2)$,
$\mathrm{Su}^*(m) \oplus \mathrm{Sl}(n,\mathbb{R})$	$\pi_1(K_1) \otimes \pi_{n-1}(K_2)$,
$\mathrm{Su}^*(m) \oplus Su(q, n-q)$	$\pi_1(K_1) \otimes \pi_1(K_2)$,

while $\rho'' = \pi_{m-1}(K_1) \otimes \pi_1(K_2)$.

The possible real forms are thus the following: (1) $G_{0C} = \mathrm{Sl}(m,\mathbb{R}) \oplus \mathrm{Sl}(n,\mathbb{R}) \oplus \mathbb{R}$. The involutive semimorphism is $CX = \bar{X}$ and it preserves Y' and Y''. With the notations of Lemma 2.4, we thus have $Ck_0 = k_0$, and hence $K_{0C} = \mathbb{R}$. (2) $G_{0C} = \mathrm{Su}^*(m) \oplus \mathrm{Su}^*(n) \oplus \mathbb{R}$ if m and n are even. The involutive semimorphism is $CX = M\bar{X}M^{-1}$ where

$$M = \begin{pmatrix} \text{antidiag } (-I_r, I_r) & \\ & \text{antidiag } (-I_s, I_s) \end{pmatrix}.$$

Again C preserves Y' and Y''; hence, $K_{0C} = \mathbb{R}$. (3) $G_{0C} = \mathrm{Su}(p, m-p) \oplus \mathrm{Su}(q, n-q) \oplus i\mathbb{R}$. The involutive semimorphism is $C_0X = -N\,{}^{\tau}\bar{X}N$ if $X \in G_0$ and $C_1X = iN\,{}^{\tau}\bar{X}N$ if $X \in G_1$, where $N = \mathrm{diag}(-I_p, I_{m-p}, -I_q, I_{n-q})$. Since C permutes Y' and Y'', it follows from Lemma 2.4 that $Ck_0 = -k_0$, and hence $K_{0C} = i\mathbb{R}$.

We now prove that C_0 may be chosen up to conjugacy by Aut G_0. Each automorphism of G_0 preserves the two ideals K_1 and K_2 and on each ideal it is of the form $\psi\theta$, where θ is inner and $\psi X = -{}^{\tau}X$. It is easy to check that the three C_0 we have chosen to define the real forms all commute with ψ; hence, any C_0' conjugate to C_0 is of the form $\theta C_0 \theta^{-1}$, where θ is inner and Lemma 2.2 then applies.

Finally, because of Lemma 2.4, all semimorphisms extending C_0 are conjugate in the first two cases. In the third case, we define an automorphism φ of G by $\varphi = \varphi_0 + \varphi_1$, where $\varphi_0 X = -N\,{}^{\tau}XN$ if $X \in G_0$ and $\varphi_1 X = -iN\,{}^{\tau}XN$ if $X \in G_1$. Then $\varphi_0 C_0 = C_0 \varphi_0$ and $\varphi_1 C_1 = -C_1 \varphi_1$, thus proving that $C_0 + C_1$ and $C_0 - C_1$ are conjugate by Aut G.

B. Real forms of $A(m, m)$

The Lie superalgebra $\mathrm{SL}(m, m)$ is not simple; it contains a one-dimensional center which consists of the scalar matrices. The quotient of $\mathrm{SL}(m, m)$ by its center is simple and is denoted by $A(m, m)$. This algebra $A(m, m)$ can also be presented in the following way:

$A(m, m)$ is the set of elements of $L(m, m)$ such that $\mathrm{tr}X_1 = \mathrm{tr}X_2 = 0$ and the bracket of two elements X and Y is the ordinary bracket of the two matrices except when both belong to G_1 in which case $[X, Y] = XY + YX - (1/m) \times \mathrm{tr}(XY)I$. The Lie subalgebra G_0 is $K_1 \oplus K_2$, where K_1 and K_2 are defined as in Sec. 3 A and are isomorphic of type $\mathscr{A}_{m-1}$.

For the semimorphisms C_0 which preserve the two ideals K_1 and K_2, the reasoning made for $A(m, n)$ applies and we obtain the real forms containing (1) $G_{0C} = \mathrm{Sl}(m, \mathbb{R}) \oplus \mathrm{Sl}(m, \mathbb{R})$. (2) $G_{0C} = \mathrm{Su}^*(m) \oplus \mathrm{Su}^*(m)$ if m is even. (3) $G_{0C} = \mathrm{Su}(p, m-p) \oplus \mathrm{Su}(q, m-q)$. (4) Besides, there is a real form in which $G_{0C} = \mathrm{Sl}(m, \mathbb{C})$ is the real Lie algebra of dimension $2(m^2-1)$. It is defined by the semimorphism $CX = P\bar{X}P$, where $P = \text{antidiag } (I_m, I_m)$, which permutes the two ideals K_1 and K_2 of G_0 and also permutes the two invariant subspaces Y' and Y'' of G_1.

In the first three cases, we prove as in Sec. 3.A that all semimorphisms extending C_0 are conjugate in Aut G. In the fourth case, the other possible semimorphism $C_0 - C_1$ is conjugate to $C_0 + C_1$ by the automorphism $\varphi = \varphi_0 + \varphi_1$, where $\varphi_0 X = -{}^{\tau}X$ if $X \in G_0$ and $\varphi_1 X = i\,{}^{\tau}X$ if $X \in G_1$.

Finally, to prove that C_0 may be chosen up to conjugacy by Aut G_0, it is sufficient to show that any automorphism of G_0 extends to an automorphism of G. Any element of Aut G_0 may be written as a product $\psi_0\theta_0\eta_0$ or $\psi_0\theta_0$ or θ_0, where θ_0 is inner, $\psi_0 X = -{}^{\tau}X$, and $\eta_0 X = PXP$. By Lemma 2.2 we know that θ_0 extends to G. Automorphisms of G extending ψ_0 and η_0 are defined, respectively, by $\psi_1 X = i\,{}^{\tau}X$ and $\eta_1 X = PXP$ if $X \in G_1$.

C. Real forms of $d(n)$

The subalgebra of $\mathrm{SL}(n, n)$ consisting of the elements X such that $X_1 = X_2$, $X_3 = X_4$, and $\mathrm{tr}\,X_3 = 0$ is not simple; the set of scalar matrices forms a one-dimensional center. The quotient of this algebra by its center is simple for $n > 2$ and is

denoted by $d(n)$. The Lie subalgebra G_0 of $d(n)$ is $\mathscr{A}_{n-1}$; the subspace G_1 is isomorphic to $\mathscr{A}_{n-1}$ as a vector space. The Lie superalgebra $d(n)$ can also be presented as the set of couples (A, B), where $A, B \in \mathscr{A}_{n-1}$ and the bracket is defined by

$$[(A,B),(A',B')] = AA' - A'A + BB' + B'B - \frac{2}{n} \times \mathrm{tr}(BB')I, AB' - B'A + BA' - A'B).$$

The representation of G_0 on G_1 is equivalent to the adjoint representation of G_0; hence, it is real for any involutive semimorphism of G_0. We obtain a real form of G for each possible real form of G_0:

(1) $G_{0C} = \mathrm{Sl}(n, \mathbb{R})$: the semimorphism is $C(A, B) = (\bar{A}, \bar{B})$.

(2) $G_{0C} = \mathrm{Su}^*(n)$ if $n = 2p$: The semimorphism is $C(A, B) = (M\bar{A}M^{-1}, M\bar{B}M^{-1})$, where $M = \mathrm{antidiag}(-I_p, I_p)$.

(3) $G_{0C} = \mathrm{Su}(p, n-p)$: The semimorphism is $C(A, B) = (-N\,{}^{\tau}\bar{A}N, iN\,{}^{\tau}\bar{B}N)$, where $N = \mathrm{diag}(-I_p, I_{n-p})$.

In each case, the other possible semimorphism $C_0 - C_1$ is conjugate to $C_0 + C_1$ by the automorphism φ of G defined by $\varphi(A, B) = (-{}^{\tau}A, i\,{}^{\tau}B)$.

To check that C_0 may be chosen up to conjugacy by $\mathrm{Aut}G_0$, we notice that any automorphism of G_0 is of the form $\theta_0 A = QAQ^{-1}$ or $\psi_0 A = -Q\,{}^{\tau}AQ^{-1}$. Automorphisms of G which extend θ_0 and ψ_0 are then defined respectively, by $\theta(A, B) = (QAQ^{-1}, QBQ^{-1})$ and $\psi(A, B) = (-Q\,{}^{\tau}AQ^{-1}, iQ\,{}^{\tau}BQ^{-1})$.

D. Real forms of $b(n)$

The Lie superalgebra $b(n)$ is the subalgebra of $L(n, n)$ which consists of the elements X such that $X_2 - {}^{\tau}X_1 = X_3 - {}^{\tau}X_3 = X_4 + {}^{\tau}X_4 = 0$ and $\mathrm{tr}X_1 = 0$. It is simple for $n > 2$.

The Lie subalgebra G_0 is $\mathscr{A}_{n-1}$ and the representation ρ of G_0 on G_1 is reducible. The two irreducible subspaces are $Y' = \{X \in b(n) | X_1 = X_4 = 0\}$ and $Y'' = \{X \in b(n) | X_1 = X_3 = 0\}$. The representation ρ' on Y' is the representation induced on the symmetric 2 tensors by the natural representation π_1 of $\mathscr{A}_{n-1}$. The representation ρ'' on Y'' is the representation induced on the skew 2 tensors by the contragredient π_{n-1}. Since Y' and Y'' have different dimensions, any semimorphism of G must preserve each of them, and thus ρ' and ρ'' must be real for G_{0C}. This implies $G_{0C} = \mathrm{Sl}(n, \mathbb{R})$ or $G_{0C} = \mathrm{Su}^*(n)$: (1) $G_{0C} = \mathrm{Sl}(n, \mathbb{R})$: The semimorphism is $CX = \bar{X}$. (2) $G_{0C} = \mathrm{Su}^*(n)$ if $n = 2p$. The semimorphism is $CX = M\bar{X}M^{-1}$, where

$$M = \begin{pmatrix} \mathrm{antidiag}(-I_p, I_p) & \\ & \mathrm{antidiag}(-I_p, I_p) \end{pmatrix}.$$

Every automorphism of G_0 may be written θ or $\psi\theta$, where θ is inner and $\psi X = -{}^{\tau}X$. Both semimorphisms C_0 chosen above to define the real forms G_{0C} commute with ψ and by Lemma 2.2. θ extends to an automorphism of G. Hence, C_0 may be chosen up to conjugacy by Aut G.

By Lemma 2.4 we know that all involutive semimorphisms which extend a given C_0 are conjugate in Aut G.

4. ORTHOSYMPLECTIC LIE SUPERALGEBRAS

On the graded vector space $V = V_0 + V_1$ with $\dim V_0 = 2m > 0$ and $\dim V_1 = n > 0$ we consider a nondegenerate bilinear form Ω which is skew on V_0, symmetric on V_1, and such that $\Omega(V_0, V_1) = \Omega(V_1, V_0) = 0$. A linear transformation f of V is said to be *orthosymplectic* if $\Omega(fx, y) + \Omega(x, fy) = 0$ except if $x \in V_0$ and $y \in V_1$ in which case $\Omega(fx, y) - \Omega(x, fy) = 0$. The set of orthosymplectic linear transformations of V is a subalgebra of $L(2m, n)$ which is simple and is denoted by $\mathrm{OSp}(2m, n)$. The only isomorphism is between $\mathrm{OSp}(2,2)$ and $A(2,1)$.

A. Real forms of $B(m, n)$ and $D(m, n)$

We have

$B(m, n) = \mathrm{OSp}(2m, 2n+1)$, where $n \geqslant 0$,

$D(m, n) = \mathrm{OSp}(2m, 2n)$, where $n \geqslant 2$.

The Lie subalgebra G_0 is $\mathscr{C}_m \oplus \mathscr{B}_n$ and $\mathscr{C}_m \oplus \mathscr{D}_n$. The representation ρ of G_0 on G_1 is irreducible and equivalent to the tensor product of the natural representations of the two simple ideals of G_0.

The natural representation of $\mathscr{C}_m$ is real for $\mathrm{Sp}(m, \mathbb{R})$ and antireal for $\mathrm{Sp}(r, m-r)$ while the natural representation of $\mathscr{B}_n$ or $\mathscr{D}_n$ is real for $\mathrm{So}(k, n-k)$ and antireal for $\mathrm{So}^*(n)$. Hence, the only real forms G_{0C} for which ρ is real are

$$G_{0C} = \mathrm{Sp}(m, \mathbb{R}) \oplus \mathrm{So}(k, n-k),$$

$$G_{0C} = \mathrm{Sp}(r, m-r) \oplus \mathrm{So}^*(n), \text{ if } n \text{ is even.}$$

We consider a basis of the vector space $V = V_0 \oplus V_1$ with respect to which the matrix of the bilinear form Ω is written $\Omega = \mathrm{diag}(\Omega_0, \Omega_1)$, where $\Omega_1 = J_n = \mathrm{antidiag}(1,1,\ldots,1)$ and $\Omega_0 = \mathrm{antidiag}(J_m, -J_m)$.

Written in this basis, the matrix X of an element of $\mathrm{OSp}(2m, n)$ satisfies $X_1\Omega_0 + \Omega_0\,{}^{\tau}X_1 = 0$, $X_2\Omega_1 + \Omega_1\,{}^{\tau}X_2 = 0$, $X_4\Omega_0 + \Omega_1\,{}^{\tau}X_3 = 0$, and $X_3\Omega_1 - \Omega_0\,{}^{\tau}X_4 = 0$. We set $\Delta_m = \mathrm{diag}(I_m, -I_m)$, $R_{k,n} = \mathrm{antidiag}(J_k, I_{n-2k}, J_k)$ and $K_{r,m} = \mathrm{diag}(I_r, -I_{2m-2r}, I_r)$: (1) $G_{0C} = \mathrm{Sp}(m, \mathbb{R}) \oplus \mathrm{So}(k, n-k)$. The involutive semimorphism of G is $CX = Q\bar{X}Q^{-1}$, where $Q = \mathrm{diag}(\Delta_m\Omega_0, iR_{k,n}J_n)$. (2) $G_{0C} = \mathrm{Sp}(r, m-r) \oplus \mathrm{So}^*(2p)$ when $n = 2p$. The involutive semimorphism of G is $CX = S\bar{X}S^{-1}$, where $S = \mathrm{diag}(K_{r,m}\Omega_0, i\Delta_p J_n)$.

In the case of $B(m, 0) = \mathrm{OSp}(2m, 1)$, the Lie subalgebra G_0 reduces to $\mathscr{C}_m$ and there is only one real form, in which $G_{0C} = \mathrm{Sp}(m, \mathbb{R})$, defined by the semimorphism $CX = Q\bar{X}Q^{-1}$, where Q is as above with $k = 0$ and $n = 1$.

Since all automorphisms of $\mathscr{C}_m$ and $\mathscr{B}_n$ are inner, we may apply Lemma 2.2 to $B(m, n)$. For $D(m, n)$, it is known that any automorphism of $\mathscr{D}_n$ is of the form $X_2 \to A_2X_2A_2^{-1}$, where A_2 is an orthogonal matrix with respect to the bilinear form Ω_1 on V_1, i.e., $A_2\Omega_1\,{}^{\tau}A_2 = \Omega_1$; so every automorphism of $G_0 = \mathscr{C}_m \oplus \mathscr{D}_n$ is of the form $X \to AXA^{-1}$, where $A = \mathrm{diag}(A_1, A_2)$ and $A_1\Omega_0\,{}^{\tau}A_1 = \Omega_0$. It is then easy to check that $X \to AXA^{-1}$ also defines an automorphism of $G = D(m, n)$. Hence, C_0 may be chosen up to conjugacy by Aut G_0.

For both of the real forms defined above, the other possible involutive semimorphism $C_0 - C_1$ is conjugate to $C = C_0 + C_1$ by the automorphism $\varphi X = C\bar{X}$ of G.

B. Real forms of $C(m)$

The Lie superalgebra $C(m) = \mathrm{OSp}(2m, 2)$ contains the Lie subalgebra $G_0 = \mathscr{C}_m \oplus \mathscr{D}_1$ and the representation ρ of G_0 on G_1 is reducible. The restriction of ρ to each of the two invariant subspaces Y' and Y'' of G_1 is the tensor product of the natural representation π_1 of $\mathscr{C}_m$ and the one-dimensional representation of $\mathscr{D}_1$. Observe that in a one-dimensional space any semilinear transformation is of positive square. If an involutive semimorphism C of G preserves Y' and Y'', the restriction of ρ to Y' and Y'' must be real for G_{0C}; hence, the representation π_1 of $\mathscr{C}_m$ must be real for that real form of $\mathscr{C}_m$.

This implies (1) $G_{0C} = \mathrm{Sp}(m, \mathbb{R}) \oplus \mathrm{So}(2, \mathbb{R})$. The semimorphism is $CX = Q\bar{X}Q^{-1}$, where Q is defined as in Sec. 4.A with $k = 0$ and $n = 2$. There exists an other real form of $C(m)$, defined by a semimorphism which permutes Y' and Y''. It contains (2) $G_{0C} = \mathrm{Sp}(r, m-r) \oplus \mathrm{So}(2, \mathbb{R})$. The semimorphism is $CX = S\bar{X}S^{-1}$, where S is defined as in Sec. 4.A with $p = 1$.

All automorphisms of $\mathscr{C}_m$ are inner, so Lemma 2.2 applies. On the other hand, the argument used in Sec. 4.A shows that all involutive semimorphisms which extend C_0 are conjugate by Aut G.

5. THE THREE ISOLATED CLASSICAL LIE SUPERALGEBRAS

A. Notations

Let S be a complex simple Lie algebra, β its Killing form, and ρ an irreducible representation of S on a vector space V. Let Ω be a nondegenerate bilinear form on V which is invariant by $\rho(S)$, i.e.,

$$\Omega(\rho(s)x, y) + \Omega(x, \rho(s)y) = 0, \quad \text{for all } s \in S,\ x, y \in V$$

Definition: The bilinear map P from V into S is defined by

$$\beta(s, P(x, y)) = \Omega(\rho(s)x, y), \quad \text{for all } s \in S.$$

This bilinear map P is symmetric (skew symmetric) if and only if Ω is skew symmetric (symmetric).

Remarks: We will use the presentation of the three isolated Lie superalgebras given in Ref. 5. In the case of $G(3)$ and $F(4)$, all automorphisms of G_0 are inner, so Lemma 2.2 applies.

The proof that the two involutive semimorphisms which extend a given C_0 are conjugate by Aut G will be given in Sec. 5.E for the three isolated Lie superalgebras.

B. Real forms of $D(1, 2, \alpha)$

The Lie subalgebra G_0 of $D(1, 2, \alpha)$ is the direct sum of three isomorphic ideals of type $\mathscr{A}_1$, which we shall denote by $S_1 \oplus S_2 \oplus S_3$. The subspace G_1 is the tensor product $V_1 \otimes V_2 \otimes V_3$ of three vector spaces of dimension 2. The representation ρ of G_0 on G_1 is the tensor product $\rho = \rho_1 \otimes \rho_2 \otimes \rho_3$, where each ρ_i is the natural representation of S_i on V_i; hence, ρ is irreducible.

Let Ω_i denote a nondegenerate skew-symmetric bilinear form on V_i which is invariant by $\rho(S_i)$. Let β_i denote the Killing form of S_i and P_i the bilinear map of V_i into S_i defined as in Sec. 5.A. Since Ω_i is skew symmetric, P_i is symmetric. The bracket of two elements of G_1 is the element of G_0 defined by

$$\begin{aligned}[v_1 \otimes v_2 \otimes v_3,\ w_1 \otimes w_2 \otimes w_3] &= \alpha_1 \Omega_2(v_2, w_2)\Omega_3(v_3, w_3)P_1(v_1, w_1) \\ &\quad + \alpha_2 \Omega_1(v_1, w_1)\Omega_3(v_3, w_3)P_2(v_2, w_2) \\ &\quad + \alpha_3 \Omega_1(v_1, w_1)\Omega_2(v_2, w_2)P_3(v_3, w_3),\end{aligned}$$

where $\alpha_i \in \mathbb{C}_0$ and $\alpha_1 + \alpha_2 + \alpha_3 = 0$.

Moreover, it is shown in Ref. 5 that two such algebras defined by the triples $(\alpha_1, \alpha_2, \alpha_3)$ and $(\alpha_1', \alpha_2', \alpha_3')$ are isomorphic if and only if there exists a permutation σ of $\{1, 2, 3\}$ such that $\alpha_i' = k\alpha_{\sigma i}$, where $k \in \mathbb{C}_0$. So we may choose the parameter α by setting $\alpha_1 = 1$, $\alpha_2 = \alpha$, and $\alpha_3 = -1 - \alpha$.

If C is an involutive semimorphism of G, the irreducible representation ρ must be real for G_{0C}. Hence, if C preserves each of the ideals S_i, there may be zero or two of the representations ρ_i which are antireal for the real forms S_{iC}, the remaining ρ_j being real for S_{jC}. If C permutes two of the ideals S_i, for instance S_2 and S_3, $\rho_2 \otimes \rho_3$ is automatically real for the real form $(S_2 \oplus S_3)_C$, so in that case ρ_1 must be real for S_{1C}.

The natural representation of $\mathscr{A}_1$ is real for $\mathrm{Sl}(2, \mathbb{R})$ and antireal for $\mathrm{Su}(2)$, and hence the only possible real forms of G_0 are as follows:

(1) $G_{0C} = \mathrm{Sl}(2, \mathbb{R}) \oplus \mathrm{Sl}(2, \mathbb{R}) \oplus \mathrm{Sl}(2, \mathbb{R})$,

(2) $G_{0C} = \mathrm{Sl}(2, \mathbb{R}) \oplus \mathrm{Su}(2) \oplus \mathrm{Su}(2)$,

(3) $G_{0C} = \mathrm{Sl}(2, \mathbb{R}) \oplus \mathrm{Sl}(2, \mathbb{C})$.

Each of these real Lie subalgebras G_{0C} is contained in a real form of G for certain values of the parameter α.

To see this, we choose a basis (e_i, e_i') in each V_i such that $\Omega_i(e_i, e_i') = 1$. Conjugation $v_i \to \bar{v}_i$ in V_i is defined with respect to the real subspace spanned by (e_i, e_i'). So we have

$$\Omega_i(\bar{v}_i, \bar{w}_i) = \overline{\Omega_i(v_i, w_i)}$$

and

$$P_i(\bar{v}_i, \bar{w}_i) = \overline{P_i(v_i, w_i)}.$$

We write the elements of G_0 as $X_1 + Y_2 + Z_3$, where X, Y, Z are 2×2 matrices with zero trace, and the subscript $i = 1, 2, 3$ indicates the ideal S_i to which each of them belongs.

1. $G_{0C} = \mathrm{Sl}(2, \mathbb{R}) \oplus \mathrm{Sl}(2, \mathbb{R}) \oplus \mathrm{Sl}(2, \mathbb{R})$

We choose $C_0(X_1 + Y_2 + Z_3) = \bar{X}_1 + \bar{Y}_2 + \bar{Z}_3$. Since the representation ρ of G_0 on G_1 is irreducible, a semilinear involution C_1 of G_1 which satisfies $\rho(C_0 g)C_1 = C_1\rho(g)$ for all $g \in G_0$ is unique up to a factor of modulus 1. So we may write $C_1(v_1 \otimes v_2 \otimes v_3) = \lambda(\bar{v}_1 \otimes \bar{v}_2 \otimes \bar{v}_3)$, where $\lambda\bar{\lambda} = 1$. If $C_0 + C_1$ is a semimorphism of G, we have

$$\begin{aligned}&[C_1(v_1 \otimes v_2 \otimes v_3),\ C_1(w_1 \otimes w_2 \otimes w_3)] \\ &\quad = C_0[v_1 \otimes v_2 \otimes v_3,\ w_1 \otimes w_2 \otimes w_3],\end{aligned}$$

which is satisfied if and only if $\lambda^2\alpha_i = \bar{\alpha}_i$ for all $i = 1, 2, 3$. This condition says that the three complex numbers α_i have

the same argument, or, if we assume $\alpha_1 = 1$ and $\alpha_2 = \alpha$, that the parameter α is real.

2. $G_{0C} = \mathrm{Sl}(2, \mathbb{R}) \oplus \mathrm{Su}(2) \oplus \mathrm{Su}(2)$

We choose $C_0(X_1 + Y_2 + Z_3) = \bar{X} - {}^{\tau}\bar{Y}_2 - {}^{\tau}\bar{Z}_3$. Let us denote by J the semilinear transformation of V_i such that $Je_i = e_i'$ and $Je_i' = -e_i$. It is easy to check that $\Omega_i(Jv_i, Jw_i) = \overline{\Omega_i(v_i, w_i)}$ and that $\rho_i(-{}^{\tau}X_i)J = J\rho(X_i)$ from which follows

$$\beta_i(X_i, P_i(Jv_i, Jw_i)) = \Omega_i(\rho_i(X_i)Jv_i, Jw_i)$$
$$= -\overline{\Omega_i(J\rho_i(X_i)Jv_i, w_i)} = \overline{\Omega_i(\rho_i(-{}^{\tau}\bar{X}_i)v_i, w_i)}$$
$$= \overline{\beta_i(-{}^{\tau}\bar{X}_i, P_i(v_i, w_i))} = \beta_i(X_i, -{}^{\tau}\overline{(P_i(v_i, w_i))});$$

hence,

$$P_i(Jv_i, Jw_i) = -{}^{\tau}\overline{(P_i(v_i, w_i))}.$$

If we set $C_1(v_1 \otimes v_2 \otimes v_3) = \lambda\bar{v}_1 \otimes Jv_2 \otimes Jv_3$, we have that

$$[C_0(X_1 + Y_2 + Z_3), C_1(v_1 \otimes v_2 \otimes v_3)]$$
$$= C_1[X_1 + Y_2 + Z_3, v_1 \otimes v_2 \otimes v_3]$$

and

$$[C_1(v_1 \otimes v_2 \otimes v_3), C_1(w_1 \otimes w_2 \otimes w_3)]$$
$$= C_0[v_1 \otimes v_2 \otimes v_3, w_1 \otimes w_2 \otimes w_3]$$

are satisfied if and only if $\lambda^2\alpha_i = \bar{\alpha}_i$ for all $i = 1, 2, 3$.

3. $G_{0C} = \mathrm{Sl}(2, \mathbb{R}) \oplus \mathrm{Sl}(2, \mathbb{C})$

We choose $C_0(X_1 + Y_2 + Z_3) = \bar{X}_1 + \bar{Z}_2 + \bar{Y}_3$. Then any extension C_1 of C_0 to G_1 may be written $C_1(v_1 \otimes v_2 \otimes v_3) = \lambda\bar{v}_1 \otimes \bar{v}_3 \otimes \bar{v}_2$ and $C_0 + C_1$ is a semimorphism if and only if $\lambda^2\alpha_1 = \bar{\alpha}_1$, $\lambda^2\alpha_2 = \bar{\alpha}_3$, and $\lambda^2\alpha_3 = \bar{\alpha}_2$. This condition says that α_2 and α_3 have the same module and that the argument of α_1 is half the sum of the arguments of α_2 and α_3. If we set $\alpha_1 = 1$ and $\alpha_2 = \alpha$, the condition becomes $\alpha + \bar{\alpha} = -1$.

We now prove that C_0 may be chosen up to conjugacy by Aut G_0. By virtue of Lemma 2.2 any inner automorphism of G_0 extends to an automorphism of G. Any automorphism φ of G_0 which permutes the ideals S_i may be written $\varphi = \psi\eta$, where η is inner and ψ maps a matrix X_i belonging to S_i onto the same matrix $X_{\sigma i}$ belonging to the ideal $S_{\sigma i} = \varphi S_i$.

When $C_0(X_1 + Y_2 + Z_3) = \bar{X}_1 + \bar{Y}_2 + \bar{Z}_3$, ψ and C_0 commute, so there is nothing to prove.

When $C_0(X_1 + Y_2 + Z_3) = \bar{X}_1 - {}^{\tau}\bar{Y}_2 - {}^{\tau}\bar{Z}_3$, either ψ and C_0 commute or $\psi C_0\psi^{-1}$ is of the same form as C_0 but with the ideals S_i written in a different order. Hence the condition on the α_i are the same.

When $C_0(X_1 + Y_2 + Z_3) = \bar{X}_1 + \bar{Z}_2 + \bar{Y}_3$, it is clear that $\psi C_0\psi^{-1}$ is the same type of semimorphism permuting two of the ideals as C_0. Thus, we would obtain with $\psi C_0\psi^{-1}$ the same condition on the α_i, namely, that two of the α_i have the same module and that the argument of the third one is half the sum of the arguments of the two others. Finally, for each G_{0C}, the extension C_1 was given up to a factor λ and we saw that λ^2 was determined by the α_i. There are thus two semimorphisms extending C_0, but we will prove in Sec. 5.E that they are conjugate by Aut G.

C. Real forms of $G(3)$

We first recall some facts about Cayley algebras. Let us denote by K the Cayley algebra over a field F ($= \mathbb{R}$ or $\mathbb{C}$) whose elements we write in the form $x = q_1 + pq_2$, where q_1, q_2 are elements of a quaternion algebra over F and whose multiplication is defined by

$$(q_1 + pq_2)(q_3 + pq_4) = (q_1q_3 + \mu q_4\bar{q}_2) + p(\bar{q}_1q_4 + q_3q_2),$$

where $0 \neq \mu \in F$.

There exists in K an involution $x = q_1 + pq_2 \to \bar{x} = \bar{q}_1 - pq_2$ such that $\overline{xy} = \bar{y}\,\bar{x}$. The trace and norm of x are $t(x) = x + \bar{x}$ and $n(x) = x\bar{x}$. The algebra K has no zero divisors if and only if $x \neq 0$ implies $n(x) \neq 0$. Let us denote by K' the subspace of traceless elements of K. The bilinear form $B(x, y) = \frac{1}{2}t(x\bar{y})$ is symmetric and nondegenerate on K'.

If L_x and R_x denote left and right multiplication by x, $D_{x,y} = [L_x, L_y] + [R_x, R_y] + [L_x, R_y]$ is a derivation of K which stabilizes K'. The Lie algebra of all derivations of K is simple and of the exceptional type $\mathscr{G}_2$; it is the linear span of the derivations of the form $D_{x,y}$. Moreover, for any derivation D of K, we have $B(Dx, y) + B(x, Dy) = 0$. If $F = \mathbb{R}$, the Cayley algebra $K(-1)$ corresponding to $\mu = -1$ has no zero divisors and the bilinear form B is definite. The Lie algebra of derivations of $K(-1)$ is the compact real form of $\mathscr{G}_2$, which we denote by $\mathscr{G}_{2,0}$.

The Cayley algebra $K(+1)$ corresponding to $\mu = 1$ has zero divisors and the bilinear form B is of signature (4,4). The Lie algebra of derivations of $K(+1)$ is the noncompact real form of $\mathscr{G}_2$, which we denote by $\mathscr{G}_{2,2}$. The complex Lie superalgebra $G(3)$ of dimension 31 is defined as follows: The Lie subalgebra G_0 is the direct sum of two ideals of respective type $\mathscr{A}_1$ and $\mathscr{G}_2$. The subspace G_1 is the tensor product $G_1 = V \otimes K'$ of a two-dimensional vector space V and a seven-dimensional vector space K' which is viewed as the set of traceless elements of a Cayley algebra over $\mathbb{C}$. The representation ρ of G_0 on G_1 is the tensor product $\rho = \rho_1 \otimes \rho_2$ of the natural representation of $\mathscr{A}_1$ on V and the fundamental representation of $\mathscr{G}_2$ by derivations on K'; hence, ρ is irreducible.

Let Ω be a skew-symmetric nondegenerate bilinear form on V, invariant by $\rho_1(\mathscr{A}_1)$ and let P be the bilinear map from V into $\mathscr{A}_1$ defined as in Sec. 5.A. The product of two elements of G_1 is then given by

$$[v \otimes x, w \otimes y] = B(x, y)P(v, w) + \frac{1}{32}\Omega(v, w)D_{x,y}.$$

If G_C is a real form of G, the representation $\rho = \rho_1 \otimes \rho_2$ must be real for G_{0C}. Since ρ_2 is real for the two real forms of $\mathscr{G}_2$, ρ_1 must be real for the real form of $\mathscr{A}_1$, which excludes Su(2). Hence, the two possible real forms of G_0 are as follows:

(1) $G_{0C} = \mathrm{Sl}(2, \mathbb{R}) \oplus \mathscr{G}_{2,0}$,

(2) $G_{0C} = \mathrm{Sl}(2, \mathbb{R}) \oplus \mathscr{G}_{2,2}$.

The real forms of G containing these real Lie algebras are obtained by considering $G_1 = V \otimes K'$, where V is a real two-dimensional vector space and K' is the subset of traceless elements of the Cayley algebra $K(-1)$ in the first case and

$K(+1)$ in the second case; the bracket is defined as in the complex algebra.

The unicity, up to isomorphism, of these two real forms results from Lemmas 2.2 and 2.3 and Proposition 5.5.

D. Real forms of $F(4)$

The Lie subalgebra G_0 of $F(4)$ is the direct sum of two ideals of respective type $\mathscr{A}_1$ and $\mathscr{B}_3$. The subspace G_1 is the tensor product $G_1 = V \otimes U$, where V is of dimension 2 and U is of dimension 8. The representation ρ of G_0 on G_1 is the tensor product $\rho = \rho_1 \otimes \rho_2$ of the natural representation of $\mathscr{A}_1$ on V and the spin representation of $\mathscr{B}_3$ on U.

Let Γ_i, $i = 1,...,7$ be seven Dirac 8×8 matrices, such that $\Gamma_i\Gamma_j + \Gamma_j\Gamma_i = 2\delta_{ij}$, $\bar{\Gamma}_i = \Gamma_i$ if i is odd, and $\bar{\Gamma}_i = -\Gamma_i$ if i is even. The Lie algebra with basis $\{\Gamma_i\Gamma_j \mid 1 \leqslant i < j \leqslant 7\}$ is simple of type $\mathscr{B}_3$. On U there exists a nondegenerate symmetric bilinear form B such that $B(\Gamma_i x, y) + B(x, \Gamma_i y) = 0$ for all $i = 1,...,7$. We define a skew-symmetric bilinear map P_2 from U into the Lie algebra $\mathscr{B}_3$ by

$$P_2(x, y) = \sum_{1 \leqslant j < k \leqslant 7} B(\Gamma_j\Gamma_k x, y)\Gamma_j\Gamma_k .$$

The bracket of two elements of G_1 is given by

$$[v \otimes x, w \otimes y] = B(x, y)P_1(v, w) - \frac{1}{24}\Omega(v, w)P_2(x, y) ,$$

where Ω is the nondegenerate skew-symmetric bilinear form on V and P_1 is the symmetric bilinear map from V into $\mathscr{A}_1$ defined in Sec. 5.A.

If C is an involutive semimorphism of G, the representation $\rho = \rho_1 \otimes \rho_2$ must be real for G_C. The natural representation of $\mathscr{A}_1$ is real for Sl(2, $\mathbb{R}$) and antireal for Su(2) and the spin representation of $\mathscr{B}_3$ is real for So(7) and So(3, 4) and antireal for So(1, 6) and So(2, 5). Hence, the only possible real forms G_{0C} are as follows:

(1) $G_{0C} = \mathrm{Sl}(2, \mathbb{R}) \oplus \mathrm{So}(3, 4)$,

(2) $G_{0C} = \mathrm{Sl}(2, \mathbb{R}) \oplus \mathrm{So}(7)$,

(3) $G_{0C} = \mathrm{Su}(2) \oplus \mathrm{So}(1, 6)$,

(4) $G_{0C} = \mathrm{Su}(2) \oplus \mathrm{So}(2, 5)$.

In the Lie algebra $\mathscr{B}_3$ we consider the four involutive semimorphisms T_α, $0 \leqslant \alpha \leqslant 3$, defined by

$$T_\alpha(\Gamma_i\Gamma_j) = Q_\alpha \overline{\Gamma_i\Gamma_j} Q_\alpha^{-1} ,$$

where $Q_0 = I$, $Q_1 = \Gamma_4$, $Q_2 = \Gamma_2\Gamma_6$, and $Q_3 = \Gamma_2\Gamma_4\Gamma_6$.

From the relation $\mathrm{tr}(\Gamma_i\Gamma_j\Gamma_k\Gamma_m) = -8\delta_{ik}\delta_{jm}$, it is easy to deduce the signature of the Killing form of the real forms of $\mathscr{B}_3$ defined by these semimorphisms. One finds

for T_0, signature (9, 12) and hence So(3, 4);

for T_1, signature (11, 10) and hence So(2, 5);

for T_2, signature (15, 6) and hence So(1, 6);

for T_3, signature (21, 0) and hence So(7).

Let us denote by T'_α the semilinear transformations of V defined by $T'_0(x) = \bar{x}$, $T'_1(x) = iQ_1\bar{x}$, $T'_2(x) = Q_2\bar{x}$, and $T'_3(x) = iQ_3\bar{x}$. Here the conjugation $x \to \bar{x}$ is defined with respect to a real subspace on which B is real. This is possible since we may take as matrix for B the charge conjugation matrix $\Gamma = \Gamma_1\Gamma_3\Gamma_5\Gamma_7$ which is real, symmetric, and satisfies $\Gamma\Gamma_i + {}^{\tau}\Gamma_i\Gamma = 0$ for all $i = 1,...,7$.

Then T'_0 and T'_3 are involutive while T'_1 and T'_2 are of square $-I$. In all four cases, it is clear that

$$T_\alpha(\Gamma_i\Gamma_j)T'_\alpha(x) = T'_\alpha(\Gamma_i\Gamma_j x) .$$

Moreover, a straightforward calculation shows that

$$B(T'_\alpha x, T'_\alpha y) = \overline{B(x, y)} ,$$

$$P_2(T'_\alpha x, T'_\alpha y) = T_\alpha P_2(x, y) .$$

As in Sec. 5.B let (e, e') be a basis of V such that $\Omega(e, e') = 1$ and let J be the semilinear transformation of V defined by $Je = e'$ and $Je' = -e$. The conjugation in V is taken with respect to the real subspace spanned by (e, e'). Using the properties of P_1 obtained in Sec. 5.B and the above remarks, it is easy to check that the following transformations are involutive semimorphisms of $F(4)$. We write $X + Y$, where $X \in \mathscr{A}_1$ and $Y \in \mathscr{B}_3$, for the elements of G_0:

(1) $C_0(X + Y) = \bar{X} + \bar{Y}$ and $C_1(v \otimes x) = \bar{v} \otimes \bar{x}$,

(2) $C_0(X + Y) = \bar{X} + T_3(Y)$ and $C_1(v \otimes x) = i\bar{v} \otimes Q_3\bar{x}$,

(3) $C_0(X + Y) = -{}^{\tau}\bar{X} + T_2(Y)$ and $C_1(v \otimes x) = Jv \otimes Q_2\bar{x}$,

(4) $C_0(X + Y) = -{}^{\tau}\bar{X} + T_1(Y)$ and $C_1(v \otimes x) = iJv \otimes Q_1\bar{x}$.

The real form of $F(4)$ defined by one of these four involutive semimorphisms contains the corresponding real Lie subalgebra G_{0C} listed previously. Lemmas 2.2 and 2.3 and Proposition 5.5 imply that these real forms are unique, up to isomorphism.

Proposition 5.5: If G is one of the Lie superalgebras $D(1, 2, \alpha)$, $G(3)$ or $F(4)$ and C is an involutive semimorphism of G, the real forms defined by $C = C_0 + C_1$ and $C = C_0 - C_1$ are isomorphic.

Proof: In the three cases, G_0 is the direct sum of two ideals, one of which is $\mathscr{A}_1$. We will set $G_0 = \mathscr{A}_1 \oplus S$ and $X + Y \in \mathscr{A}_1 \oplus S$. The subspace G_1 is a tensor product $V \otimes U$ where V is of dimension two. The representation ρ of G_0 on G_1 is the tensor product of the natural representation of $\mathscr{A}_1$ on V and of some irreducible representation of S on U. The bracket of two elements of G_1 is given by

$$[v \otimes x, w \otimes y] = B(x, y)P_1(v, w) + \Omega(v, w)P_2(x, y) ,$$

where Ω is a nondegenerate skew-symmetric bilinear form on V and P_1 is the symmetric bilinear map from V into $\mathscr{A}_1$ defined in Sec. 5.A; B is a nondegenerate symmetric bilinear form on U and P_2 is a skew-symmetric bilinear map from U into S.

We consider first the real forms containing Sl(2, $\mathbb{R}$), i.e., all real forms of $D(1, 2, \alpha)$ and $G(3)$ and two real forms of $F(4)$.

The linear transformation ψ of G defined by $\psi(X + Y) = MXM^{-1} + Y$, where $M = \mathrm{diag}(i, -i)$ and $\psi(v \otimes x) = Mv \otimes x$ is an automorphism of G. Moreover, on $\mathscr{A}_1$ and on V the semimorphism C is here conjugation and thus we have $\psi C_0 = C_0\psi$ and $\psi C_1 = -C_1\psi$, which proves the proposition in those cases. In the two remaining cases

containing, respectively, $\mathrm{Su}(2) \oplus \mathrm{So}(2, 5)$ and $\mathrm{Su}(2) \oplus \mathrm{So}(1, 6)$, the semimorphism C is of the form $C_0(X + Y) = -{}^{\tau}X + Q\bar{Y}Q^{-1}$ and $C_1(v \otimes x) = Jv \otimes Q\bar{x}$, where $Q = i\Gamma_4$ for $\mathrm{So}(2, 5)$ and $Q = \Gamma_2\Gamma_6$ for $\mathrm{So}(1, 6)$. The linear transformation ψ of G defined by $\psi(X + Y) = X + PYP^{-1}$ and $\psi(v \otimes x) = iv \otimes Px$, where $P = \Gamma_2$ or Γ_6, is an automorphism of G. Again we have $\psi C_0 = C_0\psi$ and $\psi C_1 = -C_1\psi$, which completes the proof.

[1] S. Sternberg, Some recent results on supersymmetry, Differential Geometrical Methods in Mathematical Physics, Lecture Notes in Mathematics 570 (1977), 145-176.
[2] S. Sternberg and J. Wolf, Trans. Am. Math. Soc. **238**, 1 (1978).
[3] I. Kaplansky, "Graded Lie algebras," University of Chicago Preprints, 1975.
[4] V.G. Kats (KAC), Funct. Anal. Appl. **9**, 91 (1975).
[5] M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **17**, 1626 (1976).
[6] S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962).

Ewald evaluation and Poisson summation formulas for a class of near-Coulombic lattice sums

George L. Hall
Department of Physics, North Carolina State University, Raleigh, North Carolina 27650

(Received 21 September 1979; accepted for publication 21 December 1979)

The class of lattice sums studied is defined by $F_{n,\tau} = \Sigma_\tau^{\infty\prime}[(1/\tau^n) - (1/\Omega)\int_\tau (d^3r/r^n)] - (1/\Omega)\int_0 (d^3r/r^n)$ for $1 \leqslant n \leqslant 2$, where $\{\vec{\tau}\}$ denotes a three-dimensional Bravais lattice with volume Ω per lattice point, $\{\vec{\gamma}\}$ denotes its reciprocal lattice normalized with $\exp(i\vec{\tau}\cdot\vec{\gamma}) = 1$ with volume $v = (2\pi)^3/\Omega$ per lattice point, and except for the case $n = 1$ the cellular integrations are taken over centered cells, proximity or primitive. Ewald evaluation is discussed for all cases and reciprocal relations between the $F_{n,\tau}$ and the $F_{(3-n),\gamma}$ are supplied for all n. These relations are essentially Poisson summation formulas (PSF) with modifications for the cases $n = 1,2$. The method of securing the PSF's involves a direct, and an inverse, use of an Ewald method coupled with careful analysis of all interchanges of integrations and summations. The method should be useful for similar studies on other complex sums which, like these, are not very amenable to direct application of a PSF with assurance of validity.

I. INTRODUCTION

Let $\{\vec{\tau}\}$ denote a three-dimensional Bravais[1] lattice with volume Ω per lattice point, and let $\{\vec{\gamma}\}$ denote its reciprocal lattice normalized with $\exp(i\vec{\tau}\cdot\vec{\gamma}) = 1$ with a volume $v = (2\pi)^3/\Omega$ per lattice point. I shall report two sets of results for the class of lattice sums given by

$$F_{n,\tau} = \sum_\tau^{\infty\prime}\left(\frac{1}{\tau^n} - \frac{1}{\Omega}\int_\tau \frac{d^3r}{r^n}\right) - \frac{1}{\Omega}\int_0 \frac{d^3r}{r^n}, \tag{1.1}$$

where $1 \leqslant n \leqslant 2$ and where, except for the case $n = 1$, the cellular integrations are taken over a centered cell, proximity or primitive. In the first set I show that the usual Ewald method of evaluating lattice sums does not need to be modified for the cases $1 < n \leqslant 2$ as we now know[2,3] it does for the case $n = 1$; i.e., no "correction term" is required in Ewald evaluation when $1 < n \leqslant 2$. In the second set of results I supply for the $F_{n,\tau}$ a reciprocal formula that is useful in solid state analysis and especially interesting when n has any of the values 1, 3/2, 2. The procedure used to secure the reciprocal formula should be of importance to other classes of lattice sums that share with the $F_{n,\tau}$ the feature of not being very amenable to the usual procedure for finding a Poisson summation formula (PSF).[4] If the summation of Eq. (1.1) is taken over the reciprocal lattice and we denote the sum with $F_{n,\gamma}$, then $F_{1,\tau}$ and $F_{2,\gamma}$ are reciprocally related with a correction term in the relation. Thus one at least needs to study the case $n = 2$ even if one has interest only in $n = 1$.

The extension of this study to the larger class of sums in which the τ^{-n} of Eq. (1.1) is replaced with $|(\vec{\tau} - \vec{R})|^{-n}$ will be considered elsewhere, but the transcription of the present results to the analogous sums arising in the classical Coulomb lattice (CCL) model is trivial. For the CCL sums the cellular integrations in Eq. (1.1) are appropriately replaced[3] with finite sums. The present sums may be viewed as associated with the Wigner[5] solid (WS) model; at least the case $n = 1$ arises directly[2,3] and, because of the reciprocal relation, the case $n = 2$ arises indirectly.

II. PRELIMINARY RESULTS FOR THE CASE $1 < n \leqslant 2$

For these cases we have that the summand of Eq. (1.1) satisfies

$$\frac{1}{\tau^n} - \frac{1}{\Omega}\int_\tau \frac{d^3r}{r^n} = \frac{n}{\tau^{n+1}}\frac{1}{\Omega}\int_0 \vec{r}\cdot\vec{\tau}\, d^3r - \frac{n}{2\Omega\tau^{n+2}}\int_0 ((n+2)(\vec{r}\cdot\hat{\tau})^2 - r^2) d^3r + O\left(\frac{1}{\tau^{n+3}}\right), \tag{2.1}$$

and the first term on the right vanishes for cellular integrations taken over centered cells, proximity or primitive. Thus the summation in Eq. (1.1) converges absolutely for $1 < n \leqslant 2$ and hence has a unique limit. To prepare for application of the Ewald method,[6] we note that

$$\Gamma(\tfrac{1}{2}n)/\tau^n = \int_0^\infty x^{-1+n/2}\exp(-x\tau^2)\, dx, \tag{2.2}$$

and that by following Appendix A of Ref. 2 one easily proves the lemma

$$\int_\tau \frac{d^3r}{r^n} = \frac{1}{\Gamma(\frac{1}{2}n)}\int_0^\infty x^{-1+n/2}\int_\tau \exp(-xr^2)\, d^3r\, dx, \tag{2.3}$$

where Γ denotes the gamma function. This allows writing

$$\bar{F}_{n,\tau} = F_{n,\tau} + \frac{1}{\Omega}\int_0 \frac{d^3r}{r^n} = \sum_\tau^{\infty\prime}\left(\frac{1}{\tau^n} - \frac{1}{\Omega}\int_\tau \frac{d^3r}{r^n}\right), \tag{2.4}$$

$$\bar{F}_{n,\tau} = \frac{1}{\Gamma(\frac{1}{2}n)}\sum_\tau^{\infty\prime}\int_0^\infty t^{-1+n/2}\times\left[\exp(-t\tau^2) - \frac{1}{\Omega}\int_\tau \exp(-tr^2)\, d^3r\right] dt = \Gamma(\tfrac{1}{2}n)\sum_\tau^{\infty\prime}\int_0^\infty t^{-1+n/2} a_\tau(t)\, dt, \tag{2.5}$$

where as in Ref. 2

$$a_\tau(t) = \exp(-t\tau^2)$$

 0022-2488/80/040698-04$1.00

$$\times\left[1-(1/\Omega)\int_0\exp(-2\vec{x}\cdot\vec{\tau}t-x^2t)\,d^3x\right]. \quad (2.6)$$

Then in the Appendix I show rigorously that the summation of Eq. (2.5) can be passed inside the integration without any correction term[2] arising as it does for the case $n=1$. Thus from Eq. (2.4) we have that

$$F_{n,\tau}=\frac{1}{\Gamma(\frac{1}{2}n)}\int_0^\infty t^{-1+n/2}\times\left(\sum_\tau{}'\exp(-t\tau^2)-\frac{1}{\Omega}(\pi/t)^{3/2}\right)dt. \quad (2.7)$$

Other forms that readily follow from a PSF[4] on theta functions are

$$F_{n,\tau}=\frac{1}{\Gamma(\frac{1}{2}n)}\int_0^\infty t^{-1+n/2}\left[\frac{1}{\Omega}\left(\frac{\pi}{t}\right)^{3/2}\sum_\gamma{}'\exp\left(\frac{-\gamma^2}{4t}\right)-1\right]dt$$
$$=\frac{1}{2^n\Gamma(\frac{1}{2}n)}\int_0^\infty\frac{1}{y^{1+n/2}}\left(\frac{(4\pi y)^{3/2}}{\Omega}\sum_\gamma{}'\exp(-y\gamma^2)-1\right)dy \quad (2.8)$$
$$=\frac{(4\pi)^{3/2}}{\Omega\,2^n}\frac{1}{\Gamma(\frac{1}{2}n)}\int_0^\infty y^{(1/2)(3-n)-1}\times\left[\sum_\gamma{}'\exp(-y\gamma^2)-\frac{1}{v}\left(\frac{\pi}{y}\right)^{3/2}\right]dy.$$

Equations (2.7) and (2.8) supply the preliminary results we shall need.

III. EWALD EVALUATION FOR THE CASES $1<n\leqslant 2$

Passing from Eq. (2.5) to Eq. (2.7) without rigorous proof would probably be widely accepted were it not for our having been alerted to the possibility of difficulties from the known results[2] for the case $n=1$. In any event, we now know Eq. (2.7) holds, and Ewald evaluation for the cases $1<n\leqslant 2$ is straightforward although some of the final integrals involved may not be tabulated.

The procedure is to use the integrand of Eq. (2.7) in the domain $b\leqslant t<\infty$ and that of the first form of Eq. (2.8) in the domain $0<t\leqslant b$, where b is a "separation parameter" to be chosen to balance the contributions from the two resulting series for $F_{n,\tau}$. With these domains the respective summations may be taken outside the integration, because one has uniform convergence. The best choice of b may depend somewhat on n.

The procedure is well known[6] especially from applications to Lennard–Jones sums. Also, Coldwell-Horsfall and Maradudin[7] give the procedure for evaluating by this method the integral in Eq. (2.7) when $n=1$ although this integral does not give $F_{1,\tau}$. I shall return to this case latter.

IV. POISSON SUMMATION FORMULA FOR $1<n<2$

Note that Eq. (2.8) in its last form is of the same form as Eq. (2.7) provided $(3-n)$ replaces n. Note further that $1<n<2$ implies that $1<(3-n)<2$. Accordingly, the proof that we could interchange the order of summation and integration in Eq. (2.5) to secure Eq. (2.7) now permits us to interchange the orders of these operations in Eq. (2.8) for $1<n<2$, giving the reciprocal formula for $1<n<2$

$$F_{n,\tau}=\{(4\pi)^{3/2}\Gamma[\tfrac{1}{2}(3-n)]/\Omega\,2^n\Gamma(\tfrac{1}{2}n)\}F_{3-n,\gamma}, \quad (4.1)$$

which gives a particularly symmetric result for $n=3/2$, namely

$$F_{3/2,\tau}=(2\pi)^{-3/2}vF_{3/2,\gamma}. \quad (4.2)$$

Equation (4.1) possesses reciprocal properties similar to those one secures for sums for which the PSF is applicable. Next I shall try to give a stronger justification for thinking of Eq. (4.1) as a PSF. Of course, Eq. (4.1) will have to be modified for the cases $n=1$, 2.

The PSF[4] for suitably behaved functions $g(\vec{r})$ possessing a Fourier transform

$$G(\vec{\beta})=[1/(2\pi)^3]\int g(\vec{r})\exp(-i\vec{\beta}\cdot\vec{r})\,d^3r, \quad (4.3)$$

which we need is the simple one

$$\sum_\tau{}'g(\vec{\tau})=\frac{(2\pi)^3}{\Omega}\sum_\gamma{}'G(\vec{\gamma}). \quad (4.4)$$

This may be written in the two forms

$$\sum_\tau{}'g(\vec{\tau})-\frac{1}{\Omega}\int g(\vec{r})\,d^3r=\frac{(2\pi)^3}{\Omega}\left(\sum_\gamma{}'G(\vec{\gamma})-\frac{1}{v}\int G(\vec{\beta})\,d^3\beta\right), \quad (4.5)$$

$$\sum_\tau{}'\left(g(\vec{\tau})-\frac{1}{\Omega}\int_\tau g(\vec{r})\,d^3r\right)-\frac{1}{\Omega}\int_0 g(r)\,d^3r=\frac{(2\pi)^3}{\Omega}\left[\sum_\gamma{}'\left(G(\vec{\gamma})-\frac{1}{v}\int_\gamma G(\vec{\gamma})\,d^3\beta\right)-\frac{1}{v}\int_0 G(\vec{\beta})\,d^3\beta\right], \quad (4.6)$$

which suggest relations that might hold, possibly with "correction terms", for less well-behaved functions. These rearrangements might be expected to be helpful in cases where $g(\vec{0})$ is infinite and/or the sum of the left side of Eq. (4.4) does not exist. Let us couple these somewhat vague notions to the basic idea of the Ewald procedure[6] as applied to the simple Lennard–Jones sums

$$L(k)=\sum_\tau{}'\tau^{-(3+k)},\quad 0<k. \quad (4.7)$$

The $L(k)$ converge, but slowly unless k is much larger than one, but one cannot apply the PSF directly to L, because the summand is not defined at the origin of the τ lattice. Ewald's procedure gets around this by first using a Laplace transform as in Eq. (2.2), interchanging summation and integration, and then using the PSF (or Jacobi identity) on the resulting theta functions. Thus we may view the Ewald procedure as a device for utilizing the power of the PSF for sums to which it cannot be directly applied.

Of course, a sum like $\Sigma_\tau{}'\tau^{-n}$, $1\leqslant n\leqslant 2$, does not converge (has difficulty at large τ) and its summand does not exist at the origin, but we see that Eq. (1.1) has the same form as the left side of Eq. (4.6). Thus we might expect Eq. (4.6) to hold, but we might also expect correction terms to arise since we are nearly "subtracting infinities" as we definitely would be doing if we were to use Eq. (4.5). In securing Eq. (4.1) I

have shown that Eq. (4.6) holds without corection terms for the cases $1 < n < 2$. For a tool I have used the Ewald procedure in what we may say is the direct way and then in an inverse way, being careful about the summation-integration interchanges both times. I expect this building on Ewald's procedure and this seeking possibly modified PSF's to be useful in treating other difficult sums.

Next I shall show that Eq. (4.6) holds when modified with a correction term for the cases $n = 1, 2$.

V. EWALD EVALUATION AND PSF FOR THE CASES $n = 1, 2$

From Ref. 2 we know that Eq. (2.7) does not hold for $n = 1$. We note further that because of this we cannot interchange the orders of summation and integration in the last form of Eq. (2.8) when $n = 2$. However, we have already discussed in Sec. III how Ewald evaluation is achieved for the case $n = 2$. For $n = 1$ Eq. (2.7) holds when modified with additive correction term on the right side, which we give momentarily. Since Ref. 7 gives the Ewald evaluation of the integral in Eq. (2.7) with $n = 1$, we only need here to discuss the correction term and give the PSF (modified) relating $F_{1,\tau}$ and $F_{2,\gamma}$. However, we do need to add to Ref. 2 the information concerning the larger set of correction terms discussed in Ref. 3.

Consider the case $n = 1$, which is associated with a WS[5] with a charge distribution

$$\rho(\vec{r}) = Q\left[\sum_{\tau}^{\infty}\delta(\vec{r}-\vec{\tau}) - 1/\Omega\right], \tag{5.1}$$

when the site of a point charge is taken as origin of coordinates. The energy K of one point charge interacting with all other charge is given by

$$K = Q^2 F_{1,\tau}, \tag{5.2}$$

which is a multivalued set of quantities with each number of the set corresponding to a different way of taking the summation limit in Eq. (1.1), i.e., how the uniform background charge is grouped with one or more point charges to form a composite summation unit and then how the composite units are grouped in a limiting process to cover all space. Reference 2 treats summation units formed from one point charge grouped with the uniform charge in a centered proximity cell or primitive cell and then uses finite nested groups of these at every step in the limiting process. Reference 3 extends the treatment to include larger cells of volume V and "displaced" cells centered about potential minima (maxima) if the point charges are positive (negative).

The results we need from Refs. 2 and 3 are

$$K = S + A, \quad A = Q\langle\Phi\rangle, \tag{5.3}$$

$$\langle\Phi\rangle = \frac{-2\pi}{3V}\int_V r^2\rho(r)\,d^3r, \tag{5.4}$$

where S is given by Q^2 times the right-hand sides of Eqs. (2.7) and (2.8) with n set equal to one. Prior to the appearance of Ref. 2, K was set equal to S, which lead me to call A a correction term. In Eq. (5.4) the $\rho(\vec{r})$ is given[3] for displaced cells not by Eq. (5.1) but by the same charge distribution appropriately modified to reflect an origin for $\vec{r}$ being taken at a minimum of potential (if the point charges are positive) instead of at the site of a (positive) charge.

Although $S \neq Q^2 F_{1,\tau}$ we see from the right hand sides of Eqs. (2.7) and (2.8) that

$$S = 4\pi\Omega^{-1}Q^2 F_{2,\gamma}, \tag{5.5}$$

If we further define $C = A/Q^2$, we finally have

$$F_{1,\tau} = 4\pi\Omega^{-1}F_{2,\gamma} + C, \tag{5.6}$$

reciprocally relating $F_{1,\tau}$ and $F_{2,\gamma}$, which is in the form of Eqs. (4.1) and (4.6) except for the modification introduced by the correction term C. Clearly, C and $F_{1,\tau}$ are sets of quantities while $F_{2,\gamma}$ is unique for a given lattice.

VI. SUMMARY

Equations (4.1) and (5.6) supply the reciprocal relations to be reported. For $1 < n < 2$ the relations are in the form of Eq. (4.6), a Poisson summation formula (PSF); for $n = 1, 2$ the relations are of the form of Eq. (4.6) with a correction term, i.e., a PSF with a correction term. Ewald evaluation is discussed in Secs. III and V.

In addition to any interest or importance these results for this class of near-Coulombic sums may possess for Wigner solid (WS) or near-WS's, these results provide examples for which modified PSF's hold although the basic functions are either ill behaved or are difficult to analyze. The methods I use, as laid out in Sec. III, involve a direct and an inverse use of Ewald techniques with a rigorous examination of all interchanges of summations and integrations. The methods should be useful for other classes of sums which like these appear to involve "subtraction of infinities" or which are not amenable to direct application of a PSF with assurance of validity.

APPENDIX A

Here I show that the summation of Eq. (2.5) can be passed inside the integration for the cases $1 < n \leqslant 2$, which is the basis for the proof of Eq. (2.7). The case $n = 1$, for which Eq. (2.7) does not hold without a correction term, has been analyzed for cubic lattices in Appendix A of Ref. 2 and for the general Bravais lattice in Sec. II of Ref. 2. The cases under consideration here possess no correction terms and are easier to analyze from the outset for Bravais lattices, but the dependence on n in the proof needs to be exhibited explicitly.

The first part of the present Sec. II parallels the first part of Appendix A of Ref. 2 with the present Eq. (2.6) being the same as the first form of Eq. (A9) of Ref. 2. This forms our starting point for investigating the possibility of interchanging the order of summation and integration in our Eq. (2.5).

Toward that end, write as before

$$a_\tau(t) = -\exp(-t\tau^2)\sum_{j=1}^{\infty}{}^{*}\sum_{m=0}^{j}\frac{(-i)^j}{m!(j-m)!} \times\left(\frac{1}{\Omega}\int_0(2\vec{x}\cdot\vec{\tau}t)^m(x^2t)^{j-m}\,d^3x\right), \tag{A1}$$

where the asterisk on the summation means that the terms $j = 0$ and $j = 1, m = 1$ have been omitted and where we have

used that

$$\int_0 \vec{x}\cdot\vec{\tau}\, d^3x = 0\,, \tag{A2}$$

holds for centered cells, proximity or primitive, in a Bravais lattice. Now with the understanding Eq. (A1) is replacing Eq. (2.6), we again write Eq. (2.5) for $F_{n,\tau}$

$$\bar{F}_{n,t} = \frac{1}{\Gamma(\frac{1}{2}n)} \sum_{\tau}{}' \int_0^\infty t^{(1/2)n-1} a_\tau(t)\, dt\,. \tag{A3}$$

To prove we can interchange the order of summation and integration, consider

$$\int_0^\infty t^{(1/2)n-1} \sum_{\tau}{}' a_\tau(t)\, dt$$
$$= \sum_{\tau\leqslant T}{}' \int_0^\infty t^{(1/2)n-1} a_\tau(t)\, dt + \int_0^\infty \sum_{\tau>T} a_\tau(t)\, dt\,, \tag{A4}$$

for which I shall show that the last term vanishes in the limit as $T\to\infty$.

To show that the second term in Eq. (A4) vanishes in this limit, let d be the maximum value of $|\vec{x}|$ in the centered cell at the origin. Then we have

$$|a_\tau(t)|$$
$$\leqslant \exp(-t\tau^2) \sum_{j=1}^{\infty}{}^{*} \sum_{m=0}^{j} \frac{1}{m!(j-m)!} (2d\tau t)^m (d^2 t)^{j-m}$$
$$= \exp(-t\tau^2)[\exp(2d\tau t + d^2 t) - 1 - 2d\tau t\ ]\,, \tag{A5}$$

and

$$\left|\int_0^\infty t^{(1/2)n-1} \sum_{\tau>T} a_\tau(t)\, dt\right| \leqslant \int_0^\infty t^{(1/2)n-1} \sum_{\tau<T} |a_\tau|\, dt$$
$$= \sum_{\tau>T} \int_0^\infty t^{(1/2)n-1} |a_\tau|\, dt\,, \tag{A6}$$

where the last step is by the monotone convergence theorem[8] of Lebesgue integration. The last integral is easily evaluated, and one obtains

$$\left|\int_0^\infty t^{(1/2)n-1} \sum_{\tau>T} a_\tau(t)\, dt\right|$$
$$\leqslant \sum_{\tau>T} \Gamma(\tfrac{1}{2}n)[(\tau-d)^{-n} - \tau^{-n} - nd\tau^{-(n+1)}]$$
$$= \sum_{\tau>T} O[\tau^{-(n+2)}]\,, \tag{A7}$$

which is vanishingly small as $T\to\infty$ because the lattice is three-dimensional and $1<n\leqslant 2$ implies that $3<(n+2)\leqslant 4$. Thus the last term in Eq. (A4) vanishes in the limit and we have that

$$\bar{F}_{n,\tau} = \frac{1}{\Gamma(\frac{1}{2}n)} \int_0^\infty t^{(1/2)n-1} \sum_{\tau}{}' a_\tau(t)\, dt\,, \tag{A8}$$

as required. Using Eq. (2.4) then yields Eq. (2.7).

[1] G. Weinreich, *Solids: Elementary Theory For Advanced Students* (Wiley, NY 1965).
[2] G.L. Hall, Phys. Rev. B **19**, 3921 (1979).
[3] G.L. Hall and T.R. Rice, to appear in Phys. Rev. B **15**.
[4] E.M. Stein and G. Weiss, *Fourier Analysis on Euclidean Spaces* (Princeton U., Princeton, New Jersey, 1971), p. 253.
[5] E.P. Wigner, Phys. Rev. **46**, 1002 (1934); Trans. Faraday Soc. **34**, 678 (1936).
[6] M. Born and K. Huang, *Dynamical Theory of Crystal Lattices* (Oxford U., Oxford, 1954), pp. 248, 388, and other references given in this book.
[7] R.A. Coldwell-Horsfall and A.A. Maradudin, J. Math. Phys. **1**, 395 (1960).
[8] K. Hoffman, *Analysis in Euclidean Space* (Prentice–Hall, Englewood Cliffs, New Jersey, 1975), p. 328.

Sum rules for zeros of polynomials. I

K. M. Case
The Rockefeller University, New York, New York 10021

(Received 7 May 1979; accepted for publication 26 October 1979)

It is shown that for polynomials satisfying differential equations of a particular form it is easy to generate sum rules for the powers of the zeros. All of the classical orthogonal polynomials are of this form. Examples are given for the Hermite, Laguerre, Tchebycheff, and Jacobi polynomials. In particular an explicit formula is given for the sums of all powers of Tchebycheff zeros. This same formula gives the sums for general Jacobi polynomials in the limit of large N.

I. INTRODUCTION

Many polynomials of interest for physical applications satisfy rather simple differential equations. This is true, for example, for all the classical orthogonal polynomials. Here we will show that with rather mild restrictions on the form of the differential equation it implies explicit formulas for sums of the powers of the zeros of the polynomials.

Our plan is the following: In Sec. II the differential equation is used to obtain certain relations involving sums over functions of the positions of the zeros. Section III shows how these relations imply recursion formulas from which the sums of powers of the zeros can be computed. In Sec. IV the first two such sums corresponding to an arbitrary order differential equation are given explicitly. The remaining sections are devoted to the classical orthogonal polynomials. Thus the lowest few sum rules are obtained for the Hermite, Laguerre, Tchybecheff, and Jacobi polynomials $(H_N,\ L_N^{\alpha},\ T_N,\ P_N^{(\alpha,\beta)})$.[1] It is then noted that an explicit simple formula can be obtained for all sums of powers of zeros of the Tchebycheff polynomials. This formula also gives an asymptotic formula $(N\to\infty)$ for all Jacobi polynomials. Finally a special sum rule that applies to some, but not all, of the classical polynomials is given.

II. BASIC RELATIONS

We assume our polynomials f_N satisfy a differential equation of the form

$$\sum_{i=0}^{n} g_i(x) f_N^{(i)}(x) = 0, \tag{1}$$

where the coefficient g_i of the ith derivative is a polynomial of degree no higher than i. Thus, explicitly

$$g_i(x) = \sum_{j=0}^{i} a_j^{(i)} x^j \tag{2}$$

with constant coefficients $a_j^{(i)}$.

Let us assume further that all the zeros of $f_N(x)$ are simple. Then

$$f_N(x) = (\text{const}) \prod_{l=1}^{N} (x - x_l). \tag{3}$$

Inserting this expression into the differential Eq. (1), dividing by $f_N^{(1)}(x)$ and evaluating at a zero x_{l_1} of f_N we obtain,

$$\sum_{i=2}^{n} (i!) \sum_{j=0}^{i} a_j^{(i)} \sum_{\neq}{}' \frac{(x_{l_1})^j}{(x_l - x_{l_2})(x_{l_1} - x_{l_3}),\ldots,(x_{l_1} - x_{l_i})} = -a_0^{(1)} - a_1^{(1)} x_{l_1}. \tag{4}$$

(Here $\Sigma'_{\neq}$ means a sum over all l_j except l_1 subject to the condition that all l are different.)

Now let us multiply Eq. (4) by $(x_{l_1})^r$ and sum over l_1 also. The result is

$$\sum_{i=2}^{n} (i!) \sum_{j=0}^{i} a_j^{(i)} J_{r+j}^{(i)} = -a_0^{(1)} y_r - a_1^{(1)} y_{r+1}. \tag{5}$$

The y_r are the rth moments of the zeros of f_N, i.e.,

$$y_r = \sum_{l=1}^{N} (x_l)^r \tag{6}$$

and

$$J_r^{(i)} = \sum_{\neq} \frac{(x_{l_1})^r}{(x_{l_1} - x_{l_2})(x_{l_1} - x_{l_2})\cdots(x_{l_1} - x_{l_i})}. \tag{7}$$

($\Sigma_{\neq}$ now means to sum over all l's subject to none of them being equal.) Equation (5) is our fundamental relation. The point here is that, as will be shown in the next section, the rather complicated looking $J_r^{(i)}$ are rather simply related to the y_s. Thus Eq. (5) is a recursion relation which permits us to calculate the y_s successively.

III. PROPERTIES OF THE $J_r^{(i)}$

Simple results for these are readily deduced from looking at the first few.

A. $J_r^{(2)}$

By definition

$$J_r^{(2)} = \sum_{\neq} \frac{(x_{l_1})^r}{(x_{l_1} - x_{l_2})}. \tag{8}$$

If we add and substract $(x_{l_2})^r$ in the numerator of this equation, we obtain

$$J_r^{(2)} = \sum_{\neq} \frac{[(x_{l_1})^r - (x_{l_2})^r]}{[x_{l_1} - (x_{l_2})]} + \sum_{\neq} \frac{(x_{l_2})^r}{(x_{l_1} - x_{l_2})}. \tag{9}$$

Now if the labels l_1, l_2 are interchanged in the last term on the right we see that this is just $-J_r^{(2)}$.

$$\therefore J_r^{(2)} = \frac{1}{2} \sum_{\neq} \frac{[(x_{l_1})^r - (x_{l_2})^r]}{(x_{l_1} - x_{l_2})}. \tag{10}$$

Clearly,

$$J_0^{(2)} = 0, \tag{11}$$

 0022-2488/80/040702-07$1.00

and

$$J_1^{(2)} = \tfrac{1}{2}\sum_{\neq} 1 = N(N-1)/2. \tag{12}$$

Further

$$J_2^{(2)} = \tfrac{1}{2}\sum_{\neq}(x_{l_1} + x_{l_2}) = \sum_{\neq}(x_{l_1}) = (N-1)\ y_1 . \tag{13}$$

Using the expansion

$$\frac{(x_{l_1})^r - (x_{l_2})^r}{x_{l_1} - x_{l_2}} = \sum_{s=0}^{r-1} (x_{l_1})^{r-1-s}(x_{l_2})^s ,$$

and separating out the highest moment it shows that for $r > 2$

$$J_r^{(2)} = (N - r/2)\, y_{r-1} + \tfrac{1}{2}\sum_{s=1}^{r-2} y_{r-1-s}\, y_s . \tag{14}$$

B. $J_r^{(3)}$

By definition

$$J_r^{(3)} = \sum_{\neq} \frac{(x_{l_1})^r}{(x_{l_1} - x_{l_2})(x_{l_1} - x_{l_3})} . \tag{15}$$

Using the identity

$$\frac{1}{(x_{l_1} - x_{l_2})(x_{l_1} - x_{l_3})} = \left[\frac{1}{x_{l_1} - x_{l_2}} - \frac{1}{x_{l_1} - x_{l_3}}\right]\frac{1}{(x_{l_2} - x_{l_3})} \tag{16}$$

shows that

$$J_r^{(3)} = \sum_{\neq} (x_{l_1})^r \times \left[\frac{1}{(x_{l_1} - x_{l_2})} - \frac{1}{(x_{l_1} - x_{l_3})}\right]\frac{1}{(x_{l_2} - x_{l_3})} . \tag{17}$$

But on interchanges, $l_2 \longleftrightarrow l_3$ we see the second term on the right here is the same as the first. Thus,

$$J_r^{(3)} = 2\sum_{\neq} \frac{(x_{l_1})^r}{(x_{l_1} - x_{l_2})(x_{l_2} - x_{l_3})} .$$

Now again add and substract $(x_{l_2})^r$ in the numerator. Then

$$J_r^{(3)} = 2\sum_{\neq} \frac{[(x_{l_1})^r - (x_{l_2})^r]}{(x_{l_1} - x_{l_2})(x_{l_2} - x_{l_3})} + 2\sum_{\neq} \frac{(x_{l_2})^r}{(x_{l_1} - x_{l_2})(x_{l_2} - x_{l_3})} .$$

Interchanging labels l_1 and l_2 in the second term on the right and comparing with Eq. (15) we see it is $-2J_r^{(3)}$. Hence

$$J_r^{(3)} = \frac{2}{3}\sum_{\neq} \frac{[(x_{l_1})^r - (x_{l_2})^r]}{(x_{l_1} - x_{l_2})(x_{l_2} - x_{l_3})} , \tag{18}$$

obviously $J_0^{(3)} = 0$. Further

$$J_1^{(3)} = \frac{2}{3}\sum_{\neq} \frac{1}{x_{l_2} - x_{l_3}} = 0, \tag{19}$$

since the integrand changes sign when we interchange labels l_2, l_3. For $r = 2$ we find

$$J_2^{(3)} = \frac{2}{3}\sum_{\neq} \frac{x_{l_1} + x_{l_2}}{(x_{l_2} - x_{l_3})} . \tag{20}$$

But $\sum_{\neq} x_{l_1}/(x_{l_2} - x_{l_3}) = 0$, since the sign changes on interchanging l_2, l_3, and

$$\sum_{\neq} \frac{x_{l_2}}{x_{l_2} - x_{l_3}} = \sum_{\neq} \frac{(x_{l_2} - x_{l_3})}{x_{l_2} - x_{l_3}} + \sum_{\neq} \frac{x_{l_3}}{x_{l_2} - x_{l_3}} = \sum_{\neq} 1 - \sum_{\neq} \frac{x_{l_2}}{x_{l_2} - x_{l_3}}$$

$$\therefore J_r^{(3)} = \tfrac{1}{3}\sum_{\neq} 1 = \tfrac{1}{3} N(N-1)(N-2). \tag{21}$$

Similarly we find

$$J_3^{(3)} = (N-1)(N-2)y_1 , \tag{22}$$

while for $r \geqslant 3$, $J_r^{(3)}$ is a combination of the y_s with $s \leqslant r-2$. In particular the highest moment y_{r-2} occurs linearly.

C. $J_r^{(4)}$

By similar manipulations it can be shown that

$$J_r^{(4)} \equiv \sum_{\neq} \frac{(x_{l_1})^r}{(x_{l_1} - x_{l_2})(x_{l_1} - x_{l_3})(x_{l_1} - x_{l_4})} = \sum_{\neq} \frac{[(x_{l_1})^r - (x_{l_2})^r]}{(x_{l_1} - x_{l_2})(x_{l_2} - x_{l_3})(x_{l_1} - x_{l_4})} . \tag{23}$$

From this it follows that $J_0^{(4)} = J_1^{(4)} = J_2^{(4)} = 0$ and

$$J_3^{(4)} = \tfrac{1}{4} N(N-1)(N-2)(N-3),$$
$$J_4^{(4)} = (N-1)(N-2)(N-3)y_1 . \tag{24}$$

Further for $r \geqslant 4$, $J_r^{(4)}$ depends only on the y_s for $s \leqslant r-3$. Indeed it depends *linearly* on y_{r-3}.

D. The general case

From our examples some general properties emerge.

(1) $J_r^{(i)} = 0, \quad 0 \leqslant r \leqslant n-2.$ (25)

(2) $J_{i-1}^{(i)} = \frac{1}{i}\prod_{t=1}^{i-1}(N-t) \equiv C^{(i)}(N).$ (26)

(3) $J_i^{(i)} = y_1 \prod_{t=1}^{i-1}(N-t) \equiv D^{(i)}(N)\, y_1 .$ (27)

(4) For $r \geqslant i$, $J_r^{(i)}$ depends only on the y_s for $s \leqslant r+1-i$. The dependence on y_{r+1-i} is linear.

We leave it to the reader to verify this for $J_r^{(5)}$. For this purpose the relevant representation (obtained as previously is

$$J_r^{(5)} = \frac{2}{5}\left(2\sum_{\neq} \frac{1}{(x_{l_2} - x_{l_3})(x_{l_2} - x_{l_4})} \frac{[(x_{l_1})^r - (x_{l_2})^r]}{(x_{l_1} - x_{l_2})(x_{l_1} - x_{l_5})} - \sum_{\neq} \frac{1}{(x_{l_2} - x_{l_3})(x_{l_2} - x_{l_4})} \frac{1}{(x_{l_1} - x_{l_5})}\right.$$

$$\times\left\{\frac{[(x_{l_1})^r-(x_{l_1})^r]}{x_{l_1}-x_{l_3}}-\frac{[(x_{l_1})^r-(x_{l_3})^r]}{x_{l_1}-x_{l_3}}\right\}\Bigg). \tag{28}$$

We also note one more general relation. This is

$$J^{(i)}_{i+1} = \left\{\left[N-\frac{(i+1)}{2}\right]y_2+\frac{(i-1)}{2}(y_1)^2\right\}\times\prod_{t=2}^{i-1}(N-t)$$
$$\equiv E^{(i)}(N)\,y_2+F^{(i)}(N)(y_1)^2. \tag{29}$$

IV. TWO GENERAL SUM RULES

Given the results so far we can readily write down two sum rules for polynomials which satisfy Eq. (1).

A. y_1

Let us put $r=0$ in Eq. (5). Using Eqs. (25) and (27) we have

$$\sum_{i=2}^{n}(i!)\{a^{(i)}_{i-1}J^{(i)}_{i-1}+a^{(i)}_i J^{(i)}_i\} = -a^{(1)}_0 y_0 - a^{(1)}_1 y_1 .$$

Or using Eqs. (25) and (26) and the fact that $y_0\equiv N$ we see that this is a linear equation for y_1 with the solution

$$y_1 = -a^{(1)}_0 N - \Sigma^n_{i=2} i!\,a^i_{i-1}\,C^{(i)}(N)/\times[a^{(1)}_1+\Sigma^n_{i=2} i!\,a^{(i)}_i D^{(i)}(N)]. \tag{30}$$

It is interesting to note that only the $a^{(i)}_i$ and $a^{(i)}_{i-1}$ enter.

B. y_2

If we put $r=1$ in Eq. (5), we see that we have terms linear in y_2, known constants $a^{(i)}_j$, $C^{(i)}(N)$, $D^{(i)}(N)$, $E^{(i)}(N)$, $F^{(i)}(N)$, and y_1. But now y_1 is known. Hence we can solve for y_2 with the result

$$y_2 = -a^{(1)}_0 y_1 - \sum_{i=2}^{n} i!\{a^{(i)}_{i-2}J^{(i)}_{i-1}+a^{(i)}_{i-1}J^{(i)}_i + a^{(i)}_i F^{(i)}_{(N)}(y_1)^2\}(a^{(1)}_1+\sum_{i=2}^{n}(i!)\,a^{(i)}_i E^{(i)}(N))^{-1}. \tag{31}$$

Here only $a^{(i)}_i$, $a^{(i)}_{i-1}$, $a^{(i)}_{i-2}$ play a role.

V. THE CLASSICAL ORTHOGONAL POLYNOMIALS

To see how the sum rules of the higher order work, let us specialize to the case $n=2$—a second order differential equation. Then the polynomials are described by only the five parameters $a^{(2)}_0$, $a^{(2)}_1$, $a^{(2)}_2$, $a^{(1)}_0$, $a^{(1)}_1$. Equation (5) becomes

$$2\{a^{(2)}_0 J_r + a^{(2)}_1 J_{r+1} + a^{(2)}_2 J_{r+2}\} = -a^{(1)}_0 y_r - a^{(1)}_1 y_{r+1} . \tag{32}$$

(Here we have dropped the superscript (2) on J_r since this will now *always* mean $J^{(2)}_r$.) For later use we summarize the first few here using Eqs. (11)–(14). We have: $J_0=0$,

$J_1 = N(N-1)/2$, $J_2=(N-1)y_1$,

$J_3=(N-\frac{3}{2})y_2+\frac{1}{2}(y_1)^2$,

$J_4=(N-2)y_3+y_1y_2$,

$J_5=(N-\frac{5}{2})y_4+y_1y_3+(y_2)^2/2$,

$J_6=(N-3)y_5+y_1y_4+y_2y_3$,

$J_7=(N-\frac{7}{2})y_6+y_5y_1+y_4y_2+(y_3)^2/2$,

$J_8=(N-4)y_7+y_6y_1+y_5y_2+y_4y_3$,

$J_9=(N-\frac{9}{2})y_8+y_9y_1+y_6y_2+y_5y_3+(y_4)^2/2$,

$J_{10}=(N-5)y_9+y_8y_1+y_7y_2+y_6y_3+y_5y_4$,

$J_{11}=(N-\frac{11}{2})y_{10}+y_9y_1+y_8y_2+y_7y_3+y_6y_4+(y_5)^2/2$.

Now let us specialize further to various of the classical orthogonal polynomials.[2]

A. Hermite polynomials $H_N(x)$

Here $a^{(2)}_0=1$, $a^{(1)}_1=-2$, $a^{(2)}_1=a^{(2)}_2=a^{(1)}_0=0$. Equation (32) is

$$y_{r+1}=J_r. \tag{33}$$

With $r=0$ we obtain

$$y_1=J_0=0. \tag{34}$$

(This is hardly surprising. We know the Hermite polynomials are odd or even. Hence the zeros are symmetrically distributed around the origin. All odd sums—y_{2s+1}—are zero. Our method, of course, reproduces this.) Hence let us consider only even moments.

With

$r=1$: $y_2=J_1=N(N-1)/2$;

$r=3$: $y_4=J_3=(N-\frac{3}{2})y_2=[N(N-1)/2](N-\frac{3}{2})$;

$r=5$: $y_6=J_5=(N-\frac{5}{2})y_4+(y_2)^2/2$;

or $y_6=[N(N-1)/2](N-\frac{3}{2})(N-\frac{5}{2})+N^2(N-1)^2/8$.

Continuing (with patience) an arbitrary number of these can be computed.

B. Laguerre polynomials $L^{(\alpha)}_N(x)$

Now $a^{(2)}_0=a^{(2)}_2=0$, $a^{(2)}_1=1$, $a^{(1)}_0=\alpha+1$, $a^{(1)}_1=-1$. Equation (32) becomes

$$y_{r+1}=2J_r+(\alpha+1)y_r.$$

With $r=0$ we obtain

$$y_1=2J_1+(\alpha+1)y_0=N(N-1)+N(\alpha+1)$$

or

$$y_1=N(N+\alpha).$$

Similarly with $r=1$ we obtain

$$y_2=2J_2+(\alpha+1)y_1 = (2N+\alpha-1)y_1=(2N+\alpha-1)N(N+\alpha).$$

We have also calculated y_3 and y_4. The results are

$$y_3=(2N+\alpha-2)(2N+\alpha-1)N(N+\alpha)+N^2(N+\alpha)^2$$

and

$$y_4=(2N+\alpha-3)(2N+\alpha-2)(2N+\alpha-1)N(N+\alpha)+(2N+\alpha-3)N^2(N+\alpha)^2+2(2N+\alpha-1)N^2(N+\alpha)^2.$$

C. Tchebycheff polynomials $T_N(x)$

$$a^{(2)}_0=1=a^{(1)}_1,\quad a^{(2)}_2=-1,\quad a^{(2)}_1=a^{(1)}_0=0.$$

The equation is

$$J_{r+2}+y_{r+1}/2=J_r .$$

For $r=0$ we get

$$J_2 + y_1/2 = J_0$$

or

$$(N-\tfrac{1}{2})\,y_1 = 0,$$

i.e., $y_1 = 0$. Again this is expected by symmetry. All y_{2s+1} are zero.

For reasons which will become apparent later we have calculated a fair number of the evens sums. The results are

$$y_2 = N/2, \quad y_4 = 3N/8, \quad y_6 = 5N/16, \quad y_8 = 35N/128,$$
$$y_{10} = 63N/256, \quad y_{12} = 231N/2^{10}, \quad y_{14} = 459N/2^{11}.$$

D. Jacobi polynomials $P_N^{(\alpha,\beta)}$

$$a_0^{(2)} = 1 = -a_2^{(2)} \qquad a_0^{(1)} = 0,$$
$$a_0^{(1)} = \beta - \alpha, \qquad a_1^{(1)} = \alpha + \beta + 2,$$

The relation is

$$2J_{r+2} + (\alpha+\beta+2)\,y_{r+1} = 2J_r + (\beta-\alpha)\,y_r\,. \quad (35)$$

With $r=0$ we obtain

$$2J_2 + (\alpha+\beta+2)\,y_1 = 2J_0 + (\beta-\alpha)\,y_0,$$

i.e., $(2N+\alpha+\beta)\,y_1 = N(\beta-\alpha)$ or

$$y_1 = N(\beta-\alpha)/(2N+\alpha+\beta).$$

Similarly we find

$$y_2 = [N(N-1) + (\beta-\alpha)\,y_1 - (y_1)^2]/(2N+\alpha+\beta-1).$$

Clearly when $\beta \neq \alpha$ the higher order sum rules will be very messy.

However, there is a limiting case in which things become simple. Consider the case of large N and sum rules for y_n where $n \ll N$. The coefficient of y_{r+1} in J_{r+2} is $[N-(r+2)/2]$. This we can replace by N. Similarly we can drop the term $(\alpha+\beta+2)\,y_{r+1}$ in Eq. (35). This then becomes

$$J_{r+2} = J_r + [(\beta-\alpha)/2]\,y_r\,. \quad (36)$$

We now calculate the moments to highest leading order in N.

$$r=0 \quad J_2 \sim N y_1 = J_0 + [(\beta-\alpha)/2]\,y_0 = [(\beta-\alpha)/2]\,N$$
$$\therefore y_1 \cong (\beta-\alpha)/2, \quad (37)$$

$$r=1 \quad J_3 \sim N y_2 = J_1 + [(\beta-\alpha)/2]\,y_1 \sim N^2/2$$
$$\therefore y_2 \approx N/2, \quad (38)$$

$$r=2 \quad J_4 \sim N y_3 + y_1 y_2 = J_2 + [(\beta-\alpha)/2]\,y_2$$
$$\approx N y_1 + y_1 y_2$$
$$\therefore y_3 \approx y_1. \quad (39)$$

Continuing in this way we readily find that to first approximation all odd y are equal, i.e.,

$$y_{2s+1} \approx y_1 \approx (\beta-\alpha)/2. \quad (40)$$

The equations for the even moments become uncoupled from the odd moments. They can be described so: Let $y_{2n} = Z_n N$, $Z_0 = 1$, $Z_1 = \frac{1}{2}$, and generally,

(1) if n is odd,

$$Z_n = \sum_{s=1}^{n/2-1/2} Z_{n-s}(Z_{s-1} - Z_s) + \frac{(Z_{(n-1)/2})^2}{2}; \quad (41)_1$$

(2) if n is even,

$$Z_n = \sum_{s=1}^{n/2-1} Z_{n-s}(Z_{s-1} - Z_s) + Z_{n/2} \times \left[Z_{n/2-1} - \frac{Z_{n/2}}{2}\right] \cdots .$$

Using these equations we have calculated the first few Z's. The results

$$Z_1 = \tfrac{1}{2}, \quad Z_2 = \tfrac{3}{8}, \quad Z_3 = \tfrac{5}{16}, \quad Z_4 = \tfrac{35}{128},$$
$$Z_5 = \tfrac{63}{256}, \quad Z_6 = 231/2^{10}, \quad Z_7 = 429/2^{11}. \quad (42)$$

VI. THE COMPLETE SET OF SUM RULES FOR THE $T_N(x)$

Comparing the sum rules we have calculated for the even moments of the zeros of the Tchebycheff polynomials and the even moments for the large N Jacobi polynomials, we see two very strange things.

(1) The numerical coefficients are identical.

(2) In both cases the moments are just proportional to the first power of N. (Neither higher nor lower powers occur.) That this is so for the Jacobi case is obvious. We have only carried the calculation through to the highest power of N—which is indeed 1. What is peculiar is that this is true for the Tchebycheff case.

If we look in detail, it becomes clear that the Tchebycheff moments can have the form

$$y_{2n} = Z_n N, \quad (43)$$

where the Z_n satisfy Eqs. (41) if and only if

$$Z_n = [(2n-1)/2n\,]\,Z_{n-1}\,. \quad (44)$$

Of course this equation implies

$$Z_n = \prod_{t=1}^{n} \frac{(2t-1)}{2t}. \quad (45)$$

Now given the explicit form for Z_n it is simple to see if Eqs. (41) are satisfied. They are indeed. Hence we have the rather remarkable result that for the Tchebycheff polynomials

$$y_{2n} = \sum_l (x_l)^{2n} = N \prod_{t=1}^{n} \frac{(2t-1)}{2t}. \quad (46)$$

Similarly we now know that in the large N limit the sum Jacobi polynomials are

$$y_{2n+1} = (\beta-\alpha)/2, \quad (47)$$

while the even ones are given by Eq. (46).

It is rather amusing to see just what the strange set of numbers we have encountered really are. Thus,

$$\frac{1}{2} = \frac{1}{2}, \quad \frac{3}{8} = \frac{1\cdot3}{2\cdot4}, \quad \frac{5}{16} = \frac{1\cdot3\cdot5}{2\cdot4\cdot6},$$
$$\frac{35}{128} = \frac{1\cdot3\cdot5\cdot7}{2\cdot4\cdot6\cdot8}, \quad \frac{63}{256} = \frac{1\cdot3\cdot5\cdot7\cdot9}{2\cdot4\cdot6\cdot8\cdot10},$$
$$\frac{231}{2^{10}} = \frac{1\cdot3\cdot5\cdot7\cdot9\cdot11}{2\cdot4\cdot6\cdot8\cdot10\cdot12}, \quad \frac{429}{2^{11}} = \frac{1\cdot3\cdot5\cdot7\cdot9\cdot11\cdot13}{2\cdot4\cdot6\cdot8\cdot10\cdot12\cdot14}.$$

VII. AN ADDITIONAL SUM RULE

For the case when the differential equation for our poly-

nomials is of order two (the classical case) there can be an additional sum rule. Thus, suppose $g_2(x)$ is not constant. If we divide Eq. (4) by $g_2(x_l)$ we obtain

$$2\sum_{l_2\neq l_1}\frac{1}{x_{l_1}-x_{l_2}}=\frac{-a_0^{(1)}-a^{(1)}x_{l_1}}{g_2(x_{l_1})} \tag{48}$$

Summing this over l_1 gives

$$\sum_l \frac{a_0^{(1)}+a^{(1)}x_l}{g_2(x_l)}=0, \tag{49}$$

since the summand on the left of Eq. (48) is antisymmetric. This statement can be empty. Thus in the Tchebycheff case the result states

$$\sum_l \frac{x_l}{1-x_l^2}=0,$$

which is obvious since the xeros are symmetrical around the origin. However, there are cases when the relation is nontrivial. We give two examples.

A. Laguerre polynomials $L_N^{\alpha}(x)$

The relation of Eq. (49) is

$$\sum_l \frac{\alpha+1-x_l}{x_l}=0$$

or

$$y_{-1}\equiv\sum_l\frac{1}{x_l}=\frac{1}{\alpha+1}\sum_l 1=\frac{N}{\alpha+1}. \tag{50}$$

B. Jacobi polynomials $P_N^{(\alpha,\beta)}(x),(\alpha\neq\beta)$

Then

$$(\beta-\alpha)\sum_l\frac{1}{1-x_l^2}=(\alpha+\beta+2)\sum_l\frac{x_l}{1-x_l^2}. \tag{51}$$

VIII. CONCLUSION

It has been found that for polynomials satisfying particular forms of differential equations sum rules for powers of the zeros can be found. These include *all* of the classical orthogonal polynomials. In addition for the latter the sum rules are particularly easy to obtain.

Especially striking are the explicit sum rules for all powers of zeros of Tchebycheff polynomials. These also imply such for zeros of Jacobi polynomials of large order.

ACKNOWLEDGMENTS

I would like to thank the Institute for Advanced Study for hospitality during part of the time this work was being done. This work was supported by the National Science Foundation under Grant No. NSF MCS78 20455.

APPENDIX

The results we have obtained are closely related to properties of the zeros of classical polynomials recently discussed by Ahmed *et al.*[3] The methods we have used show very clearly why their relations hold and ways that they can be generalized. We sketch this here.

Consider any polynomial with simple zeros which satisfies the second-order differential equation

$$g_2(x)\frac{d^2y}{dx^2}+g_1(x)\frac{dy}{dx}+g_0(x)y=0,$$

where g_2 and g_1 are polynomials of degree 2 and 1, respectively. Thus,

$$g_2(x)=a_0^{(2)}+a_1^{(2)}x+a_2^{(2)}x^2, \tag{A1}$$

and

$$g_1(x)=a_0^{(1)}+a_1^{(1)}x.$$

(This class includes all the classical orthogonal polynomials.) Then we have the *theorem*:

$$[g_2(x_l)]^s\sum_{m\neq s}(x_l-x_m)^{-s}=f_s(x_l),\quad s=1,2,\cdots, \tag{A2}$$

where the $f_s(x_l)$ are simple polynomials of degree at most s.

The proof is constructive. It shows us how to find explicit forms for the f_s.

The first of these relations has been given in our main text. It is

$$g_2(x_l)\sum_l(x_l-x_m)^{-1}=f_1(x_l), \tag{A3}$$

where

$$f_1(x_l)=-\tfrac{1}{2}g_1(x_l).$$

Since this equation served to determine all sums of powers of the x_l's it, at least implicitly, determines the x_l and hence all their properties. In particular *all* relations of the form of Eq. (A2) should follow from Eq. (A3). This is indeed true. In essence what we find is that if we take the sth power of Eq. (A3) and separate out terms we get Eqs. (A2).

More generally we can proceed so: Suppose we know all the f up to and including f_s. Then if $r+t=s+1$ we can find f_{s+1} by separating out terms in the identity

$$[g_2(x_l)]^{r+t}\sum_{\substack{m\neq l\\ n\neq l}}(x_l-x_m)^{-s}(x_l-x_n)^{-t}=f_r(x_l)f_t(x_l), \tag{A4}$$

and using Eqs. (A2) for $f_1,f_2,\ldots,f_s$.

The simplest, but not necessarily the most efficient, approach is to compute the f successively. Thus, suppose Eqs. (A2) are known for $f_1,f_2,\ldots,f_s$. Multiply Eq. (A2) by Eq.(A3) we obtain

$$[g_2(x_l)]^{s+1}\sum_{\substack{m\neq l\\ n\neq l}}(x_l-x_m)^{-s}(x_l-x_n)^{-1}=f_1(x_l)f_s(x_l). \tag{A5}$$

If we separate out the term $n=m$, we find

$$f_{s+1}(x_l)+T_0^{(s)}(x_l)=f_1(x_l)f_s(x_l), \tag{A6}$$

where we define[4]

$$T_r^{(s)}\equiv[g_2(x_l)]^{s+1}\sum_{\neq}(x_l-x_m)^{s-r}(x_l-x_n)^{-1}\times(x_m-x_n)^{-r}. \tag{A7}$$

The $T_r^{(s)}$ satisfy very simple recursion relations. Thus, using the identity of Eq. (16) it follows that

$$T_r^{(s)} = R_r^{(s)} - T_{r+1}^{(s)}, \tag{A8}$$

with

$$R_r^{(s)} = [g_2(x_l)]^{s+1} \sum_{\neq} (x_l - x_m)^{-(s-r)}(x_m - x_n)^{-(r+1)}. \tag{A9}$$

The recursion relation of Eq. (A8) terminates rather differently depending on the parity of s.

(1) If s is even:

$$T_{s-1}^{(s)} \equiv [g_2(x_l)]^{s+1} \sum_{\neq} (x_l - x_m)^{-1}(x_l - x_n)^{-1} \times (x_m - x_n)^{s-1} = 0, \tag{A10}$$

since the sign changes when we interchange the dummy labels m and n.

(2) If s is odd:

$$T_{s-1}^{(s)} = [g_2(x_l)]^{s+1} \sum_{\neq} (x_l - x_m)^{-1}(x_l - x_n)^{-1} \times (x_m - x_n)^{-(s-1)}$$

$$= [g_2(x_l)]^{s+1} \sum_{\neq} \{(x_l - x_m)^{-1} - (x_l - x_n)^{-1}\} \times (x_m - x_n)^{-s}$$

$$= 2R_{s-1}^{(s)}. \tag{A11}$$

Thus from Eqs. (A6)–(A11) we see that calculating $f_{s+1}(x_l)$ is reduced to finding the $R_r^{(s)}$. But these are readily reduced to polynomials in x_l. Indeed

$$R_r^{(s)} = [g_2(x_l)]^{s+1} \sum_{\substack{m\neq l \\ n\neq m}} (x_l - x_m)^{-(s-r)}(x_m - x_n)^{-(r+1)} - [g_2(x_l)]^{s+1} \sum_{m\neq l} (x_l - x_m)^{-(s-r)} \times (x_m - x_l)^{-(r+1)},$$

where we have added and subtracted the term $n = l$ in Eq. (A9). We note that the second term on the right is just $(-1)^r f_{s+1}(x_l)$. In the first term let us put $g_2(x_l)^{r+1}$ under the summation sign and add and subtract $g_2^{r+1}(x_m)$. Our result is

$$R_r^{(s)} = (-1)^r f_{s+1}(x_l) + [g_2(x_l)]^{s-r} \sum_{\substack{m\neq l \\ n\neq m}} \frac{\{[g_2(x_l)]^{r+1} - [g_2(x_m)]^{r+1}\}}{(x_l - x_m)^{s-r}(x_m - x_n)^{r+1}} + [g_2(x_l)]^{s-r} \sum_{\substack{m\neq l \\ n\neq m}} \frac{[g_2(x_m)]^{r+1}}{(x_l - x_m)^{s-r}(x_m - x_n)^{r+1}}.$$

Now using Eq. (A2) we can do the sum over m in the last term. This yields

$$R_r^{(s)} = (-1)^r f_{s+1}(x_l) + [g_2(x_l)]^{s-r} \sum_{\substack{m\neq l \\ n\neq m}} \frac{\{[g_2(x_l)]^{r+1} - [g_2(x_m)]^{r+1}\}}{(x_l - x_m)^{s-r}(x_m - x_n)^{r+1}} + [g_2(x_l)]^{s-r} \sum_{m\neq l} \frac{f_{r+1}(x_m)}{(x_l - x_m)^{s-r}}.$$

(It is readily checked that all terms in the sums here are polynomials.)

Combining our equations together we now have f_{s+1} expressed in terms of those of lower order. Thus:

(1) *s even*

$$sf_{s+1}(x_l) = f_1(x_l) f_s(x_l) - \sum_{r=0}^{s-2} (-1)^r [g_2(x_l)]^{s-r} \sum_{\substack{m\neq l \\ n\neq m}} \frac{\{[g_2(x_l)]^{r+1} - [g_2(x_m)]^{r+1}\}}{(x_l - x_m)^{s-r}(x_m - x_n)^{r+1}} - \sum_{r=0}^{s-2} (-1)^r [g_2(x_l)]^{s-r} \sum_{m\neq l} \frac{f_{r+1}(x_m)}{(x_l - x_m)^{s-r}}.$$

(2) *s odd*

$$(s+2) f_{s+1}(x_l) = f_1(x_l) f_s(x_l) - \sum_{r=0}^{s-2} (-1)^r [g_2(x_l)]^{s-r} \sum_{\substack{m\neq l \\ n\neq m}} \frac{\{[g_2(x_l)]^{r+1} - [g_2(x_m)]^{r+1}\}}{(x_l - x_m)^{s-r}(x_m - x_n)^{r+1}} - \sum_{r=0}^{s-2} (-1)^r [g_2(x_l)]^{s-r} \times \sum_{m\neq l} \frac{f_{r+1}(x_m)}{(x_l - x_m)^{s-r}} - 2g_2(x_l) \sum_{\substack{m\neq l \\ n\neq m}} \frac{\{[g_2(x_l)]^s - [g_2(x_m)]^s\}}{(x_l - x_m)(x_m - x_n)^s} - 2g_2(x_l) \sum_{m\neq l} \frac{f_s(x_m)}{x_l - x_m}.$$

When the coefficients in Eq. (A1) are quite general, the computation of the f_s can be somewhat tedious. However, we have done this for the first three. The result is

$$f_1(x_l) = -g_1(x_l)/2,$$

$$3f_2(x_l) = -[f_1(x_l)]^2 - a_1^{(1)}(N-1) g_2(x_l) + 2[a_1^{(2)} + 2a_2^{(2)}(x_l)] f_1(x_l) - a_2^{(2)} N(N-1) g_2(x_l),$$

$$2f_3(x_l) = -\tfrac{1}{2}a_1^{(1)} f_1(x_l) g_2(x_l) - \tfrac{1}{2}(a_1^{(2)} + 2a_2^{(2)} x_l) f_1^2(x_l) - 3f_2(x_l) - a_2^{(2)} f_1(x_l) g_2(x_l).$$

Of course when we specialize to particular polynomials the expressions become much simpler. The simplest of these seem to be the Hermite polynomials. Then the formulas are:

(1) *s even*

$$sf_{s+1}(x_l) = f_1(x_l)f_s(x_l) - \sum_{r=0}^{s-2}(-1)^r \sum_{m\neq l} \frac{f_{r+1}(x_m)}{(x-x_m)^{s-r}};$$

(2) *s odd*

$$(s+2)f_{s+1}(x_l) = f_1(x_l)f_s(x_l) - \sum_{r=0}^{s-2}(-1)^r \sum_{m\neq l} \frac{f_{r+1}(x_l)}{(x-x_m)^{s-r}} - 2\sum_{m\neq l}\frac{f_s(x_m)}{x_l - x_m}.$$

The first few f_s are then

$$f_1(x_l) = x,$$
$$f_2(x_l) = \tfrac{1}{3}[2(N-1) - (x_l)^2],$$
$$f_3(x_l) = x_l/2,$$
$$f_4(x_l) = [f_2(x_l)/5][2 + f_2(x_l)],$$
$$f_5(x_l) = (x_l/6)\{1 + f_2(x_l)\}.$$

Two remarks:

(A) Combining the results of this Appendix with those of the main text we obtain new sum rules. Thus summing Eq. (A2) with respect to l we find

$$\sum_{\substack{ml \\ m\neq l}} \frac{[g_2(x_l)]^s}{(x_l - x_m)^s} = \sum_l f_s(x_l), \quad s = 1,2,\cdots. \tag{A12}$$

However, since the $f_s(x_l)$ are polynomials, we know the sum on the right side.

For example, for the Hermite polynomials we have seen that

$$\sum_l (x_l)^2 = N(N-1)/2,$$
$$\sum_l (x_l)^4 = [N(N-1)/2](N - \tfrac{3}{2}). \tag{A13}$$

This gives the two sum rules:

$$\sum_{\substack{ml \\ m\neq l}} (x_l - x_m)^{-2} = N(N-1)/2$$

and

$$\sum_{\substack{ml \\ m\neq l}} (x_l - x_m)^{-4} = [N(N-1)/2](N - \tfrac{7}{10}).$$

(B) While we have restricted ourselves here to polynomials, it is clear that many of our results can be extended to more general functions. Thus while we have used the special differential Eq. (A1), we have not used the fact that $g_0(x)$ is constant. Thus if the solution of the differential equation is an integral function with simple zeros essentially all our methods remain unchanged. A typical example is given by the Bessel functions.[3]

ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation under Grant No. NSF MCS78 20455.

[1] We use the notation of Milton Abramowitz and Irene A. Stegun, *Handbook of Mathematical Function* (U.S. Department of Commerce, National Bureau of Standards, Appl. Math. Series 55, 1964).

[2] The form chosen for the differential equations is that given in Ref. 1 on page 781.

[3] S. Ahmed, M. Brushi, F. Calogero, M.A. Olshanetsky, and A.M. Perelomov (Preprint 10, October, 1978, submitted to Nuovo Cimento.)

[4] By $\Sigma_{\neq}$ we mean here $\sum_{\substack{m\neq l \\ n\neq l \\ m\neq n}}$.

Sum rules for zeros of polynomials. II

K. M. Case
The Rockefeller University, New York, New York 10021

(Received 4 June 1979; accepted for publication 26 October 1979)

Sum rules for zeros of polynomials which satisfy differential equations with polynomial coefficients of order no higher than the derivative they multiply are extended. First, higher order coefficients are considered. Second, sum rules for negative powers are obtained. Finally, the results are extended to a class of integral functions.

I. INTRODUCTION

Previously[1] it has been shown that under simple (but widely applicable) conditions it is possible to obtain sum rules for the powers of the zeros of polynomials by completely elementary means. Explicitly the assumptions were:

(1) The zeros are simple.

(2) The polynomials satisfy a differential equation of second or higher order.

(3) The coefficients of the ith derivative is a polynomial of degree not greater than i.

Here we look at what happens when the assumptions are relaxed. While it will be apparent that our results are quite general, we will for simplicity (and since it is true for most of the obvious applications) restrict most of the discussion to the case where the differential equation is of second order.

The plan is as follows: In Sec. II we briefly summarize the results of Ref. 1, denoted as (I). The following section is devoted to the Lamé polynomials. This is one of the simplest set of polynomials in which the coefficients of the various derivatives are of higher order than the derivative. It will be shown that the sums of the zeros are all determined in terms of one sum. A discussion of methods to determine this remaining sum is deferred until later. Next we want to generalize our results to integral functions with simple Weierstrass expansions. Since the zeros of these have infinity as a point of accumulation the analog of the formulas of Sec. II would all be infinite. Hence in Sec. IV formulas for the sums of inverse powers of the zeros are developed. The results for the classical orthogonal polynomials are somewhat between those of Secs. II and III. Thus for the Hermite, Laguerre, Tchybecheff, Legendre, and Gegenbauer polynomials the sums of the inverse powers of the zeros are all completely determined from our fundamental identity. However, for the general Jacobi polynomial one sum still has to be found. Again discussion of the way to find this is deferred.

In Sec. V integral functions are discussed. The Bessel and Trigonometric functions are dealt with in detail. Again one sum remains to be determined. It is shown how to do this.

Section VI is devoted to more complicated sum rules. Applied to the trigonometric functions this yields many amusing relations involving sums over inverse powers of differences of squares of integers.

II. REVIEW

Assume the polynomials satisfy the differential equation

$$g_2(x)f^{(2)}(x) + g_1(x)f^{(1)}(x) + g_0(x)f = 0, \tag{1}$$

where

$$g_2(x) = a_0^{(2)} + a_1^{(2)}x + a_2^{(2)}x^2,$$

$$g_1(x) = a_0^{(1)} + a_1^{(1)}x,$$

$$g_0 = \text{const}.$$

If x_l, $l = 1, 2, \ldots, N$, denote the zeros of f we have

$$f = C\prod_{m=1}^{N}(x - x_m). \tag{2}$$

Note: The zeros are all distinct. If not f, and $f^{(1)}$ would both vanish at x_l and from Eq. (1) it would follow that f is identically zero. Inserting this expansion for f in Eq. (1), evaluating at x_l and dividing by $f^{(1)}(x_l)$ yields

$$g_2(x_l)\sum_{m\neq l}\frac{1}{x_l - x_m} = \frac{-g_1(x_l)}{2}. \tag{3}$$

Multiplying by $(x_l)^r$ and summing over l shows that

$$a_0^{(2)}J_r + a_1^{(2)}J_{r+1} + a_2^{(2)}J_{r+2} = (-a_0^{(1)}y_r - a_1^{(1)}y_{r+1})/2. \tag{4}$$

Here

$$J_r = \sum_{\substack{ml\\ m\neq l}}\frac{(x_l)^r}{x_l - x_m},$$

and

$$y_r = \sum_{l=1}^{N}(x_l)^r,$$

The main point here is that J_r is expressible in terms of y_{r-1}, and y_s, $s < r-1$. Thus $J_0 = 0$, $J_1 = N(N-1)/2$, $J_2 = (N-1)y_1$, and for $r > 2$

$$J_r = (N - r/2)y_{r-1} + \tfrac{1}{2}\sum_{s=1}^{r-2}y_{r-1-s}y_s, \tag{5}$$

This folows from the identity

$$J_r = \sum_{m\neq 1}\frac{(x_l)^r - (x_m)^r + (x_m)^r}{x_l - x_m}$$

$$= \sum_{m\neq 1}\frac{(x_l)^r - (x_m)^r}{x_l - x_m} - J_r$$

or

 0022-2488/80/040709-06$1.00

$$J_r = \tfrac{1}{2}\sum_{\substack{ml \\ m\neq l}} \frac{(x_l)^r - (x_m)^r}{x_l - x_m}, \tag{6}$$

and explicitly doing the division. Thus Eqs. (4) are a set of recursion relations which yield the sum rules.

III. LAMÉ POLYNOMIALS

These polynomials[2] are a relatively simple set which satisfy second-order differential equations but not the restriction on the coefficients that we have assumed. Thus consider the equation

$$\frac{d^2f}{dz^2} + \{h - (n(n+1)k^2 sn^2(z,k)\}f = 0, \tag{7}$$

where sn is the Jacobian elliptic function with modulus k. The variable change $x = sn^2 z$ reduces Eq. (7) to

$$g_2(x)f^{(2)} + g_1(x)f^{(1)} + g_0(x)f = 0, \tag{8}$$

where now

$$g_2(x) = 4x(1-x))1 - k^2x),$$
$$g_1(x) = 2\{3k^2x^2 - 2(1+k^2)x + 1\},$$

and

$$g_0(x) = h - n(n+1)k^2x,$$

Thus here the coefficient polynomials are each of one degree higher than the derivative they multiply.

The Lamé polynomials are defined as follows: We look for solutions of Eq. (8) in the form

$$f(x) = x^{\rho/2}(1-x)^{\sigma/2}(1-k^2x)^{\tau/2}F_N(x). \tag{9}$$

Here F_N is to be a polynomials of degree N. There are eight possibilities since ρ, σ, τ can individually take on the values 0 and 1. We will consider only the two simplest, namely

$$\sigma = \tau = 0.$$

A number of questions must be answered. What are the possible values of n (given ρ, σ, τ)? Given n what is N? What is h?

A. $\rho = \sigma = \tau = 0$

Here $F_N(x)$ itself satisfies Eq. (7). Proceeding as in Sec. (II) we obtain

$$k^2J_{r+3} - (1+k^2)J_{r+2} + J_{r+1} = -\tfrac{1}{4}[3k^2y_{r+2} - 2(1+k^2)y_{r+1} + y_r]. \tag{10}$$

Clearly this relation, by itself, cannot completely determine the y_s. Indeed it does not even contain h or n, what does it determine? Let us put $r = 0$ in Eq. (10). Then

$$k^2J_3 - (1+k^2)J_2 + J_1 = -\tfrac{1}{4}\{3k^2y_2 - 2(1+k^2)y_1 + N\}. \tag{11}$$

But from Sec. II we find: $J_1 = N(N-1)/2$, $J_2 = (N-1)y_1$, $J_3 = (N - 3/2)y_2 + \tfrac{1}{2}(y_1)^2$. Thus Eq. (11) gives us y_2 in terms of y_1 and N. Similarly with $r = 1$, Eq. (10) yields y_3 in terms of N, y_1, and y_2 (which is now known in terms of N, y_1). Proceeding we see that Eq. (10) gives all the y_s in terms of N and y_1.

To determine N and y_1, we note that F can be written in the form

$$F_N = C\prod_{l=1}^{N}(x - x_l). \tag{12}$$

To the terms of the two highest powers in x, we have

$$F_N = C\{x^N - ax^{N-1} + \cdots\} \tag{13}$$

with

$$a = \sum_{l=1}^{N} x_l \equiv y_1. \tag{14}$$

Substituting the expression for F_N in the differential equation and equating to zero the coefficient of x^{N+1} yields

$$N = n/2, \tag{15}$$

i.e., n is even and N is one half of it. Similarly if we equate to zero the coefficient of x^N we obtain

$$y_1 = [h - 4N^2(1+k^2)]/[2k^2(1-4N)]. \tag{16}$$

Therefore if h and n are given we have all the sums of positive powers of the zeros. However, one of the main problems for the Lamé polynomials is to determine the possible values of h. For this we need slightly more information about the polynomials. This would be along the lines discussed in Sec. V.

B. $\rho = 1, \sigma = \tau = 0$

Here F_N satisfies the equation

$$xg_2F'' + (xg_1 + g_2)F' + (xg_0 + g_1/2 - g_2/4x)F = 0. \tag{17}$$

Again proceeding as in Sec. II we obtain

$$x_l g_2(x_l)\sum_{m\neq l}\frac{1}{x_l - x_m} = -\frac{1}{2}[x_l g_1(x_l) + g_2(x_l)]. \tag{18}$$

Multiplying by $(x_l)^r$ and summing over l gives

$$4k^2J_{4+r} - 4(1+k^2)J_{3+r} + 4J_{2+r} = -[5k^2y_{3+r} - 4(1+k^2)y_{2+r} + 3y_{1+r}]. \tag{19}$$

Putting $r = -1$ yields

$$k^2J_3 - (1+k^2)J_2 + J_1 = -\tfrac{1}{4}[5k^2y_2 - 4(1+k^2)y_1 + 3N]. \tag{20}$$

We see this again gives y_2 in terms of y_1 and N. [The relation is, however, different from that implied by Eq. (11).] Putting $r = 1$ then 2, then 3, ..., in Eq. (19). gives all the y_s in terms of y_1 and N_1. Equating the coefficient of x^{N+2} in Eq. (17) to zero yields

$$N = (n-1)/2, \tag{21}$$

thus n must be odd. From the coefficient of x^{N+1} we obtain

$$y_1 = \{[1+k^2][4N(N+1)+1] - h\}/2k^2(1+4N), \tag{22}$$

Hence, again if n, h are given we have the sum rules for the positive powers of the zeros.

IV. SUM RULES FOR THE NEGATIVE POWERS OF THE ZEROS

We consider these for two reasons:

(1) One application of the sum rules for positive powers is as a means of bounding the largest zero. If we can obtain such for the inverse powers, we can bound the smallest zero.

(2) We will want to apply the techniques developed to certain integral functions. The zeros of these have a point of accumulation at infinity. The analog of the positive sum rules for these functions are the trivial statements that $y_s = \infty$.

Consider now polynomials satisfying Eq. (1) where, for the moment, the g_0, g_1, g_2 are arbitrary polynomials. Proceeding as before we obtain

$$g_2(x_l)\sum_{m\neq l}\frac{1}{x_l - x_m} = -\frac{1}{2}g_1(x_l). \tag{23}$$

If we multiply this by $(x_l)^{-r}$ and sum over l, the right-hand side will involve sums over the negative powers of x_l.. The right-hand side will be a combination of terms J_{-r} where

$$J_{-r} = \sum_{\substack{ml\\ m\neq l}}\frac{1}{(x_l)^r(x_l - x_m)}. \tag{24}$$

Now we would like to show that just as the J_s for $s \gtrless 0$ are expressible in terms of the y_s for $s \gtrless 0$, the $J_{-r} r \gtrless 0$ can be expressed in terms of the y_{-r}. Then our resulting equation becomes a recursion relation from which we might hope to be able to compute the y_{-r}. We state the result as a theroem.

Theorem:

(a) If r is even,

$$J_{-r} = \frac{r}{2}y_{-(r+1)} - \sum_{s=0}^{r/2-1} y_{-(r-s)}y_{-(s+1)}. \tag{25}$$

(b) If r is odd,

$$J_{-r} = \frac{r}{2}y_{-(r+1)} - \sum_{s=0}^{(r-3)/2} y_{-(r-s)}y_{-(s+1)} - \tfrac{1}{2}(y_{-(r+1)/2})^2. \tag{26}$$

Proof:

Let

$$K^s{}_{-r} = \sum_{\neq}\frac{1}{(x_l)^{r-s}(x_m)^s(x_l - x_m)}. \tag{27}$$

Then

$$J_{-r} = K^0{}_{-r} \tag{28}$$

and

$$\begin{aligned} K^s{}_{-r} &= \sum_{\neq}\frac{(x_m)}{(x_l)^{r-s}(x_m)^{s+1}(x_l - x_m)} \\ &= \sum_{\neq}\frac{x_m - x_l + x_l}{(x_l)^{r-s}(x_m)^{s+1}(x_l - x_m)} \\ &= -\sum_{\neq}\frac{1}{(x_l)^{r-s}(x_m)^{s+1}} + K^{s+1}{}_{-r}. \end{aligned} \tag{29}$$

Applying this identity repeatedly, we arrive at two different $K^s{}_{-r}$ depending on the parity of r.

(a) *r even*

We come to

$$K^{r/2}{}_{-r} \equiv \sum_{\neq}\frac{1}{(x_l)^{r/2}(x_m)^{r/2}(x_l - x_m)} = 0, \tag{30}$$

Since the summand is antisymmetric under the interchange $l \longleftrightarrow m$.

(b) *r odd*

We come to

$$K^{(r-1)/2}{}_{-r} = -\sum_{\neq}\frac{1}{(x_l)^{(r+1)/2}(x_m)^{(r+1)/2}} + K^{(r+1)/2}{}_{-r}. \tag{31}$$

But one readily sees that

$$K^{(r+1)/2}{}_{-r} = -K^{(r-1)/2}{}_{-r}.$$

$$\therefore K^{(r-1)/2}{}_{-r} = -\frac{1}{2}\sum_{\neq}\frac{1}{(x_l)^{(r+1)/2}(x_m)^{(r+1)/2}}. \tag{32}$$

Combining our results we see that

$$J_{-r} = -\sum_{\neq}\sum_{s=0}^{r/2-1}\frac{1}{(x_l)^{r-s}(x_m)^{s+1}}, \quad r \text{ even}, \tag{33}$$

and

$$J_{-r} = -\sum_{\neq}\sum_{s=0}^{(r-3)/2}\frac{1}{(x_l)^{r-s}(x_m)^{s+1}} - \frac{1}{2}\sum_{\neq}\frac{1}{(x_l)^{(r+1)/2}(x_m)^{(r+1)/2}}. \tag{34}$$

Finally on noting that

$$\sum_{\neq}\frac{1}{(x_l)^p(x_m)^q} = y_{-p}y_{-q} - y_{-(p+q)}, \tag{35}$$

we obtain Eqs. (25) and (26).

Examples

Hermite polynomials—$H_N(x)$

The differential equation is[3]

$$H_N^{(2)} - 2x\,H_N^{(1)} + 2N\,H_N = 0, \tag{36}$$

while the fundamental identity is

$$x_l = \sum_{m\neq l}\frac{1}{x_l - x_m}, \quad l = 1, 2, \ldots, N. \tag{37}$$

Note:

(a) The zeros of the Hermite polynomials are located symmetrically around the origin. Hence the sums of odd powers are zero.

(b) In order to avoid dividing by zero we must treat the case where one of the roots is zero separately. Thus:

N even

Zero is not a zero of the polynomial. We can use Eq. (37) as it stands. Dividing by $(x_l)^r$ and summing over l gives the recurrence relation

$$J_{-r} = y_{-(r-1)}. \tag{38}$$

$$\begin{aligned} &r = 1 \\ &J_{-1} = \tfrac{1}{2}y_{-2} = y_0 = N \\ &\therefore y_{-2} = 2N. \end{aligned} \tag{39}$$

$$\begin{aligned} &r = 3 \\ &J_{-3} = \tfrac{3}{2}y_{-4} + \tfrac{1}{2}(y_{-2})^2 = y_{-2}. \end{aligned}$$

Using Eq. (39), we then find

$$y_{-4} = y_{-2}(y_{-2} + 2)/3 = (4N/3)(N + 1). \tag{40}$$

Successively putting the odd integers for r, we obtain all the sums y_{-2n}.

N odd

Now restrict x_l to the nonzero zeros of the Hermite polynomial, i.e.,

$$x_l = \sum_{\substack{m=1 \\ m\neq l}}^{N-1} \frac{1}{x_l - x_m} + \frac{1}{x_l}, \quad l = 1, 2, \ldots, N-1, \quad x_l \neq 0. \tag{41}$$

The recursion relation becomes

$$J_{-r} + y_{-(r+1)} = y_{-(r-1)}. \tag{42}$$
$$r = 1$$
$$J_{-1} + y_{-2} = y_0.$$

But

$$J_{-1} = \tfrac{1}{2} y_{-2},$$

and

$$y_0 = \sum_{l=1}^{N-1} 1 = N-1,$$
$$\therefore y_{-2} = \tfrac{2}{3}(N-1). \tag{43}$$
$$r = 3$$
$$J_{-3} + y_{-4} = y_{-2}.$$

Then since

$$J_{-3} = \tfrac{1}{2} y_{-4} - (y_{-2})^2/2$$

we obtain

$$y_{-4} = \tfrac{4}{15}(N-1)(N+2). \tag{44}$$

Again we see all the sums y_{-2n} are completely determined.

2. Laguerre polynomials—$L_N^{(\alpha)}(x)$

The differential equation is

$$x[L_N^{(\alpha)}]'' + [\alpha + 1 - x][L_N^{(\alpha)}]' + NL_N^{(\alpha)} = 0. \tag{45}$$

This leads to the identity

$$x_l \sum_{m\neq l} \frac{1}{x_l - x_m} = \frac{-[\alpha + 1 - x_l]}{2}. \tag{46}$$

The recursion relation becomes

$$J_{-(r-1)} = [-(\alpha+1)/2]\, y_{-r} + \tfrac{1}{2} y_{-(r-1)}. \tag{47}$$
$$r = 1$$
$$J_0 = [-(\alpha+1)/2]\, y_{-1} + \tfrac{1}{2} y_0.$$

But $J_0 = 0$, $y_0 = N$ and therefore

$$y_{-1} = N/(\alpha+1). \tag{48}$$

Similarly putting $r = 2$ we find

$$y_{-2} = [N/(\alpha+1)^2(2-\alpha)][N+\alpha+2]. \tag{49}$$

Proceeding with higher r we find all y_{-r}.

3. Lamé polynomials

For simplicity we restrict ourselves to the first case ($\rho/ = \sigma = \tau = 0$). Then Eq. (8) is the differential equation. The fundamental identity is

$$\{k^2 (x_l)^3 - (1+k^2)x_l^2 + x_l\} \sum_{m\neq l} \frac{1}{x_l - x_m} = -\tfrac{1}{4}[3k^2 (x_l)^2 - 2(1+k^2)x_l + 1]. \tag{50}$$

Assuming none of the x_l are zero we obtain as recursion relation

$$k^2 J_{-3r} - (1+k^2)J_{2-r} + J_{1-r} = -\tfrac{1}{4}[3k^2 y_{2-r} - 2(1+k^2)y_{1-r} + y_{-r}]. \tag{51}$$

With $r = 1$ we obtain

$$k^2 J_2 - (1+k^2)J_1 + J_0 = -\tfrac{1}{4}[3k^2 y_1 - 2(1+k^2)y_0 + y_{-1}]. \tag{52}$$

Now $J_0 = 0$, $J_1 = N(N-1)/2$, $J_2 = (N-1)y_1$, $y_0 = N$. In Sec. III we saw that if we know h, we know y_1. Thus we can use Eq. (52) to determine y_{-1}. The result is

$$y_{-1} = h/2. \tag{53}$$

Given y_{-1} we can put $r = 2$ in Eq. (51) and find y_{-2} and then successively all y_{-r}. To generalize later, it is perhaps best to turn the argument around. All the sum rules (positive and negative) for the Lamé polynomials are expressible in terms expressible in terms of h which is essentially y_{-1}.

4. Jacobi polynomials—$P_N^{(\alpha,\beta)}$

Here the results are slightly more complicated. The sum rules for the positive powers are like those for the Hermite and Laguerre polynomials. For the negative powers the rules are like the Hermite and Laguerre cases for some values of the parameters and like those of the Lamé polynomials for other values.

We take the equation in the form

$$(1-x^2)[P_N^{(\alpha,\beta)}]'' + [\beta - \alpha - (\alpha+\beta+2)x] \times [P_N^{(\alpha,\beta)}]' + N(N+\alpha+\beta+1)P_N^{(\alpha,\beta)} = 0. \tag{54}$$

The fundamental identity is

$$[1 - x_l^2] \sum_{m\neq l} \frac{1}{x_l - x_m} = \frac{1}{2}[(\alpha - \beta) + (\alpha + \beta + 2)x_l]. \tag{55}$$

Let us, for simplicity, assume none of the x_l are zero. (If one is, we proceed as for odd Hermite polynomials.) Equation (55) then gives rise to the recursion relation

$$J_{-r} - J_{(-r-2)} = [(\alpha-\beta)/2]\, y_{-r} + [(\alpha+\beta+2)/2]\, y_{-(r-1)}. \tag{56}$$

With $r = 1$ this becomes

$$J_{-1} - J_1 = [(\alpha-\beta)/2] y_{-1} + [(\alpha+\beta+2)/2] y_0. \tag{57}$$

But $J_{-1} = \tfrac{1}{2} y_{-2} - \tfrac{1}{2}(y_{-1})^2$, $J_1 = N(N-1)/2$, $y_0 = N$. Now we must distinguish two cases:

(a) $\beta = \alpha$

This is the case for many of the commonly occuring polynomials. For example:

The Gegenbauer Polynomials $C_N^{\alpha+(1/2)} P_N^{(\alpha,\alpha)}$,

The Tchebycheff Polynomials $\begin{cases} T_N = P_N^{(-1/2,-1/2)} \\ U_N = P_N^{(1/2,1/2)} \end{cases}$,

The Legendre Polynomials $P_N = P_N^{(0,0)}$.

In this case the Eq. (54) is invariant under reflection. The polynomials will either be even or odd. Hence the zeros will occur in pairs symmetrical around the origin and all $y_{-(2n+1)}$ will be zero. In particular Eq. (57) then yields

$$y_{-2} = N[N + 2\alpha + 1]. \tag{58}$$

Further setting $r = 3, 5, 7, ,,,,$ then enables us to calculate all the y_{-2n}. The situation is just as for the Hermite and Laguerre Polynomials.

(b) $\beta \neq \alpha$

In this case Eq. (57) merely gives us y_{-2} in terms of y_{-1}. Similarly putting $r = 2$ in Eq. (56) enables us to express y_{-3} in terms of y_{-1}. In general then the sum rules express the sums of the negative powers in terms of y_{-1}. The situation is much like that for the Lamé polynomials. The remaining information must be obtained by other means as discussed later.

V. SUM RULES FOR ZEROS OF INTEGRAL FUNCTIONS[4]

The simplicity of the sum rules for polynomials satisfying differential equations with polynomial coefficients suggest we look for such for the simplest generalization. For simplicity we again restrict our attention to second-order differential equations. Thus again we consider functions satisfying Eq. (1). However, we do not assume them to be polynomials but rather functions with the simple Weierstrass expansion[5]

$$f(z) = Cz^{\nu}\prod_{l=1}^{\infty}[1 - z^2/x_l]. \tag{59}$$

We now obtain sum rules for the x_l—which are the squares of the non-zero zeros of f.

Insert the expansion of Eq. (59) into Eq. (1), evaluate at $z_l \equiv (x_l)^{1/2}$ and divide by $f'[(x_l)^{1/2}]$. The result is

$$g_2(z_l)\sum_{l}\frac{1}{x_l - x_m} = \frac{-\{(2\nu+1)g_2(z_l) + z_l g_1(z_l)\}}{x_l}. \tag{60}$$

We consider three simple examples.

(1) The Bessel functions—$j_\nu(z)$:
These satisfy

$$z^2\frac{d^2}{dz^2}j_\nu + z\frac{d}{dz}j_\nu + (z^2 - \nu^2)j_\nu = 0,$$

i.e., $g_2 = z^2$, $g_1 = z$, and have an expansion of the form[5] of Eq. (59). Then Eq. (60) is

$$x_l\sum_{m\neq l}\frac{1}{x_l - x_m} = \frac{-(\nu+1)}{2}. \tag{61}$$

(2), (3) The trigonometric functions—cosΠz, sinΠz:
These satisfy

$$\frac{d^2 f}{dz^2} + \Pi^2 f = 0,$$

i.e., $g_2 = 1, g_1 = 0$, and have an expansion of the form of Eq. (59) with $\nu = 0$ and 1, respectively.[6] For these Eq. (60) is

$$\sum_{m\neq l}\frac{1}{x_l - x_m} = \frac{-(2\nu+1)}{4x_l}.$$

Clearly all three cases are included in the relation

$$\sum_{m\neq l}\frac{1}{x_l - x_m} = \frac{\lambda}{x_l}. \tag{63}$$

If we divide by $(x_l)^r$ and sum over l, we obtain the recursion relation

$$J_{-r} = \lambda y_{-(r+1)}. \tag{64}$$

For $r = 1$ this is: $J_{-1} = \lambda y_{-2}$. From this we conclude that

$$y_{-2} = (y_{-1})^2/(1 - 2\lambda). \tag{65}$$

Thus y_{-2} is given in terms of y_{-1}. Similarly proceeding to higher order we see all y_{-r} will be expressible in terms of the y_{-1}. The result is analogous to that for the Lamé polynomials or the *general* Jacobi polynomials. Clearly we need to know (a little) more about the functions.

Thus suppose we know the first two terms in the expansion of $f(z)$ in the vicinity of the origin. From Eq. (59) we know this is of the form

$$f(z) \approx Cz^{\nu}\{1 - az^2 + \cdots\}, \tag{66}$$

but explicitly

$$a = \sum_{l}\frac{1}{x_l} \equiv y_{-1}. \tag{67}$$

However, we know
(1) for the Bessel functions $a = 1/4(\nu+1)$;
(2) for cosΠz, $a = \Pi^2/2$;
(3) for sinΠz, $a = \Pi^2/6$.
Then for the Bessel functions we obtain

$$y_{-1} = [2^2(\nu+1)]^{-1}, y_{-2} = [2^4(\nu+2)(\nu+1)^2]^{-1},$$
$$y_{-3} = [2^5(\nu+2)(\nu+2)(\nu+1)^3]^{-1}.$$

For cosΠz we have: $y_{-1} = \Pi^2/2, y_{-2} = \Pi^4/6, y_{-3} = \Pi^6/15$, and for sin$\Pi z$:

$$y_{-1} = \sum_{n=1}^{\infty}\frac{1}{n^2} = \frac{\Pi^2}{6},$$
$$y_{-2} = \sum_{n=1}^{\infty}\frac{1}{n^4} = \frac{\Pi^4}{90},$$
$$y_{-3} = \sum_{n=1}^{\infty}\frac{1}{n^6} = \frac{\Pi^6}{(21)(45)}.$$

VI. ADDITIONAL SUM RULES

We have tried to emphasize that our fundamental identity and hence the sum rules arise from fact that the functions satisfy differential equations with polynomial coefficients. Now if we take our differential equation and differentiate it any number of times we will have an equation of the same type. This should give rise to additional identities satisfied by the zeros. What are these? In (I) we showed that additional identities can be obtained by multiplying the fundamental one by itself any number of times, resumming terms and using the lower order identities repeatedly. The general result was identities of the form

$$[g_2(x_l)]^s\sum_{m\neq l}(x_l - x_m)^{-s} = f_s(x_l), \tag{68}$$

where the f_s were readily determined polynomials. It turns out that these identities are exactly the same as those obtained by differentiating the original differential equation.

The derivation of these additional identities given in (I) is quite correct for all polynomial cases. It is suspect in the case of the integral functions of Sec. V because of convergence questions. Thus by summing Eq. (63) over l and noting that the summand on the left is odd one would tend to conclude that

$$\sum_{m\neq l}(x_l)^{-1} = 0$$

which we know is not true. The problem is one of interchanging orders of summation. Accordingly we give here a deriva-

tion of the second identity which is appropriate for the functions of Sec. V. We give this in the form of a theorem.

Theorem: If

$$\sum_{m\neq l}(x_l - x_m)^{-1} = \lambda / x_l, \tag{69}$$

then

$$(x_l)^2 \sum_{m\neq l}(x_l - x_m)^{-2} = \frac{\lambda}{3}\left[(2-\lambda) - 2y_{-1}x_l\right]. \tag{70}$$

Proof: Multiply Eq. (69) by itself. We obtain

$$\sum_{\substack{m\neq l\\ n\neq l}} \frac{1}{(x_l - x_m)(x_l - x_n)} = \lambda^2/(x_l)^2. \tag{71}$$

Separate out the term $m = n$. Then[7]

$$\sum_{\neq} \frac{1}{(x_l - x_m)(x_l - x_n)} + \sum_{m\neq l} \frac{1}{(x_l - x_m)^2} = \frac{\lambda^2}{(x_l)^2}. \tag{72}$$

Use the identity

$$\begin{aligned}\sum_{\neq} &\frac{1}{(x_l - x_m)(x_l - x_n)} \\ &= \sum_{\neq}\left(\frac{1}{x_l - x_m} - \frac{1}{x_l - x_n}\right)\frac{1}{(x_m - x_n)} \\ &= 2\sum_{\neq}\frac{1}{(x_l - x_m)(x_m - x_n)}.\end{aligned}$$

In the last sum let us add and subtract the term $n = l$. The Eq. (72) becomes

$$\begin{aligned}&2\sum_{\substack{m\neq l\\ n\neq m}}\frac{1}{(x_l - x_m)(x_m - x_n)} \\ &+3\sum_{m\neq l}\frac{1}{(x_l - x_m)^2} = \frac{\lambda^2}{(x_l)^2}.\end{aligned} \tag{73}$$

Using Eq. (69) we see that

$$\begin{aligned}&\sum_{\substack{m\neq l\\ n\neq m}}\frac{1}{(x_l - x_m)(x_m - x_n)} \\ &= \lambda\sum_{m\neq l}\frac{1}{x_m(x_l - x_n)} \\ &= \lambda\sum_{m\neq l}\left(\frac{1}{x_l - x_m} + \frac{1}{x_m}\right)\frac{1}{x_l} \\ &= \frac{\lambda^2}{x_l^2} + \frac{\lambda}{x_l}\sum_{m\neq l}\frac{1}{x_m} = \frac{\lambda^2}{(x_l)^2} - \frac{\lambda}{(x_l)^2} + \frac{\lambda y_{-1}}{x_l}.\end{aligned} \tag{74}$$

Inserting Eq. (74) into Eq. (73) yields

$$\sum_{m\neq l}\frac{1}{(x_l - x_m)} = \frac{\lambda}{3}\left(\frac{2-\lambda}{(x_l)^2} - \frac{2y_{-1}}{x_l}\right). \tag{75}$$

Multiplying by $(x_l)^2$ yields the theorem.

By exactly the same procedure we obtain identities involving $\sum_{m\neq l}(x_l - x_m)^{-s}$. We note that the right-hand side of Eq. (75) involves only negative powers of x_l. If we sum over l, the right-hand side is thus known sums (y_{-1}, y_{-2}). Thus we have the sum rule

$$\sum_{\substack{ml\\ m\neq l}}(x_l - x_m)^{-2} = \frac{\lambda}{3}\left[(2-\lambda)y_{-2} - 2(y_{-1})^2\right]. \tag{76}$$

For our examples these results are:

(1) Bessel functions

Identity[4]:

$$(x_l)^2\sum_{m\neq l}\frac{1}{(x_l - x_m)^{-2}} - \frac{1}{12}\left[(\nu+1)(\nu+2) - x_l\right].$$

Sum rule: $\sum_{\substack{ml\\ m\neq l}}\frac{1}{(x_l - x_m)^{-2}} = \frac{1}{2^6(2+\nu)}$.

(2) Cosine

Identity:

$$(x_l)^2\sum_{m\neq l}\frac{1}{(x_l - x_m)^{-2}} = -\frac{1}{12}\left\{\frac{9}{4} - \Pi^2 x_l\right\}.$$

Sum Rule: $\sum_{\substack{ml\\ m\neq l}}\frac{1}{(x_l - x_m)^2} = \frac{\Pi^4}{96}$.

(3) Sine

Identity:

$$(x_l)^2\sum_{m\neq l}\frac{1}{(x_l - x_m)^2} = -\frac{1}{4}\left(\frac{11}{4} - \frac{\Pi^2}{3}x_l\right).$$

Sum rule: $\sum_{m\neq l}\frac{1}{(x_l - x_m)^{-2}} \equiv \sum_{m\neq l}\frac{1}{(l^2 - m^2)^2} = \frac{\Pi^4}{160}$.

Remark: The sum rules for the odd powers are merely

$$\sum_{m\neq l}(x_l - x_m)^{-(2n+1)} = 0,$$

$n > 0$.

The convergence is sufficiently rapid so that the antisymmetry argument can be used.

VII. CONCULSION

The sum rules for zeros of polynomials satisfying differential equations have been extended in three ways.

(1) Higher order polynomials coefficients are treated.

(2) Sum rules for negative powers of the zeros are obtained.

(3) The sum rules are extended to zeros of integral functions of a particular form.

ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation under Grant No. NSF MCS78 20455.

[1] K.M. Case, J. Math. Phys. **21**, 702 (1980). We refer to this paper as (I).

[2] We follow the summary of F.M. Arscott and I.M. Khabaza, *Tables of Lamé* Polynomials (Macmillan, New York, 1962).

[3] We use the notation of Milton Abramowitz and Irene A. Stegun, *Handbook of Mathematical Functions* (U.S. Dept. of Commerce, National Bureau of Standards, Appl. Math. Series 55, 1964).

[4] The results stated here for the Bessel functions have been obtained previously, c.f., S. Ahmed, M. Brushi, F. Calogero, M.A. Olshanetsky, and A.M. Perelomov/ (Preprint 10, Oct. 1978, submitted to Nuovo Cimento.)

[5] G.N. Watson, *A Treatise on the Theroy of Bessel Functions* (Macmillan, New York, 1945, p. 498.

[6] E.C. Zitchmarsh, *The Theory of Functions* (Oxford University, London, 1939), p. 114.

[7] $\Sigma_{\neq}$ means omit $m = l$, $n = l$, $n = m$.

A connection between nonlinear evolution equations and ordinary differential equations of P-type. I

M. J. Ablowitz
Mathematics and Computer Science, Clarkson College, Potsdam, New York 13676

A. Ramani[a)]
Program in Applied Mathematics, Princeton University, Princeton, New Jersey 08540

H. Segur
Aeronautical Research Associates of Princeton, Inc., 50 Washington Road, P.O. Box 2229, Princeton, New Jersey 08540

(Received 6 February 1979; accepted for publication 9 August 1979)

We develop here two aspects of the connection between nonlinear partial differential equations solvable by inverse scattering transforms and nonlinear ordinary differential equations (ODE) of P-type (i.e., no movable critical points). The first is a proof that no solution of an ODE, obtained by solving a linear integral equation of a certain kind, can have any movable critical points. The second is an algorithm to test whether a given ODE satisfies necessary conditions to be of P-type. Often, the algorithm can be used to test whether or not a given nonlinear evolution equation may be completely integrable.

I. INTRODUCTION

The development of the inverse scattering transform (IST; e.g., see Ref. 1) has shown that certain nonlinear evolution equations possess a number of remarkable properties, including the existence of solitons, an infinite set of conservation laws, and an explicit set of action-angle variables. It has been noted[2] that there is a connection between these nonlinear partial differential equations (PDE's) solvable by IST and nonlinear ordinary differential equations (ODE's) without movable critical points. (Some definitions: a *critical point* is a branch point or an essential singularity in the solution of the ODE. It is *movable* if its location in the complex plane depends on the constants of integration of the ODE. A family of solutions of the ODE without movable critical points has the P-property: here P stands for Painlevé. The ODE is of P-type if all of its solutions have this property.) The reductions of some of the equations solvable by IST to the classical Painlevé transcendents have been discussed in the literature.[2-6] In Ref. 3 we announced a number of results which indicate that this connection to ODE's of P-type is yet another remarkable property of these special nonlinear PDE'S. The purpose of this paper is to develop some of those announced results in more detail.

In particular, it was conjectured[3] that:

Every nonlinear ODE obtained by an exact reduction of a nonlinear PDE of IST class is of P-type.

Here a nonlinear PDE is of IST class if nontrivial solutions of it can be found by solving a linear integral equation of the Gelfand–Levitan–Marchenko form. No proof of this conjecture is available yet, but in Sec. II of this paper we prove a more restricted result in this direction. It is known that under scaling transformations certain nonlinear PDE's of IST class reduce to ODE's. Moreover, the solutions of these ODE's may be obtained by solving linear integral equations.[2] We show in Sec. II that every such family of solutions has the P-property.

The reader should note that this conjecture relates to ODE's obtained from equations solved *directly* by IST. There are many examples of equations solved only indirectly by IST; the sine–Gordon equation is perhaps the best known example. An ODE obtained from an equation solved indirectly by IST need not be of P-type, but it must be related through a simple transformation to an ODE that is. We discuss this point in more detail in our following paper.[7]

One consequence of this conjecture is an explicit test of whether or not a given PDE may be of IST class; namely, reduce it to an ODE, and determine whether the ODE is of P-type. To this end, we identify in Sec. III certain necessary conditions that an ODE must satisfy to be of P-type and describe an explicit algorithm to determine whether an ODE meets these necessary conditions. Examples are given to illustrate the main ideas. In many cases, this algorithm seems to be simpler than the α-method of Painleve and his coworkers, [8] which also determines whether an ODE satisfies necessary conditions to be of P-type. It is similar to the method of Kovalevskya,[9] who made major contributions to the theory of the motion of a rigid body about a fixed point after first determining the choices of parameters for which the equations of motion had no movable critical points. It is interesting to note that she found the complete solution whenever the equations were of P-type, and that no solutions are known to this day when they are not of P-type.

II. LINEAR INTEGRAL EQUATIONS AND ODE's OF P-TYPE

In this section we demonstrate exact reductions of nonlinear PDE's of IST class to ODE's such that solutions of the ODE may be obtained *via* a linear integral equation with a nonsingular kernel. Then we prove that any family of solutions obtained in this way necessarily has the P-property, by applying Fredholm's theory of linear integral equations.[10]

Consider a nonlinear PDE of IST class, with two independent variables, x and t. Solutions of the PDE may be

[a)]Permanent address: L.P.T.H.E., Université Paris-Sud 91405, Orsay, France.

 0022-2488/80/040715-07$1.00

obtained *via* a linear integral equation of the form

$$\mathscr{K}(x,y;t) = \mathscr{F}(x+y;t) + \int_x^\infty \mathscr{K}(x,z;t)\mathscr{N}(x;z,y;t)\,dz, \tag{2.1}$$

where $\mathscr{N}$ is explicitly given in terms of $\mathscr{F}$ (possibly with one or more integrals from x to ∞). For instance,

$$\mathscr{N}(x;z,y;t) = \mathscr{F}(z+y;t), \tag{2.2a}$$

$$\mathscr{N}(x;z,y;t) = \pm \int_x^\infty \mathscr{F}(z+u;t)\mathscr{F}(u+y;t)\,du, \tag{2.2b}$$

$$\mathscr{N}(x;z,y;t) = \pm \int_x^\infty \bar{\mathscr{F}}(z+u;t)\mathscr{F}(u+y;t)\,du, \tag{2.2c}$$

where $\bar{\mathscr{F}}(x;t) = [\mathscr{F}(x^*;t^*)]^*$ and $(*)$ means complex conjugate,

$$\mathscr{N}(x;z,y;t) = \pm i \int_x^\infty \partial_z \bar{\mathscr{F}}(z+u;t)\mathscr{F}(u+y;t)\,du, \tag{2.2d}$$

$$\mathscr{N}(x;z,y;t) = \int_x^\infty \int_x^\infty \mathscr{F}(z+u;t)\mathscr{F}(u+v;t) \times \mathscr{F}(v+y;t)\,du\,dv, \tag{2.2e}$$

etc.

If $\mathscr{F}$ satisfies a suitable linear PDE and decays rapidly enough as $x \to +\infty$ for the integral equation to be meaningful and have a unique solution $\mathscr{K}$, then

$$\mu(x;t) = \mathscr{K}(x,x;t) \quad \text{or} \quad \mu = \frac{d}{dx}\mathscr{K}(x,x;t)$$

satisfies the original nonlinear PDE. This is the "direct" method,[11,12,7] first introduced by Zakharov and Shabat; it involves no reference to a scattering problem.

Among the PDE's of IST class, we restrict our attention to those that admit self-similar solutions of the scaling type. Thus, we assume that the integral equation is compatible with the ansatz

$$\mathscr{F}(x+y;t) = \phi(x+y;t)F[\psi(x,t)+\psi(y,t)], \tag{2.3a}$$

$$\mathscr{K}(x,y;t) = \phi(x+y;t)K[\psi(x,t),\psi(y,t)], \tag{2.3b}$$

where ϕ and ψ are known functions, and that this ansatz reduces the linear PDE for $\mathscr{F}$ to a linear ODE for F. If the integral equation in this form can be solved, its solution is the self-similar solution of the nonlinear PDE; i.e.,

$$u(\xi) = K(\xi,\xi) \quad \left[\text{or } \frac{d}{d\xi}K(\xi,\xi)\right]$$

satisfies a nonlinear ODE that is an exact reduction of the PDE. As an example, we note that a one-parameter family of solutions of Painlevé's second equation,

$$w'' = 2w^3 + zw, \tag{2.4}$$

was constructed in this way.[2] More examples will be given in.[7] If it could be proved, the conjecture stated above would assert that any ODE obtained in this way must be of P-type, i.e., its general solution has no movable singularities other than poles.

Next, we prove that there is a family of solutions of the nonlinear ODE which have no movable singularities other than poles, provided that there are solutions of the corresponding linear ODE that decay rapidly enough.

First we rewrite the integral Eq. (2.1) as

$$K(\xi;\eta) = F(\xi+\eta) + \int_\xi^\infty K(\xi,\zeta)N(\xi;\zeta,\eta)\,d\zeta, \tag{2.5}$$

where N is given in terms of F. Here F satisfies a linear homogeneous ODE, and its general solution has no movable singularities whatsoever; its only singularities are fixed and their positions can be determined by direct examination of the equation.

Among those solutions F, some may decay at $+\infty$ fast enough for the integral equation to make sense and be solvable. Since its ODE is linear and homogeneous, the set of such solutions is a vector subspace. From these solutions, a family of solutions of the nonlinear ODE can be obtained by solving the integral equation.

So far, the variable x has been considered to be real. Now, we allow x to be complex valued. To do this we must define the integral more carefully. The same applies to the integral(s) that define N in terms of F.

For each ξ, choose a contour $\mathscr{C}(\xi)$ with ξ as one endpoint, and going to $+\infty$ while staying in a strip of finite width along the positive real axis. If F and N have (fixed) singularities, it may be necessary to make cuts in the complex plane, starting at the singular points of F, and to demand that for any ξ the contour $\mathscr{C}(\xi)$ crosses no cuts. But these cuts are fixed and delimit a fixed domain $\mathscr{A}$ in which N depends on ξ only and is analytic.

Now fix ξ and a contour $\mathscr{C}(\xi)$. Consider the following Fredholm integral equation:

$$K(\lambda;\eta) = F(\xi+\eta) + \lambda \oint K(\lambda;\zeta)N(\xi;\zeta,\eta)\,d\zeta. \tag{2.6}$$

Assume that the decay of F when its argument goes to infinity in the strip defined previously is fast enough for the following conditions to be satisfied.

$$|N(\xi;\zeta,\eta)| < M(\xi;\eta) \text{ for all } \zeta, \text{ such that} \tag{2.7a}$$

$$\left|\oint M(\xi;\eta)d\eta\right| = \mathscr{M}(F) < \infty, \tag{2.7b}$$

$$\left|\oint F(\xi+\eta)M(\xi;\eta)\,d\eta\right| < +\infty. \tag{2.7c}$$

Then the powerful results of Fredholm theory apply (cf. Appendix). Namely, the integral equation has a unique solution

$$K(\lambda;\eta) = F(\xi+\eta) + \lambda \oint F(\xi+\eta) \times \mathscr{N}(\mathscr{C}(\xi);\lambda;\zeta,\eta)\,d\zeta, \tag{2.8a}$$

with

$$\mathscr{N}(\mathscr{C}(\xi);\lambda;\zeta,\eta) = \mathscr{D}_2(\mathscr{C}(\xi);\lambda;\zeta,\eta)/\mathscr{D}_1(\mathscr{C}(\xi);\lambda). \tag{2.8b}$$

$\mathscr{D}_1$ and $\mathscr{D}_2$ are given as series in powers of λ

$$\mathscr{D}_i = \sum_{n=0}^{\infty} \lambda^n \mathscr{D}_i^{(n)}, \quad i = 1,2,$$

where the $\mathscr{D}_i^{(n)}$ are multiple integrals over $\mathscr{C}$ of polynomials in N. Fredholm theory states that the radius of convergence of these series in λ is infinite. Since N is analytic in the fixed domain $\mathscr{A}$, the $\mathscr{D}_i^{(n)}(\mathscr{C}(\xi);\lambda;\zeta,\eta)$ are analytic in ζ and η in the domain $\mathscr{A}$. Moreover, as long as $(\mathscr{C})$ does not cross a cut, the $\mathscr{D}_i^{(n)}$ do not depend on $\mathscr{C}$ but only on its endpoint ξ and this dependence is analytic for ξ in $\mathscr{A}$.

Hence $\mathscr{D}_1(\xi;\lambda)$ and $\mathscr{D}_2(\xi;\lambda;\zeta,\eta)$ are analytic in all their arguments, in λ in the entire complex plane, and in ξ,ζ,η in a fixed domain $\mathscr{A}$. Since

$$\lim_{\xi\to\infty} \mathscr{D}_1(\xi,\lambda) = 1$$

in the strip for any λ, $\mathscr{D}_1(\xi,\lambda)$ is not identically 0 in ξ and its zeros in $\mathscr{A}$ are isolated.

There is therefore a family of solutions $u(\xi)$ of the nonlinear ODE given by

$$u(\xi) = K(\xi,\xi) = F(2\xi) + \frac{1}{\mathscr{D}_1(\xi;1)} \times \oint F(\xi+\zeta)\mathscr{D}_2(\xi;1;\zeta,\xi)\,d\xi, \quad (2.9)$$

where F belongs to some vector subspace. The only singularities of the solution u, apart from the fixed singularities of F, come from the zeros of $\mathscr{D}_1(\xi,1)$. Since $\mathscr{D}_1$ is analytic in $\mathscr{A}$, these movable singularities are indeed poles.

III. SINGULAR POINT ANALYSIS

For an ODE to be of P-type, it is necessary that it have no movable branch points, either algebraic or logarithmic. We now describe an algorithm to determine whether a nonlinear ODE (or system of ODE's) admits movable branch points. The presentation is simplified considerably if we make two assumptions.

(i) The nth order system of ODE's has the form

$$\frac{d}{dz} w_j = F_j(z;w_1,w_2,\ldots,w_n), \quad j = 1,\ldots,n, \quad (3.1)$$

where each F_j is analytic in z and rational in its other arguments. An important special case is a nth order ODE

$$\frac{d^n w}{dz^n} = F\left(z;w,w',\ldots,\frac{d^{n-1}w}{dz^{n-1}}\right), \quad (3.2)$$

where F is analytic in z and rational in its other arguments.

(ii) The dominant behavior of the function in a sufficiently small neighborhood of the (movable) singularity is algebraic, i.e.,

$$w_j \sim \alpha_j(z-z_0)^{p_j} \text{ as } z\to z_0. \quad (3.3)$$

This does not exclude logarithmic branch points; it does exclude branch points in which the dominant behavior is logarithmic (but see Example 6). Neither of these restrictions is essential to the method, but we make no attempt to remove them in this paper.

An ODE without movable branch points might still admit movable essential singularities. This method does not identify essential singularities, and therefore it provides only necessary conditions for an ODE to be of P-type. That these conditions are not also sufficient may be seen from the ODE,[8]

$$w'' = (w')^2[(2w-1)/(w^2+1)].$$

Its general solution is

$$w = \tan\ln(Az+B),$$

where A and B arbitrary. The solution has a movable essential singularity at $z = -B/A$, but the equation nevertheless satisfies the necessary conditions we will describe.

Let us consider first the case of a single ODE, (3.2) and assume that the function becomes infinite at the singularity. There are basically three steps to the algorithm.

A. Find the Dominant Behavior

Look for a solution of (3.2) in the form

$$w \sim \alpha(z-z_0)^p, \quad (3.4)$$

where $\mathrm{Re}(p) < 0$ and z_0 is arbitrary. Substituting (3.4) into (3.2) shows that for certain values of p, two or more terms in the equation may balance (depending on α), and the rest can be ignored as $z\to z_0$. For each such choice of p, the terms which can balance are called the *leading terms*. Requiring that the leading terms do balance (usually) determines α.

Example 1:

$$w''' + ww'' - 2w^3 + \lambda w^2 + \mu w = 0. \quad (3.5)$$

There are two possible choices:

(i) $p = -1$, $\alpha = 3$, the leading terms are w''' and ww''.

(ii) $p = -2$, $\alpha = 3$, the leading terms are ww'' and $-2w^3$. ■

Example 2:

$$w''' + aww'' + b(w')^2 + cw^4 + dww' + ew^3 + fw^2 + gw = 0. \quad (3.6)$$

The only possible choice is $p = -1$, and α may be any of the three roots of

$$-6 + (2a+b)\alpha + c\alpha^3 = 0.$$

The first four terms in (3.6) are all leading terms. ■

Example 3:

$$w'' = 2w(w')^2/(w^2-1). \quad (3.7)$$

The only possible choice is $p = -1$, but α is entirely unrestricted. [This particular equation is of P-type; its general solution is

$$w = \tanh(Az+B),$$

for arbitrary A and B. It has no movable critical points.] ■

A given equation may have several choices of p. If *any* of the possible p's is not an integer, and if (3.4) actually is asymptotic near z_0, then it represents the dominant behavior in the neighborhood of a movable algebraic branch point of order p. The existence of such a branch point means that the equation is not of P-type.

To prove that (3.4) is asymptotic, define a new variable

$$v = w^{1/p}, \quad (3.8)$$

and rewrite (3.2) in terms of v. By construction, v vanishes at z_0, and v' is finite. We must show from its ODE that $v(z)$ is

analytic at z_o. Then it follows from (3.8) that w has a branch point of order p at z_o.

Expample 4:

$$w'' + 10w^4 = 0. \tag{3.9}$$

The only possible choice is $p = -2/3$, so we define

$$w = v^{-2/3}.$$

The equation for v is

$$\tfrac{3}{5}vv'' = (v')^2 - (\tfrac{3}{2})^2. \tag{3.10}$$

There is a solution of (3.10) that is regular at z_0 if

$$v(z_o) = 0, \quad v'(z_o) = \pm 3/2, \text{ and } v''(z_o) \text{ is finite.}$$

Then $v(z)$ is analytic at z_o, and (3.9) has a movable branch point of order $-2/3$. It is not of P-type. ■

Even if a given equation has a movable algebraic branch point, a simple transformation may turn it into one of P-type.

Example 5:

$$ww'' = \tfrac{5}{2}(w')^2. \tag{3.11}$$

The only possible choice is $p = -2/3$, so we define

$$w = v^{-2/3}.$$

The equation for v is

$$v'' = 0.$$

This equation is linear and therefore of P-type. ■

If all possible p's are integers, then for each p, (3.4) *may* represent the first term in the Laurent series, valid in a deleted neighborhood of a movable pole. In this case, a solution of (3.2) is

$$w(z) = (z - z_o)^p \sum_{j=0}^{\infty} a_j (z - z_o)^j, \quad 0 < |z - z_o| < R. \tag{3.12}$$

Here z_o is an arbitrary constant. If $(n-1)$ of the coefficients $\{a_j\}$ are also arbitrary, these are the n constants of integration of the ODE, and (3.12) is the general solution in the deleted neighborhood. The powers at which these arbitrary constants enter are called *resonances*.

B. Find the Resonances

For each (p,α) from step A, construct a simplified equation that retains only the leading terms of the original equation. Substitute

$$w = \alpha(z - z_o)^p + \beta(z - z_o)^{p+r} \tag{3.13}$$

into the simplified equation. To leading order in β, this equation reduces to

$$Q(r)\beta(z - z_o)^q = 0, \quad q \geqslant p + r - n. \tag{3.14}$$

If the highest derivative of the original equation is a leading term, $q = p + r - n$, and $Q(r)$ is a polynomial of order n. If not, $q > p + r - n$, and the order of the polynomial $Q(r)$ equals the order of highest derivative among the leading terms ($< n$).

The roots of $Q(r)$ determine the resonances. [$Q(r) = 0$ corresponds to the "indicial equation" in the method of Frobenius for finding solutions of a linear ODE near a regular singular point.]

(i) One root is always (-1). It represents the arbitrariness of x_o.

(ii) If α is arbitrary in step 1, another root is (0).

(iii) Ignore any roots with $\mathrm{Re}(r) < 0$, because they violate the hypothesis that $(z - z_o)^p$ is the dominant term in the expansion near z_o.

(iv) Any root with $\mathrm{Re}(r) > 0$, but r not a real integer, indicates a (movable) branch point at $z = z_o$. There is no need to continue the algorithm, but it remains to prove that the equation actually has such a branch point (cf. Example 4).

(v) If for every possible (p,α) from step A, all of the roots of $Q(r)$ (except -1 and possibly 0) are positive real integers, then there are no algebraic branch points. Proceed to step C to check for logarithmic branch points.

(vi) To represent the general solution of the nth order ODE in the neighborhood of a movable pole, $Q(r)$ must have $(n-1)$ nonnegative distinct roots, all real integers. If for *every* (p,α) from step A, $Q(r)$ has fewer than $(n-1)$ such roots, then none of the local solutions is general. This suggests that (3.4) misses an essential part of the solution.

Example 6:

$$w'' + 4ww' + 2w^3 = 0. \tag{3.15}$$

The only dominant algebraic behavior is

$$w \sim \alpha(z - z_o)^{-1}, \tag{3.16}$$

where α satisfies

$$2 - 4\alpha + 2\alpha^2 = 0. \tag{3.17}$$

Substituting (3.13) into (3.15) leads to

$$r^2 + r = 0.$$

The root (-1) corresponds to the arbitrariness of x_o. The root (0) corresponds not to the *arbitrariness* of α but to the fact that $\alpha = 1$ is a double root of (3.17). Thus, (3.16) cannot be the first term in a Laurent series of a *general* solution of (3.15). In fact, under the transformation

$$w = f(\ln(z - z_o))/(z - z_o),$$

(3.15) becomes

$$f'' + (4f - 3)f' + 2f(f - 1)^2 = 0. \tag{3.18}$$

Clearly, (3.18) has a variety of regular solutions, including those for which $f(0) = 1$. Every nonexponential solution of (3.18) corresponds to a movable logarithmic branch point of (3.15). ■

C. Find the Constants of Integration

For a given (p,α) from step A, let $r_1 \leqslant r_2 \leqslant \cdots \leqslant r_s$ denote the positive integer roots of $Q(r)$; ($s \leqslant n-1$). Substitute

$$w = \alpha(z - z_o)^p + \sum_{j=1}^{r_s} a_j (z - z_o)^{p+j} \tag{3.19}$$

into the full equation, (3.2). [This portion of the analysis is similar to a certain portion of the method of Frobenius for linear problems. One might think of substituting (3.12), but little is gained in most nonlinear ODE's by continuing beyond the largest root of $Q(r)$.] The coefficient of $(z - z_o)^{p+j-n}$, which must vanish identically, is

$$Q(j)a_j - R_j(z_o, \alpha, a_1, \ldots, a_{j-1}) = 0. \tag{3.20}$$

(i) For $j<r_1$, (3.20) determines a_j.

(ii) For $j=r_1$, (3.20) becomes

$$0\cdot a_{r_1} - R_{r_1}(z_o,\alpha,a_1,\ldots,a_{r_1-1}) = 0.$$

If

$$R_{r_1}(z_o,\alpha,a_1,\ldots,a_{r_1-1})\neq 0, \tag{3.21}$$

then (3.20) cannot be satisfied. There is no solution of the form (3.19), and we must introduce logarithmic terms into the expansion. Replace (3.19) with

$$w = \alpha(z-z_o)^p + \sum_{j=1}^{r-1} a_j(z-z_o)^{p+j} + [a_{r_1} + b_{r_1}\ln(z-z_o)](z-z_o)^{p+r_1} + \cdots. \tag{3.22}$$

Now the coefficient of $\{(z-z_o)^{p+r_1-n}\ln(z-z_o)\}$ is

$$Q(r_1)b_{r_1} = 0,$$

but b_{r_1} is determined by demanding that the coefficient of $(z-z_o)^{p+r_1-n}$ vanish; a_{r_1} is arbitrary. Continuing the expansion (3.22) to higher orders introduces more and more logarithmic terms. Thus, (3.21) signals a (movable) logarithmic branch point. Using methods similar to those in Example 4, one may prove that this series is asymptotic. Then it follows that the equation is not of P-type.

(iii) If it happens that (3.21) is false (i.e., $R_{r_1}=0$), then a_{r_1} is an arbitrary constant of integration. Proceed to the next coefficient.

(iv) Any resonance that is a multiple root of $Q(r)$ represents a (movable) logarithmic branch point with an arbitrary coefficient. If the assumed representation is asymptotic, the equation is not of P-type.

Example 7:

$$w'''' = 27w''w^2 + 21w'w^3 - 9w^5. \tag{3.23}$$

One approximate solution is

$$w\sim\alpha(z-z_o)^{-1},$$

with $\alpha=1$. There are three other choices of α possible. Substituting

$$w\sim(z-z_o)^{-1} + \beta(z-z_o)^{r-1},$$

yields

$$Q(r) = (r+1)(r-1)^2(r-9) = 0. \tag{3.24}$$

At the first resonance, $r=1$, the approximate solution is

$$w\sim(z-z_o)^{-1} + c_1 + c_2\{22\ln(z-z_o) + 27[\ln(z-z_o)]^2\} + \cdots, \tag{3.25}$$

where (z_0,c_1,c_2) are all arbitrary constants. If (3.25) is asymptotic as $z\to z_o$, then (3.23) is not of P-type. ■

(v) At each nonresonant power, (3.20) determines a_j. At each resonance, either

$$R_{r_1}\neq 0,$$

logarithmic terms must be introduced into (3.19), and the equation is not of P type, or

$$R_{r_1} = 0,$$

and a_{r_1} is a arbitrary constant of integration.

(vi) If no logarithms are introduced at any of the resonances, one could in principle compute all of the terms in the series. However, because the recursion relations are nonlinear, it is usually not feasible to determine the region of convergence of the series, as one does in a linear problem. An alternative is to prove directly from the ODE that each arbitrary constant, (a_j), is the coefficient of an analytic function (cf. Ref. 8, §14.41).

(vii) If no logarithms are introduced at any of the resonances for all possible (p,α) from step A, then the equation has met the necessary conditions, under the assumption that $p<0$ in (3.4). Cases for which $p>0$ are treated below. This completes the algorithm.

To this point, we have assumed that the function becomes infinite at the singularity. Other possibilities, where the function remains finite while some derivative becomes singular, may be treated either directly (with $p>0$) or by considering the ODE to be a system of first-order ODE's which we consider next.

The basic steps of the algorithm for a system of first-order ODE's, (3.1), are not essentially different from what we have already discussed.

1. Dominant behavior of the system

Substitute

$$w_j\sim\alpha_j(z-z_o)^{p_j}, \quad j=1,\ldots,n, \tag{3.26}$$

into (3.1) and determine both the $\mathbf{p}(=\{p_j\}_{j=1}^n)$ for which there is a balance of leading terms, and what the leading terms are. Ordinarily, the $\{\alpha_j\}$ are not entirely determined, but must satisfy k relations, with $k\leqslant n$. The algorithm stops unless the only possible p_j's are integers.

2. Resonances of the system

For each p, construct a simplified equation from (3.1) that retains only the leading terms. Substitute into this simplified equation

$$w_j = \alpha_j(z-z_o)^{p_j} + \beta_j(z-z_o)^{p_j+r}, \quad j=1,\ldots,n, \tag{3.27}$$

with the same r for every w_j. To leading order in β, this becomes

$$[Q(r)]\beta = 0,$$

where $[Q]$ is an $n\times n$ matrix, whose elements depend on r. The resonances are the nonnegative roots of

$$\det[Q(r)] = 0, \tag{3.28}$$

a polynomial of order $\leqslant n$. One root is always (-1); zero may also be a root, with multiplicity depending on how many α_j's were determined in step 1. The algorithm stops unless all of the resonances are integers.

3. The constants of integration

Substitute into (3.1)

$$w_j\sim\alpha_j(z-z_o)^{p_j} + \sum_{k=1}^{r_s} a_{j_k}(z-z_o)^{p_j+k}, \tag{3.29}$$

where r_s is the largest resonance. The coefficient of each power of $(z-z_o)$, which must vanish, has the form of a matrix generalization of (3.20). Its treatment is identical to the previous case, and we omit the details.

Because we are interested in the analytic properties of functions, special attention must be given to equations (or systems) where the complex conjugate of an unknown function appears explicitly. The algorithm still can be applied, with these changes.

(i) Write the ODE for w, and the ODE for w^*.

(ii) Treat $v = w^*$ as a new variable. Then apply the algorithm to this system of two ODE's, without assuming *any* relation between v and w. Because the equation for v is the formal complex conjugate of that for w, if the initial conditions at a point of the *real axis* are formally complex conjugate, then the solutions will satisfy

$$v(z) = [w(z^*)]^*. \tag{3.30}$$

The behavior of v at a singular point z_o is related to the behavior of w at z_o^*, but *not* at z_o unless z_o is real. If among the possible leading behaviors, there are some where v and w differ, it only means that if complex conjugate initial data is given, there can be no such singularities on the real axis.

Example 8. For real x,d,

$$w'' = xw + 2|w|^2 w + d. \tag{3.31}$$

For real z, this implies

$$w'' = zw + 2w^2 w^* + d,$$

$$(w^*)'' = zw^* + 2(w^*)^2 w + d.$$

Therefore, consider

$$\begin{aligned} w'' &= zw + 2w^2 v + d, \\ v'' &= zv + 2v^2 w + d. \end{aligned} \tag{3.32}$$

1. Substitute

$$w \sim \alpha_1 (z - z_o)^p, \quad v \sim \alpha_2 (z - z_o)^q.$$

The only possibility is

$$p = q = -1, \quad \alpha_1 \alpha_2 = 1. \tag{3.33}$$

2. The simplified equations are

$$w'' \sim 2w^2 v, \quad v'' \sim 2v^2 w.$$

Substitute

$$w \sim \alpha (z - z_o)^{-1} + \beta_1 (z - z_o)^{r-1},$$

$$v \sim (1/\alpha)(z - z_o)^{-1} + \beta_2 (z - z_o)^{r-1},$$

and find

$$Q(r) = [(r-1)(r-2) - 4\alpha]^2 - 4 = 0, \tag{3.34}$$

with roots $r = -1,0,3,4$. The first two roots correspond to the arbitrary constants x_o and α.

3. Set $\xi = z - z_o$, and substitute into (3.32)

$$\begin{aligned} w &\sim \alpha \xi^{-1} + \sum_0^3 a_n \xi^n, \\ v &\sim \frac{1}{\alpha} \xi^{-1} + \sum_0^3 b_n \xi^n. \end{aligned} \tag{3.35}$$

Then find that

$$a_o = b_o = 0,$$

$$a_1 = -\alpha z_o/6, \quad b_1 = -z_o/6\alpha.$$

The first resonance occurs at the next order, where logarithmic terms arise unless

$$d = \alpha^2 d. \tag{3.36}$$

But α is arbitrary, so it follows from (3.36) that (3.32) is not of P-type unless

$$d = 0. \tag{3.37}$$

With (3.37), we find that b_2 is arbitrary, and

$$a_2 = \alpha/2 - \alpha^2 b_2.$$

The last resonance arises at the next order, where b_3 is arbitrary and

$$a_3 = \alpha^2 b_3.$$

It follows that (3.32) satisfies the necessary conditions to be of P-type if and only if d vanishes. With (3.37), we have four free constants, and (3.35) apparently represents the general solution of (3.32) in a deleted neighborhood of the movable pole at z_o. ■

ACKNOWLEDGMENTS

We are grateful for many illuminating discussions with Martin Kruskal. This work was partially supported by A.F.O.S.R. and by the U.S. Army Research Office.

APPENDIX

Here we show that the conditions in (2.7) are sufficient to apply the results of Fredholm theory to the linear integral equation in (2.6). For fixed ξ and contour $\mathscr{C}(\xi)$, let $(s_1, s_2, \ldots, s_n)$ denote a finite ordered sequence of points on $\mathscr{C}$. We need suitable bounds on Fredholm determinants of the form:

$$\Delta_n = \begin{vmatrix} N(s_1,s_1) N(s_1,s_2) & \cdots & N(s_1,s_n) \\ N(s_2,s_1) N(s_2,s_2) & \cdots & N(s_2,s_n) \\ \vdots & & \\ N(s_n,s_1) N(s_n,s_2) & \cdots & N(s_n,s_n) \end{vmatrix}, \tag{A1}$$

where the ξ dependence of N was suppressed. The bound comes from a complex generalization of Hadamard's Lemma [10]

$$\begin{aligned} |\Delta_n|^2 &\leqslant \prod_{\nu=1}^{n} \sum_{\mu=1}^{n} |N(s_\mu, s_\nu)|^2 \\ &\leqslant \prod_{\nu=1}^{n} n \sup_{\text{son } \mathscr{C}} |N(s, s_\nu)|^2 \\ &\leqslant n^n \prod_{\nu=1}^{n} M^2(S_\nu), \end{aligned}$$

using (2.7a). Therefore,

$$|\Delta_n| \leqslant n^{n/2} \prod_{\nu=1}^{n} M(s_\nu). \tag{A2}$$

Fredholm's first series has the form

$$\mathscr{D}_1(\mathscr{C}(\xi); \lambda) = 1 + \sum_{n=1}^{\infty} \frac{(-\lambda)^n}{n!} \oint \cdots \oint \Delta_n \, ds_1 \cdots ds_n. \tag{A3}$$

The magnitude of the nth term in this series does not exceed

$$\frac{(-\lambda)^n}{n!} \left(n^{n/2} \prod_{\nu=1}^{n} \oint M(s_\nu) \, ds_\nu \right) \leqslant n^{n/2} (c_1 \lambda)^n c_2, \tag{A4}$$

where c_1, c_2 are constants. It follows that the series converges for all values of λ and for all (ξ). Proof of the convergence of

Fredholm's second series follows similar lines, and these two series define the resolvent kernel. Then (2.7c) is needed to solve (2.6).

[1]M.J. Ablowitz, D.J. Kaup, A.C. Newell, and H. Segur, Stud. Appl. Math. **53**, 249 (1974).
[2]M.J. Ablowitz and H. Segur, Phys. Rev. Lett. **38**, 1103 (1977).
[3]M.J. Ablowitz, A. Ramani, and H. Segur, Lett. Nuovo Cimento **23**, 333 (1978).
[4]G.L. Lamb, Rev. Mod. Phys. **43**, 99 (1977).
[5]R. Rosales, Proc. R. Soc. London A **361**, 265 (1978).
[6]R. Nakach, Tech. Report #2, School of Elect. Eng., Chalmers Univ. Techn., Göteborg, Sweden (1977).
[7]M.J. Ablowitz, A. Ramani, and H. Segur, "A Connection between nonlinear evolution equations and ordinary differential equations of P-type. II", to appear in J. Math. Phys.
[8]E.L. Ince, *Ordinary Differential Equations* (Dover, New York, 1944).
[9]V.V. Golubev, *Lectures on Integrals of Equations of Motion* (State Publ. House, Moscow, 1953).
[10]W. Pogorzelski, *Integral Equations and Their Applications* (Pergamon, New York, 1966), Vol. I.
[11]V.E. Zakharov and A.B. Shabat, Funct. Anal. Appl. **8**, 226 (1974).
[12]H. Cornille, J. Math. Phys. **18**, 1855 (1977).

On Mathieu equation with damping

D. Y. Hsieh
Division of Applied Mathematics, Brown University, Providence, R. I. 02912

(Received 20 March 1979; accepted for publication 27 April 1979)

A direct variational method is applied to the linear and nonlinear Mathieu equation with damping. It is found that the nature of the periodic solutions and the characteristic curves are modified due to the presence of the damping. A threshold value of β is required to overcome the damping for the existence of the periodic solutions. Stability analyses for the periodic solutions are also carried out.

I. INTRODUCTION

In a previous paper,[1] hereafter referred as MI, a variational method was applied to study the linear and nonlinear Mathieu equation. We shall now extend this method to the study of the Mathieu equations with damping. Since in most physical problems some form of dissipation is always present, it is of realistic significance to include damping in the study of the Mathieu equations. The equation we shall study is to be taken as

$$\frac{d^2x}{dt^2} + \nu\frac{dx}{dt} + (\alpha + \beta\cos 2t)x + rx^3 = 0, \quad \nu > 0. \tag{1}$$

Following a previous study,[2] the last equation is equivalent to the variational formulation

$$\Delta J + \Delta I = 0, \tag{2}$$

$$J = \int_0^t \left[\left(\frac{dx}{dt}\right)^2 - \alpha x^2 - \beta x^2\cos 2t - \tfrac{1}{2}rx^4\right]dt, \tag{3}$$

and

$$\Delta I = -\int_0^t \left[\nu\left(\frac{dx}{dt}\right)\Delta x\right]dt. \tag{4}$$

We shall first study the linear equation, i.e., when $r = 0$, and then the nonlinear equation. The periodic solutions corresponding to the mode $m = 1$ and their stability are investigated in detail as in previous studies by the direct variational method.

II. LINEAR MATHIEU EQUATION

Let us take the asymptotic trial solution as in MI

$$x = \frac{A_0}{2} + \sum_{k=1} A_k\cos kt + \sum_{k=1} B_k\sin kt, \tag{5}$$

and substitute in (3) and (4). For this linear case, we have $r = 0$. The approximate expression of the functional J for large t is again the same as given in MI, while

$$\Delta I \simeq -\nu\int_0^t dt\left\{\frac{1}{4}\frac{dA_0}{dt}\Delta A_0 + \frac{1}{2}\sum_{m=1}\left[\left(\frac{dA_m}{dt} + mB_m\right)\times\Delta A_m + \left(\frac{dB_m}{dt} - mA_m\right)\Delta B_m\right]\right\}. \tag{6}$$

The variation with respect to A_m and B_m by Eq. (2) then leads to

$$\frac{d^2A_0}{dt^2} + \alpha A_0 + \beta A_2 + \nu\frac{dA_0}{dt} = 0, \tag{7}$$

$$-\frac{d}{dt}\left(\frac{dA_1}{dt} + B_1\right) + \left(A_1 - \frac{dB_1}{dt}\right) + \left[-\alpha A_1 - \frac{\beta}{2}(A_1 + A_3)\right] - \nu\left(\frac{dA_1}{dt} + B_1\right) = 0, \tag{8}$$

$$-\frac{d}{dt}\left(\frac{dA_m}{dt} + mB_m\right) + m\left(mA_m - \frac{dB_m}{dt}\right) + \left[-\alpha A_m - \frac{\beta}{2}(A_{m-2} + A_{m+2})\right] - \nu\left(\frac{dA_m}{dt} + mB_m\right) = 0, \quad m \geqslant 2, \tag{9}$$

$$-\frac{d}{dt}\left(\frac{dB_1}{dt} - A_1\right) + \left(\frac{dA_1}{dt} + B_1\right) + \left[-\alpha B_1 + \frac{\beta}{2}(B_1 - B_3)\right] - \nu\left(\frac{dB_1}{dt} - A_1\right) = 0, \tag{10}$$

and

$$-\frac{d}{dt}\left(\frac{dB_m}{dt} - mA_m\right) + m\left(mB_m + \frac{dA_m}{dt}\right) + \left[-\alpha B_m - \frac{\beta}{2}(B_{m-2} + B_{m+2})\right] - \nu\left(\frac{dB_m}{dt} - mA_m\right) = 0, \quad m \geqslant 2. \tag{11}$$

The equilibrium solutions of this system of equations represent the periodic solutions of the original differential equation. When $\nu = 0$, they are the usual Mathieu functions. With nonvanishing ν, the equations for the characteristic curves for various modes will be modified. Let us now consider in particular the modes identified with the Mathieu functions $\{ce_1, se_1\}$, i.e., with $m = 1$. We shall study the case where both β and ν are small in the same order, i.e., $\beta = 0(\epsilon)$, $\nu = 0(\epsilon)$, and $\epsilon \ll 1$.

To the order of $0(\epsilon)$, we can simply take $x = A_1\cos t + B_1\sin t$ in (5), and obtain only Eqs. (8) and (10) with A_3 and B_3 missing. As in MI, the terms with second derivatives with respect to time are neglected, since A_1 and B_1 are assumed to be slowly varying functions of time. Thus we obtain

$$-2\frac{dB_1}{dt} - \nu\frac{dA_1}{dt} + \left[(1-\alpha) - \frac{\beta}{2}\right]A_1 - \nu B_1 = 0, \tag{12}$$

 0022-2488/80/040722-04$1.00

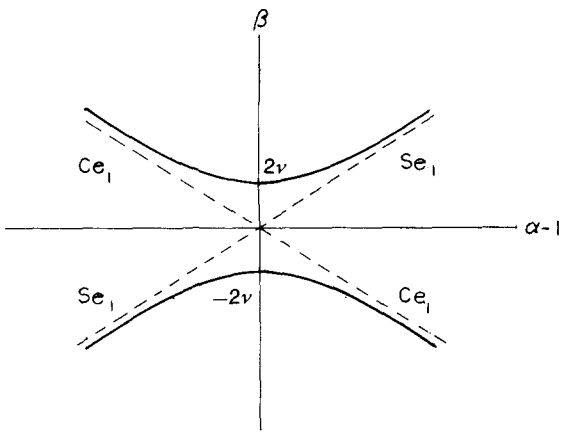


FIG. 1. Characteristic curves for the mode $m = 1$.

and

$$2\frac{dA_1}{dt} - \nu\frac{dB_1}{dt} + \left[(1-\alpha) + \frac{\beta}{2}\right]B_1 + \nu A_1 = 0. \quad (13)$$

The above system has equilibrium solutions if the following condition is satisfied:

$$\left(\frac{\beta}{2}\right)^2 - (\alpha - 1)^2 = \nu^2. \quad (14)$$

For these equilibrium solutions, we have

$$A_1^{(0)} = \frac{1}{\nu}\left[(\alpha - 1) - \frac{\beta}{2}\right]B_1^{(0)}.$$

Thus the null solution $A_1^{(0)} = B_1^{(0)} = 0$ is a particular case, and otherwise $A_1^{(0)}$ and $B_1^{(0)}$ are both nonzero except when $\nu = 0$. This situation is quite different from the case of Mathieu equation without damping. For that case, as given in MI, three branches of equilibrium solutions are permissible, i.e., the null solution, the modes se_1 and ce_1. The characteristic curves for the modes se_1 and ce_1 are shown in dotted lines in Fig. 1. Now, due to the presence of damping, we have two branches of the equilibrium solutions; each is a mixture of modes corresponding to se_1 and ce_1. The characteristic curves as given by Eq. (14) are shown in solid lines in Fig. 1. It is worth noting that the minimum value of β necessary to overcome the damping for periodic solution to be possible is 2ν.

To study the stability of the equilibrium solutions, let us take

$$A_1 = A_1^{(0)} + a(t) \quad \text{and} \quad B_1 = B_1^{(0)} + b(t).$$

Then the equations governing $a(t)$ and $b(t)$ are the same as Eqs. (12) and (13) with a and b replacing A_1 and B_1 respectively. Thus, after eliminating b, we obtain

$$(4+\nu^2)\frac{d^2a}{dt^2} + 2\nu(1+\alpha)\frac{da}{dt} + \left[(\alpha-1)^2 - \left(\frac{\beta}{2}\right)^2 + \nu^2\right]a = 0. \quad (15)$$

Using the equilibrium condition (14), we thus have

$$(4+\nu^2)\frac{d^2a}{dt} + 2\nu(1+\alpha)\frac{da}{dt} = 0. \quad (16)$$

Therefore, the system is always stable if and only if ν is positive. We may recall that when the damping is absent the modes $\{se_1, ce_1\}$ are not stable as shown in MI.

It is significant that the introduction of a slight damping, which is invariably present in any real physical system, will stabilize the periodic solutions. When ν is small, then periodic solutions are practically the Mathieu functions se_1 and ce_1 when $(\alpha - 1)$ is not small. Only when $(\alpha - 1)$ is small do they become a mixture of se_1 and ce_1 of comparable magnitude.

It is of interest to note that the investigation of the stability of the plane free surface of a liquid in vertical periodic motion will lead to the study of Mathieu equations. An experimental characteristic curve has been obtained by Benjamin and Ursell.[3] It is significant that the curve is indeed like a hyperbola with a finite threshold value of β as shown in Fig. 1.

Without going into details, it may be shown that a similar modification occurs for the mode $m = 2$. To the order of $0(\epsilon)$, the characteristic curves are given by the equation

$$\tfrac{5}{48}\beta^2 = -2(\alpha - 4) + [9(\alpha - 4)^2 + 20\nu^2]^{1/2}. \quad (17)$$

In Fig. 2, the characteristic curves for the mode $m = 2$ are shown schematically in solid lines. The dotted lines are the characteristic curves when $\nu = 0$. The characteristic curves are again symmetric in β.

III. NONLINEAR MATHIEU EQUATIONS.

We shall concentrate on the mode $m = 1$ for the study of the nonlinear Mathieu equation (1). Take again the asymptotic trial solution $x = A_1 \cos t + B_1 \sin t$ and neglect the second derivative terms; we obtain the following equations corresponding to (12) and (13):

$$-2\frac{dB_1}{dt} - \nu\frac{dA_1}{dt} + \left[(1-\alpha) - \frac{\beta}{2} - \frac{3}{4}r(A_1^2 + B_1^2)\right] \times A_1 - \nu B_1 = 0, \quad (18)$$

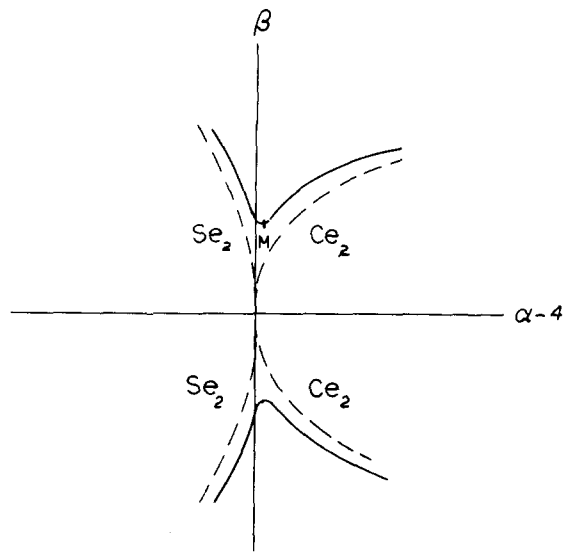
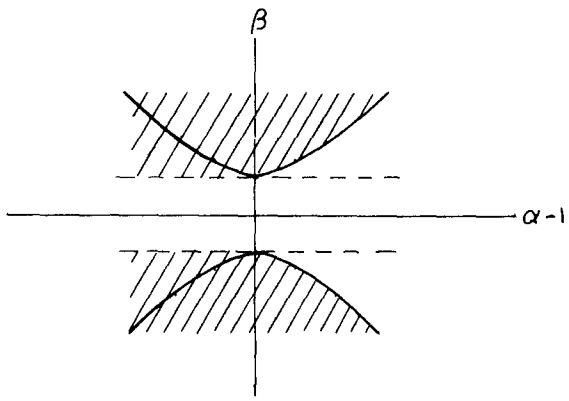
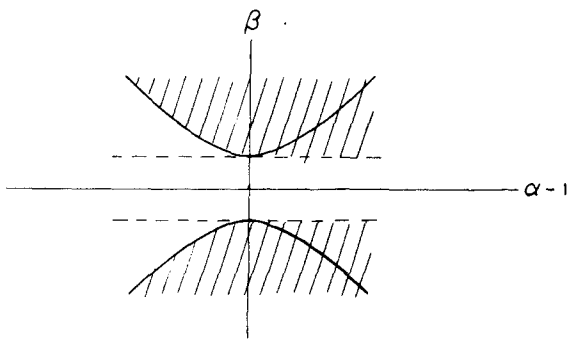


FIG. 2. Characteristic curves for the mode $m = 2$. The coordinates of M are $\alpha - 4 = (4/3)\nu$ and $\beta = 4(2\nu)^{1/2}$.



(a) r > 0



(b) r < 0

FIG. 3. Permissible region of the branch S. Solid curves are given by $(\alpha-1)^2=(\delta/2)^2$.

and

$$2\frac{dA_1}{dt}-\nu\frac{dB_1}{dt}+\left[(1-\alpha)+\frac{\beta}{2}-\frac{3}{4}r(A_1^2+B_1^2)\right]B_1+\nu A_1=0. \tag{19}$$

Besides the null solution, the equilibrium solutions of the above system satisfy the following condition:

$$\left[(1-\alpha)-\frac{3}{4}r(A_1^2+B_1^2)\right]^2-\left(\frac{\beta}{2}\right)^2+\nu^2=0. \tag{20}$$

It is clear that the real solutions are permissible only if

$$\left|\frac{\beta}{2}\right|\geqslant\nu. \tag{21}$$

Thus, a threshold value of β when damping is present is again required for the nonlinear case. Equation (20) leads to

$$A_1^2+B_1^2=\frac{4}{3r}\left[(1-\alpha)\pm\left(\frac{\beta^2}{4}-\nu^2\right)^{1/2}\right]. \tag{22}$$

Let us denote the branch with (+) sign in front of the square root by S, and the other branch with (−) sign by U. The permissible region in the (α,β) plane for branches S and U are shown in Figs. 3 and 4.

Let us consider first the case $r>0$, and denote

$$\frac{\delta}{2}=\left(\frac{\beta^2}{4}-\nu^2\right)^{1/2}. \tag{23}$$

We shall consider the branches S and U separately.

(i) Branch S. The branch S of the equilibrium solutions as found from (18) and (19) is given by

$$A_{1S}^2=\frac{8\nu^2}{3r\beta(\beta+\delta)}\left(1-\alpha+\frac{\delta}{2}\right), \tag{24}$$

$$B_{1S}^2=\frac{2(\beta+\delta)}{3r\beta}\left(1-\alpha+\frac{\delta}{2}\right), \tag{25}$$

and

$$B_{1S}=-\frac{1}{2\nu}(\beta+\delta)A_{1S}. \tag{26}$$

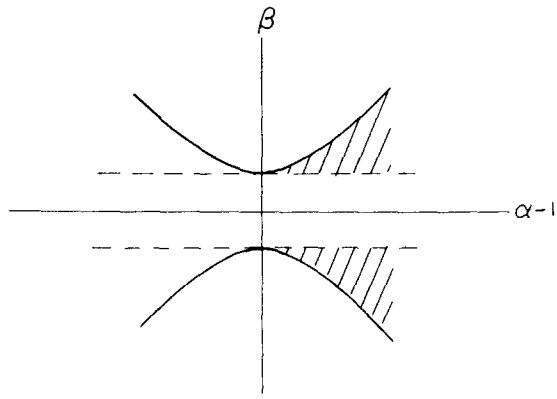
The stability of the branch S can be investigated by taking

$$A_1=A_{1S}+a(t),\quad B_1=B_{1S}+b(t),$$

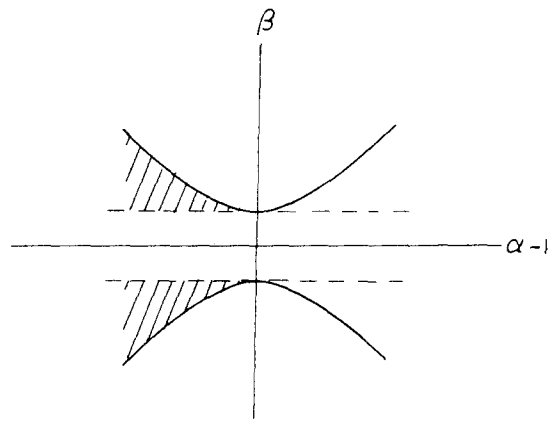
and substituting into Eqs. (18) and (19). Thus, we obtain

$$-2\frac{db}{dt}-\nu\frac{da}{dt}+\left[(1-\alpha)-\frac{\beta}{2}-\frac{3}{4}r(3A_{1S}^2+B_{1S}^2)\right]\times a-\left(\frac{3}{2}rA_{1S}B_{1S}+\nu\right)b=0, \tag{27}$$

$$2\frac{da}{dt}-\nu\frac{db}{dt}+\left[(1-\alpha)+\frac{\beta}{2}-\frac{3}{4}r(A_{1S}^2+3B_{1S}^2)\right]b$$



(a) r > 0



(b) r < 0

FIG. 4. Permissible region of the branch U. Solid curves are given by $(\alpha-1)^2=(\delta/2)^2$.

$$+\left(\nu - \frac{3}{2} r A_{1S} B_{1S}\right) a = 0. \tag{28}$$

Using relations (22)–(26) and after eliminating b, we can obtain from (27) and (28) the following equation:

$$(4+\nu^2)\frac{d^2a}{dt^2} + 2\nu(3-\alpha+\delta)\frac{da}{dt} + \delta[2(1-\alpha)+\delta]a = 0. \tag{29}$$

Let us denote $L = 4+\nu^2$, $M = \nu(3-\alpha+\delta)$, and $N = \delta[2(1-\alpha)+\delta]$. Then

$$a(t) = c_1 e^{\mu_1 t} + c_2 e^{\mu_2 t},$$

where

$$\mu_{1,2} = \frac{1}{L}[-M \pm (M^2 - LN)^{1/2}].$$

Now $L > 0$, and for $(\alpha - 1) = 0(\epsilon)$, and $\delta = 0(\epsilon)$ we have $M > 0$. Therefore, both μ_1 and μ_2 have negative real parts if $N > 0$, while μ_1 will have positive real parts if $N < 0$. Since δ is positive, the equilibrium branch S is stable if and only if

$$(1-\alpha) + \delta/2 > 0. \tag{30}$$

The condition (30) is exactly the same as the permissibility conditions for the branch S. Therefore, the branch S is stable whenever it is permissible, i.e., the branch S is a stable branch.

(ii) Branch U. The corresponding relations for the branch U can be obtained from the branch S by changing δ to $(-\delta)$. Thus, corresponding to (29), the stability equation now becomes

$$(4+\nu^2)\frac{da}{dt} + 2\nu(3-\alpha-\delta)\frac{da}{dt} - \delta[2(1-\alpha)-\delta]a = 0. \tag{31}$$

The equilibrium branch U is thus stable if and only if

$$(1-\alpha) - \delta/2 < 0. \tag{32}$$

The condition (32) is exactly opposite to the permissibility condition for the branch U. Thus, the branch U is an unstable branch.

For the case $r < 0$, similar analyses can be carried out. Let us denote the equilibrium branch S by

$$-\frac{3}{4} r(A_1^2 + B_1^2) = (\alpha - 1) + \frac{\delta}{2}, \tag{33}$$

and the equilibrium branch U by

$$-\frac{3}{4} r(A_1^2 + B_1^2) = (\alpha - 1) - \frac{\delta}{2}. \tag{34}$$

Then it can be shown in a corresponding manner that the branch S is stable, while the branch U is unstable. The permissible regions in the (α,β) plane for these branches are again shown in Figs. 3 and 4.

To sum up, the detailed study on the periodic solutions of the mode $m = 1$ has shown that a threshold value of β is required to overcome the damping, that the periodic solutions are stable and are in general a mixture of the modes corresponding to se_1 and ce_1. For the nonlinear case, it is found that only one equilibrium branch of the two permissible solutions is stable. Numerical computations have been carried out for some initial value problems. They tend to support at least qualitatively the main results from the analyses.

ACKNOWLEDGMENT

The computational assistance of Paul Hsieh is appreciated.

[1] D.Y. Hsieh, J. Math. Phys. **19**, 1147–51 (1978).
[2] D.Y. Hsieh, J. Math. Phys. **16**, 275–80 (1975).
[3] T.B. Benjamin and F. Ursell, Proc. Roy. Soc. A **225**, 505–15 (1954).

On the separability of the sine-Gordon equation and similar quasilinear partial differential equations. II. Dependent- and independent-variable transformations

Anthony D. Osborne and Allan E.G. Stuart
Department of Mathematics, The City University, Northampton Square, London EC1V OHB, England

(Received 4 June 1979; accepted for publication 11 September 1979)

In an earlier paper, we investigated the separability of the sine-Gordon equation (SGE), and of similar quasilinear partial differential equations, under transformations of the dependent variable (i.e, of the codomain). We found, in particular, that there is a general class of dependent-variable transformations which leads to separable forms of the SGE. In this paper, we extend our previous analysis to include independent- as well as dependent-variable transformations (i.e., transformations of both the domain and codomain) and treat, in detail, constant coefficient equations of the first and second orders. We illustrate our method by applying it to the SGE and find combinations of domain and codomain transformations which reduce the equation to separable forms. Some of these transformations lead to known solutions of the SGE, but others give new solutions expressible in terms of a fifth Painlevé transcendent. Our method can, in principle, be used to map out the space of separable solutions of the SGE and other similar second-order equations, but it does have limitations. A discussion of these limitations and suggestions for possible improvements are also given.

1. INTRODUCTION

In the theory of linear partial differential equations one of the more important ways of obtaining families of solutions depending on arbitrary parameters is by the "separation of variables". The application of this technique reduces the task of solving a separable linear partial differential equation to that of solving an uncoupled system of linear ordinary differential equations which, of course, is much easier. Furthermore, the existence of a superposition principle enables us to use Fourier series expansions to fit a great variety of initial and boundary conditions either exactly or approximately.

The success of this method in the class of linear partial differential equations of mathematical physics leads us, quite naturally, to ask whether the technique can be extended to the classes of those quasilinear and nonlinear partial differential equations which are currently being used as mathematical models in various fields of science. However, in asking this question we should keep in mind that the problems of nonlinear theory are much more difficult and varied than those of linear theory and that, as a consequence, any nonlinear extension of the separability technique is bound to be less versatile and less general than its linear counterpart. There are several reasons for this. First, the absence of a superposition principle means that we do not have the possibility of constructing general solutions out of the special separable ones. In effect, this implies that unless some other way is found of systematically generating general solutions from special solutions, then the separable solutions will only form an isolated subset of the set of all solutions. Second, the reduction of a nonlinear partial differential equation to a nonlinear ordinary differential equation does not necessarily mean that the problem becomes any easier. If the resulting ordinary differential equation is one we know little or nothing about, then the advantage gained may be minimal. Third, information obtained from other methods developed for solving nonlinear equations, such as inverse scattering, Bäcklund transformations, etc., suggest that, in general, the set of separable solutions may be small or even trivial. Nevertheless, there is enough evidence to indicate that there are many interesting equations with nontrivial sets of separable solutions and, in these favorable cases, the development of systematic methods for separating the equations should prove useful. Furthermore, a detailed study of these methods and solutions may give us some insight into the structure of the equations and into the structure and formation of soliton solutions when they exist.

Now one of the more important equations which does have a variety of separable solutions is the sine-Gordon equation (SGE) and early indications of this fact were given in papers by Lamb[1] and Zagrodzinski.[2] We followed up their leads and, in a previous paper,[3] showed how one can systematically derive separable forms for the SGE, and similar quasilinear equations, through the use of dependent-variable (i.e., codomain) transformations. However, in linear theory it is well-known that the separability or otherwise of an equation depends very often on the coordinate system in which it is expressed. Thus, the results presented in Ref. 3 are somewhat incomplete, since we did not make use of all the degrees of freedom available to us. In this paper, we attempt to remedy this deficiency by extending our analysis to include transformations of the domain as well as the codomain. We also consider a wider class of equations, although our explicit demonstrations are restricted to the SGE and a hyperbolic variant of Liouville's equation.

The plan of the rest of the paper and a summary of our main findings are as follows. In Sec. 2, we discuss the separability of first-order quasilinear equations with constant coefficients and specify the transformations which lead to separable forms. We find that the only separable form of a

 0022-2488/80/040726-09$1.00

quasilinear equation is linear and show how an infinite number of solutions of the equation can be obtained by using a combination of both dependent- and independent-variable transformations. Second-order quasilinear equations with constant coefficients are considered in Sec. 3, where we show how to find the dependent-variable transformations which lead to separable forms; the coefficients of the derivatives in the equation being critical. Independent-variable transformations leading to constant-coefficient equations and their "interactions" with the dependent-variable transformations are then discussed. Section 2 ends with an application of this analysis to a hyperbolic variant of Liouville's equation. In Sec. 4, we turn our attention to the SGE and its elliptic variant. Using the theorems developed in Sec. 3, we find combinations of domain and codomain transformations which reduce the equations to separable forms. Some of these forms lead to known solutions of the SGE, but others give new solutions some of which are expressible in terms of a fifth Painlevé transcendent.

The appearance of Painlevé transcendents in connection with the SGE is not surprising. Recent work by Ablowitz and Segur[4] and by Ablowitz, Ramani, and Segur,[5] has shown that there is a close connection between Painlevé equations and nonlinear evolution equations solvable by the inverse scattering method. For example a second Painlevé equation was studied in connection with the modified Korteweg–de Vries equation,[4] while a special case of the third Painlevé transcendent was obtained from the SGE.[4,6] The Painlevé equations themselves are now under intensive study and many of their basic properties and solutions are fairly well known.[7,8]

Our method can, in principle, be used to map out the space of separable solutions of the SGE and other similar second-order equations, but there are limitations. A discussion of these limitations, together with our suggestions for possible improvements are presented in our concluding remarks in Sec. 5.

Throughout this paper we shall be working with a two-dimensional domain $U \subset R^2$ and shall, for the most part, assume that the dependent variable is the image of a continuous map from U to the real line which is at least twice differentiable on U

2. FIRST-ORDER EQUATIONS WITH CONSTANT COEFFICIENTS

Since the solutions of any quasilinear partial differential equation of the first order can be obtained from the solutions of a system of first order ordinary differential equations using the theory of characteristics,[9] the results given in this section may not be of much practical use. Furthermore, they are somewhat limited since we only deal with constant-coefficient equations and restrict ourselves to the class of separable solutions. However, the aim of our paper is to present results on the separability of quasilinear equations and, from this point of view, the inclusion of this section has some advantages. These are: (i) it helps to make our presentation systematic; (ii) it shows up important differences between first-order and higher-order equations: (iii) in certain cases, we feel that the ordinary differential equations obtained by our method are easier to solve than those obtained for the characteristics; and (iv) the results in themselves are not uninteresting. It is for these reasons and also for completeness that we have included the first-order case.

Consider, therefore, a general, first-order quasilinear equation in two independent variables and with constant coefficients

$$a\phi_x + b\phi_y = f(\phi), \tag{2.1}$$

where $a,b \in \mathbb{R}$ and the subscripts denote partial derivatives. Scaling the independent variables enables us to write (2.1) as

$$\phi_x + \phi_y = f(\phi). \tag{2.2}$$

From the general test for separability, given in Ref. 3, one can easily deduce that (2.2) is separable if and only if $f(\phi)$ is linear. If $f(\phi)$ is nonlinear, then we need a codomain transformation, $\phi \to \psi = g^{-1}(\phi)$ say, to put (2.2) in separable form. Assuming that $f(g)$ is analytic in ψ then gives us the following theorem about the possible forms of g.

Theorem 2.1: The only dependent-variable transformations which lead to separable versions of (2.2) are necessarily of the form

$$\ln(\alpha\psi^{\beta}) = \int \frac{dg}{f(g)}, \tag{2.3}$$

where ψ is the new dependent variable and $\alpha,\beta \in \mathbb{R} - \{0\}$ are constants. Furthermore, these separable forms are always linear. □

Proof: Applying the dependent-variable transformation $\phi = g(\psi)$ to (2.2) we obtain

$$(\psi_x + \psi_y)\, g'(\psi) = f[g(\psi)]. \tag{2.4}$$

We now assume $\psi(x,y) = X(x)Y(y)$ and substitute the series expansions $X' = \Sigma_0^{\infty} a_n X^{n+\lambda}$ and $Y' = \Sigma_0^{\infty} b_n Y^{n+\rho}$ into (2.4). This turns (2.4) into an identity and equating coefficients leads to a set of recurrence relations for the a_n, b_n, λ and ρ. A consistent solution of these recurrence relations then gives us $\lambda = \rho = 0$ and the reduced equation

$$A\psi g' = f(g), \tag{2.5}$$

where $A = (a_1 + b_1) \in \mathbb{R} - \{0\}$. Integration of (2.5) leads to (2.3) and hence gives the first result. If we now substitute the necessary condition (2.5) back into (2.4) we obtain the separable form

$$\psi_x + \psi_y = A\psi, \quad A \in \mathbb{R} - \{0\}. \tag{2.6}$$

This is a linear partial differential equation for ψ and hence shows us that at least some of the separable versions of (2.2) are linear.

To prove the uniqueness of (2.6), we consider further transformations of the dependent-variable ψ which maintain separability. According to what has been proved above [i.e., (2.3)], these further transformations must necessarily take the form $\psi \to \theta^{\delta}$ with $\delta \in \mathbb{R} - \{0\}$. However, as is easily checked, a transformation of this type leaves (2.6) invariant modulo a change in the value of the constant A. Thus, (2.6) gives the only separable forms of (2.2). This completes the proof of the theorem. ■

We now consider domain transformations of (2.2) which lead to separable forms with constant coefficients.

Thus, if we let $\xi = \xi(x,y)$ and $\eta = \eta(x,y)$ be differential maps of the initial domain, then, in terms of these new coordinates, (2.2) becomes

$$(\xi_x + \xi_y)\phi_\xi + (\eta_x + \eta_y)\phi_\eta = f(\phi). \tag{2.7}$$

For constant coefficients we must have

$$\xi_x + \xi_y = \text{constant} = a, \tag{2.8a}$$

and

$$\eta_x + \eta_y = \text{constant} = b, \tag{2.8b}$$

and the general solutions of these equations are given by

$$F(x-y, ax-\xi) = 0, \quad G(x-y, bx-\eta) = 0, \tag{2.9}$$

where F and G are arbitrary functions with $\partial F/\partial(ax-\xi)$ and $\partial G/\partial(bx-\eta) \neq 0$. We therefore have an infinite set of possible transformations which enable us to get from (2.2) back to an equation of the form (2.1), but with a different coordinate system on the domain. In particular, if we put $a = b = 1$, then (2.2) remains invariant under the transformations represented by (2.9). Since neither (2.1) nor (2.2) is a separable form, it is clear that a domain transformation on its own will not separate the equation. However, combining transformations of the type (2.9) with the dependent-variable transformation (2.3) enables us to get a separable form and, in this way, any number of solutions of the original equation may be found. The example given below will make this clear.

Example: Consider the equation

$$\phi_x + \phi_y = \phi^2. \tag{2.10}$$

Now let $F = -\sin(x-y) + ax - \xi = 0$ and $G = \sin(x-y) - (x-y) + bx - \eta = 0$. Choosing $a = b = 1$ then gives us the transformation

$$\xi = x - \sin(x-y), \quad \eta = y + \sin(x-y),$$

and, under this transformation, (2.10) becomes

$$\phi_\xi + \phi_\eta = \phi^2. \tag{2.11}$$

Using theorem 2.1 we let $\phi = -1/\ln\psi$ and obtain the linear equation

$$\psi_\xi + \psi_\eta = \psi. \tag{2.12}$$

The separable solutions of (2.12) are given by

$$\psi(\xi,\eta) = \exp[A\xi + (1-A)\eta + B] \quad (A,B \text{ constants}),$$

and going back to the original variables gives us one set of solutions of (2.10) as

$$\phi(x,y) = [(2A-1)\sin(x-y) - A(x-y) - y - B]^{-1}. \tag{2.13}$$

On the other hand, if we make the obvious linear transformation

$$F = -(x-y) + 2x - \xi = 0 \quad \text{and} \quad G = (x-y) - \eta = 0 \quad (\text{i.e., } \xi = x+y, \eta = x-y),$$

then (2.10) reduces to

$$2\phi_\xi = \phi^2. \tag{2.14}$$

This can be integrated straightaway to give the solutions

$$\phi(x,y) = -2/[(x+y) + f(x-y)], \tag{2.15}$$

where f is an arbitrary function. It is easy to check that reducing (2.14) to a separable form via (2.3), also leads to the solution (2.15). Note, however, that (2.15) and (2.13) are, in general, different solutions. They only coincide for the special case when $A = \frac{1}{2}$ and f is the constant function.

From the above example, it is clear that the usefulness of our method depends to a great extent on the choice of a coordinate system. The same difficulty arises in the method of characteristics where one has to choose an initial curve through which the integral surface has to pass. This can easily be seen by trying, for example, to obtain the solutions (2.13) of (2.10) using the method of characteristics. However, in both cases, if one has some information about the solution one is trying to obtain, such as periodicity, asymptotic form, etc., then these choices may be made somewhat easier.

To summarize, the main result of this section is that first-order, quasilinear partial differential equations of the type (2.1), may be solved by reducing them to a linear equation of the form (2.6) using a combination of domain and codomain transformations. However, we can only deal with a restricted class of equations, whereas, as mentioned earlier, the method of characteristics is a general technique applicable to *any* quasilinear, first-order equation. As a result, our approach can only be of limited practical use. On the other hand, most of the quasilinear equations of mathematical physics are at least of second order and in these cases, unlike the first-order case, there is no general method of solution. A systematic procedure for finding the separable forms of second-order equations should therefore prove useful. The rest of the paper deals with such equations subject to the condition that they have constant coefficients.

3. SECOND-ORDER EQUATIONS WITH CONSTANT COEFFICIENTS

A general, quasilinear partial differential equation of the second order, defined with constant coefficients over a two-dimensional domain is

$$a\phi_{xx} + b\phi_{yy} + c\phi_{xy} + d\phi_x + e\phi_y = f(\phi), \tag{3.1}$$

where $a,b,c,d,e \in \mathbb{R}$ and $\phi = \phi(x,y)$. Using the general technique for separability,[3] it can be shown, after some manipulation, that one can obtain Lemma 3.1.

Lemma 3.1: Equation (3.1) is not separable as it stands unless $f(\phi) = \phi$, or $a = b = d = e = 0$ and $f(\phi) = \phi^\lambda$, $\lambda \in \mathbb{R}$.

Remark: Note that this is an extension of Lemma 3.1 of Ref. 3 and that, in both these Lemmas, our definition of separability excludes direct procedures which involve further differentiations of (3.1). There are reasons for this exclusion and these are best seen via an example.

Example: In (3.1), let $f(\phi) = \phi^3 - \phi$ and assume that $\phi(x,y) = X(x)Y(y)$ not identically zero. Now carry out the following sequence of operations: (i) substitute for ϕ and its partial derivatives in (3.1) and divide by XY; (ii) partially differentiate the resulting expression with respect to x; (iii) partially differentiate again, but this time with respect to y. Then, if $c \neq 0$, the equation separates to give

$$cX' = \alpha X^3 + \beta X \Rightarrow c^2 X'' = (3\alpha X^2 + \beta)(\alpha X^3 + \beta X),$$

$$\alpha Y' = Y^3 + \gamma Y \Rightarrow \alpha^2 Y'' = (3Y^2 + \gamma)(Y^3 + \gamma Y),$$

where α is the separation constant and β and γ are constants of integration. However, this separation is only compatible

with our choice of $f(\phi)$ when $a = b = 0, c = -\beta d, e = -\beta$, and $\alpha = \gamma\beta$. This means, for example, that the method will not lead to a consistent separation of the equation $\phi_{xy} = \phi^3 - \phi$. The situation gets worse when $c = 0$. In this case, the only separations one can obtain are those for which either X or Y is a constant function; which is equivalent to having either $c = a = d = 0$ or $c = b = e = 0$ in (3.1).

This example demonstrates quite clearly that a direct procedure, based on further differentiations, only works on special cases of (3.1) rather than on (3.1) itself. Furthermore, it is obvious that the method is only feasible when $f(\phi)$ is a polynomial of low order or has special forms; for example, $\phi_{xx} + \phi_{yy} = \phi(1 + \ln\phi)$ separates directly under further differentiation. However, if $f(\phi)$ is a high-order polynomial, or a transcendental function, then the method becomes virtually useless. It is for these reasons that we have excluded it from Lemma 3.1.

Returning to the main theme of this section, since (3.1) is not generally separable as it stands, we first look for dependent-variable transformations which lead to separable forms. A method for finding these transformations is given by the following theorem.

Theorem 3.1: (a) If at least one of a,b,d and e is nonzero, then a necessary condition for (3.1) to reduce to a separable form under the codomain transformation $\phi \to \psi = g^{-1}(\phi)$, is that $g(\psi)$ satisfies the ordinary differential equation

$$\psi(Ag' + B\psi g'') = f(g), \tag{3.2}$$

where A and B are functions of the expansion coefficients and the coefficients of (3.1).

(b) If a,b,d and e are zero, then the differential equation for g takes the form

$$C\psi^m(g' + \psi g'') = f(g), \tag{3.3}$$

where C is a constant similar to A and B above, but m is an arbitrary nonnegative integer. □

Proof: Applying the transformation $\phi = g(\psi)$ to (3.1), we obtain

$$(a\psi_{xx} + b\psi_{yy} + c\psi_{xy} + d\psi_x + e\psi_y)g' + (a\psi_x^2 + b\psi_y^2 + c\psi_x\psi_y)g'' = f(g). \tag{3.4}$$

We now let $\psi = X(x)Y(y)$ and, assuming that $f(g)$ is analytic in ψ, substitute the series expansions $X' = \Sigma_0^\infty a_n X^n$ and $Y' = \Sigma_0^\infty b_n Y^n$ into (3.4). This gives us the identity

$$\begin{aligned}&\left[aY\left(\sum_0^\infty a_n X^n\right)\left(\sum_1^\infty n a_n X^{n-1}\right) + bX\left(\sum_0^\infty b_n Y^n\right)\right.\\ &\quad\times\left(\sum_1^\infty n b_n Y^{n-1}\right) + c\left(\sum_0^\infty a_n X^n\right)\left(\sum_0^\infty b_n Y^n\right)\\ &\quad\left.+ dY\left(\sum_0^\infty a_n X^n\right) + eX\left(\sum_0^\infty b_n Y^n\right)\right]g'\\ &+\left[aY^2\left(\sum_0^\infty a_n X^n\right)^2 + bX^2\left(\sum_0^\infty b_n Y^n\right)^2\right.\\ &\quad\left.+ cXY\left(\sum_0^\infty a_n X^n\right)\left(\sum_0^\infty b_n Y^n\right)\right]g''\\ &\quad= f(g).\end{aligned} \tag{3.5}$$

We now have three cases:

(i) Assume $c = 0$. Since g is a function of XY, g', g'' and $f(g)$ must all be functions of XY. Thus, a necessary condition for (3.5) to be an identity is that the coefficients of X^nY^n on the left-hand side (lhs) are consistent with the coefficients of $(XY)^n$ on the right-hand side (rhs). Hence, on eliminating the zeroes on the lhs, we obtain the reduced equation

$$XY[a(2a_0a_2 + a_1^2) + b(2b_0b_2 + b_1^2) + da_1 + eb_1]g' + X^2Y^2[a(2a_0a_2 + a_1^2) + b(2b_0b_2 + b_1^2)]g'' = f(g). \tag{3.6}$$

Letting

$$A = a(a_1^2 + 2a_0a_2) + b(b_1^2 + 2b_0b_2) + da_1 + eb_1,$$

and

$$B = A - da_1 - eb_1,$$

then takes (3.6) to (3.2) as required.

(ii) Assume $c \neq 0$, but $a = b = d = e = 0$. The identity (3.5) is now reduced to

$$c\left(\sum_0^\infty a_n X^n\right)\left(\sum_0^\infty b_n Y^n\right)(g' + XYg'') = f(g). \tag{3.7}$$

It is easy to see that (3.7) is consistent if and only if

$$\left(\sum_0^\infty a_n X^n\right)\left(\sum_0^\infty b_n Y^n\right) = a_m b_m (XY)^m,$$

for some nonnegative integer m. Thus, (3.7) becomes

$$ca_m b_m \psi^m(g' + \psi g'') = f(g), \tag{3.8}$$

which is the same as (3.3) with $C = ca_m b_m$.

(iii) Assume $c \neq 0$ and that at least one of the a,b,d,e is not zero. Then, combining the results of (i) and (ii) above, we see that (3.5) will only be consistent as an identity if $X' = a_1X$ and $Y' = b_1Y$. This leads to the equation

$$XY(aa_1^2 + bb_1^2 + ca_1b_1 + da_1 + eb_1)g' + X^2Y^2(aa_1^2 + bb_1^2 + ca_1b_1)g'' = f(g), \tag{3.9}$$

and putting

$$A = aa_1^2 + bb_1^2 + ca_1b_1 + da_1 + eb_1, \tag{3.10a}$$

$$B = A - da_1 - eb_1, \tag{3.10b}$$

then reduces (3.9) to (3.2) as required. ■

Remark: Note the part played by the mixed derivative in the above proof. This will obviously become important when we consider the influence of domain transformations on the separability of (3.1).

From Theorem 3.1, we see that a codomain transformation, leading to a separable form, can be found for any equation of the type (3.1) provided, of course, that (3.2) or (3.3) can be solved for g. That these ordinary differential equations do indeed have some solutions for a general f, is shown by the following theorem.

Theorem 3.2: If $A = B$ in (3.2) [$\Rightarrow d = e = 0$ in (3.1)], or $m = 1$ in (3.3), then these equations may be integrated to give

$$\ln(\alpha\psi) = \int dg/h(g), \tag{3.11}$$

where $h(g) = \pm[\beta + (2/\gamma)\int f\,dg]^{1/2}$ and α,β,γ are constants. □

Proof: Putting $A = B$ in (3.2), or $m = 1$ in (3.3), reduces both these equations to the common form

$$\gamma\psi(g' + \psi g'') = f(g), \quad \gamma \in \mathbb{R}. \tag{3.12}$$

Multiplying (3.12) by $2g'$ turns the lhs into the derivative of $(\psi g')^2$ and, hence, leads to the first integral

$$(\psi g')^2 = \beta + (2/\gamma)\int f\,dg = [h(g)]^2, \quad \beta \in \mathbb{R}. \tag{3.13}$$

Separating the variables in (3.13) and integrating again gives the desired result. ■

If $A \neq B$ in (3.2), or $m \neq 1$ in (3.3), then the possibility of obtaining analytic (i.e., closed form) solutions for these equations depends, to a large extent, on the form of $f(g)$ and specific cases, in which $f(g) = e^g$ and $f(g) = \sin g$, will be dealt with below and in Sec. 4. However, if we assume that solutions can be found for general $f(g)$, then the following results should be useful.

Theorem 3.3: (i) One set of solutions of (3.1), for arbitrary values of the coefficients a,b,c,d and e, is always

$$\phi(x,y) = g[\exp(a_1 x + b_1 y + k)], \tag{3.14}$$

where k is an arbitrary constant, a_1 and b_1 satisfy (3.10) and g is a solution of (3.2).

(ii) If c is the only nonzero coefficients in (3.1), then one set of solutions is always

$$\phi(x,y) = g\left\{[(m-1)^2(a_m x + k_1)(b_m y + k_2)]^{-1/(m-1)}\right\}, \tag{3.15}$$

where k_1 and k_2 are arbitrary constants, $a_m b_m = C/c$ [Eq. (3.8)] and g is a solution of (3.3) with $m \neq 1$. □

Proof: (i) Assume $X' = a_1 X$ and $Y' = b_1 Y$, and substitute into (3.5). The identity reduces to (3.2) and hence must be satisfied since g is, by definition, a solution of (3.2). Thus, our assumed expressions for X' and Y' are consistent and straightforward integration and substitution, into $\phi = g(\psi)$, leads to (3.14).

(ii) Assume $X' = a_m X^m$ and $Y' = b_m Y^m$ (no summation), and substitute into (3.5). The identity reduces to (3.3) and thus, by an argument similar to that in (i) above, shows the consistency of the assumed expressions for X' and Y'. Solving for X and Y and substituting into $\phi = g(\psi)$, then gives (3.15). ■

Remark: If $c \neq 0$ in case (i) above, then the solutions (3.14) are the only ones obtainable by this method.

We now turn our attention to independent-variable transformations of (3.1) which keep it in the set of constant-coefficient equations. In our discussion of the first-order case, given in Sec. 2 we found that this requirement of keeping within the constant coefficient set does not seriously limit the class of allowable transformations. Thus, although constrained by (2.9), the first-order class still includes a large variety of nonlinear transformations; as shown, for example, in our solution of Eq. (2.10). The situation in the second-order case, however, is much more restrictive. In fact, we have the following theorem.

Theorem 3.4: The only domain transformations of (3.1) which lead to constant coefficient equations, are those of the linear type

$$\xi = A_1 x + B_1 y + C_1, \quad \eta = A_2 x + B_2 y + C_2, \tag{3.16}$$

where $A_1,B_1,C_1,A_2,B_2,C_2 \in \mathbb{R}$ and $A_1 B_2 \neq A_2 B_1$. □

Proof: Let $\xi = \xi(x,y)$ and $\eta = \eta(x,y)$ be twice differentiable maps of the initial domain with continuous second derivatives. Then, in terms of these new coordinates, (3.1) becomes

$$\begin{aligned}&(a\xi_x^2 + b\xi_y^2 + c\xi_x\xi_y)\phi_{\xi\xi} + (a\eta_x^2 + b\eta_y^2 + c\eta_x\eta_y)\phi_{\eta\eta}\\ &+ (2a\xi_x\eta_x + 2b\xi_y\eta_y + c\xi_x\eta_y + c\xi_y\eta_x)\phi_{\xi\eta}\\ &+ (a\xi_{xx} + b\xi_{yy} + c\xi_{xy} + d\xi_x + e\xi_y)\phi_\xi\\ &+ (a\eta_{xx} + b\eta_{yy} + c\eta_{xy} + d\eta_x + e\eta_y)\phi_\eta = f(\phi).\end{aligned} \tag{3.17}$$

Consider the coefficient of $\phi_{\xi\xi}$. If this is to be a constant, say α, then we must have

$$a\xi_x^2 + b\xi_y^2 + c\xi_x\xi_y = \alpha, \quad \alpha \in \mathbb{R}. \tag{3.18}$$

Letting $\xi_x = p$ and $\xi_y = q$, enables us to write (3.18) as

$$ap^2 + bq^2 + cpq - \alpha = F(p,q) = 0. \tag{3.18a}$$

Now this is a nonlinear equation of the first order and can be solved by considering its characteristic system of ordinary differential equations[9]

$$\frac{dx}{F_p} = \frac{dy}{F_q} = \frac{d\xi}{pF_p + qF_q} = \frac{-dp}{F_x + pF_\xi} = \frac{-dq}{F_y + qF_\xi}. \tag{3.19}$$

Using F, as given by (3.18a), reduces these to

$$dx : dy : d\xi : dp : dq = (2ap + cq) : (2bq + cp) : 2\alpha : 0 : 0, \tag{3.20}$$

and hence we have $dp = dq = 0$, giving $p = \xi_x = A_1$, a constant, and $q = \xi_y = B_1$, another constant. If $\alpha \neq 0$, then straightforward integration of these partial derivatives of ξ gives us the solution of (3.18) as $\xi = A_1 x + B_1 y + C_1$, with

$$\alpha = aA_1^2 + bB_1^2 + cA_1B_1. \tag{3.21}$$

However, if $\alpha = 0$, i.e., $\phi_{\xi\xi}$ does not occur in (3.17), then (3.18) has the general solution $\xi = G(A_1 x + B_1 y + C_1)$, where G is an arbitrary function.

We now consider the coefficient of ϕ_ξ in (3.17). This must also be a constant, say δ, and so we have

$$a\xi_{xx} + b\xi_{yy} + c\xi_{xy} + d\xi_x + e\xi_y = \delta, \quad \delta \in \mathbb{R}. \tag{3.22}$$

If we now assume $\alpha = 0$ and substitute the general solution of (3.18) into (3.22), the latter reduces to

$$(dA_1 + eB_1)G' = \delta.$$

Thus, G is the identity function or else $\delta = 0$ [i.e., ϕ_ξ as well as $\phi_{\xi\xi}$ do not occur in (3.17)]. In either case, we have

$$dA_1 + eB_1 = \delta. \tag{3.23}$$

In a similar manner, the constancy of the coefficients, β and ϵ say, of $\phi_{\eta\eta}$ and ϕ_η in (3.17), leads to the general solution $\eta = H(A_2 x + B_2 y + C_2)$, where H is an arbitrary function if $\beta = \epsilon = 0$ and the identity function if either β or ϵ is nonzero. In both cases (H arbitrary or H identity), the relationships between the coefficients are

$$aA_2^2 + bB_2^2 + cA_2B_2 = \beta, \quad \beta \in \mathbb{R}, \tag{3.24}$$

$$dA_2 + eB_2 = \epsilon, \quad \epsilon \in \mathbb{R}. \tag{3.25}$$

Finally, we look at the coefficient of the mixed derivative $\phi_{\xi\eta}$ in (3.17). Since this has to be a constant, γ say, we have

$$2a\xi_x\eta_x + 2b\xi_y\eta_y + c(\xi_x\eta_y + \xi_y\eta_x) = \gamma, \quad \gamma\in\mathbb{R}. \tag{3.26}$$

If we assume $\alpha = \beta = \delta = \epsilon = 0$, so as to preserve the arbitrariness of G and H, then γ cannot be zero. Otherwise, the lhs of (3.17) would vanish and we would have no differential equation. Thus, assuming $\gamma\neq 0$ and substituting for ξ and η in (3.26), gives

$$(2aA_1A_2 + 2bB_1B_2 + C(A_1B_2 + A_2B_1))\, G'H' = \gamma. \tag{3.27}$$

However, the arguments of G and H are independent and so $G'H'$ can only be constant if G' and H' are separately constants. Thus, we must have $G = H =$ the identity function and

$$\gamma = 2aA_1A_2 + 2bB_1B_2 + c(A_1B_2 + A_2B_1). \tag{3.28}$$

To sum up, we have shown that if any one of the coefficients of (3.17) is nonzero, then G and H are restricted to be identity functions. As at least one of these coefficients must be nonzero for (3.17) to exist as a differential equation, this proves the theorem. ■

Corollary: A domain transformation of the linear type applied to (3.1) gives an equation of the same class with coefficients $\alpha,\beta,\gamma,\delta,\epsilon$ given by (3.21), (3.24), (3.28), (3.23) and (3.25) respectively. □

Thus, it is clear that the requirement of keeping within the class of constant-coefficient equations severely restricts the usefulness of domain transformations in the second-order case. The freedom we had with the first-order equations, to change solution functions by changing domain functions in a fairly arbitrary manner while still remaining within the class of constant-coefficient equations, is no longer there, and the most we can possibly do to (3.1), is to reduce the lhs to a form in which only the coefficient of the mixed derivative is nonzero. However, things are not as bad as they seem, since this reduction enables us to use the infinite number of separating codomain transformations defined by (3.3), rather than the finite number, defined by (3.2), which would have applied to (3.1) in its unreduced form. Thus, by using a domain transformation, even though it is restricted to be of linear type, we should, in principle, be able to generate more solutions. The following example will illustrate this.

Example: Consider the hyperbolic variant of Liouville's equation[10]

$$\phi_{xx} - \phi_{yy} = e^{\phi}. \tag{3.29}$$

By Lemma 3.1, the equation is not separable as it stands and, by Theorem 3.2, the separating codomain transformations are

$$\pm \ln(\alpha\psi) = \int dg/[\beta + (2/\gamma)e^g]^{1/2}, \tag{3.30}$$

where $\alpha,\beta,\gamma\in\mathbb{R}$. If, for conveience, we choose $|\beta| = 4$ and $|\gamma| = 1$, then, by Eq. (3.14) of Theorem 3.3, the corresponding real solutions of (3.29) are given by

$$\phi(x,t) = \ln\{2\sec^2[\delta \pm \mu x \pm (\mu^2 - 1)^{1/2}y]\} \quad \text{for } \beta < 0 \text{ and } \gamma > 0, \tag{3.31a}$$

and

$$\phi(x,t) = \ln\{2\,\mathrm{sech}^2[\delta \pm \mu x \pm (\mu^2 + 1)^{1/2}y]\} \quad \text{for } \beta > 0 \text{ and } \gamma < 0, \tag{3.31b}$$

where $\delta,\mu\in\mathbb{R}$.

We now consider linear, domain transformations of (3.29). From Theorems 3.2 and 3.3, we know that as long as either $\phi_{\xi\xi}$ or $\phi_{\eta\eta}$ occur in the transformed equations, then (3.31) are the only solutions obtainable by our method. To eliminate these unwanted derivatives, we go to characteristic coordinates $\xi = x + y$, $\eta = x - y$ and transform (3.29) into

$$4\phi_{\xi\eta} = e^{\phi}. \tag{3.32}$$

This time the second part of Theorem 3.1 applies and the codomain transformations are given by Eq. (3.3), i.e.,

$$C\psi^m(g' + \psi g'') = e^g, \tag{3.33}$$

where $C\in\mathbb{R}$ and m is a nonnegative integer. If $m = 1$, the solutions of (3.33) are once again (3.30) and so we have the same solutions of (3.29) as before. However, if $m\neq 1$, then new solutions are generated. For example, letting $C = 1$ and $W = e^g$ in (3.33) changes it into

$$W'' = \frac{(W')^2}{W} - \frac{W'}{\psi} + \frac{W^2}{\psi^{m+1}}. \tag{3.34}$$

If we now replace ψ by $\theta = \psi^{1-m}/(1-m)^2$, then (3.34) reduces to

$$\frac{d^2W}{d\theta^2} = \frac{1}{W}\left(\frac{dW}{d\theta}\right)^2 - \frac{1}{\theta}\frac{dW}{d\theta} + \frac{W^2}{\theta}, \tag{3.35}$$

which is the defining equation for one of the third Painlevé transcendents.[11] Hence, by using the second part of Theorem 3.3, we obtain solutions of (3.29) expressible in terms of a Painlevé transcendent.

The above example shows how combinations of domain and codomain transformations lead to distinct separable forms of an originally nonseparable equation. However, the process does not have to stop there. By applying the same analysis to those separable forms which are also quasilinear (or linear) and have constant coefficients, one may be able to obtain other *distinct* separable forms and so on, and, in this way, the totality of separable solutions of the original equation may be built up. However, there is a word of caution. It may happen that the domain and codomain transformations are commutative, in the sense that a different order of application leads to the same solutions, on the other hand, certain nonanalytic solutions may be missed by changing the order. An explicit example of this "repeated analysis" is given in Sec. 4, where we apply it to the SGE and its elliptic variant.

4. THE SGE AND ITS ELLIPTIC VARIANT

We now turn our attention to the SGE and its elliptic variant. Using the theorems developed in Sec. 3, we look for combinations of domain and codomain transformations which reduce these equations to separable forms. We shall need some of the results of our earlier study[3] and for convenience, as well as completeness, we list these below.

Thus, consider the SGE and its elliptic variant which may be written together as

$$\phi_{xx} + \epsilon\phi_{yy} = \sin\phi, \quad \epsilon = \pm 1. \tag{4.1}$$

By Lemma 3.1, these equations are not separable as they stand and, using Theorems 3.1 and 3.2, the codomain transformations leading to separable forms are given by

$$\ln(\alpha\psi) = \pm \int dg/[\beta - (2/\gamma)\cos g]^{1/2}, \tag{4.2}$$

where $\psi = g^{-1}(\phi)$ is the new dependent variable and $\alpha,\beta,\gamma \in \mathbb{R} - \{0\}$. The solution to (4.2) is

$$\phi = g(\psi) = 2\cos^{-1}\operatorname{sn}\left(\frac{\ln(\alpha\psi)}{k\gamma^{1/2}}, k\right), \tag{4.3}$$

where sn is a Jacobian elliptic function[12] with modulus $0 < k \leqslant 1$. Applying this transformation to (4.1) gives us the separable forms

$$-\gamma^{1/2}\psi(\psi_{xx} + \epsilon\psi_{yy}) + [\gamma^{1/2} + k\operatorname{sn}(u,k)\operatorname{sn}(u+K,k)] \times(\psi_x^2 + \epsilon\psi_y^2) = k\gamma\psi^2 \operatorname{sn}(u,k)\operatorname{sn}(u+K,k), \tag{4.4}$$

where $u = \ln(\alpha\psi)/k\gamma^{1/2}$ and K is the complete elliptic integral of the first kind.

The nonelliptic (i.e., $k = 1$) limit of (4.3) is

$$\psi = (1/\alpha)\tan^{n}(\phi/4), \tag{4.5}$$

where $-n = \gamma^{1/2}$, and for the usual case, $\alpha = n = 1$, the corresponding separable form of (4.1) is

$$(1+\psi^2)(\psi_{xx} + \epsilon\psi_{yy}) - 2\psi(\psi_x^2 + \epsilon\psi_y^2) = \psi(1-\psi^2). \tag{4.6}$$

Equations (4.4) and (4.6) were dealt with, in some detail, in Ref. 3.

We now consider the domain transformations $\xi = A_1 x + B_1 y + C_1$ and $\eta = A_2 x + B_2 y + C_2$, $A_1B_2 \neq A_2B_1$, acting on (4.1). The transformed equation is

$$(A_1^2 + \epsilon B_1^2)\phi_{\xi\xi} + 2(A_1A_2 + \epsilon B_1B_2)\phi_{\xi\eta} + (A_2^2 + \epsilon B_2^2)\phi_{\eta\eta} = \sin\phi, \tag{4.7}$$

and A_1, B_1, A_2 and B_2 may now be chosen to give four different combinations of coefficients. These are: (i) only the coefficient of $\phi_{\xi\eta}$ is zero, which is the case detailed above; (ii) only the coefficient of either $\phi_{\xi\xi}$ or $\phi_{\eta\eta}$ is zero; (iii) none of the coefficients are zero; and (iv) only the coefficient of $\phi_{\xi\eta}$ is nonzero. As mentioned above, case (i) has been dealt with in Ref. 3 and leads to separable forms [Eqs. (4.4) and (4.6) above] which have a fairly wide variety of solutions.[13] Cases (ii) and (iii) may be considered together and, from Theorems 3.1 and 3.2, the codomain transformations leading to separable forms are again given by (4.3), just as in case (i). However, the only separable solutions, $\psi(\xi,\eta) = E(\xi)H(\eta)$, obtainable from these forms are those for which $E' = a_1E$ and $H' = b_1H$. We have thus lost the variety of solutions we had in the first case, indicating, as was implicit in the proof of Theorem 3.1, that combining the mixed derivative with the other derivatives would lead to a restriction on the set of separable solutions. Case (iv), on the other hand, is different. Here we have the mixed derivative on its own and, as indicated by Theorem 3.1, we would expect to get at least as large a set of solutions as we had for case (i). We therefore consider case (iv) in some detail.

Thus, let $\xi = x + y$, $\eta = x - y$ for $\epsilon = -1$ and $\xi = x + iy$, $\eta = x - iy$ for $\epsilon = +1$. Then, (4.7) becomes

$$4\phi_{\xi\eta} = \sin\phi \tag{4.8}$$

and, from the second part of Theorem 3.1, the codomain transformations leading to separable forms of (4.8) are given by the solutions of

$$C\psi^m(g' + \psi g'') = \operatorname{sing}, \tag{4.9}$$

where $C \in \mathbb{R}$ and m is an arbitrary nonnegative integer. If we choose $m = 1$, then (4.9) reduces to (4.2) and we obtain the same restricted class of solutions that we had in cases (ii) and (iii) above. However, as in the case of Liouville's equation [Eq. (3.29) of Sec. 3], if $m \neq 1$ then new solutions are generated. Thus, we have Theorem 4.1.

Theorem 4.1: The solution of (4.9) for $m \neq 1$ may be given in terms of a fifth Painlevé transcendent, say W, and these lead to solutions of (4.8) of the form

$$\cos(\phi/2) = \frac{1 + W\{[(a\xi + K_1)(b\eta + K_2)]^{1/2}\}}{1 - W\{[(a\xi + K_1)(b\eta + K_2)]^{1/2}\}}, \tag{4.10}$$

which is independent of m and where $a,b,K_1,K_2 \in \mathbb{R}$. □

Proof: Since the solution of (4.9), when $m = 1$, is the inverse trigonometric function of an elliptic function, we first make the substitution $g \to h = \cos(g/2)$ in (4.9). After some manipulation, this leads to the equation

$$h'' = -\frac{h}{1-h^2}(h')^2 - \frac{1}{\psi}h' - \frac{h(1-h^2)}{C\psi^{m+1}}. \tag{4.11}$$

Now (4.11) is an equation of algebraic class and is of the form necessary for the absence of movable critical points in its solution.[10,11] To reduce it to standard form, we proceed in two stages. Firstly, we note that the coefficient of $(h')^2$ has poles at $h = \pm 1$ and so we make the linear, fractional transformation of the dependent variable $h \to W = (h-1)/(h+1)$. This gives us

$$W'' = \left(\frac{1}{2W} + \frac{1}{W-1}\right)(W')^2 - \frac{1}{\psi}W' + \frac{1}{C\psi^{m+1}}\frac{2W(1+W)}{1-W}, \tag{4.12}$$

where the coefficient of $(W')^2$ now has one of the eight distinct forms (listed in Ref. 11) necessary for the solutions of the equation to have fixed critical points. We next make the independent-variable substitution $\psi \to \theta = \alpha\psi^{(1-m)/2}$, where $\alpha^2 = 4/C(1-m)^2$, to obtain

$$\frac{d^2W}{d\theta^2} = \left(\frac{1}{2W} + \frac{1}{W-1}\right)\left(\frac{dW}{d\theta}\right)^2 - \frac{1}{\theta}\frac{dW}{d\theta} + 2W\left(\frac{1+W}{1-W}\right), \tag{4.13}$$

which is the canonical form of the defining equation for one of the fifth Painlevé transcendents.[14] Thus, the solution to (4.9), with $m \neq 1$, is

$$g = 2\cos^{-1}\left(\frac{1 + W(\alpha\psi^{(1-m)/2})}{1 - W(\alpha\psi^{(1-m)/2})}\right), \tag{4.14}$$

and this transformation, acting on (4.8), reduces it to a separable form which has the same separable solutions as the equation

$$4\psi_{\xi\eta} = \psi^m \quad \text{for} \quad m \neq 1. \tag{4.15}$$

Hence, from the second part of Theorem 3.3 [i.e., Eq. (3.15)], one set of solutions of (4.8) is

$$\phi = 2\cos^{-1}\left(\frac{1 + W[C^{-1/2}(a_m\xi + \delta_1)^{1/2}(b_m\eta + \delta_2)^{1/2}]}{1 - W[C^{-1/2}(a_m\xi + \delta_1)^{1/2}(b_m\eta + \delta_2)^{1/2}]}\right), \tag{4.16}$$

which is independent of $m \neq 1$ and where a_m, b_m are expansion coefficients, with $4a_m b_m = C$, and $\delta_1, \delta_1 \in \mathbb{R}$. Relabeling the constants then gives us Eq. (4.10) and hence completes the proof of the theorem. ■

The above analysis shows how combinations of domain and codomain transformations enable us to reduce the nonseparable equations (4.1), (4.7), and (4.8) to the separable forms (4.4) and (4.6), and to a form which has the same separable solutions as (4.15). The question now arises as to whether there exist transformations which, when applied to these reduced equations, give other *distinct* separable forms. Unfortunately, none of the reduced equations has constant coefficients and thus our theorems do not apply. On the other hand, the separable form of (4.8) has the same separable solutions as (4.15) and the latter is a quasilinear equation with constant coefficients. Thus, it may be of some interest to apply the analysis to this equation.

Hence, consider a codomain transformation $\psi \to \theta = h^{-1}(\psi)$ applied to (4.15). From Theorem 3.1 we see that in order to maintain separability h must satisfy the equation

$$C\theta^n(h' + \theta h'') = h^m, \tag{4.17}$$

where $m \neq 1$ and n is an arbitrary nonnegative integer. Now if $n \neq 1$, then the solutions to (4.17) are merely of the form $\psi = h(\theta) = \theta^s$, with $s = (n-1)/(m-1)$,and hence we do not obtain anything essentially different from the solutions of (4.15). However if $n = 1$, then we do obtain something new. In this case, we can use Theorem 3.2 to write the solutions of (4.17) as

$$\ln(A\theta) = \int dh/p(h), \tag{4.18}$$

where $(m+1)Cp^2(h) = (m+1)BC + 2h^{m+1}$ and A and B are constants of integration. Under the substitutions $(m+1)BCY^{m+1} = 2h^{m+1}$ and $\alpha = (1/\sqrt{B})[(m+1)BC/2]^{1/(m+1)}$, (4.18) becomes

$$\ln(A\theta) = \alpha \int dY/(1 + Y^{m+1})^{1/2}, \tag{4.19}$$

where the integral on the rhs is, in general a hyperelliptic integral of the first kind.[15] Thus, the codomain maps which maintain the separability of (4.15) are, in general, hyperelliptic functions. We use the term "in general", because for $m = 0$ or 1 the integral is elementary and for $m = 2, 3, 5$ or 7 is reducible to one of the elliptic type.

Example $m = 5$: In this case, under the substitution

$$\cos Z = [1 + (\sqrt{3} - 1)Y^2]/[1 + (\sqrt{3} + 1)Y^2],$$

(4.19) can be written as

$$\beta \ln(A\theta) = \int \frac{dZ}{(1 - k^2 \sin^2 Z)^{1/2}}, \tag{4.20}$$

where $k^2 = (2 + \sqrt{3})/4$ and $\beta \in \mathbb{R}$. Thus, we have $\mathrm{cn}(\beta \ln(A\theta), k) = \cos Z$ and hence, from the first part of Theorem 3.3, one set of solutions of (4.15), with $m = 5$, is given by

$$\psi = \gamma[1 - \mathrm{cn}(a_1\xi + b_1\eta + K, k)]^{1/2} \times [(1 + \sqrt{3})\,\mathrm{cn}(a_1\xi + b_1\eta + K, k) + (\sqrt{3} - 1)]^{-1/2}, \tag{4.21}$$

where $k^2 = (2 + \sqrt{3})/4$ and $\gamma, a_1, b_1, K \in \mathbb{R}$. We emphasize, however, that (4.21) does *not* lead to a corresponding solution of the SGE. This is because the only solutions of (4.15) which do lead to solutions of the SGE [via the Painlevé transcendent transformation (4.14)] are those in which ψ is directly separable, i.e., $\psi(\xi, \eta) = E(\xi)H(\eta)$. The solution (4.21) is not of this form. On the other hand, the above analysis of (4.15) does show the usefulness of our method even in those cases in which the given equation is already in separable form.

5. CONCLUDING REMARKS

In this paper, which is an extension of our earlier work,[3] we have studied the separability and the existence of separable solutions of constant coefficient, quasilinear partial differential equations of the first and second orders, with particular reference to the SGE. By using combinations of domain and codomain transformation we have reduced the original equations to separable forms and have developed criteria for determining the separating codomain transformations in each case. In the particular case of the SGE, we found that these separating transformations are expressed in terms of a Jacobian elliptic function if the equation is in laboratory coordinates and in terms of a fifth Painlevé transcendent if it is in characteristic coordinates. An interesting point here is that the hyperbolic Liouville equation, which has an exponential rather than a sine function on the rhs, is separated by a third Painlevé transcendent when it is in characteristic coordinates. Another point to bear in mind, in connection with the elliptic transformation, is that the first and second Painlevé transcendents are asymptotically related to the elliptic functions in a manner somewhat analogous to the relationship between Bessel functions and circular functions.

We have also applied our analysis to hyperbolic equations in which the rhs is a power of the dependent variable. In this case, the equations are already in separable form when expressed in characteristic coordinates. However, by using the "right" codomain transformations, we have shown how one can get to other distinct separable forms and, in this way, obtain solutions which cannot be obtained from a *direct* separation of the original equations.

Our method can be extended to deal with quasilinear equations of third and higher orders, provided they have constant coefficients, and to equations with variable coefficients where these coefficients are low-order polynomials in the dependent variable; such as occurs, for example, in the Korteweg–de Vries equation. However, although the application of the method in these cases is fairly straightforward, the difficulty of solving the resulting ordinary, differential equations for the separating codomain transformations in-

creases rapidly with the order and complexity of the equations discussed.

The major limitation of our method, as applied to quasilinear equations, is our inability to deal with the most general quasilinear equations. That is, equations with variable coefficients where these coefficients are functions of the domain coordinates as well as of the dependent variable; for example, we cannot work in polar coordinates. In the case of first-order equations this is not a problem, since the variety of allowed domain transformations, even in the case of constant coefficients, is such that one can always absorb the coefficients by a redefinition of the variables. However, in the case of second- and higher-order equations this is no longer possible and we do have a problem. The only way out that we see at the moment is to use the differentiable manifold approach. That is, to work in "patches" of the domain, over which the coefficients are sensibly constant, and then to join these patches up using some form of analytic continuation. We shall be reporting on our efforts in this direction in due course.

[1]G.L. Lamb, Rev. Mod. Phys. **43**, 99 (1971).
[2]J. Zagrodzinski, Phys. Lett. A **57**, 213 (1976).
[3]A. Osborne and A.E.G. Stuart, J. Math. Phys. **19**, 1573 (1978).
[4]M.J. Ablowitz and H. Segur, Phys. Rev. Lett. **38**, 103 (1977).
[5]M.J. Ablowitz, A. Ramani, and H. Segur, Lett. Nuovo Cimento **23**, 333 (1978).
[6]P. Kaliappan and M. Lakshmanan, J. Phys. A **12**, L249 (1979).
[7]B.M. McCoy, C.A. Tracy, and T.T. Wu, J. Math. Phys. **18**, 1058 (1977).
[8]H. Airault, Stud. Appl. Math. **61**, 31 (1979).
[9]R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1962), Vol. 2, Chap. 2.
[10]H.T. Davis, *Introduction to Nonlinear Differential and Integral Equations* (Dover, New York, 1962), pp. 20 and 377.
[11]E.L. Ince, *Ordinary Differential Equations* (Dover, New York, 1956), p. 345.
[12]H. Hancock, *Theory of Elliptic Functions* (Dover, New York, 1958).
[13]A.C. Bryan, C.R. Haines and A.E.G. Stuart, Report IDE/78, Dept. of Mathematics, The City University, London, England (1978).
[14]There are six Painlevé transcendents, but only the first three were discovered by Painlevé. The last three were subsequently added by Gambier. References to the original works of both Painlevé and Gambier can be found in the books by Davis and Ince cited above.
[15]P.F. Byrd and M.D. Friedman, *Handbook of Elliptic Integrals for Engineers and Physicist* (Springer, Berlin, 1954).

Construction of the phase of the scattering amplitude by iteration of bounds from unitarity integral

S. K. Chan
Applied Mathematics Department, University of Western Ontario, London, Ontario, Canada

I. A. Sakmar[a)]
Applied Mathematics Department, University of Western Ontario, London, Ontario, Canada and Université de Nice, Laboratoire de Physique Théorique, Parc Valrose 06034 NICE CEDEX, France

(Received 26 October 1979; accepted for publication 24 December 1979)

Given the modulus of the amplitude a convergent iteration scheme is given to obtain the phase from the unitarity equation. The method is based on improving the upper and lower bounds on the phase. Uniqueness of the solution is investigated and numerical results for a specific case are given.

I. INTRODUCTION

During the last decade the unitarity equation has been investigated as an integral equation for the phase, when the modulus function is known.[1-11] The two questions one would like to answer are the necessary and sufficient conditions for the existence and uniqueness of the solutions.

A sufficiency condition for the existence was given ten years ago and as far as we know there has not been much progress on this aspect of the problem since. That is, we still do not know the necessary conditions for the existence of the solution.

A sufficiency condition for the uniqueness of the solution was first given in Ref. 2. This condition was improved in Ref. 3. It was also conjectured that a better sufficiency condition for the uniqueness of the solution is $\sin\mu < 1$ where $\sin\mu$ is the supremum of the unitarity integral with cosine factors majorized by 1. Unfortunately it has not been possible to prove this conjecture up to now. However there has been some improvements[4,7,9,10,11] on the uniqueness aspect of the solution. Angle independent[4,7,9,11] and angle dependent[10] conditions were given for the regions not covered by the conditions of the Refs. 2 and 3.

In this paper we use both the angle dependent upper and lower bounds to set up an iteration scheme which continuously improves both bounds. We show that the range between them is continuously narrowed so that there is guaranteed convergence to the solution from above and below if the initial bounds satisfy a certain condition. This condition insures thus the uniqueness of the solution.

In Sec. II we give the unitarity integral and summarize the main results known so far. We then give the lower bounds on the phase, set up the iteration scheme and discuss its implications.

In Sec. III we apply the iteration to the $\pi^+ p$ scattering which is considered for the purposes of the problem at hand as a spinless scalar particle scattering described by a single amplitude. The convergence feature as well as the comparison of the existing uniqueness conditions are discussed with the help of this numerical example.

Finally in Sec. IV we summarize and discuss our results.

[a)]Research partially supported by the National Research Council of Canada.

II. THE UNITARITY INTEGRAL

The unitarity equation in the case of spinless (scalar) particle scattering can be considered an integral equation for the phase when the modulus is known at all angles

$$\sin\alpha(z) = \frac{q}{2\pi}\int_{-1}^{+1}\int \frac{|f(x)||f(y)|}{|f(z)|} \times \cos[\alpha(x)-\alpha(y)]\,\frac{\theta(K)}{\sqrt{K}}\,dx\,dy\,. \tag{1}$$

Here $|f(x)|$, $|f(y)|$, and $|f(z)|$ are the moduli of the scattering amplitude, the square of which gives the differential cross section. q is the c.m. wave number, $\theta(K)$ the step function, and $K = 1 - x^2 - y^2 - z^2 + 2xyz$. The integral is taken over the ellipse $K > 0$. x,y and z are the cosines of the scattering angles. $\alpha(z)$ is the phase of the scattering amplitude which we want to find. By dividing the amplitude by q one can make it dimensionless. Thus in the rest of the paper we shall redefine f and omit q. With this Eq. (1) becomes

$$\sin\alpha(z) = \frac{1}{2\pi}\int\int \frac{|f(x)||f(y)|}{|f(z)|} \times \cos[\alpha(x)-\alpha(y)]\,\frac{\theta(K)}{\sqrt{K}}\,dx\,dy\,. \tag{2}$$

The phase is between 0 and $\pi/2$ (see Ref. 3). It is obvious that an angle-dependent upper bound of $\sin\alpha(z)$ is obtained by majorizing the cosine by 1:

$$\sin\mu(z) = \frac{1}{2\pi}\int\int \frac{|f(x)||f(y)|}{|f(z)|}\,\frac{\theta(K)}{\sqrt{K}}\,dx\,dy\,. \tag{3}$$

We shall call the supremum and infinum of (3) in the interval $-1 < z < 1$

$$\sin\mu = \sup \sin\mu(z), \quad -1 < z < 1, \tag{4}$$

$$\sin\nu = \inf \sin\mu(z), \quad -1 < z < 1. \tag{5}$$

A lower bound of $\sin\alpha(z)$ will be shown by $\gamma(z)$ and will be derived later [Eq. (14)]. We now give the known sufficiency conditions for the existence and uniqueness of the solutions of the Eq. (2).

Existence-sufficiency condition:

$$\sin\mu < 1 \quad \text{(Refs. 2,3)}. \tag{6}$$

Uniqueness-sufficiency conditions:

 0022-2488/80/040735-05$1.00

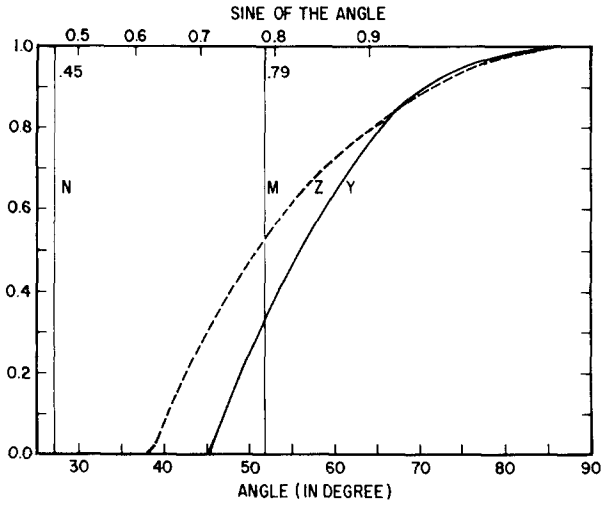


FIG. 1. Comparison of conditions (7)–(10). The curves Y and Z are the left-hand sides of conditions (10) and (9), respectively. The lines M and N are the left-hand sides of conditions (8) and (7), respectively.

$$\sin\mu < \frac{1}{\sqrt{5}} \ (= 0.4472) \quad \text{(Ref. 2)}, \tag{7}$$

$$\sin\mu < 0.79 \quad \text{(Ref. 3)}, \tag{8}$$

$$\sin\nu > \sin\mu - \frac{\cos\mu}{\sqrt{4\tan^2\mu - 1}} \quad \text{(Ref. 4)}, \tag{9}$$

$$\sin\nu > \frac{\sin\mu - \cos\mu}{1 - \sin\mu\cos\mu} \quad \text{(Ref. 9)}, \tag{10}$$

$$\sup \frac{q}{2\pi} \int\int [\tan\mu(y) - \tan\gamma(x)]\, 2|f(y)| \times \frac{\theta(K)}{\sqrt{K}}\, dx\, dy < 1 \quad \text{(Ref. 10)}. \tag{11}$$

Equations (7)–(10) are shown in Fig. 1. Up to $\mu = 68.76°$ or $\sin\mu = 0.9320$, Eq. (10) is better than Eq. (9). Above this value up to $\sin\mu = 1$, Eqs. (9) and (10) are approximately equal with (9) slightly better than (10). At 90° they coincide.

Equation (1) was obtained with the help of the angle-dependent lower bound given by one of us in a recent paper.[9] In a similar way we can find an angle-dependent upper bound. We repeat the calculation of the upper bound (lower bound) using the previous lower bound (upper bound). The numerical solution of the unitarity equation will be obtained as the lower and upper bounds approach each other. If there is no strip between the upper and lower bounds the first time, then the solution is unique for $\sin\mu < 1$ and even for $\sin\mu > 1$ if in this second case the additional assumption is made that $\alpha < 90°$. If the first time

$$\inf\mu_1(z) < \sup\gamma_1(z), \tag{12}$$

then, because $\inf\mu(z)$ is improved downwards and $\sup\gamma(z)$ is improved upwards, it will be true that

$$\inf\mu_n(z) < \sup\gamma_n(z), \quad \text{for all } n.$$

From this it follows that there will always be a region where there is no overlap between the ordinate differences $\mu(x) - \gamma(x)$ and $\mu(y) - \gamma(y)$. Hence it will be possible to improve the bounds until they merge. Thus the solution must be unique. What we said here will be clearer when we discuss below the actual iteration process.

The process is as follows:

(A) In the first iteration the angle dependent upper bound is

$$\sin\mu_1(z) = \frac{1}{2\pi} \int\int \frac{|f(x)||f(y)|}{|f(z)|} \frac{\theta(K)}{\sqrt{K}}\, dx\, dy > \sin\alpha(z), \tag{13}$$

and $\mu_1(z) > \alpha(z)$.

This upper bound was obtained by majorizing the cosine factors on the right-hand side of Eq. (2) $[\cos[\alpha(x) - \alpha(y)] = 1]$ which is equivalent to minimizing its argument $\alpha(x) - \alpha(y) = 0$. (Note that $0 < \alpha < \pi/2$).

(B) We find the angle-dependent lower bound[9] from $\mu_1(z)$. We define

$$\sin\gamma_{1,1}(z) = \frac{1}{2\pi} \int\int \frac{|f(x)||f(y)|}{|f(z)|} \times \cos[\max(A,B)] \frac{\theta(K)}{\sqrt{K}}\, dx\, dy, \tag{14}$$

where $A = \mu_1(x) - 0$ and $B = \mu_1(y) - 0$. $\sin\gamma_{1,1}(z)$ is obtained by maximizing the angle (i.e., minimizing the cosine factor) on the right-hand side of Eq. (2). Thus

$$\sin\gamma_{1,1}(z) < \sin\alpha(z),$$

or

$$\gamma_{1,1}(z) < \alpha(z),$$

and $\gamma_{1,1}(z)$ is the angle-dependent lower bound. Next we improve this lower bound using $\mu_1(z)$ and $\gamma_{1,1}(z)$

$$\sin\gamma_{1,2} = \frac{1}{2\pi} \int\int \frac{|f(x)||f(y)|}{|f(z)|} \times \cos[\max(A,B)] \frac{\theta(K)}{\sqrt{K}}\, dx\, dy.$$

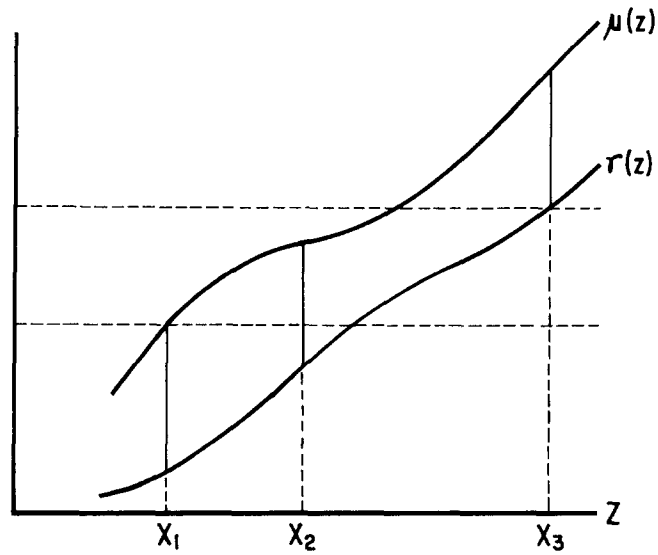


FIG. 2. Overlapping and nonoverlapping cases. There is overlapping between $(\mu(x_1),\gamma(x_1))$ and $(\mu(x_2),\gamma(x_2))$, but no overlapping between $(\mu(x_1),\gamma(x_1))$ and $(\mu(x_3),\gamma(x_3))$.

TABLE I. $\pi^+ p$ scattering $p_{CM} = 0.168$ GeV/c, $p_{lab} = 0.211$ GeV/c.

z	$\sin\mu_1(z)$	$\sin\gamma_1(z)$	$\sin\mu_2(z)$	$\sin\gamma_2(z)$	$\sin\gamma_4(z)$	$\sin\mu_5(z)$	LHS of Eq. (11)
1.00	0.60598	0.60598	0.60598	0.60598	0.60598	0.60598	0
0.90	0.62784	0.62182	0.62490	0.62417	0.62433	0.62434	0.22294
0.80	0.66898	0.65816	0.66273	0.66160	0.66185	0.66185	0.30561
0.70	0.71182	0.69622	0.70212	0.70063	0.70096	0.70096	0.36323
0.60	0.75430	0.73397	0.74112	0.73928	0.73968	0.73969	0.40766
0.50	0.79323	0.76832	0.77660	0.77445	0.77491	0.77492	0.44185
0.40	0.82436	0.79530	0.80455	0.80213	0.80265	0.80265	0.46925
0.30	0.84322	0.81072	0.82071	0.81808	0.81864	0.81865	0.49141
0.20	0.84639	0.81145	0.82187	0.81911	0.81970	0.81970	0.50781
0.10	0.83296	0.79676	0.80728	0.80447	0.80509	0.80508	0.51967
0	0.80493	0.76863	0.77894	0.77618	0.77677	0.77677	0.52846
−0.10	0.76639	0.73098	0.74083	0.73818	0.73874	0.73875	0.53292
−0.20	0.72197	0.68818	0.69741	0.69491	0.69544	0.69545	0.53476
−0.30	0.67566	0.64402	0.65252	0.65020	0.65069	0.65070	0.53351
−0.40	0.63028	0.60107	0.60877	0.60666	0.60711	0.60711	0.52854
−0.50	0.58752	0.56088	0.56778	0.56587	0.56627	0.56628	0.52117
−0.60	0.54820	0.52417	0.53027	0.52857	0.52893	0.52893	0.51180
−0.70	0.51258	0.49111	0.49644	0.49494	0.49525	0.49526	0.49992
−0.80	0.48059	0.46160	0.46623	0.46490	0.46513	0.46518	0.49260
−0.90	0.45199	0.43526	0.43938	0.43817	0.43843	0.43843	0.50573
−1.00	0.43836	0.42328	0.42715	0.42600	0.42624	0.42624	0.55147
$\sin\mu$	0.846386		0.821868			0.819705	
$\sin\gamma$	0.438365		0.427149			0.426244	
RHS of (9)	0.669868		0.611386			0.606122	
RHS of (10)	0.571365		0.474221			0.465459	
Cond. (6)	Met		Met			Met	
Cond. (8)	Violated		Violated			Violated	
Cond. (9)	Violated		Violated			Violated	
Cond. (10)	Violated		Violated			Violated	Cond. (11) met.

Here $A = \mu_1(x) - \gamma_{1,1}(y)$ and $B = \mu_1(y) - \gamma_{1,1}(x)$. Again

$$\sin\alpha(z) > \sin\gamma_{1,2} > \sin\gamma_{1,1}(z),$$

or

$$\alpha(z) > \gamma_{1,2}(z) > \gamma_{1,1}(z).$$

We repeat this process, each time replacing $\gamma_{1,n-2}$ by $\gamma_{1,n-1}$, to find $\gamma_{1,n}$ until the limit is reached. We call

$$\sin\gamma_1(z) = \lim_{n\to\infty} \sin\gamma_{1,n}(z).$$

We have now

$$\alpha(z) > \sin\gamma_1(z) > \cdots > \sin\gamma_{1,1}(z).$$

(C) In a similar way we find the angle-dependent upper bound. However in this case we majorize $\cos[\alpha(x) - \alpha(y)]$ by using $\gamma_1(z)$ and $\mu_1(z)$.

$$\sin\mu_{2,1}(z) = \frac{1}{2\pi}\int\int \frac{|f(x)||f(y)|}{|f(z)|} \times \cos[\max(A,B,0)]\,\frac{\theta(K)}{\sqrt{K}}\,dx\,dy.$$

Here $A = \gamma_1(x) - \mu_1(y)$ and $B = \gamma_1(y) - \mu_1(x)$.

Note that both A and B can not be positive simultaneously.

Since $\gamma_1(x) < \alpha(x) < \mu_1(x)$, $\gamma_1(y) < \alpha(y) < \mu_1(y)$, if both $\gamma_1(x) - \mu_1(y)$ and $\gamma_1(y) - \mu_1(x)$ are negative there is an overlap between the regions $(\mu_1(x), \gamma_1(x))$ and $(\mu_1(y), \gamma_1(y))$ (see Fig. 2). Otherwise there is no overlap. In the case there is no overlap the minimum of $\alpha(x) - \alpha(y)$ is the positive one of the pair $\gamma_1(x) - \mu_1(y)$ and $\gamma_1(y) - \mu_1(x)$ and it is not zero. Obviously

$$\sin\mu_{2,1}(z) < \sin\mu_1(z),$$

or

$$\mu_{2,1}(z) < \mu_1(z).$$

So the angle-dependent upper bound has been improved. We also repeat this process with $\mu_{2,n-1}$ replacing $\mu_{2,n-2}$, to find an improved upper bound until the limit has been reached. We call

$$\mu_2(z) = \lim_{n\to\infty} \mu_{2,n}(z),$$

we thus have

$$\mu_1(z) > \mu_{2,1}(z) > \mu_{2,2}(z) \cdots > \mu_2(z) > \alpha(z).$$

In the case there is overlap everywhere, that is if

$$\inf\mu(x) > \sup\gamma(x), \quad -1 < x < 1,$$

the possible minimum of $\alpha(x) - \alpha(y)$ is zero. There is no better majorization of the cosine factor under the integral than one. The angle-dependent upper bound cannot be improved further. $\mu_1(z)$ and $\gamma_1(z)$ are the best angle-dependent upper and lower bounds and the solution of the integral equation cannot be obtained by improving them.

We note that if there is overlap at some pair x_i and x_j and no overlap at other points (see Fig. 2) $\mu_1(z)$ can still be improved and the method is applicable.

(D) We repeat steps (B) and (C). In step (B) the new

TABLE II. π^+p scattering $p_{CM} = 0.176$ GeV/c, $p_{lab} = 0.221$ GeV/c.

z	$\sin\mu_1(z)$	$\sin\gamma_1(z)$	$\sin\mu_2(z)$	$\sin\gamma_2(z)$	$\sin\mu_{25}(z)$	$\sin\gamma_{25}(z)$	LHS of Eq. (11)
1.00	0.69233	0.69233	0.69233	0.69233	0.69233	0.69233	0
0.90	0.71759	0.66719	0.71729	0.68659	0.70929	0.70927	0.73736
0.80	0.76576	0.69484	0.76393	0.71888	0.74915	0.74915	0.97520
0.70	0.81722	0.72692	0.81319	0.75514	0.79225	0.79225	1.13657
0.60	0.87019	0.76150	0.86362	0.79353	0.83690	0.83689	1.24891
0.50	0.92151	0.79493	0.91224	0.83072	0.88005	0.88004	1.33900
0.40	0.96652	0.82383	0.95458	0.86294	0.91750	0.91750	1.40795
0.30	0.99947	0.84375	0.98511	0.88526	0.94404	0.94403	1.45930
0.20	1.01491	0.85022	0.99859	0.89328	0.95474	0.95473	1.49566
0.10	1.00974	0.84117	0.99210	0.88461	0.94696	0.94695	1.51999
0.00	0.98457	0.81717	0.96634	0.85972	0.92141	0.92140	1.53608
−0.10	0.94349	0.78170	0.92536	0.82276	0.88196	0.88196	1.53777
−0.20	0.89233	0.73948	0.87488	0.77821	0.83400	0.83399	1.53162
−0.30	0.83673	0.69516	0.82039	0.73104	0.78271	0.78270	1.51437
−0.40	0.78098	0.65185	0.76600	0.68478	0.73186	0.73185	1.48312
−0.50	0.72779	0.61154	0.71432	0.64141	0.68389	0.68389	1.44341
−0.60	0.67860	0.57837	0.66670	0.60216	0.64003	0.64002	1.38901
−0.70	0.63397	0.54351	0.62366	0.56729	0.60064	0.60064	1.31785
−0.80	0.59392	0.51588	0.58533	0.53676	0.56576	0.56576	1.23704
−0.90	0.55821	0.49164	0.55156	0.50996	0.53502	0.53502	1.17037
−1.00	0.54115	0.47999	0.53685	0.49765	0.52216	0.52215	1.32860
$\sin\mu$	1.01491		0.99859		0.95473		
$\sin\nu$	0.54115		0.53685		0.52215		
RHS of (9)			0.99718		0.90781		
RHS of (10)			0.99843		0.91798		
Cond. (6)	Violated		Met		Met		
Cond. (8)	Violated		Violated		Violated		
Cond. (9)	Violated		Violated		Violated		Cond. (11)
Cond. (10)	Violated		Violated		Violated		Violated

$\mu_n(x)$ will replace $\mu_{n-1}(x)$ and in step (C) the new $\gamma_n(x)$ will replace $\gamma_{n-1}(x)$.

Using this iteration method the angle-dependent upper and lower bounds are continuously improved. Finally $\mu_n(z)$ and $\gamma_n(z)$ will converge to the same value, which is the solution of the integral equation.

III. ITERATION OF π^+p SCATTERING

We now apply this method to a specific case. For this purpose π^+p scattering is considered a scalar particle scattering neglecting the spin. The scalar amplitude is taken as the square root of the elastic differential cross section and we have reconstructed the elastic differential cross section from Lovelace phase shifts. We have applied the iteration process at two energies. The results for $p_{CM} = 0.168$GeV/c are presented in the Table I. In this case

$$\sup \sin\mu(z) < 1, \quad -1 < z < 1,$$

so that the sufficiency condition for the existence of the solution is met. After the 4th iteration the difference between $\mu_5(z)$ and $\gamma_4(z)$ is less than 10^{-5}. Thus the solution has been obtained with this accuracy.

We notice that all uniqueness sufficiency conditions (Eqs. 7–10) are violated except the last one (Eq. (11).

In Table II. we show the numerical results for $p_{CM} = 0.176$GeV/c. In this case

$$\sup \sin\mu(z), \quad -1 < z < 1,$$

is slightly above one. We chose this energy on purpose to show the convergence in a case when $\sin\mu > 1$. However one has to make here the additional assumption that $0 \leqslant \alpha \leqslant \pi/2$. Thus in the majorizations under the integral we take $\alpha = \pi/2$ whenever $\sin\mu(z) > 1$. The convergence of both $\mu_n(z)$ and $\gamma_n(z)$ is much slower than in the first case.

But it is interesting to note that $\sup \sin\mu(z)$ has been brought down to below one and $\mu_n(z)$ and $\gamma_n(z)$ converge to the same value (with five decimal accuracy) after 25 iterations. All the sufficiency conditions for existence and uniqueness are violated, which once more demonstrates the need for necessary conditions. It is also worth noticing that the solution is unique once the assumption is made that $0 \leqslant \alpha \leqslant \pi/2$. We have checked other examples and it is found that the larger the angle region is for which $\sin\mu(z) > 1$, the slower is the convergence.

If $\sin\mu(z) > 1$ for all z (as in the case near N_{33} resonance energy) the method can not be applied. The other case when the iteration cannot be further pursued was already mentioned, namely when

$$\inf \sin\mu(z) > \sup\gamma(z), \quad -1 < z < 1,$$

the first time in the iteration.

IV. CONCLUSION

Subject to the condition

$$\inf \sin\mu(z) < \sup\gamma(z), \quad -1 < z < 1,$$

the iteration method based on improving the upper and lower bounds of the phase of the amplitude guarantees both the

existence and uniqueness of the solution of the unitarity integral equation.

If one restricts onself to the first quadrant for the phase, the iteration gives the same results even when $\sin\mu > 1$. We do not know whether the condition (12) is satisfied by all functions or only by certain functions. In the first case it would be equivalent to the aforementioned condition $\sin\mu < 1$. However it is at least to us not obvious that Eq. (12) should be satisfied by any arbitrary function. It therefore looks like Eq. (12) is one of the best available uniqueness conditions so far for the solution of the unitarity integral equation.

ACKNOWLEDGMENTS

One of the authors (I.A. Sakmar) would like to thank Professor Robert L. Warnock for stimulating his interest in the topic of nonlinear integral equations and for discussions the author had with him. He would also like to thank the University of Nice and Professor Michel Le Bellac and Professor François Rocca for the hospitality and support extended to him at the Laboratoire de Physique Théorique at Nice.

[1]N.P. Klepikov, Zh. Eksp. Teor. Fiz. **47**, 757 (1964) [Sov. Phys. JETP **20**, 505 (1965)].
[2]R.G. Newton, J. Math. Phys. **9**, 2050 (1968).
[3]A. Martin, Nuovo Cimento A **59**, 131 (1969).
[4]D. Atkinson, P.W. Johnson, and R.L. Warnock, Commun. Math. Phys. **28**, 133 (1972).
[5]D. Atkinson, P.W. Johnson, and R.L. Warnock, Commun. Math. Phys. **33**, 221 (1973).
[6]D. Atkinson, M. Kaebebeke, and M. De Roo, J. Math. Phys. **16**, 685 (1975).
[7]H. Cornille, Lett. Nuovo Cimento **13**, 273 (1975).
[8]M. Tortorella, J. Math. Phys. **16**, 823 (1975).
[9]I.A. Sakmar, J. Math. Phys. **19**, 2124 (1978).
[10]I.A. Sakmar, J. Math. Phys. **20**, 11 (1979).
[11]I.A. Sakmar, Phys. Rev. Lett. **42**, 1252 (1979).
[12]S.A. Almehed and C. Lovelace, CERN Report No. CERN Th. 1408, 1971 (unpublished).

Separating variables in two-way diffusion equations

Nathaniel J. Fisch
Plasma Physics Laboratory, Princeton University, Princeton, New Jersey 08544

Martin D. Kruskal
Department of Astrophysical Sciences and Program in Applied Mathematics, Princeton University, Princeton, New Jersey 08544

(Received 10 October 1979; accepted for publication 4 January 1980)

It is shown that solutions to a class of diffusion equations of the two-way type may be found by a method akin to separation of variables. The difficulty with such equations is that the boundary data must be specified partly as initial and partly as final conditions. In contrast to the one-way diffusion equation, where the solution separates only into decaying eigenfunctions, the two-way equations separate into both decaying and growing eigenfunctions. Criteria are set forth for the existence of linear eigenfunctions, which may not be found directly by separating variables. A speculation with interesting ramifications is that the growing and decaying eigenfunctions are separately complete in an appropriate half of the problem domain. This conjecture is not proved, although it does enjoy some numerical support.

I. INTRODUCTION

Many physical systems may be described by what might be called a two-way diffusion equation, which we write in the form

$$h(\theta)\frac{\partial f(x,\theta)}{\partial x} = \frac{\partial}{\partial \theta} D(\theta)\frac{\partial}{\partial \theta} f(x,\theta) \tag{1}$$

in the domain $a<\theta<b$ and $0<x<L$, with $D(\theta)$ assumed positive. If $h(\theta)$ is also positive, then Eq. (1) represents the usual diffusion equation, which is well posed when initial conditions are given at $x=0$ and boundary conditions are given at $\theta=a$ and $\theta=b$. However, in the event that $h(\theta)$ changes sign in the interval (a,b), Eq. (1) then describes diffusion towards increasing x where h is positive and diffusion towards decreasing x where h is negative. Hence, we have the nomenclature "two-way" diffusion, which is also found in the literature as "forward–backward" diffusion. These equations are then well-posed only when initial conditions are given where h is positive and final conditions (i.e., at $x=L$) are given where h is negative. Consideration of these equations occurs in the literature as early as 1913.[1]

More complicated variations of Eq. (1) may be envisioned, for example, when h and D depend on x as well as θ. However, we restrict our consideration to h, D, and boundary conditions that are independent of x, so that Eq. (1) may be approached by the method of separation of variables. Also, for simplicity, we will assume, except in Sec. VIII, that h has only isolated zeros. The goal of this paper is, in part, to examine various subtleties arising in separating variables in the two-way diffusion equation.

In practice the restriction on h and D does not exclude most cases of interest arising in physics applications. For example, particles impinging with velocity $|\mathbf{v}|$ upon an infinite slab of randomly located, small-angle, elastic point-scatterers are governed by the diffusion equation

$$\left(\frac{\partial}{\partial t} + \mathbf{v}\cdot\frac{\partial}{\partial \mathbf{r}}\right) f = \alpha \Delta_v f, \tag{2}$$

where f is the particle phase space density, α is a constant, and Δ_v is the angular Laplacian operator in velocity space. The steady-state distribution of particles along the axis of the slab is described by[2]

$$\cos\theta\frac{\partial f}{\partial x} = \frac{1}{\sin\theta}\frac{\partial}{\partial \theta}\sin\theta\frac{\partial}{\partial \theta} f, \tag{3}$$

where θ is the angle the velocity vector makes with the x axis, x being the distance along the slab axis normalized to $|\mathbf{v}|/\alpha$. The range of θ is $(0,\pi)$, so that Eq. (3) is a special case of the two-way type considered in Eq. (1).

The two-dimensional analogue of Eq. (2) occurs when particles are scattered instead by randomly located rod scatterers, the rods being oriented in one direction and parallel to the face of the slab. Instead of Eq. (3), we then get

$$\sin\theta\frac{\partial f}{\partial x} = \frac{\partial^2}{\partial \theta^2} f, \quad -\pi<\theta<\pi, \tag{4}$$

where θ is now the angle the velocity vector makes with the direction perpendicular to both the slab axis and the rod orientation. The distance along the slab axis, x, is now normalized to $\alpha/v_\perp$, where $v_\perp$ is the particle velocity perpendicular to the rod orientation and is conserved now during scattering events. Equations (3) and (4) govern what is called diffuse reflection. Proper boundary conditions would specify the incident particle distribution [corresponding, say, in Eq. (4) to $\theta>0$] at $x=0$ and would specify that only outgoing particles are present at $x=L$. A recent problem of interest governed by Eq. (4) is the scattering of plasma waves in a tokamak by random density fluctuations aligned in rodlike fashion along the magnetic field.[3]

Much of the research on two-way diffusion equations has centered around a special case of Eq. (1), namely Eq. (3), which was derived by Bothe.[2] Bethe *et al.*[4] treated this equation by using separation of variables and finding the dominant behavior from the lowest eigenfunctions. A numerical check of those conclusions has been provided by Stein and Bernstein.[5] Beals[6] analytically proved the existence of a solution to Eq. (3) and, furthermore, proved that it could be represented by the eigenfunction expansion proposed by

 0022-2488/80/040740-11$1.00

Bethe *et al.* Considerations of other two-way equations occur in Refs. 7–11.

Our approach to the general equation, Eq. (1), is to use separation of variables as in Ref. 4. Here, our concern is that the resulting eigenvalue equation is not governed by the usual Sturm–Liouville theorems, a situation that we seek to remedy. The paper is organized as follows. In Sec. II we examine when, as Bethe *et al.* found in their case, the separation of variables solutions must be supplemented by an aditional singular eigenfunction. In Sec. III we show how this singular eigenfunction may be derived from a limiting case of completely separable (nonsingular) two-way equations. In Sec. IV we show that no other singular types of solutions are possible for this class of equations. In Secs. V and VI we prove a completeness theorem on the interval (a,b) for the eigenfunctions obtained by separating variables. In Sec. VII we conjecture a further completeness property of these eigenfunctions and we appeal to, among other things, a numerical computation that lends support to the conjecture. In Sec. VIII we show how to extend our considerations to the case that h vanishes over an interval. In Sec. IX we conclude with a summary of the salient findings, including ramifications of the proved completeness theorem and the conjectured completeness property.

Before concluding this introductory section, we wish to point out that the uniqueness of the solution to Eq. (1), if it exists, is an easy matter to show via the usual energy integral. Since Eq. (1) is linear, it is satisfied by the difference, $\phi = f_1 - f_2$, of any two supposed solutions. We multiply Eq. (1) by ϕ and integrate over x and θ. Upon integrating by parts in θ on the right-hand side, the surface terms vanish for suitable boundary conditions at $\theta = a$ and $\theta = b$ [see Eq. (7)], implying that ϕ vanishes, hence also uniqueness.

II. SEPARATION OF VARIABLES

To solve Eq. (1) by the method of separation of variables, we attempt an expansion

$$f = \sum_k c_k \phi_k(x) u_k(\theta), \tag{5}$$

where c_k is a constant, $\phi_k(x) = \exp(kx)$, and u_k satisfies the eigenvalue equation

$$\mathscr{L}_k u_k(\theta) \equiv \left[\frac{d}{d\theta} D(\theta) \frac{d}{d\theta} - kh(\theta) \right] u_k(\theta) = 0. \tag{6}$$

We assume that the boundary conditions are given such that Eq. (6) is self-adjoint, i.e., the boundary conditions are such that

$$D(a)\left(\frac{du_j(a)}{d\theta} u_k(a) - \frac{du_k(a)}{d\theta} u_j(a)\right) = D(b)\left(\frac{du_j(b)}{d\theta} u_k(b) - \frac{du_k(b)}{d\theta} u_j(b)\right). \tag{7}$$

Self-adjointness assures the existence of an orthogonality relation between the u_k with weighting function h, i.e.,

$$\int_a^b h(\theta) u_k(\theta) u_l(\theta) = 0, \quad \text{if } k \neq l. \tag{8}$$

The question arises whether the u_k found from Eq. (6) comprise a complete set of eigenfunctions. Completeness is not assured by the usual Sturm–Liouville[12] theorems, which do not apply when h vanishes in the interval (a,b). In fact, the u_k are not, in general, complete. For example, suppose that, as in Eq. (4), we pick $D = 1$, $h = \sin\theta$, and we try to use the eigenfunctions calculated from Eq. (6) to describe $f(\theta)$ such that

$$\int_{-\pi}^{\pi} f(\theta) \sin\theta \, d\theta = J \neq 0. \tag{9}$$

Since all the u_k are orthogonal to $\sin\theta$, any function represented as a linear combination of the u_k must have $J = 0$. Hence, an f characterized by Eq. (9) is not representable, implying that the set $\{u_k\}$ is not complete. It turns out, however, that supplementing the set $\{u_k\}$ with $\sin\theta$ does produce a complete set of functions. This assertion will be proved in Sec. V.

More generally, suppose that there exists a function g obeying the boundary conditions of the eigenvalue Eq. (6), and satisfying

$$\frac{d}{d\theta} D(\theta) \frac{d}{d\theta} g(\theta) + h(\theta) = 0. \tag{10}$$

The conditions for the existence, which is not assured, of such a g are well known.[12] When g exists, there may be a solution to Eq. (1) of the form $x - g(\theta)$, which we refer to as the linear or diffusion solution. This solution, which is not obtained by means of product separation of variables like the other solutions, but by sum separation, may be used to complete the u_k, so that the union of the u_k and the linear solution can represent any function of θ at a given, i.e., constant, x.

Since the linear-in-x part of the diffusion solution must also satisfy the boundary conditions, in fact at every x, only a subset of the self-adjoint boundary conditions allow diffusion solutions. It may be seen that for diffusion solutions to exist in well-posed problems, the self-adjoint boundary conditions must be restricted such that for some constant η

$$D(a) \frac{\partial f(x,a)}{\partial \theta} = D(b) \frac{\partial f(x,b)}{\partial \theta} \equiv \eta \tag{11}$$

and either

$$f(x,a) = f(x,b), \quad \text{if } \eta \neq 0, \tag{12}$$

or suitable conditions hold on allowable f if $\eta = 0$. In this latter category falls the first example of Sec. I, i.e., Eq. (3), where $D(a) = D(b) = 0$ and f is assumed to be nonsingular. In the former category, $\eta \neq 0$, falls Eq. (4), the example discussed in this section, where periodic boundary conditions of f are assumed. In this case we have $g = \sin\theta$, so that $x - \sin\theta$ is the diffusion solution that completes the u_k.

The rule is that when h is orthogonal to all the u_k, so that it cannot be expanded in the u_k, then the solution $x - g(\theta)$ exists and may be used to represent h. Furthermore, in such a case, g is orthogonal, with weighting function h, to all the u_k except u_0. This may be demonstrated by multiplying Eq. (10) by u_k, integrating twice by parts the left-hand side, and finally substituting from Eq. (6). Invoking the orthogonality of h and u_k on the right-hand side, which is ob-

tained by virtue of Eq. (11), then gives the desired orthogonality property for g,

$$k\int_a^b h(\theta)\, g(\theta) u_k(\theta) = 0\,, \tag{13}$$

which is nontrivial for $k \neq 0$.

III. DEGENERACY OF THE ZEROTH EIGENVALUE

In the previous section we noted that the diffusion solution is orthogonal to all but u_0, the $k = 0$ eigenfunction. This observation naturally leads us to suspect that there may be a particularly close connection between the diffusion solution and u_0. In this section we explore this connection and derive the diffusion eigenfunction in a natural way from the limiting form of a separable equation.

For convenience, we consider a specific example, although the conclusions are general. Therefore, considering the example in Sec. I, we attempt to break the degeneracy in Eq. (4) by posing instead

$$(\epsilon + \sin\theta)\frac{\partial f}{\partial x} = \frac{\partial^2 f}{\partial \theta^2}\,, \tag{14}$$

again with periodic boundary conditions and in the interval $(-\pi,\pi)$. For $\epsilon \neq 0$, the boundary conditions are incompatible with the existence of a diffusion solution, i.e., no g can solve Eq. (10). It naturally follows, then, to ask how the diffusion solution can arise in the limit $\epsilon \to 0$.

For $\epsilon \neq 0$, we can solve Eq. (14) by separation of variables, i.e.,

$$f = \sum_k a_k \exp(kx)\phi_k(\theta)\,, \tag{15}$$

where ϕ_k satisfies the eigenvalue equation

$$k(\epsilon + \sin\theta)\phi_k = \phi_k''\,, \tag{16}$$

the prime denoting differentiation with respect to θ.

We wish to prove now that there is an eigenvalue k of order ϵ, for ϵ small. This eigenvalue is in addition to the eigenvalue $k = 0$. To find this eigenvalue, we assume $k \sim \epsilon$ and formally expand

$$\phi = \phi_0 + \phi_1 + \phi_2 + \cdots\,, \tag{17}$$

where $\phi_n \sim \epsilon^n \sim k^n$. Inserting Eq. (17) into (16), we find to zeroth order

$$\phi_0'' = 0\,, \tag{18}$$

whereupon invoking periodicity boundary conditions, we find $\phi_0 = C_0$, where C_0 is a constant. The order-ϵ equation now gives

$$\phi_1'' = C_0 k \sin\theta\,, \tag{19}$$

and, again invoking periodicity boundary conditions, we find

$$\phi_1 = -C_0 k \sin\theta\,, \tag{20}$$

which is of order ϵ, as supposed. So far k has not been determined. However, the order ϵ^2 equation may now be written as

$$\phi_2'' = k\,[(\sin\theta)\phi_1 + \epsilon\phi_0] = kC_0(\epsilon - k\sin^2\theta)\,, \tag{21}$$

which has a solution for periodicity boundary conditions only if the consistency condition

$$\int_{-\pi}^{\pi} (\epsilon - k\sin^2\theta)\, d\theta = 0 \tag{22}$$

is satisfied. From Eq. (22), we determine $k = 2\epsilon$. This allows us to find

$$\phi_2 = -C_0(\tfrac{1}{2}\epsilon^2 \cos 2\theta + C_2)\,, \tag{23}$$

where $C_2 \sim \epsilon^2$ is a constant that may be lumped in with C_0.

Thus, we have found an eigenvalue of order ϵ to Eq. (16), and the corresponding term in Eq. (15) is

$$\begin{aligned} a_k \exp(kx)\phi_k &= a_k(1 + kx + \tfrac{1}{2}k^2x^2 + \cdots)(1 - k\sin\theta \\ &\quad - \tfrac{1}{8}k^2\cos 2\theta + \cdots) \\ &= a_k[1 + kx - k\sin\theta + O(k^2)]\,, \end{aligned} \tag{24}$$

where we have expanded the exponential term in kx, assuming that x is $O(1)$ when retaining terms. The eigenfunction in Eq. (24) is additional to the eigenfunction corresponding to $k = 0$ exactly, which is a constant. Thus, we may use the $k = 0$ eigenfunction to substract out the constant part of Eq. (24), i.e., a_k, the remaining part also satisfying Eq. (14). We write this remaining part as $a_k k\,[x - \sin\theta + O(k)]$ and take the limit $k \to 0$ while taking $a_k \sim 1/k$. Thus, higher-order terms in k drop out and we are left with the diffusion eigenfunction, $x - \sin\theta$, as found in Sec. II.

Finally, we note that before removing the constant term in Eq. (24) and taking the limit $\epsilon \to 0$, all eigenfunctions are orthogonal with respect to the weighting function, $\epsilon + \sin\theta$. Taking the limit $\epsilon \to 0$ may be viewed as the merging of two eigenvalues or a degeneracy in the zeroth eigenvalue. Thus, the two eigenfunctions remain orthogonal to the remaining eigenfunctions. They may also be made orthogonal to each other at any x, but not simultaneously at all x.

These arguments may be applied to the general case, where h is altered perturbatively in such a manner that g no longer exists. This removes the degeneracy in the lowest eigenvalue, and taking the limit of zero perturbation recovers the diffusion solution in a manner entirely analogous to the case presented. That the resulting set of eigenfunctions, the u_k plus the diffusion solution, is complete is still not assured, although this property is now somewhat motivated by the observations on the degeneracy as a limiting case. In the next section we show that other degeneracies, not in the zeroth eigenvalue, are impossible. The proof of the completeness property is reserved for Sec. V.

IV. SIMPLICITY OF THE NONZERO EIGENVALUES

The results of the previous section concerning the degeneracy of the zeroth eigenvalue point naturally to the possibility of degeneracy in the nonzero eigenvalues also. If such a degeneracy were to occur, then by analogy to the degeneracy already studied, we may expect solutions of the form

$$f = g_1(\theta)x\exp(kx) + g_2(\theta)\exp(kx)\,. \tag{25}$$

Substituting into Eq. (1), we see that g_1 and g_2 must satisfy

$$\mathscr{L}_k g_1(\theta) = 0\,, \tag{26a}$$

$$\mathscr{L}_k g_2(\theta) = h(\theta) g_1(\theta)\,. \tag{26b}$$

We assume boundary conditions that allow diffusion solutions, i.e.,

$$f(a) = f(b)\,, \tag{27a}$$

$$D(a)\frac{\partial f(a)}{\partial\theta} = D(b)\frac{\partial f(b)}{\partial\theta}, \tag{27b}$$

although we prove now that such solutions cannot occur for $k \neq 0$.

Note that g_1 satisfies the usual eigenvalue equation, i.e., Eq. (6), and thus has the usual property

$$k\int_a^b hg_1^2\,d\theta = -\int_a^b D\left(\frac{dg_1}{d\theta}\right)^2 d\theta. \tag{28}$$

Now we also find

$$\int_a^b (g_2\mathscr{L}_k g_1 - g_1\mathscr{L}_k g_2)\,d\theta = -\int_a^b hg_1^2\,d\theta = 0, \tag{29}$$

where the second equality is obtained because the left-hand side of Eq. (29) vanishes upon application of the boundary conditions after the obvious integration by parts. Equation (29), which is a necessary condition for any degenerate solution, implies that h must change sign. Furthermore, Eqs. (28) and (29) together imply that

$$\int_a^b D\left(\frac{dg_1}{d\theta}\right)^2 d\theta = 0, \tag{30}$$

which, in turn, implies that g_1 must be constant, since D is not zero at any interior points. However, if g_1 is constant, then k must be zero to satisfy Eq. (26a). Hence, all nonzero eigenvalues are nondegenerate or simple, and degenerate solutions can occur only for $k = 0$, i.e., they are of the type considered in the previous section. The necessary and sufficient conditions for their occurrence are that the boundary conditions permit the constant solution to be an allowable eigenfunction and that Eq. (29) hold, which may now be written simply as

$$\int_a^b h(\theta)\,d\theta = 0. \tag{31}$$

Although we have shown that merging of roots may occur only at $k = 0$, we have not yet limited the number of roots which may merge, i.e., the order of the degeneracy. For example, if n roots were to merge at $k = 0$, then a solution of the form

$$f = x^n + x^{n-1}g_1(\theta) + \cdots + xg_{n-1}(\theta) + g_n(\theta) \tag{32}$$

could exist, where the coefficient of x^n must obviously be independent of θ as taken. We show now that multiple degeneracies of the form of Eq. (32) cannot exist for $n > 1$. For Eq. (32) to satisfy Eq. (1), the g_n must satisfy

$$nh = \frac{d}{d\theta}\frac{Ddg_1}{d\theta}, \tag{33a}$$

$$(n-l)hg_l = \frac{d}{d\theta}\frac{Ddg_{l+1}}{d\theta}, \quad 1 \leqslant l \leqslant n, \tag{33b}$$

and obey boundary conditions of the type given by Eqs. (27). From Eq. (33a) we find that

$$\int_a^b D\left(\frac{dg_1}{d\theta}\right)^2 d\theta = -n\int_a^b hg_1\,d\theta = 0, \tag{34}$$

where the second equality arises from use of Eq. (33b) when $n \neq 1$ and implies that $dg_1/d\theta = 0$, which is impossible in view of Eq. (33a). Hence there cannot exist any solutions of the type given by Eq. (32) for $n > 1$.

In summary the rule is that there is only one possible degeneracy, namely, the merging of two roots at $k = 0$. Corresponding to this degeneracy, there can be at most one independent diffusion solution of the form $x - g(\theta)$. [Any other diffusion solution, say $x - y(\theta)$, is a linear combination of the above solution and the u_0 eigenfunction, since $g - y$ satisfies Eq. (6) with $k = 0$.]

V. PROOF OF CLOSEDNESS

In the next two sections, we prove that any function defined on the interval (a,b) and obeying the boundary conditions at $\theta = a$ and $\theta = b$ may be written as a linear combination of the u_k and g, which are found from

$$\mathscr{L}_k u_k(\theta) = 0, \tag{35a}$$

$$\mathscr{L}_0 g(\theta) = h(\theta), \tag{35b}$$

where $\mathscr{L}_k$ is defined in Eq. (6) and the boundary conditions are given by Eq. (11) and either Eq. (12) or a suitable replacement as discussed in Sec. II. The completeness proof that we offer is motivated by the method of Kneser,[13] which was useful in proving completeness for proper Sturm–Liouville problems. First we prove a "closedness" property, i.e., if any function is orthogonal to all of the eigenfunctions, then it must be zero. The completeness property follows from the closedness property. Kneser's idea, which may be found within the context of subsidiary theorems in Ref. 12, is to construct a series solution to a related inhomogeneous problem. Information concerning the closedness of the eigenfunctions of the homogeneous problem then follows from the convergence properties of the series.

Motivated in this manner, we first consider the inhomogeneous equation

$$\frac{d}{d\theta}D(\theta)\frac{d}{d\theta}v(\theta) + kh(\theta)v(\theta) + p(\theta) = 0, \tag{36}$$

with boundary conditions imposed on v that allow diffusion solutions to the related homogeneous equation, i.e., Eq. (36) with $p = 0$. We construct a solution to Eq. (36) by means of the expansion

$$v(\theta) = v_0 + kv_1 + \cdots + k^n v_n + \cdots, \tag{37a}$$

where the terms of the series are found from

$$\frac{d}{d\theta}D\frac{d}{d\theta}v_0 + p = 0, \tag{37b}$$

$$\frac{d}{d\theta}D\frac{d}{d\theta}v_n + hv_{n-1} = 0, \quad n \geqslant 1, \tag{37c}$$

with the v_n obeying the same boundary conditions that v obeys.

The radius of convergence of the series solution is bounded by $\rho = |l|$, where l is the smallest (in absolute value) eigenvalue of the homogeneous system for which

$$\int_a^b p(\theta)u_l(\theta)d = 0 \tag{38}$$

does not hold. If no such l exists, then $\rho = \infty$. For $|k| < \rho$, the series will converge (e.g., Sec. 11.3 in Ref. 12 with minor modification assures this) so long as successive v_n may be found unambiguously. The condition on the existence of the series' coefficients v_n imposes restrictions on p in addition to

those already imposed by Eq. (38). The analysis now departs somewhat from that for a proper Sturm–Liouville problem, for which the existence of the v_n is assured without further restrictions on p.

First, we note that upon integrating Eqs. (37b) and (37c) over the interval (a,b) and applying the boundary condition in Eq. (11), we obtain

$$\int_a^b p(\theta)\, d\theta = 0\,, \tag{39a}$$

$$\int_a^b h(\theta) v_n(\theta)\, d\theta = 0\,, \tag{39b}$$

the latter relation holding for all n. The restriction on p given by Eq. (39a) is necessary for the convergence of the series anywhere and is equivalent to Eq. (38) with $l = 0$, where then u_l is the zeroth eigenfunction, which is a constant when the boundary conditions allow diffusion solutions. Nevertheless, the necessity of satisfying Eq. (39a) means that v_0 cannot be fully determined from Eq. (37b) alone. Similarly, the compatibility condition, Eq. (39b), means that Eq. (37c) alone is insufficient to determine v_n unambiguously. It may be suspected, however, that the compatibility conditions provide sufficient additional restrictions to determine the v_n. We now set out to prove that, in fact, this is so.

Multiplying Eq. (37c) by g and integrating yields

$$\begin{aligned}\int_a^b ghv_{n-1}\, d\theta &= -\int_a^b g\left(\frac{d}{d\theta} D \frac{d}{d\theta} v_n\right) d\theta \\ &= \int_a^b \left(\frac{dv_n}{d\theta}\right) D \frac{dg}{d\theta}\, d\theta \\ &= -\int_a^b v_n \left(\frac{d}{d\theta} D \frac{dg}{d\theta}\right) d\theta \\ &= -\int_a^b v_n h\, d\theta\,, \end{aligned} \tag{40}$$

where, upon each integration by parts, the boundary terms vanished by virtue of the assumed boundary conditions. The last equality was written on the basis of substitution from Eq. (35b). Now by virtue of Eq. (39b), the right-hand side of Eq. (40) vanishes, from which we obtain the orthogonality of g and the v_n with weighting function h.

Although g is orthogonal to the v_n, it is not orthogonal to the constant function, as may be shown by multiplying Eq. (35b) by g and integrating once by parts to obtain

$$\int_a^b hg\, d\theta = -\int_a^b D\left(\frac{dg}{d\theta}\right)^2 d\theta \neq 0\,. \tag{41}$$

The inequality above obtains because D does not pass through zero and g cannot be constant.

It may be seen that each v_n is determined from Eq. (37c) up to an arbitrary additive constant, say of the form

$$v_n = C + R(\theta)\,, \tag{42}$$

where $R(\theta)$ is a known function and C is an unknown constant. We show now that Eqs. (40) and (41) are sufficient to determine C and resolve the ambiguity in the v_n. Substituting for v_n from Eq. (42) into Eq. (40) and using Eqs. (39b) and (41), we construct

$$C = \int_b^a Rhg\, d\theta \left[\int_a^b D\left(\frac{dg}{d\theta}\right)^2 d\theta\right]^{-1}. \tag{43}$$

The point is that C is always determined because the denominator cannot vanish. It may be noted that simply substituting v_n from Eq. (42) directly into the compatibility condition, Eq. (39b), would not determine C since h has zero area.

It should be noted that, in the above, nonzero k has been tacitly assumed in the application of the orthogonality relations, i.e., Eqs. (39b) and (40). When k vanishes, these relations, and hence Eq. (43) also, do not necessarily hold and C cannot be uniquely determined. In fact, when $k = 0$, the solution to Eq. (36) is not unique and can be determined only up to an arbitrary constant. Similarly, an additional constraint is needed to uniquely determine the solution to Eq. (37b), which is identical to Eq. (36) with $k = 0$. Here, nearly any additional constraint will do; we could, for example, assume that Eq. (43) holds also for $k = 0$ in which case a unique solution to Eq. (36) exists, even for $k = 0$, and may be constructed by means of the series for $|k| < \rho$.

Finally, multiplying Eq. (37b) by g, integrating, and manipulating the subsequent expression in a manner analogous to Eq. (40), we obtain

$$\int_a^b gp\, d\theta = 0\,, \tag{44}$$

which represents an additional constraint on p if the series is to have a nonzero radius of convergence. It should be appreciated that this condition is essentially anticipated by the resemblance of Eq. (44) to Eq. (38) in view of the discussion given in Sec. III.

To recapitulate, we have so far shown that if Eqs. (39a) and (44) hold, then the series solution converges in a finite interval about $k = 0$, namely for $|k| < \rho$. The series coefficients may be determined uniquely order by order by means of Eqs. (37) and (43). We now proceed to exploit the fact that if $\rho = \infty$, i.e., if Eq. (38) holds for all l, then v is an entire function of k.

We consider two cases of Eq. (37c), say

$$\frac{d}{d\theta} D \frac{d}{d\theta} v_n + h v_{n-1} = 0\,, \tag{45a}$$

$$\frac{d}{d\theta} D \frac{d}{d\theta} v_{m+1} + h v_m = 0\,. \tag{45b}$$

Multiplying Eq. (45a) by v_{m+1} and Eq. (45b) by v_n, subtracting, and integrating, we obtain

$$\int_a^b (v_{m+1} v_{n-1} - v_n v_m) h\, d\theta = 0\,, \tag{46}$$

where we made use of the boundary condition, Eq. (11). We may note that since n and m were arbitrarily chosen, the definition

$$W_q \equiv \int_a^b v_n v_m h\, d\theta\,, \quad n + m = q\,, \tag{47}$$

is unambiguous.

For proper Sturm–Liouville problems, the proof proceeds somewhat more simply as W_q is always non-negative for q even, since h is non-negative. This simplification does not occur in our case, where h does, in fact, pass through zero. We can, however, show that W_q is always non-negative for q odd. We do so by multiplying Eq. (45a) by v_n and

integrating to obtain

$$0 \leqslant \int_a^b D\left(\frac{dv_n}{d\theta}\right)^2 d\theta = \int_a^b h v_n v_{n-1}\, d\theta \equiv W_{2n-1} \,. \tag{48}$$

We may also consider the quantity

$$\delta \equiv \int_a^b \left[\alpha\left(\frac{dv_n}{d\theta}\right) + \beta\left(\frac{dv_m}{d\theta}\right)\right]^2 d\theta$$
$$= \alpha^2 W_{2n-1} + 2\alpha\beta W_{m+n-1} + \beta^2 W_{2m-1} \,, \tag{49}$$

where α and β are arbitrary constants. Since δ is non-negative for any choice of α and β, we must have

$$W_{m+n-1}^2 \leqslant W_{2n-1} W_{2m-1} \,. \tag{50}$$

Let q be an odd integer. From Eq. (50) we see that if $W_q = 0$ for any q, then all the W_q must be zero, except possibly for W_1. We now, in fact, show that W_1 must then also be zero, although this is not implied by Eq. (50). Using instead Eq. (48) with $n = 2$, we see that $W_3 = 0$ implies that $dv_2/d\theta = 0$, which, in turn, implies from Eq. (45a) with $n = 2$ that $hv_1 = 0$, which requires that W_1 vanish too. Hence, either all the W_q vanish or none do. Suppose first that none do. Since by Eq. (48) $W_q \geqslant 0$, Eq. (50) implies the inequalities

$$\frac{W_3}{W_1} \leqslant \frac{W_5}{W_3} \leqslant \frac{W_7}{W_5} \leqslant \cdots \leqslant \frac{W_{2n+1}}{W_{2n-1}} \leqslant \cdots \,, \tag{51}$$

which in turn implies that

$$W_{2n+1} \geqslant T^n W_1 \,, \tag{52}$$

where we have defined

$$T \equiv W_3/W_1 > 0 \,. \tag{53}$$

Now when v is an entire function of k [i.e., when p is orthogonal to all the u_l and g, with a constant as the weighting function, as in Eqs. (38) and (44)], then so is the quantity

$$\int_a^b v v_0 h\, d\theta = W_0 + kW_1 + \cdots + k^n W_n + \cdots \,. \tag{54}$$

The assumption $W_q \neq 0$, however, leads to a contradictory statement, since the sum of the subseries of all the odd terms of the series in Eq. (54),

$$\sum_{m=0}^{\infty} k^{2m+1} W_{2m+1} \geqslant k W_1 \sum_0^{\infty} (k^2 T)^m \,, \tag{55}$$

clearly diverges for some k. Hence, we have proved by contradiction that $W_q = 0$ and, in particular, that $W_1 = 0$, which implies that

$$0 = W_1 \equiv \int_a^b v_1 v_0 h\, d\theta = \int_a^b D\left(\frac{dv_1}{d\theta}\right)^2 d\theta \,, \tag{56}$$

which means that v_1 must be a constant. This latter statement, in turn, implies from Eq. (37c) that $hv_0 = 0$, which finally implies from Eq. (37b) that $p = 0$.

What we have proved is that if p is orthogonal to all the u_k and g, then $p = 0$. This closedness property is shown in the next section to imply completeness, i.e., that any suitable p may be constructed as a linear combination of the u_k and g in the manner

$$p(\theta) = \sum_k c_k u_k(\theta) + c_g g(\theta) \,, \tag{57}$$

which is the completeness relation that we seek. The constants c_k and c_g may then be determined easily from the orthogonality properties of the eigenfunctions. We multiply Eq. (57) by either h, hu_l, or hg and integrate to obtain, respectively,

$$c_g \int_a^b hg\, d\theta = \int_a^b hp\, d\theta \,, \tag{58a}$$

$$c_l \int_a^b h u_l^2\, d\theta = \int_a^b h p u_l\, d\theta \,, \quad l \neq 0 \,, \tag{58b}$$

$$c_0 \int_a^b hg\, d\theta = \int_a^b p\, gh\, d\theta - c_g \int_a^b g^2 h\, d\theta \,. \tag{58c}$$

The constants are all determined from Eqs. (58) since all the integrals on the left-hand sides have been shown not to vanish. The case $l = 0$ is excluded in Eq. (58b) for just that reason, since h has zero area, and instead use is made of Eq. (58c) to determine c_0.

VI. PROOF OF COMPLETENESS

In order to prove that the closedness of the eigenfunctions implies their completeness, we begin by pointing out that the asymptotic behavior of large eigenvalues is given by

$$k_n \sim n|n| \,, \quad n = 0, \pm 1, \pm 2, \cdots \,, \tag{59}$$

where n indexes the k_n and spans $(-\infty, \infty)$. In contrast, for Sturm–Liouville problems, $k_n \sim n^2$ where n spans only $(0, \infty)$. The validity of Eq. (59) can be demonstrated by means of matched asymptotic expansions. For example, suppose that $H \equiv h/D$ passes through zero only once in the interval (a,b), say at $\theta = 0$, with finite positive slope $C > 0$. We can then solve Eq. (6) away from $\theta \approx 0$ by means of a WKB expansion, which matches onto Airy functions near $\theta = 0$. Doing so for $k \to +\infty$, we find $u_k = \alpha u_A(\theta) + \beta u_B(\theta)$, where α and β are constants to be determined from the boundary conditions, and

$$u_A(\theta) = \begin{cases} \dfrac{1}{2\sqrt{\pi}} H^{-1/4} \exp\left(-k^{1/2}\displaystyle\int_0^\theta [H(\theta')]^{1/2}\, d\theta'\right), & \theta > 0, \\ \mathrm{Ai}(k^{1/3} C^{1/2} \theta), & \theta \approx 0, \\ \dfrac{1}{\sqrt{\pi}} [-H]^{-1/4} \sin\left(k^{1/2}\displaystyle\int_\theta^0 [-H(\theta')]^{1/2}\, d\theta' + \dfrac{\pi}{4}\right), & \theta < 0, \end{cases} \tag{60a}$$

$$u_B(\theta) = \begin{cases} \dfrac{1}{\sqrt{\pi}} H^{-1/4} \exp\left(k^{1/2}\displaystyle\int_b^\theta [H(\theta')]^{1/2}\, d\theta'\right), & \theta > 0, \\ 0, & \theta \lesssim 0. \end{cases} \tag{60b}$$

The point is that in order to satisfy self-adjoint boundary conditions, $k^{1/2}$ is determined only up to multiples of $2\pi\Gamma$, where Γ is a finite constant, whence Eq. (59) follows. For example, for periodicity boundary conditions, we have

$$\frac{\beta}{\alpha} \simeq \left[\frac{-H(b)}{H(a)}\right]^{1/4} \sin\left(k^{1/2}\int_a^0 (-H)^{1/2}\,d\theta + \frac{\pi}{4}\right), \tag{61}$$

and k is determined from

$$\tan\left(k^{1/2}\int_a^0 (-H)^{1/2}\,d\theta + \frac{\pi}{4}\right) = \left[\frac{-H(b)}{H(a)}\right]^{1/2}. \tag{62}$$

We now turn to the question of uniform convergence. Consider the functions which represent the decomposition of f into orthogonal modes, i.e.,

$$F_n(\theta) = u_n(\theta)\int_a^b f h u_n\,d\theta, \tag{63}$$

where u_n is the eigenfunction corresponding to k_n and is normalized according to

$$\int_a^b h u_n^2 = 1. \tag{64}$$

Note that from the asymptotic representation of the u_n for large n, we see that $H^{1/4}u_n$ is uniformly bounded for all n and θ.

Suppose that f obeys the same boundary conditions as the u_n and is twice differentiable. From Eq. (6) we have

$$\begin{aligned} F_n &= k_n^{-1}u_n\int_a^b f\frac{d}{d\theta}D\frac{d}{d\theta}u_n\,d\theta \\ &= k_n^{-1}u_n\int_a^b u_n\frac{d}{d\theta}D\frac{d}{d\theta}f\,d\theta. \end{aligned} \tag{65}$$

By virtue of the boundedness of $H^{1/4}u_n$ and the integrability of the weak singularity (i.e., since $u_n \sim H^{-1/4}$) in the integrand on the right side of Eq. (65), it is seen that

$$H^{1/4}F_n \sim 1/k_n \sim 1/n^2, \quad \text{as } n\to\pm\infty, \tag{66}$$

where we now made use of Eq. (59). Hence, the series of partial sums of $H^{1/4}F_n$ is absolutely and uniformly convergent in the interval (a,b).

We are now in a position to demonstrate that closedness implies at least a mild type of completeness, i.e., completeness in that any continuous twice-differentiable function f may be uniformly approximated. This restrictive or narrow property is needed before we relax the conditions on f to include all continuous functions. Proceeding in this vein, we define the function

$$\begin{aligned} \phi(\theta) &\equiv f(\theta) - \sum_k u_k(\theta)\int_a^b u_k(\theta')h(\theta')f(\theta')\,d\theta' \\ &\equiv f(\theta) - \sum_k u_k(\theta)a_k. \end{aligned} \tag{67}$$

Consider the quantity

$$\begin{aligned} \int_a^b h u_l \sum_k u_k a_k &= \int_a^b h u_l |h|^{-1/4}\sum_k |h|^{1/4}u_k a_k \\ &= \sum_k a_k\int_a^b h u_l u_k = a_l, \end{aligned} \tag{68}$$

where reversing the order of integration and summation to obtain the second equality is clearly allowable since the series is uniformly convergent by virtue of Eq. (66). Thus, we have for every u_l

$$\int_a^b \phi h u_l\,d\theta = \int_a^b f h u_l - a_l = 0, \tag{69}$$

which implies from the closedness of the u_l that $\phi h = 0$. Hence, from Eq. (67) we see that the u_k can represent f everywhere except possibly at the isolated point $\theta = 0$, which is the mild completeness property we sought to prove. It is now possible to follow relatively standard but tedious procedures to show that the above property implies that, in fact, any continuous f, not necessarily twice-differentiable, may be uniformly approximated by the u_n (except where $\theta = 0$). This last step of the completeness proof is given in Appendix A.

Note that this proof relied on a finite first derivative of h at $\theta = 0$ in order to write down the asymptotic expansion in terms of Airy functions. Nevertheless, it is easy to show that similar conclusions may be drawn when this derivative vanishes or when more general (but self-adjoint) boundary conditions are used. The case of h vanishing over an interval is discussed in Sec. VIII.

We now show that the proved completeness property on the interval (a,b) guarantees that if a solution to Eq. (1) exists, then it can be expanded in the separation-of-variables eigenfunctions. For f may then be put in the form

$$f = \sum_k A_k(x)u_k(\theta) + B(x)[x - g(\theta)] \tag{70}$$

and it suffices to show that the A_k and B are in fact given by

$$A_k(x) = c_k e^{kx} \tag{71a}$$

$$B = -c_g, \tag{71b}$$

where the c_k and c_g are constants found, say, from evaluating Eqs. (58) at $x = 0$. Equations (58) may now be used to establish Eqs. (71). For example, using Eq. (58b) and assuming the u_l have been normalized, we have for $k \neq 0$

$$A_k(x) = \int_a^b f h u_k\,d\theta, \tag{72}$$

and differentiating by x gives

$$\begin{aligned} \frac{d}{dx}A_k(x) &= \int_a^b h u_k\frac{df}{dx}\,d\theta \\ &= \int_a^b u_k\frac{d}{d\theta}D\frac{d}{d\theta}f\,d\theta \\ &= \int_a^b f\frac{d}{d\theta}D\frac{d}{d\theta}u_k\,d\theta \\ &= \int_a^b k f h u_k\,d\theta \\ &= kA_k(x), \end{aligned} \tag{73}$$

where in deriving the third equality we integrated twice by parts. Thus, Eq. (71a) follows for $k \neq 0$. In a similar manner, $A_0(x)$ and $B(x)$ may be shown to be constant as required.

VII. COMPLETENESS ON THE HALF-INTERVAL

In Sec. VI, we showed that the u_k are complete on the interval (a,b). This implied that the expansion by separation

of variables could be used to represent f everywhere, when f exists.

However, the existence of f for arbitrary initial and final conditions indicates that the u_k possess a far stronger property than completeness on the interval (a,b). By taking the limit $d\to\infty$, it can be seen that only the non-growing-in-x eigenfunctions can contribute at $x=0$, while only the non-decaying eigenfunctions can contribute at $x=d$. Thus, assuming that h vanishes only once, say, $h(0)=0$, we see that the exponentially decaying eigenfunctions, supplemented by the constant and linear eigenfunctions, should be complete on the interval where h is positive, say in $(0,b)$. Similarly, the growing eigenfunctions, also supplemented by the constant and linear eigenfunctions, should be complete in the interval $(a,0)$. (There is no problem in switching the order of taking the limit $d\to\infty$ with taking the limit of the number of eigenfunctions becoming infinite, since the higher order eigenfunctions, decaying or growing most rapidly, certainly cannot contribute at both boundaries.)

Thus, it may be seen that existence of the solution and its representability by the eigenfunction expansion should imply the further completeness property of the u_k, that "half" the u_k are complete in the interval $(0,b)$ while the other half are complete in the interval $(a,0)$. Furthermore, it may presumably be shown that the converse is also true, i.e., if the u_k possess this completeness property, then f must exist. This should follow from a construction of f by means of a convergent, iterative scheme, where the boundary conditions are alternatively satisfied by the decaying set of u_k at $x=0$ and the growing set at $x=d$.

Unfortunately, we have not been able to prove the completeness of the eigenfunctions on the half-interval, which would have, from the above argument, presumably provided a general and independent proof of the existence of solutions to Eq. (1). We must therefore content ourselves, at present, with the reverse argument. Thus, in special cases, where we may rely on other proofs for existence, then we can infer that the associated eigenfunctions possess a completeness property on the half-inteval.

To demonstrate how the eigenfunctions may be used to construct the solution, and to provide additional support for the conjecture that they are complete on the half-interval, we numerically consider one example of Eq. (1), namely

$$\sigma(\theta)\frac{\partial f}{\partial x}=\frac{\partial^2 f}{\partial\theta^2},\quad -\pi<\theta<\pi,\ 0<x<d, \tag{74}$$

where

$$\sigma(\theta)=\begin{cases}1, & \theta>0,\\ -1, & \theta<0,\end{cases} \tag{75}$$

and periodicity boundary conditions are assumed. This particular example was chosen for numerical analysis because the eigenfunctions are particularly simple. Since $\sigma(\theta)$ has zero area, we expect a linear eigenfunction of the form $x-u_{00}(\theta)$, where

$$u_{00}(\theta)=\begin{cases}\theta(\pi-\theta), & \theta>0,\\ \theta(\pi+\theta), & \theta<0.\end{cases} \tag{76}$$

The eigenfunctions that are even about $\pi/2$ in the interval $(0,\pi)$ are of the form, for $n>0$,

$$u_n(\theta)=\begin{cases}\cos[\lambda_n(\theta-\pi/2)], & \theta>0,\\ C_n\cosh[\lambda_n(\theta+\pi/2)], & \theta<0,\end{cases} \tag{77}$$

where

$$C_n=[\cos(\lambda_n\pi/2)][\cosh(\lambda_n\pi/2)]^{-1} \tag{78}$$

and λ_n solves

$$\cos\lambda_n\pi=\operatorname{sech}\lambda_n\pi\,. \tag{79}$$

For $n<0$, i.e., for the decaying eigenfunctions of the form $u_n(\theta)\exp(-\lambda_n x)$, we have $u_n(\theta)=u_{-n}(-\theta)$. Similarly, there are eigenfunctions that are odd about $\theta=\pi/2$ in the interval $(0,\pi)$. Since we will assume even boundary conditions, the odd functions need not be considered here.

Part of the reason for giving these eigenfunctions in detail is that as $n\to\infty$ the u_n exponentially fast approach $\cos[(2n+\frac{1}{2})(\theta-\pi/2)]$. If the u_n are complete on $(0,\pi)$, which we will demonstrate numerically, it is expected that the $\cos[(2n+\frac{1}{2})(\theta-\pi/2)]$, supplemented by a constant or roughly constant function, must also be complete on $(0,\pi)$. Nevertheless, despite its simple form, we have been unable to analytically demonstrate that the set $\cos[(2n+\frac{1}{2})(\theta-\pi/2)]$ is complete, which presumably might be an easier task than to demonstrate that property for the u_n.

In Fig. 1 we show the results of numerically fitting the eigenfunction expansion to boundary data using the method of least squares. The boundary conditions are that $f(\theta<0, x=d)=0$ and $f(\theta>0, x=0)$ has the Gaussian type of dependence shown in Fig. 1(a). Here, we have taken $d=10$. We find that the root-mean-square difference between the eigenfunction expansion and the given bundary data converges to zero as $1/N$, where N is the number of eigenfunctions employed. This type of convergence is indicative of the presence of Gibb's phenomenon in this problem. Indeed, some such phenomenon is expected since the solution on the boundary cannot be analytic over the whole interval $(-\pi,\pi)$. For example, at $x=d$, there cannot exist a smooth fit of any non-zero solution in the region $\theta>0$ to the zero data given in the region $\theta<0$. (Parenthetically, we remark that in analogy with the examples given in Sec. I, the zero data correspond to a condition of only outgoing particles

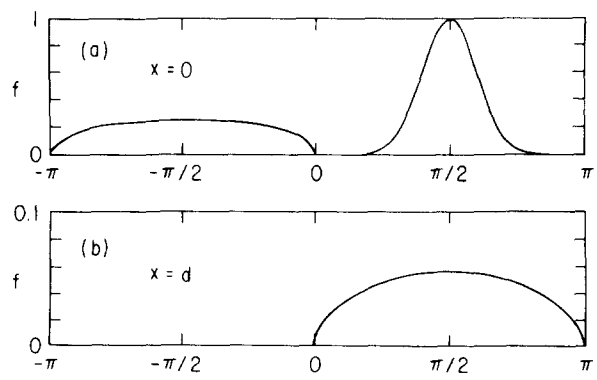


FIG. 1. Data on the boundaries. (a) $f(\theta)$ vs θ on the boundary at $x=0$; for $\theta>0$, the data is given, whereas for $\theta<0$ the data is computed, using 250 eigenfunctions. (b) $f(\theta)$ vs θ on the boundary at $x=10$; for $\theta<0$, the data is given, whereas for $\theta>0$ the data is computed. Note change of scale.

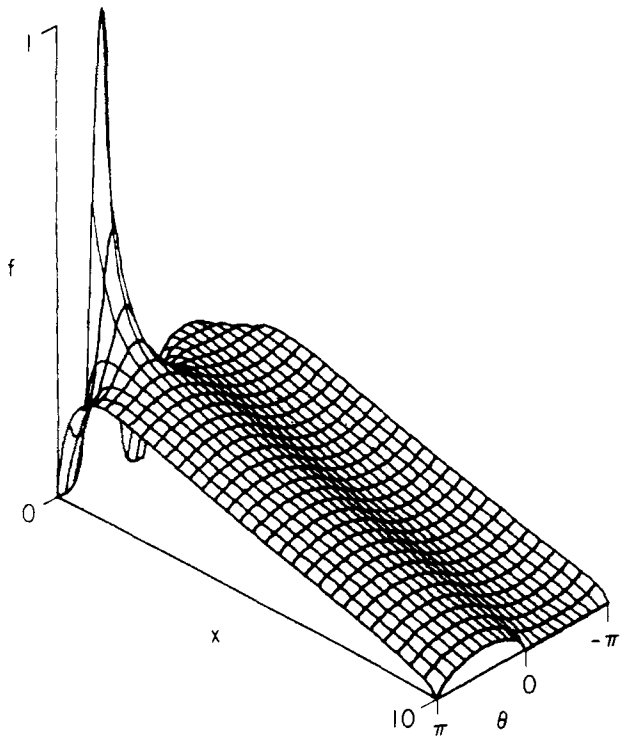


FIG. 2. Surface of $f(\theta,x)$ using the boundary data given in Fig. 1.

present at one end of the collision region.) In Fig. 2 we show, using a large number of eigenfunctions, f as a function of both x and θ. Note that, except near the boundaries, f is described mainly by the linear solution.

Numerical results for large d are similar. For very large d, only the decaying eigenfunctions remain finite. Thus, there is numerical evidence, at least in this case, for the conjectured completeness property on the half-interval.

A remark, appropriate before closing this section, is that the eigenfunctions u_n, which are supposed complete on $(0,\pi)$, pass through zero n times in that interval and not at all in $(-\pi,0)$. The reverse holds true for those eigenfunctions complete on the other half-interval.

VIII. WHEN $h(\theta)$ VANISHES OVER AN INTERVAL

When $h(\theta)$ vanishes over an interval, it turns out that the closedness and completeness properties of the u_k do not extend over the full interval (a,b). Instead, it may be shown that these properties extend over the interval (a,b) with the exclusion of the subintervals over which h vanishes. For example, in deducing the closedness property following Eq. (56), we relied on the vanishing of hv_0. Were h to vanish over an interval, then we could conclude only that p vanishes where h does not. Similarly, the conjectures regarding properties on the half-interval would then apply only where h is either positive or negative.

That the completeness property fails to apply where $h = 0$ presents no difficulty in describing the solution to Eq. (1) by means of an eigenfunction expansion. In fact, it can be shown via an energy integral (in the manner described at the end of Sec. I) that specifying boundary conditions at $x = 0$ where $h > 0$ and at $x = d$ where $h < 0$ implies uniqueness of the solution. Hence, nothing can be specified where $h = 0$ anyway, which is eminently consistent with an eigenfunction expansion wherein the eigenfunctions are complete on the interval that excludes the set where $h = 0$.

IX. SUMMARY AND CONCLUSIONS

In a manner of summary, we remark that the important features in this work are the proof of completeness of the $u_k(\theta)$ on (a,b), the criteria for the existence of a diffusion solution, and the conjecture regarding the completeness of half of the $u_k(\theta)$ on the interval $(a,0)$ or $(0,b)$.

The completeness of the eigenfunctions $u_k(\theta)$ on the interval (a,b), which was proved in Secs. V and VI, guarantees that when a solution to Eq. (1) exists it may be expanded in the form given by Eq. (5). It should be pointed out, however, that it may not always be practical to employ the expansion in numerically solving Eq. (1), since the u_k themselves may be hard to compute and, once computed, do not enjoy useful orthogonality properties for properly posed boundary value problems. Nevertheless, we have seen that when the u_k are easily found, as in the example given in Sec. VII, the eigenfunction expansion is certainly convenient.

The criteria for the existence of the diffusion solution allow the gathering of partial information about the solution without obtaining it completely. For example, the existence of a diffusion solution to Eq. (3) implies the well-known fact that the amount of sunlight reaching earth through a nonabsorbing cloud layer drops off only as the reciprocal of the layer thickness rather than with an exponential dependence on it.

Finally, the conjecture regarding the half-interval problem perhaps has the farthest reaching implications of all and its verification is worthy of future investigation. The correctness of the conjecture should imply an independent proof of the existence of a solution to the related two-way diffusion equation. There is the added academic interest in that the conjecture relates to functions which do not satisfy a Sturm–Liouville equation on the interval upon which the completeness property is supposed. This is in contrast to the other findings here, which may be viewed somewhat as a supplement or extension to the standard Sturm–Liouville theory.

The proof of the conjecture regarding completeness on the half-interval has now been provided by R. Beals.[14]

ACKNOWLEDGMENT

The work of one of the authors (NJF) was supported by the United States Department of Energy Contract No. EY-76-C-02-3073.

APPENDIX A

In this appendix we show that the capability of the eigenfunction expansion to represent continuous twice-differentiable f implies the capability to also represent any f that is merely continuous. The proof of this step, which supplements the completeness proof in Sec. VI, follows Ref. 12, Sec. 11.52. The idea is to compare the expansion in the u_n with a uniformly convergent expansion in a known set of complete orthogonal functions, χ_n. For simplicity, we con-

sider the intervals $(a,0)$ and $(0,b)$ separately. Assume that the χ_n are defined on $(0,b)$ and are orthogonal with respect to the weighting function $\nu(\theta)>0$. We define the χ_n to be zero for $\theta<0$. Then we can define the partial sums

$$S_N(\theta)\equiv\int_a^b f(t)\sum_{-N}^{N}u_n(\theta)u_n(t)h(t)\,dt, \tag{A1a}$$

$$\sigma_N(\theta)\equiv\int_a^b f(t)\sum_{0}^{N}\chi_n(\theta)\chi_n(t)\nu(t)\,dt, \tag{A1b}$$

where $\sigma_N(\theta)$ is known to converge uniformly to $f(\theta)$ in the interval $(0,b)$, and we would like to prove the same for $S_N(\theta)$. Thus, we will try to show that

$$S_N(\theta)-\sigma_N(\theta)\to 0 \tag{A2}$$

uniformly in $(0,b)$ as $N\to\infty$.

To expedite matters, we define the function

$$\Phi_N(\theta,t)\equiv\sum_{n=0}^{N}[u_n(\theta)u_n(t)+u_{-n}(\theta)u_{-n}(t)-\chi_n(\theta)\chi_n(t)], \tag{A3}$$

where, for simplicity of notation, the diffusion solution, if it exists, is not written explicitly but is understood to be included in the summation. Now if $G(\theta)$ is continuous and twice differentiable, then

$$\int_a^b\Phi_N(\theta,t)G(t)\to 0 \tag{A4}$$

uniformly in $(0,b)$ as $N\to\infty$. This is a consequence of the uniform convergence of S_N to such G (proved in Sec. VI). The same property holds for the σ_N by assumption. Equation (A4) is obtained since their difference must then uniformly converge to zero.

Consider the sequence of functions

$$G_1, G_2, \ldots, G_n,$$

such that the G_n are continuous, twice-differentiable functions that uniformly converge to f in $(0,b)$. The G_n could be, for example, n-term polynomial or Fourier approximations to f. We may then write

$$S_N(\theta)-\sigma_N(\theta)=\int_a^b\Phi_N(\theta,t)[f(t)-G_m(t)]\,dt+\int_a^b\Phi_N(\theta,t)G_m(t)\,dt. \tag{A5}$$

If $\Phi_N(\theta,t)$ is uniformly bounded for all N,θ, and t, then the first integral in Eq. (A5) may be made arbitrarily small by taking m large enough. Then, by Eq. (A4), the second integral can be made arbitrarily small by taking N large enough. Hence, Eq. (A2) follows.

It remains to show that $\Phi_N(\theta,t)$ is, in fact, uniformly bounded. For simplicity, we begin with a specific example, namely that considered in Eq. (74). We take the χ_n to be $\cos[2n(\theta-\pi/2)]$. Since only functions even about $\pi/2$ are considered here, it suffices to examine the interval $(0,\pi/2)$. Consider one of the contributions to Φ_N, which may be written as

$$A\equiv\sum_{N_0}^{N}\cos[(2n+\tfrac{1}{2})\theta]\cos[(2n+\tfrac{1}{2})t]-\cos(2n\theta)\cos(2nt)$$

$$=\sum_{N_0}^{N}\tfrac{1}{2}\{\cos[(2n+\tfrac{1}{2})\psi]-\cos(2n\psi)\}+\tfrac{1}{2}[\cos(2n+\tfrac{1}{2})z-\cos(2nz)], \tag{A6}$$

where $\psi\equiv\theta+t$ and $z\equiv\theta-t$. It is assumed that N_0 is a sufficiently large eigenvalue, so that the asymptotic representation of the u_n as $\cos(2n+\frac{1}{2})\theta$ is valid. We note, for example, that

$$\sum_{N_0}^{N}\cos(2n+\tfrac{1}{2})\psi-\cos 2n\psi=\mathrm{Re}\left(\frac{e^{i\psi/2}-1}{e^{2N_0i\psi}-1}(e^{2Ni\psi}-1)\right), \tag{A7}$$

which is obviously uniformly bounded. Thus A is bounded.

The other contribution to $\Phi_N(\theta,t)$ arises from the decaying eigenfunctions in the interval $(0,b)$, and may be written as

$$B\equiv\sum_{N_0}^{N}u_{-n}(\theta)u_{-n}(t)\sim\sum_{N_0}^{N}\exp[-(2n+\tfrac{1}{2})(\theta+t)], \tag{A8}$$

which is uniformly bounded for θ in any *closed* interval in $(0,\pi/2)$ that does not include zero. Since $\Phi_N(\theta,t)=A+B+$ (a finite number of terms), it follows that $\Phi_N(\theta,t)$ is similarly uniformly bounded in any closed interval not including zero.

This property of $\Phi_N(\theta,t)$ implies, through Eq. (A5), that $S_N(\theta)$ uniformly converges to any continuous $f(\theta)$ in any closed interval in $(0,\pi/2)$ not including the origin. The exclusion of the origin [i.e., where $h(\theta)=0$] from the completeness proof is expected, as in other aspects of this problem. In a similar manner, it can be shown that $S_n(\theta)$ converges to $f(\theta)$ in the rest of the interval $(-\pi,\pi)$. Note that for periodicity boundary conditions, the endpoints, $\theta=\pm\pi$, must be excluded in the same manner and for the same reason that the point $\theta=0$ is excluded from the interval over which the completeness holds.

The general case in which h passes through zero at $\theta=0$ is handled similarly. From the asymptotic expansion of the u_n, given by Eqs. (60), it can be seen that the u_n asymptotically approach sinusoidal or decaying functions. The Airy function behavior occurs only near the origin, so that for any closed interval not including the origin, it is possible to begin the summation at a large enough N_0, as in Eqs. (A7) and (A8), that the asymptotic behavior is valid. From the example given above, it is clear that the Φ_N would be similarly bounded. For one may, for example, pick the χ_n such that

$$\frac{d}{d\theta}D\frac{d}{d\theta}\chi_n+k_n|h|\chi_n=0, \tag{A9}$$

with Sturm–Liouville-type boundary conditions at the endpoints of either interval, $(0,b)$ or $(a,0)$. The eigenfunctions χ_n would then be asymptotically sinusoidal, but shifted from the u_n, a case known from Eq. (A7) to have the requisite properties.

[1] M. Gevrey, J. Math. Sér. **6**, 9 (1913) and 10 (1914).
[2] W. Bothe, Zeits f. Phys. **54**, 161 (1929).
[3] A. Sen and N.J. Fisch, Massachusetts Institute of Technology MIT—PRR 78/16 (1978).

[4]H.A. Bethe, M.E. Rose, and L.P. Smith, Proc. Amer, Philos. Soc. **78**, 573 (1938).
[5]D. Stein and I.B. Bernstein, Phys. Fluids **19**, 811 (1976).
[6]R. Beals, J. Math. Anal. Appl. **58**, 32 (1977).
[7]R. Beals, J. Funct. Anal. (to be published).
[8]C.D. Pagani, Rend. Inst. Lombardo **103**, 618 (1969).
[9]W. Fleming, U.S. Army Math. Res. Center, Univ. of Wisc., M.R. C. Rep. (1963).
[10]M.S. Baovendi and P. Grisvard, J. Funct. Anal. **2**, 352 (1968).
[11]C.D. Pagani and G. Talenti, Ann. Mat. Pura Appl. **90**, 1 (1971).
[12]E.L. Ince, *Ordinary Differential Equations* (Dover, New York, 1956).
[13]A. Kneser, Mathematische Ann. **58**, 136 (1904).
[14]R. Beals, J. Math. Phys. (to be published).

Cotangent bundle approach to noninertial frames

B. DeFacio
Applied Mathematics Group, Ames Laboratory DOE, Iowa State University, Ames, Iowa 50011
Department of Physics, University of Missouri, Columbia, Missouri 65211 [a]

D.G. Retzloff
Department of Chemical Engineering, University of Missouri, Columbia, Missouri 65211

(Received 20 October 1978; accepted for publication 9 November 1979)

The most general possible noninertial acceleration in special relativity is formulated with differential forms in the cotangent bundle. We show that the Lie derivative plays the same role in the cotangent bundle that the covariant derivative plays in the tangent bundle. We also show that a cotangent bundle analog of Fermi–Walker transport can be based upon the, "cotangent-geodesic" equation, $\mathcal{L}_u \omega = 0$. This gives a generalization of the work by Kiehn on classical Hamiltonian mechanics to special relativity.

1. INTRODUCTION

Recently DeFacio, Dennis and Retzloff[1] used the presymmetry of Ekstein and others[2–5] to derive an exact, closed-form expression for noninertial acceleration and Fermi–Walker transport in special relativity. They used invariant methods, including the covariant derivative, to formulate the entire problem in the tangent bundle. We will show here that an alternate formulation exists in the cotangent bundle. In addition to giving an interesting use of some different modern geometrical techniques, there are other reasons for our study. One reason is that in many studies such as by Havas and Plebański,[6] a formal $\partial/\partial t$ operator is used, which they point out in general, is our Lie derivative $\mathcal{L}_u$. In this case, our work shows explicitly what terms are involved in general noninertial frames over manifolds which are trivial bundles. Another reason is that Estabrook and collaborators[7–11] and Corones *et al.*[12–14] showed that related methods lead to useful local relations, the prolongation structures, which are important in the understanding of classical solitary waves. Finally, a series of papers by Kiehn[15–17] have used differential forms and Lie derivatives to formulate Hamilton's equation of motion and periods on manifolds for a variety of nonrelativistic problems.

There has been other recent work on noninertial acceleration by Ni and Zimmerman[18] and Li and Ni[19,20] which extend the early work by Ni[21] which was discussed by Misner, Thorne, and Wheeler.[22] These authors[18–21] were studying a more general problem, specifically general noninertial observers in gravitational fields. Therefore some of their studies include general relativity effects, whereas ours do not. Nevertheless, they find it necessary to make some simplifying assumptions and through second order noninertial and gravitational effects decouple.[18] Thus, through second order our different studies are comparable and calculationally our methods are simpler (when they are applicable). However Li and Ni[20] have shown that at third order, nonzero coupling terms exist between the gravitational and inertial effects and the methods of DeFacio, Dennis, and Retzloff[1] and the present paper are not sufficiently general to explore such phenomenon. A method is being developed to study such coupling terms and it will be reported separately.[23]

Let (L, g) denote the usual Lorentz space–time. Then L is the manifold of four-dimensional space–time points and $g = \text{diag}(-1, -1, -1, 1)$ is the metric. We denote the tangent bundle as $T(L)$ and the cotangent bundle as $T^*(L)$. A general manifold may not admit a metric structure. Of course, if a Riemannian or pseudo-Riemannian metric exists for any space–time (M, g) of interest, it induces a duality relation between $T(M)$ and $T^*(M)$. Even in this case, fully invariant formulations may look quite different from one another. We wish to understand this complementarity better. Even more important are those cases in which no metric exists because now $T(M)$ and $T^*(M)$ are inequivalent. When $T(M)$ and $T^*(M)$ are inequivalent, it is an interesting open question as to which one is more useful and when. One set of examples are the symplectic structures[24] associated with Lagrangian theories. These symplectic structures are formulated in the cotangent bundle. Another set of examples is the relativistic string models which Tucker *et al.*[25] have shown to admit a natural formulation in the cotangent bundle largely because of the simplicity of the pull-back mapping there. Related but different studies by Tucker *et al.* have shown that Yang–Mills[26] and supersymmetric gauge field theories[27] can be expressed by these methods. Finally, bifurcation questions seem most easily answered in the cotangent bundle because they are associated with the vanishing of the normal (here Lie derivative) of an n-surface. The reader can find additional discussions beyond the references of this paper in the series of monographs by Herman.[28]

In order to use presymmetry in conjunction with the invariant methods of differential geometry it is necessary to lift the Lorentz manifold into its cotangent bundle. This is achieved by identifying each event of $P \in L$ with its form $\sigma_P \in T^*_P(L)$ as

$$\sigma_P = (\exp^*_P)^{-1} q. \tag{1.1}$$

For a complete form field the form $\phi^* \sigma$ transforms to σ under the action of the map ϕ and the relevant diagram

[a] Permanent address.

 0022-2488/80/040751-07$1.00

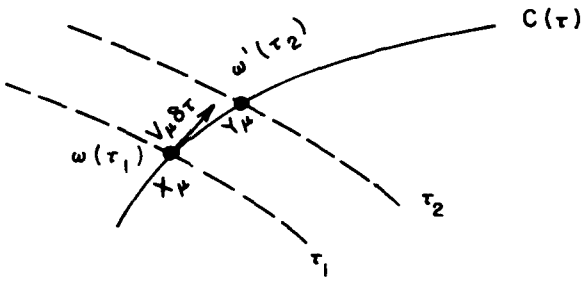


FIG. 1. Parallel transport of a 1-form ω, along the curve $C(\tau)$ in a manifold M.

$$\begin{array}{ccccc} & & \overset{\phi}{*} & & \\ & T^*(L) & \leftarrow & T^*(L) & \\ \exp^* & \downarrow & & \downarrow & \exp^*, \\ & L & \rightarrow & L & \\ & & \underset{\phi}{} & & \end{array} \tag{1.2}$$

commutes provided ϕ is a diffeomorphism. In this instance the effect of ϕ on the observation of an event is given by

$$\overset{*}{\phi}{}^{-1}\sigma = (\exp^*_{\phi(P)})^{-1}\phi\exp^*_P\sigma(P), \tag{1.3}$$

and guarantees the existence of a causality condition for $\phi \in A_G$, the presymmetry group.

It is obvious from (1.2) that for nonhomeomorphic maps only retrodictive analysis is deterministic. Predictive analysis does not exist in this case and physical descriptions based on the tangent bundle may not exist or will be difficult to formulate. The cotangent bundle description of the physics through Cartan's theory of differential forms is well defined for these situations. Dissipative systems are one realization of these ideas which has been thoroughly treated in the context of retrodictive analysis by Kiehn.[17]

The organization of this paper is as follows. In Sec 2 we will show that the action of the Lie derivative on 1-forms in the cotangent bundle generates a cotangent bundle parallel transport. These results do not seem to be available in the literature. Then the cotangent bundle version of Newton's second law will follow directly. In Sec. 3 we will use these methods to give a geometrical realization of the acceleration group and, therefore, the presymmetry. The connection coefficients $\Gamma^{\alpha}_{\beta\gamma}$ which are needed for the Cartan matrix are given in Sec. 3 as part of that analysis. In Sec. 4 we use this formalism to give the most general Fermi–Walker transport in special relativity and in Sec. 5 our conclusions are given.

2. THE LIE DERIVATIVE AND PARALLEL TRANSPORT

It is well known[22] that the motion of a freely falling particle in an inertial frame is given by the parallel transport of the local tetrad frame of the particle. The geodesic equations express this transport irrespective of whether the physical motion is described in the tangent or cotangent manifold. Therefore in order to prove that the Lie derivative acting upon some particular form describes the freely falling particle it is necessary to show that:

(i) The actions of the Lie derivative on the form of interest parallel transports that form; and

(ii) The resulting equations describing that transport when expressed upon a coordinate chart are geodesic equations.

Two theorems will be given next which establish these two propositions. Since we only need 1-forms to describe the freely falling motion, the theorems and proofs will be restricted to these cases.

Theorem 1: If ω is a 1-form on a reductive homogeneous manifold which contains a curve C parameterized by a proper time τ as shown in Fig. 1, with a tangent vector $\mathbf{v}$ which is a Killing vector, i.e., $\mathbf{v}$ is an element of the invariant principal fiber bundle of linear frames, then

$$\mathscr{L}_{\mathbf{v}}\omega = 0$$

describes the parallel transport of ω along C in the canonical linear connection.

Proof: As shown in Fig. 1 the curve C, parametrized by τ and ω, is an arbitrary 1-form defined on C. Let $\omega(\tau_1)$ denote the value of ω at $\mathbf{x}$ on C, and let $\omega'(\tau_2) = \exp(\Delta\tau)^*\omega(\tau)$, where $\Delta\tau = \tau_2 - \tau_1$, be the value of ω at $\mathbf{y}$ and C where

$$y^\mu = x^\mu + u^\mu, \quad u^\mu = v^\mu = a\tau, \quad \text{and} \quad v^\mu = \frac{dx^\mu}{d\tau}. \tag{2.1}$$

Parallel transport of ω along $C(\tau)$ is defined by

$$\lim_{\Delta\tau\to 0}\left(\frac{\exp(\Delta\tau v)^*\omega(\tau) - \omega(\tau_1)}{\Delta\tau}\right) \equiv \mathscr{L}_{\mathbf{v}}\omega = 0. \tag{2.2}$$

Consider the representations of $\omega'(\tau_2)$ and $\omega(\tau_1)$ on the coordinate atlases (u, x^μ) and (u, y^μ) as shown in Fig. 1. This gives

$$\omega(\tau_1) = A_\mu(x^\nu)dx^\mu, \quad \omega'(\tau_2) = A'_\mu(y^\nu)dy^\nu. \tag{2.3}$$

If A'_μ's are smooth enough and y^ν is near x^μ we can expand $A'_\mu(y^\nu)$ in a Taylor's series about x^μ as

$$A'(y^\nu) = A_\mu(x^\nu) + \left(\frac{\partial A_\mu}{\partial x^\nu}\right)u^\nu + \cdots. \tag{2.4}$$

But Eq. (2.1) implies that

$$\frac{\partial u^\mu}{\partial x^\nu} = \left(\frac{\partial v^\mu}{\partial x^\nu}\right)\delta\tau \quad \text{and} \quad du^\mu = \left(\frac{\partial u^\mu}{\partial x^\nu}\right)dx^\nu. \tag{2.5}$$

Also we can rewrite Eq. (2.2) using this to obtain

$$\begin{aligned} \lim_{\delta\tau\to 0}\Bigg\{\frac{1}{\delta\tau}\Bigg[&\left(\frac{\partial A_\mu(x^\nu)}{\partial x^\nu}\right)v^\nu\delta\tau\, dx^\mu \\ &+ A_\mu(x^\nu)\left(\frac{\partial v^\mu}{\partial x^\nu}\right)\delta\tau\, dx^\nu + O[(\delta\tau)^2]\Bigg]\Bigg\} = 0 \\ &= i(\mathbf{v})d\omega + di(\mathbf{v})\omega \\ &= [i(\mathbf{v})d + di(\mathbf{v})]\omega \\ &= \mathscr{L}_{\mathbf{v}}\omega = 0. \end{aligned} \tag{2.6}$$

Q.E.D.

Theorem 2: If $\{\sigma_i \mid i = 1,2,3,4\}$ is a set of basis forms in the cotangent bundle and $\vec{e}_0$ is the tangent vector to the curve, then the inertial frame equations of motion can be written as

$$\mathscr{L}_{\vec{e}_0}\sigma_i = 0. \tag{2.7}$$

Proof: In Fig. 2 a four-dimensional $\vec{e}_0$ is shown. Consider a particle which is constrained to move on a surface N, and let P be a point on the curve traced out by the motion of the

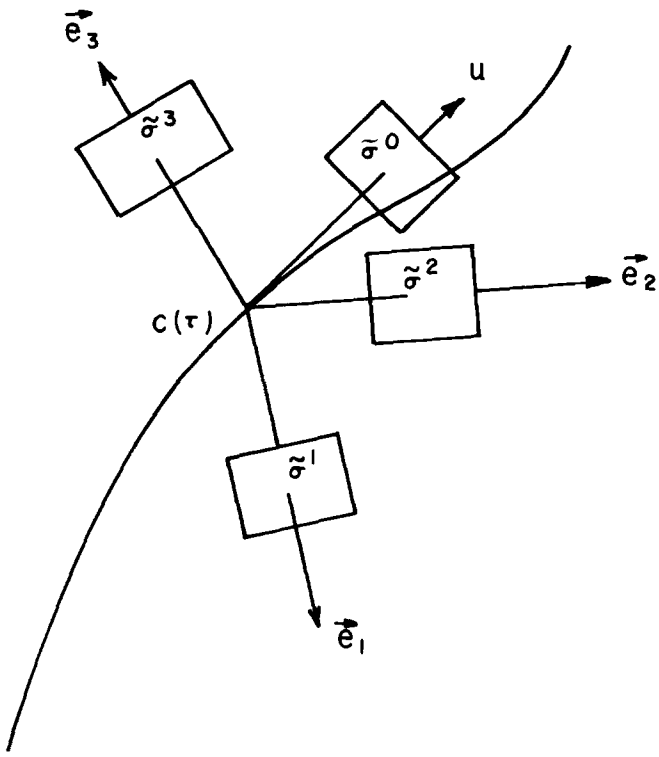


FIG. 2. The world line of accelerating observer showing the vector tetrad $\{\mathbf{u},\vec{\mathbf{e}}_i\}$ and its corresponding dual $\{\tilde{\sigma},\tilde{\sigma}^i\}$.

particle. In standard geometry notation, see Flanders,[29] Cartan,[30] O'Neill,[31] Herman,[28] or Misner, Thorne, and Wheeler,[22] one has

$$dP = \sum_i \sigma^i \vec{e}_i, \quad d\vec{e}_i = \sum_j \omega_{ij}\vec{e}_j,$$
$$d\sigma^i = \sum_j \sigma^j \omega_{ij}, \quad \omega_{ij} = \sum_k \Gamma^{\ i}_{jk}\sigma^k. \tag{2.8}$$

Using the parts of Eq. (2.8) to write out Eq. (2.7) yields

$$\mathcal{L}_{\vec{e}_0\sigma^i} = i(\vec{e}_0)d\sigma^i + di(\vec{e}_0)\sigma^i = 0$$
$$= \left[i(\vec{e}_0)\sum_j\sum_k \Gamma^{\ i}_{jk}\sigma^j\sigma^k + di(\vec{e}_0)\sigma^i\right]. \tag{2.9}$$

But on a coordinate atlas (u,x^i),

$$\sigma^i = \frac{dx^i}{d\lambda}d\lambda,$$

where λ is the arc length function which parameterizes the curve C, so Eq. (2.9) becomes

$$\left[\left(\frac{dx^j}{d\lambda}\right)\left(\frac{dx^k}{d\lambda}\right)\Gamma^{\ i}_{jk}d\lambda + d\left(\frac{dx^i}{d\lambda}\right)\right]$$
$$= 0$$
$$= \left[\left(\frac{d^2x^i}{d\lambda^2}\right) + \Gamma^{\ i}_{jk}\left(\frac{dx^j}{d\lambda}\right)\left(\frac{dx^k}{d\lambda}\right)\right]d\lambda. \tag{2.10}$$

Therefore, for $d\lambda \neq 0$, we must have

$$\left[\left(\frac{d^2x^i}{d\lambda^2}\right) + \Gamma^{\ i}_{jk}\left(\frac{dx^j}{d\lambda}\right)\left(\frac{dx^k}{d\lambda}\right)\right] = 0. \tag{2.11}$$

Q.E.D.

This allows a Newton–Cartan theory of gravity[15,22,28,31] to be cast in terms of

$$\mathcal{L}_{\vec{e}_0}\sigma^i = 0, \tag{2.12}$$

providing

$$\Gamma^{\ j}_{00} = \frac{\partial\phi}{\partial x_j} \quad \text{and} \quad \Gamma^{\ \alpha}_{\beta\gamma} = 0 \quad (\forall\,\beta,\alpha\neq 0). \tag{2.13}$$

Q.E.D.

3. INVARIANT METHODS AND THE ACCELERATION GROUP

In this section modern invariant methods, especially differential forms, will be used to study general noninertial frames. This will represent a generalization of recent work by Kiehn[15–17] who has developed a (nonrelativistic) Hamiltonian theory of dissipative systems. We will use Kiehn's notation and will work in the cotangent bundle as he did. A geometrical interpretation and physical construction of N-particles in noninertial frames will be given. The exact expression for general noninertial acceleration will be postponed until Sec. 4.

Let $C: L \to R^1$ be the trajectory curve of an accelerated observer in a Lorentz manifold L. At each proper time t, an observer at $C(t)$ has two local laboratory frames, one the tangent manifold and the other is the cotangent manifold. In the tangent manifold $T(L)$, the local frame consists of three orthonormal unit spatial vectors and a unit timelike vector which is recorded by a standard clock at $C(\tau)$. In the cotangent manifold $T^*(L)$, the local frame consists of three orthonormal unit spatial forms and a unit timelike form which is also measured by a standard clock at $C(\tau)$. This yields two vector spaces $T_{C(\tau)}(L)$ and $T^*_{C(\tau)}(L)$ which we call the "laboratory frame" of an accelerating observer. These laboratory frames are distinct from the manifold, because following Ekstein *et al.*[2–5,1] we distinguish between the "space–time point" and the clock and ruler "measurement of the space–time point" in particle mechanics. At each instantaneous proper time in special relativity, a pair of isometric isomorphisms

$$L \to T_{C(\tau)}(L), \quad L \to T^*_{C(\tau)}(L),$$

exist by the postulates of special relativity.[1]

The timelike velocity vector $\mathbf{u}\in T_{C(\tau)}$ is tangent to $C(\tau)$ and is given by

$$C_*\left(\frac{d}{d\tau}\right) \equiv \mathbf{u} = \langle dC(\tau), \mathbf{e}_0\rangle, \tag{3.1a}$$

where

$$\langle \mathbf{u},\mathbf{u}\rangle = -1. \tag{3.1b}$$

The accelerated observer carries a triad of spatial vectors $\vec{e}_i \in T_{C(\tau)}(L)$ for each $i = 1,2,3$. These vectors satisfy

$$\langle \vec{e}_i(\tau), \vec{e}_j(\tau)\rangle = \delta_{ij}, \quad \langle \mathbf{u}(\tau), \vec{e}_i(\tau)\rangle = 0. \tag{3.2}$$

At each instant t, the set of $\{\vec{e}_i \mid i = 1,2,3\}$'s span a three-dimensional vector space $S_{C(\tau)}$ which is orthogonal to $\mathbf{u}$ as shown in Fig. 2. Physically $S_{C(\tau)}$ is a spatial rest frame of the observer at $C(\tau)$. The set of $\vec{e}_i$'s are unique modulo a spatial SO(3) rotation which reflects the Galilian principle of relativity.[1,6] Overarrows denote vectors which lie entirely in $S_{C(\tau)}$.

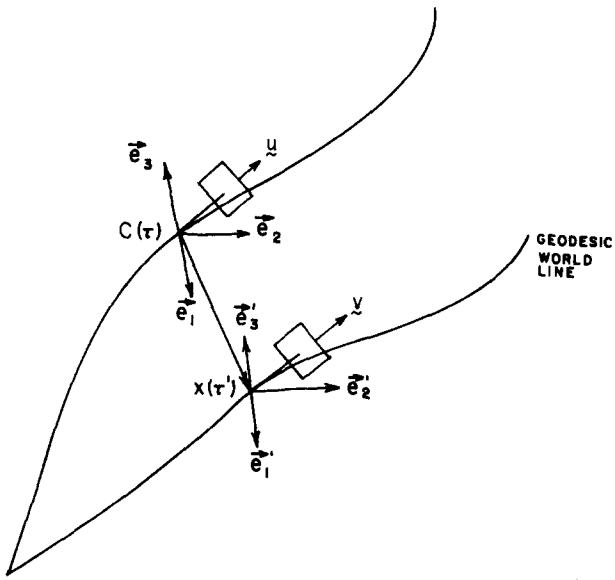


FIG. 3. The geometric concept of null simultaneity in relationship to the accelerating observer.

The 1-forms $\{\sigma^i | i = 1,2,3\}$ form an orthogonal basis for the cotangent bundle, which is dual to the orthonormal tetrad of vectors $\{\mathbf{u},\vec{e}_i | i = 1,2,3\}$ which form a basis for the tangent bundle. Both sets are shown in Fig. 2. The 1-forms are defined such that $\forall\, i,j = 1,2,3$ one has

$$\langle \sigma^0(\tau),\vec{e}_i(\tau)\rangle = 0, \quad \langle \sigma^0(\tau),\mathbf{u}(\tau)\rangle = 1,$$
$$\langle \hat{\sigma}^i(\tau),\vec{e}_j(\tau)\rangle = \delta_{ij}, \quad \langle \hat{\sigma}^i(\tau),\mathbf{u}(\tau)\rangle = 0. \tag{3.3}$$

At each instant τ, the set $\{\hat{\sigma}^i(\tau) | i = 1,2,3\}$ spans a three-dimensional vector space $S^*_{C(\tau)}$ which is orthogonal to $\hat{\sigma}^0$ as shown in Fig. 2. Overcarats denote forms which lie entirely in $S^*_{C(\tau)}$. The forms $\hat{\sigma}^i$ are therefore spacelike forms.

The tetrad

$$[\mathbf{e}] = \begin{pmatrix} \mathbf{u} & 0 & 0 & 0 \\ 0 & \vec{e}_1 & 0 & 0 \\ 0 & 0 & \vec{e}_2 & 0 \\ 0 & 0 & 0 & \vec{e}_3 \end{pmatrix}, \tag{3.4}$$

is a frame field. When the associated frame matrix as a matrix of functions is of maximal rank, it can be shown that the derivative of the field frame is closed even if the associated frame matrix does not posses a global inverse. This closure property for an immersive map can be written as

$$d[\mathbf{e}] = [\mathbf{e}]\mathscr{C}_R, \tag{3.5}$$

where $\mathscr{C}_R$ is the right Cartan matrix of differential forms.

Proposition: The right Cartan matrix $\mathscr{C}_R$ is antisymmetric.

Proof: Using Eqs. (3.1b) and (3.4) yields

$$[\mathbf{e}]^\dagger\cdot[\mathbf{e}] = \begin{pmatrix} -1 & 0 \\ 0 & \mathbb{1} \end{pmatrix} \equiv \mathscr{I}, \tag{3.6}$$

where $\mathbb{1}$ is a 3×3 unit matrix. Using the product of $[\mathbf{e}]^\dagger$ and Eq. (3.5) gives

$$[\mathbf{e}]^\dagger d[\mathbf{e}] = [\mathbf{e}]^\dagger\cdot[\mathbf{e}]\mathscr{C}_R = \mathscr{I}\cdot\mathscr{C}_R. \tag{3.7}$$

Since Eq. (3.6) clearly implies that $\mathscr{I}^2 =$ four-dimensional unit matrix, further multiply Eq. (3.7) by $\mathscr{I}$ to obtain

$$\mathscr{I}^2\cdot\mathscr{C}_R = \mathscr{C}_R = \mathscr{I}\cdot[\mathbf{e}]^\dagger\cdot d[\mathbf{e}], \tag{3.8}$$

and take the adjoint of Eq. (3.8) to get

$$\mathscr{C}^\dagger_R = d[\mathbf{e}]^\dagger\cdot[\mathbf{e}]\cdot\mathscr{I}. \tag{3.9}$$

But

$$d\{[\mathbf{e}]^\dagger\cdot[\mathbf{e}]\} = d[\mathbf{e}]^\dagger\cdot[\mathbf{e}] + [\mathbf{e}]^\dagger d[\mathbf{e}] = 0, \tag{3.10a}$$

by the properties of the d-operator. This gives that

$$d[\mathbf{e}]^\dagger\cdot[\mathbf{e}] = -[\mathbf{e}]d[\mathbf{e}]^\dagger. \tag{3.10b}$$

Putting Eq. (3.10b) into Eq. (3.9) completes the proof because

$$\mathscr{C}^\dagger_R = d[\mathbf{e}]^\dagger\cdot[\mathbf{e}]\cdot\mathscr{I} = -\mathscr{I}\cdot[\mathbf{e}]^\dagger\cdot d[\mathbf{e}] = -\mathscr{C}_R. \tag{3.11}$$

Q.E.D.

The trajectory of an accelerated observer $C(t)$ can be written as

$$dC(\tau) = \mathbf{u}\sigma^0 + \vec{e}_i\hat{\sigma}^i = [\mathbf{e}]\cdot[\boldsymbol{\sigma}]. \tag{3.12}$$

This is to use the $d[d(\cdot)] = 0$ property to obtain

$$0 = d\{dC(\tau)\} = d[\mathbf{e}]\cdot[\boldsymbol{\sigma}] + [\mathbf{e}]\cdot d[\boldsymbol{\sigma}]$$
$$= [\mathbf{e}]'\{d[\boldsymbol{\sigma}] + \mathscr{C}_R\cdot[\boldsymbol{\sigma}]\}. \tag{3.13}$$

This implies that

$$d[\boldsymbol{\sigma}] = -\mathscr{C}_R\cdot[\boldsymbol{\sigma}]. \tag{3.14}$$

The components of $\mathscr{C}_R$ are given by

$$(\mathscr{C}_R)^\mu_\nu = \Gamma^\mu{}_{\nu\lambda}\sigma^\lambda, \tag{3.15}$$

where the values of $\Gamma^\mu{}_{\nu\lambda}$ are tabulated in Table I.

The timelike velocity form $\hat{u}\in T^*_{C(\tau)}$ corresponding to the orthogonal tetral $\{\sigma^0,\hat{\sigma}^i | i = 1,2,3\}$ is defined as

$$C^*_*\left(\frac{d}{d\tau}\right) \equiv \hat{u} \equiv \mathscr{L}_\mathbf{u} C(\tau) = \sigma^0. \tag{3.16}$$

This tetrad propagates along the 0-form $C(t)$ according to the equations

$$\mathscr{L}_\mathbf{u}\hat{\sigma}^i \equiv i(\mathbf{u})d\sigma^i + di(\mathbf{u})\hat{\sigma}^i = -\Omega^i{}_j\sigma^j, \tag{3.17a}$$

and

$$\mathscr{L}_\mathbf{u}\hat{\sigma}^0 \equiv i(\mathbf{u})d\sigma^0 + di(\mathbf{u})\sigma^0 = -\Omega^0{}_j\sigma^j, \tag{3.17b}$$

where

$$\Omega^\mu{}_\nu = -\Gamma^\mu{}_{\nu 0} = a^\mu u_\nu - a^\nu u_\mu + u_\alpha\omega_\beta\epsilon^{\alpha\beta\mu}{}_\nu. \tag{3.17c}$$

The invariant acceleration $\hat{a}$ of $C(\tau)$ is

$$\hat{a} = \mathscr{L}_\mathbf{u}\cdot\mathscr{L}_\mathbf{u} C(\tau). \tag{3.18}$$

The angular momentum form of $C(\tau)$ is denoted as $\hat{\omega}$ and both $\hat{a}$ and $\hat{\omega}$ are contained in $S^*_{C(t)}$. We can prove that $\hat{a}\in T^*_{C(\tau)}$ by observing that

$$0 = \mathscr{L}_u\langle \mathbf{u},\mathbf{u}\rangle = 2\langle \mathbf{u}, \mathscr{L}_u u\rangle = 2\langle \mathbf{u},\hat{a}\rangle. \tag{3.19}$$

The propagation equations in Eq. (3.19) preserve the orthogonality relations in Eq. (3.2) and (3.3) because Ω is antisymmetric. Acting on either basis set $\{\sigma^0,\hat{\sigma}^i\}$ or $\{\mathbf{u},\vec{e}_i\}$ Ω generates an active infinitesmal Lorentz transformation.[1] Expanding Eq. (3.17) gives

$$\mathscr{L}_\mathbf{u}\hat{\sigma}^i = a_i\sigma^i - \epsilon^j{}_{ik}\omega_j\sigma^k, \tag{3.20}$$

TABLE I. The Cartan matrix coefficients for the proper reference frame of an accelerating observer.

$\Gamma^{\alpha}{}_{\beta\eta}$

$\alpha=0$					$\eta=0$				
$\beta \backslash \eta$	0	1	2	3	$\beta \backslash \alpha$	0	1	2	3
0	0	0	0	0	0	0	0	0	0
1	a^1	0	0	0	1	0	0	ω^3	$-\omega^2$
2	a^2	0	0	0	2	0	$-\omega^3$	0	ω^1
3	a^3	0	0	0	3	0	ω^2	$-\omega^1$	0

All coefficients not tabulated are zero.

and

$$\mathscr{L}_{\mathbf{u}}\sigma^0 = \hat{a}. \tag{3.21}$$

In Eq. (3.20) the first term on the right-hand side is required to preserve the relation

$$\langle \hat{u},\sigma^i\rangle = \langle \sigma^0,\sigma^i\rangle = 0,$$

and the second term represents a spatial rotation with angular momentum ω in the rest frame of $C(t)$.

The same problem with time ordering which we found in the tangent bundle[1] occurs in the cotangent bundle. A cotangent bundle version of *null simultaneity* is needed here.

Definition 1: A point $X\in L$ is called *spatially simultaneous* with $C(\tau)$, written as $x\#C(\tau)$, if there exists some $r\in S^*_{C(t)}$ for which

$$X = C(\tau) + r. \tag{3.22}$$

Definition 1: A point $X\in L$ is called *spatially simultaneous* with $C(\tau)$, written as $X\#C(\tau)$, if there exists some $r\in S^*_{C(t)}$ for which

$$S_0[C(\tau),L] = \{X\in L,\ X\#Z(\tau_0)\}. \tag{3.23}$$

Definition 3: A point $X\in L$ is called *null simultaneous* with $C(\tau)$, written as $X\,0\#0C(t)$, iff

$$X = C(\tau) + y, \tag{3.24}$$

where

$$\langle y,\hat{\sigma}^0\rangle = \langle y,\hat{u}\rangle < 0.$$

The geometry of null simultaneity is depicted in Fig. 3. As argued in Ref. 1 the decomposition in Eq. (3.24) is unique whereas Eq. (3.22) is nonunique. It is necessary to use the presymmetry structure in an essential way because the preservation of equivalence classes of both observation procedures and state preparation procedures with the same expectation values is the basis of the argument.[1]

4. COTANGENT FORMULATION OF FERMI-WALKER TRANSPORT

To obtain the equation of motion of a free particle relative to a noninertial frame, let $X(\tau')$ be the world line of a free particle in an inertial frame as shown in Fig. 3. The velocity form $\hat{v}$ of this particle is

$$\hat{v} = \mathscr{L}_{\vec{e}_0'}X(\tau') = \mathscr{L}_{\mathbf{v}}X(\tau'). \tag{4.1}$$

Suppose τ is the proper time in some noninertial frame which corresponds to τ' in the inertial frame, so we can use Definition 3 to write

$$X[\tau(\tau')] = C(\tau) + r(\tau). \tag{4.2}$$

Therefore,

$$\begin{aligned}\hat{v} = \mathscr{L}_{\mathbf{v}}X(\tau') &= \left(\frac{d\tau'}{d\tau}\mathscr{L}_{\mathbf{u}}[C(\tau) + r(\tau)]\right)\\ &= \Gamma\mathscr{L}_{\mathbf{u}}[C(\tau) + r(\tau)]\\ &= \Gamma[\hat{u} + \mathscr{L}_{\mathbf{u}}r(\tau)],\end{aligned} \tag{4.3}$$

TABLE II. The interpretation of the most general noninertial acceleration $\ddot{r}_i$. Special relativity corrections are to all orders since Eq. (4.14) is exact.

Effect	Term in $\ddot{r}$
1. "Gravitational" Doppler red shift correction to lowest order	$(a_i r_i/c^2)/a_j$
2. "Gravitational" Doppler red shift correction to all orders	$\hat{R} \equiv (1 + a^i r_i/c^2)^{-1}$
3. Gravitational magnetic" correction	$\epsilon_i{}^{jk}\omega_j r_k a_i \dot{r}_i/c^2$
4. "Gravitational electric" correction	$\dot{r}_i a_i \dot{r}_i/c^2$
5. Inertial acceleration with all orders gravitational red shift	$-a_i(1 + a^i r_i/c^2)^{-1}$
6. Corolis acceleration	$2\epsilon_i{}^{jk}\omega_j \dot{r}_k$
7. Centripetal acceleration	$-\epsilon_i{}^{jk}\omega_j \epsilon_k{}^{lm}\omega_l r_m$
8. Rotational coordinate acceleration due to $-\epsilon_i{}^{jk}\dot{\omega}_j r_k$	$\dot{\omega}_j \neq 0$
9. "Gravitational magnetic" correction with all orders of gravitational Doppler shift correction	$(\epsilon_i{}^{jk}\omega_j r_k a_i \dot{r}_i/c^2)\hat{R}$
10. "Gravitational electric" correction, with all orders of gravitational Doppler shift corrections	$(\dot{r}_i a_i r_i/c^2)\hat{R}$
11. Translational coordinate acceleration due to $\dot{a} \neq 0$	$\dot{a}_i r_i/c^2$
12. Translational coordinate acceleration due to $\dot{a} \neq 0$, with all orders of gravitational Doppler shift correction	$(a_i r_i/c^2)\hat{R}$

where $\Gamma = (d\tau'/d\tau)$. Because $X(\tau')$ is the world line of a free particle in an inertial frame, its acceleration vanishes, i.e.,

$$\mathcal{L}_v \hat{v} = 0. \tag{4.4}$$

Theorem 2 is adequate to show that Eq. (4.4) is a fully invariant cotangent bundle geodesic equation. From arguments parallel to Ref. 1 for $\hat{r} \in S^*_{C(\tau)}$

$$\mathcal{L}_u \hat{r} = \mathcal{L}_u r_i \sigma^i = \dot{r}\sigma^i - r_i (\mathcal{C}_R)^i_\mu \hat{\sigma}^\mu, \tag{4.5}$$

which implies that $\dot{\hat{r}} \in S^*_{C(\tau)}$. Using Eq. (4.5) in Eq. (4.3) gives

$$\hat{v} = \Gamma(\tau)[(1 + a^i r_i)\sigma^0 + \dot{r}_i \sigma^i + \epsilon_i{}^{jk}\omega_j r_k \hat{\sigma}^i]. \tag{4.6}$$

The fact that $\hat{v}$ is a function of t, or $\hat{u}$, only implies that

$$\mathcal{L}_{\dot{\vec{r}} + \vec{\omega}\times\vec{r}} \hat{v} = 0. \tag{4.7}$$

In terms of the convenient variable

$$\vec{w} = \Gamma(t)\{(1 + \vec{a}\cdot\vec{r})\mathbf{u} + \dot{\vec{r}} + \vec{\omega}\times\vec{r}\}, \tag{4.8}$$

Eq. (4.4) can be rewritten as

$$0 = \mathcal{L}_v \hat{v} = \mathcal{L}_{\Gamma\vec{w}} \hat{v} = \Gamma(1 + \vec{a}\cdot\vec{r})\mathcal{L}_u \hat{v}. \tag{4.9}$$

From this we obtain

$$\begin{aligned} 0 = \mathcal{L}_v \hat{v} = \dot{\Gamma}[(1 + a^i r_i)\sigma^0 + (\dot{r}_i + \epsilon_i{}^{jk}\omega_j r_k)\hat{\sigma}^i] \\ + \Gamma\{(a^i \dot{r}_i + a^i \dot{r}_i)\sigma^0 \\ + [\ddot{r}_i + \epsilon_i{}^{jk}(\dot{\omega}_j r_k + \omega_j \dot{r}_k)]\hat{\sigma}^i\} \\ + \Gamma[(1 + a^i r_i)a + (a^i \dot{r}_i + \epsilon_i{}^{jk} a^i \omega_j r_k)\sigma^0 \\ + \epsilon_i{}^{jk}(\omega^j \dot{r}_k + \epsilon_k{}^{lm}\omega_j \omega_l r_m)\hat{\sigma}^i]. \end{aligned} \tag{4.10}$$

Separating Eq. (4.10) into components parallel to σ^0

$$\dot{\Gamma}(1 + a^i r_i) + \Gamma(\dot{a}^i r_i + a^i \dot{r} + \epsilon_i{}^{jk} a^i \omega_j r_k) = 0, \tag{4.11}$$

and components perpendicular to σ^0

$$\begin{aligned} &\dot{\Gamma}(\dot{r}_i + \epsilon_i{}^{jk}\omega_j r_k) + \Gamma(\ddot{r}_i + \epsilon_i{}^{jk}\omega_j r_k + 2\epsilon_i{}^{jk}\omega_j \dot{r}_k) \\ &+ \Gamma(\dddot{r}_i + \epsilon_i{}^{jk}\omega_j r_k + 2\epsilon_i{}^{jk}\omega_j \dot{r}_k + \epsilon_i{}^{jk}\omega_j \epsilon_k{}^{lm}\omega_l r_m) \end{aligned} \tag{4.12}$$

gives

$$\dot{\Gamma} = -\Gamma[\dot{a}_i r_i + 2a^i \dot{r}_i + \epsilon_i{}^{jk} a^i \omega_j r_k](1 + a^i r_i)^{-1}. \tag{4.13}$$

Substitution of Eq. (4.13) into Eq. (4.11) allows one to solve for the noninertial acceleration

$$\begin{aligned} \ddot{r}_i = &-a_i(1 + a_i r_i/c^2) - 2\epsilon_i{}^{jk}\omega_j \dot{r}_k - \epsilon_i{}^{jk}\dot{\omega}_j r_k \\ &+ \epsilon_i{}^{jk}\omega_j \epsilon_k{}^{lm}\omega_l r_m + \frac{(r_i + \epsilon_i{}^{jk}\omega_j r_k)}{(1 + a^i r_i/c^2)} \frac{\dot{a}^i r_i}{c^2} \\ &+ \frac{2a^i r_i}{c^2} + \frac{\epsilon_i{}^{jk} a^i \omega_j r_k}{c^2}, \end{aligned} \tag{4.14}$$

where factors of c, the speed of light in vacuuo, have been reinstated. This is in complete agreement with Ref. 1 and Li and Ni, and agrees with Ni and Zimmerman in the correct limit as shown in Ref. 1b. The special cases then follow now in the cotangent bundle. In Table II, an interpretation of each term of Eq. (4.14) is given. A Newtonian physicist doing very accurate mechanics experiments involving relativistic particles on earth would "discover" these additional forces with these interpretations. We remark that Table II, unlike related tables in Refs. 18–22, has the special relativity corrections or "inertial corrections" to all orders.

This completes the discussion of the most general noninertial acceleration in the cotangent bundle; or equivalently shows that a cotangent bundle Fermi–Walker transport can be defined which is simply related to the usual tangent bundle Fermi–Walker transport. These two ways of looking at parallel transport are dual to one another when a metric exists.

5. CONCLUSIONS

We have used differential forms together with cotangent bundle techniques and implicitly presymmetry to formulate a cotangent bundle Fermi–Walker transport. The most general possible noninertial acceleration is in exact agreement with an earlier tangent bundle study.

The Lie derivative was shown to play the same role in $T^*(L)$ as the covariant derivative played in $T(L)$. Interestingly it was invariance together with physics [the particle trajectory $X(\tau')$ was free in an inertial frame] which gave the key result; the *cotangent geodesic equation*

$$\mathcal{L}_v \hat{v} = 0.$$

providing $\mathbf{v}$ is a suitable Killing vector. However, this is less "natural" than the covariant derivative ∇ since the choice of $\mathbf{v}$ must furnish the algebraic connection structure which ∇ carries.

Finally, if the metric were not defined then it might be, with 50% probability, that only the cotangent bundle parallel transport is well defined. In this case, the value of this formulation is obvious.

Even though initial and intermediate steps looked very different from Ref. 1, the latter stages became idential. Thus, we have shown that another set of mathematical techniques, certain Lie derivatives and cotangent bundles, are well suited to the study of noninertial acceleration.

ACKNOWLEDGMENTS

This work was supported in part by the Ames Laboratory DOE at Iowa State University in Ames, Iowa 50011 where BDF was a summer visitor during the summer of 1978. The hospitality of that Laboratory is gratefully acknowledged. BDF thanks James Corones and C.L. Hammer of Iowa State University and J. Ely Shrauner of Washington University, St. Louis for useful conversations. DGR thanks R.M. Kiehn of the University of Houston and both authors thank John K. Beem of the University of Missouri and David Lerner of the University of Kansas Mathematics Departments for their useful comments.

[1] B. Defacio, P.W. Dennis, and D.G. Retzloff, Phys. Rev. D **18**, 2813 (1978); Phys. Rev. D **20**, 570 (1979).
[2] H. Ekstein, Phys. Rev. **184**, 1315 (1969); Phys. Rev. D **1**, 1851 (E) (1970).
[3] H.P.W. Gottlieb, J. Phys. A: Math. Gen. **9**, 417 (1976).
[4] Y. Avashai, H. Ekstein, and J.E. Moyal, J. Math. Phys. **13**, 1139 (1972).
[5] Y. Avashai and H. Ekstein, Commun. Math. Phys. **37**, 193 (1974); Y. Avashai and H. Ekstein, Phys. Rev. D **7**, 983 (1973).
[6] P. Havas and J. Plebański, J. Math. Phys. **19**, 482 (1978); see, e.g., Ref. 22.
[7] F.B. Estabrook, in *Bäcklund Transformations, the Inverse Scattering Method, Solitons and Their Applications*, edited by R.M. Miura, Springer-Verlag Lecture Notes in Mathematics (Springer, New York, 1976), Vol. 515.
[8] B.K. Harrison and F.B. Estabrook, J. Math. Phys. **12**, 653 (1971).
[9] H.D. Wahlquist and F.B. Estabrook, J. Math. Phys. **16**, 1 (1975).
[10] F.B. Estabrook, Phys. Rev. D **8**, 2740 (1973).

[11]F.B. Estabrook and H.D. Wahlquist, SIAM Rev. **17**, 201 (1975).
[12]J. Corones, J. Math. Phys. **17**, 756 (1976).
[13]J. Corones, J. Math. Phys. **18**, 163 (1977).
[14]J. Corones, B.L. Markovski, and V.A. Rizov, J. Math. Phys. **18**, 2207 (1977).
[15]R.M. Kiehn, J. Math. Phys. **15**, 9 (1974).
[16]R.M. Kiehn, J. Math. Phys. **18**, 614 (1977).
[17]R.M. Kiehn, Int. J. Eng. Sci. **14**, 749 (1976); **13**, 941 (1975).
[18]W.-T. Ni and M. Zimmerman, Phys. Rev. D **17**, 1473 (1978).
[19]W.-Q. Li and W.-T. Ni, J. Math. Phys. **20**, 1473 (1979); **20**, 1925 (1979).
[20]F.B. Estabrook and H.G. Wahlquist, J. Math. Phys. **5**, 1629 (1964).
[21]W.-T. Ni, Astrophys. J. **176**, 769 (1972); Chinese J. Phys. **15**, 51 (1971).
[22]C.W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).
[23]D.G. Retzloff, B. DeFacio, and P.W. Dennis, (in preparation, J. Math. Phys.).
[24]A. Weinstein, *Lectures on Symplectic Manifolds* (Am. Math. Soc., Providence, Rhode Island, 1976).
[25]R.W. Tucker, Rutherford Laboratory Report (February 1976, unpublished) and P.A. Collins, J.F. Hopkinson, and R.W. Tucker, Nucl. Phys. B **103**, 67 (1976).
[26]R.W. Tucker and P.S. Howe (unpublished).
[27]P.S. Howe and R.W. Tucker, TH-2524-CERN Preprint (June 1978).
[28]R. Herman, *Interdisciplinary Mathematics* (Math. Sci., Brookline, Massachusetts, 1974–1979), Vols. I–XIX.
[29]H. Flanders, *Differential Forms* (Academic, New York, 1963).
[30]E. Cartan, *Systems differentiels exteriers et leurs applications geometriques* (Act. Sci. et Ind. 994, Paris, 1945).
[31]B. O'Neil, *Elementary Differential Geometry* (Academic, New York, 1966).

Inverse problem for the reduced wave equation with fixed incident field

V. H. Weston
Department of Mathematics, Purdue University, West Lafayette, Indiana 47907

(Received 27 September 1979; accepted for publication 30 November 1979)

The inverse problem for the reduced wave equation $\Delta u + k^2 n^2(x)u = 0$ is examined when $n(x)$ is assumed to be real continuous and equal to unity outside some prescribed compact domain D_0 in R^3. The case where a finite set of measurements of the scattered or total field (generated by a single incident wave at a fixed frequency) at points outside D_0 is considered. It is shown that the inverse problem can be expressed in terms of a system of linear functional equations, plus a quadratic nonlinear integral equation. By imposing an additional criterion for uniqueness of the functional equations and placing certain restrictions on the size of the measured data, a solution of the system can be generated by successive approximations. It is shown that the iterative scheme yields correction terms to the solution that are obtained by the common linearized approximation where $n(x)$ is assumed to be a very small perturbation of a known quantity and only linear terms are retained.

INTRODUCTION

The inverse problem of the reduced wave equation

$$\Delta u + k^2 n^2(x)u = 0, \quad x \in R^3, \tag{1}$$

associated with time dependence $\exp(-i\omega t)$ will be examined where the index of refraction $n(x)$ is assumed to be unity outside some region of compact support. To be more precise, D_0 will be a given compact region in R^3. The index of refraction $n(x)$ will be assumed to satisfy the conditions

(i) $n(x)$ is real and continuous,

(ii) $\text{supp}[n(x) - 1] \subset D_0$.

The inverse problem consists of the determination of $n(x)$ from suitable scattering measurements.

The emphasis here will be on the limited but realistic problem where a finite set of scattering measurements is made on either the scattered field u^s or the total field $u = u^i + u^s$, for a fixed frequency or wave number k and a fixed incident field u^i. Once this problem is properly understood, one can treat the problem where sets of measurements are made for a set of different incident fields. As was pointed out in "Non-linear Approach to Inverse Scattering," [1] the inverse problem under investigation here is nonlinear.

For subsequent analysis the following mathematical conventions and notations will be introduced.

A fixed compact domain D containing D_0 will be employed. D will be sufficiently large so that a continous function $\eta(x)$ with support in D can be defined so that $\eta(x) \equiv 1$ for $x \in \bar{D}_0$ and $\eta = 0$ on the boundary of D.

Both the Banach space $C(\bar{D})$ of real continuous functions on $\bar{D}$ with norm

$$\|u\|_c = \max_{x \in \bar{D}} |u(x)|$$

and the Hilbert space $\mathscr{L}_2(D)$ of square integral functions over D with norm

$$\|u\|_2 = \left(\int_D u^2 \, dx \right)^{1/2}$$

will be employed. If $\mathbb{K}$ is bounded linear operator that maps $C(\bar{D})$ into $C(\bar{D})$, its norm will be denoted by $\|\mathbb{K}\|_c$. If $\mathbb{K}$ maps $\mathscr{L}_2(D)$ into $\mathscr{L}_2(D)$, the corresponding operator norm will be denoted by $\|\mathbb{K}\|_2$. If, however, at the same time $\mathbb{K}$ maps $\mathscr{L}_2(D)$ into $C(\bar{D})$, the norm in this case will be given by $\|\mathbb{K}\|$, with

$$\|\mathbb{K}\| = \sup \|\mathbb{K}u\|_c / \|u\|_2 .$$

DIRECT SCATTERING

As a preliminary, a few results on the direct scattering problem will be presented. From Leis[2] it is shown that with the assumptions (i) and (ii) imposed on $n(x)$, the direct scattering problem

$$\Delta u + k^2 n^2(x)u = 0,$$

$$u = u^i + u^s,$$

$$\frac{\partial u^s}{\partial |x|} - iku^s \sim o(|x|^{-1}), \quad \text{as } |x| \to \infty,$$

has a unique solution with $u \in C^2(D)$, provided that the incident wave u^i is continuous in D (all sources are external to D). For plane-wave incidence $u(x)$ will be C^2 for all $x \in R^3$.

For later analysis the formulation of the direct-scattering problem as a perturbation about some known index of refraction $n_*(x)$ satisfying assumptions (i) and (ii) is required. Let $G(x,y,n_*)$ represent the Green's function satisfying the equation

$$\Delta G + k^2 n_*^2(x) G = -\delta(x - y) \tag{2}$$

and the radiation condition.

Set

$$v(x) = n^2(x) - 1, \tag{3}$$

then on rewriting Eq. (1) in the form

$$\Delta u^s + k^2 n_*^2(x) u^s = k^2 [v_*(x) - v(x)] u^s - k^2 v(x) u^i,$$

one obtains the result

$$u(x) = u_*(x) + k^2 \int_D G(x,y,v_*) [v(y) - v_*(y)] \, u(y) \, dy, \tag{4}$$

 0022-2488/80/040758-07$1.00

where

$$u_*(x) = u^i(x) + k^2 \int_D G(x,y,v_*)\, v_*(y)\, u^i(y)\, dy \tag{5}$$

is the total field produced by the wave $u^i(x)$ incident upon the scatterer characterized by the index of refraction $n_*(x)$. The direct-scattering problem consists of solving Eq. (4) for $u(x)$ when the kernel in the integral operator and $u_*(x)$ are known.

The Green's function has the decomposition

$$G(x,y,v_*) = G_0(x,y) + \tilde{G}(x,y,v_*), \tag{6}$$

where

$$G_0(x,y) = \frac{1}{4\pi} \frac{e^{ik|x-y|}}{|x-y|} \tag{7}$$

is the free-space Green's function (corresponding to $n_* \equiv 1$ or $v_* \equiv 0$) and $\tilde{G}(x,y,v_*)$ is a continuous function of x and y satisfying the integral equation

$$\tilde{G}(x,y,v_*) = k^2 \int_D G_0(x,y')\, G_0(y',y)\, v_*(y')\, dy' + k^2 \int_D G_0(x,y')\, \tilde{G}(y',y,v_*)\, v_*(y')\, dy' .$$

When $|x| \to \infty$, $\tilde{G}(x,y,v_*)$ has the asymptotic form

$$\tilde{G}(x,y,v_*) \sim \frac{e^{ik|x|} k^2}{|x| 4\pi} \int_D e^{-ik(x\cdot x')}\, G(x',y,v_*)\, v_*(x')\, dx' .$$

Hence from Eqs. (5) and (6) it follows that for $|x| \to \infty$, in the scattering direction given by the vector k^s,

$$k^s = (k/|x|)\, x, \tag{8}$$

$$G(x,y,v_*) \sim (e^{ik|x|}/4\pi|x|)\, u_*(y, -k^s), \tag{9}$$

where $u_*(y, -k^s)$ represents the total field generated by a *plane* wave propagating in the direction $-k^s$ on the scatterer with index of refraction $n_*(y)$.

The far-field behavior of $u^s(x)$ in the scattered direction k^s is represented in the form

$$u^s(x) \sim (e^{ik|x|}/|x|)\, g(k^s, v), \tag{10a}$$

where from Eqs. (4) and (9) the complex scattering amplitude $g(k^s, v)$ is given by

$$g(k^s, v) = g(k^s, v_*) + (k^2/4\pi) \times \int_D u_*(y, -k^s)[v(y) - v_*(y)]\, u(y)\, dy. \tag{10b}$$

FORMULATION OF THE INVERSE PROBLEM

The inverse problem associated with scattering by a fixed incident wave $u^i(x)$ generated by sources exterior to $\bar{D}$ (with a plane incident wave as a special case) will be formulated where measurements are made of the scattered or total field in the near or far field.

For measurements made in the near field at the set of points x_l, $l = 1,2,...,N$, one obtains from Eq. (4) the relations

$$k^2 \int_D G(x_l,y,v_*)[v(y) - v_*(y)]\, u(y)\, dy = b_l , \tag{11a}$$

where the complex numbers b_l correspond to the difference in the measured value of $u(x_l)$ and the calculated value of $u_*(x_l)$ corresponding to the index of refraction $n_*(x)$.

For measurements made in the far field in the scattering directions given by the vectors k_l^s, $l = 1,2,...,N$, one obtains from Eq. (10b) the relations

$$(k^2/4\pi) \int_D u_*(y, -k_l^s)[v(y) - v_*(y)]\, u(y)\, dy = b_l , \tag{11b}$$

where b_l is the difference in the measured value of the complex scattering amplitude $g(k_l^s, v)$ and the calculated value $g(k_l^s, v_*)$ associated with index of refraction n_*.

The functional relations (11a) and (11b) contain the unknown quantity $(v - v_*)\, u$. The total field u itself is a function of v and is obtained by solving integral equation (4). Hence equations (11a) and (11b) are nonlinear functional equations of v.

The inverse problem can be stated as follows: find the real continuous function $v(x)$ vanishing on the boundary of D, such that Eqs. (11a) and (4) or (11b) and (4) are satisfied.

Simplification of the problem is achieved by transforming the nonlinear functional equations into a system of linear functional equations. This is achieved by setting

$$(v - v_*)\, u = w, \tag{12}$$

in which case Eqs. (11a) and (11b) have the general form

$$\int_D \tilde{h}_l(y)\, w(y)\, dy = b_l, \quad l = 1,2,...,N, \tag{13}$$

$\tilde{h}_l(y)$ are known functions. Equation (4) reduces to the form

$$w = (v - v_*)[u_* + \mathbb{G}w], \tag{14}$$

where $\mathbb{G}$ is the integral operator with kernel $k^2 G(x,y,v_*)$. It can be seen that the system of linear functional equations can be used to determine $w(x)$ (not uniquely) and in turn Eq. (14) is used to determine $v(x)$ as follows,

$$v - v_* = w/[u_* + \mathbb{G}w] . \tag{15}$$

But in order for $v(x)$ to be a bounded real quantity, the following additional constraints on $w(x)$ need be imposed:

$$\mathrm{Im}\bar{w}[u_* + \mathbb{G}w] = 0, \tag{16}$$

$$|u_* + \mathbb{G}w|/|w| \geq c > 0, \tag{17}$$

where c is some suitable constant. The bar indicates the complex conjugate.

Thus an alternative formulation of the inverse problem is as follows: find the complex quantity w with continuous real and imaginary parts vanishing on the boundary of D, and which satisfies relations (13), (16), *and* (17). *Then $v(x)$ is found from Eq.* (15).

It is obvious that this latter inverse formulation is equivalent to the previous one since once $w(x)$ and $v(x)$ are found, $u(x)$ is recoverable from Eq. (12) for values of $x \in D' = \mathrm{supp}(v - v_*)$. Equation (15) can then be rewritten to yield the following integral equation which is satisfied for $u(x)$ for $x \in D'$:

$$u(x) = u_*(x) + k^2 \int_{D'} G(x,y,v_*)[v(y) - v_*(y)]\, u(y)\, dy.$$

This is sufficient since this equation can then be used to define $u(x)$ *for* $x \notin D'$.

To facilitate analysis of the inverse problem as given by Eqs. (13), (16), and (17), it is easiest to work with real quantities. But first the following implicit assumption will be imposed both upon $v_*(x)$ and the incident field:

(iii) $1/|u_*(x)| \leqslant c_0 < \infty$, for all $x \in D$.

This condition implies that the magnitude of the total field corresponding to v_* is bounded from zero. This assumption is satisfied in the case $v_* \equiv 0$ and plane-wave incidence $u^i(x) = \exp(ik^i \cdot x)$, since then $u_* = u^i$.

With the above assumption (iii), real and continuous functions $\varphi(x)$ and $\psi(x)$ are then defined by setting

$$\bar{w} u_* = \varphi(x) - i\psi(x). \tag{18}$$

Next define the operator L by the relation

$$Lf = k^2 \int_D G(x,y,v_*)[u_*(x)\,\bar{u}_*(y)]^{-1} f(y)\,dy \tag{19}$$

and decompose it by separating its kernel into real and imaginary parts yielding the respective operators such that

$$Lf = L_R f + iL_I f. \tag{20}$$

Equation (16) can now be reduced to the following form involving real variables only:

$$\psi = \psi(L_I \psi - L_R \varphi) + \varphi(L_I \varphi + L_R \psi). \tag{21}$$

The constraining inequality (17) reduces to

$$(1 + L_R \varphi - L_I \psi)^2 + (L_I \varphi + L_R \psi)^2 \geqslant c^2(\varphi^2 + \psi^2)/|u_*|^4. \tag{22}$$

The corresponding expressions for the system of linear functional equations (13) needs to be considered. These can be written in the form

$$\int_D h_l(y)(\varphi + i\psi)\,dy = b_l, \quad l = 1,2,...,N. \tag{23}$$

For measurements in the near field at the points x_l,

$$h_l(y) = k^2 G(x_l,y,v_*)\,\eta(y)[\bar{u}_*(y)]^{-1}, \tag{24}$$

and in the far field in the scattered direction k^s,

$$h_l(y) = (k^2/4\pi)\,u_*(y - k_l^s)\,\eta(y)[\bar{u}_*(y)]^{-1}. \tag{25}$$

Note that the factor $\eta(y)$, which has the property $\eta(y) \equiv 1$ for $y \in \operatorname{supp}[v(y) - v_*(y)]$, has been introduced for mathematical convenience of further technical analysis.

Finally, the appropriate expression for $v(x)$ corresponding to Eq. (15), which allows one to compute $v(x)$ from the solutions φ,ψ of Eqs. (21), (22), and (23) needs to be obtained. From Eqs. (18) and (19) it follows directly from (15) that

$$v(x) = v_*(x) + (\varphi + i\psi)/|u_*|^2[1 + L(\varphi + i\psi)].$$

Hence, from expression (21), the following simpler form is developed:

$$v(x) = v_*(x) + (1/|u_*|^2)\,\varphi\,[1 + L_R \varphi - L_I \psi]^{-1}. \tag{26}$$

The general approach that will be employed to investigate the solution of systems (21), (22), and (23) will be to first solve the system of linear functional equations (23), and obtain an expression for φ in terms of ψ and the data. The resulting expression for φ will then be substituted into Eq. (21), yielding a quadratic integral equation for ψ, the solution of which will be investigated. The resulting solution pair φ,ψ will then be examined to see if they satisfy Eq. (22). It will be seen that at certain steps in the analysis additional criteria will have to be imposed to insure uniqueness.

SOLUTION OF THE LINEAR FUNCTIONAL EQUATIONS

In the solution of the system of linear functional equations (23) the following assumption will be imposed upon the complex quantities $h_l(y)$,

(iv) The $2N$ real continuous functions $\{\operatorname{Re} h_l(y),\ \operatorname{Im} h_l(y)\}_{l=1}^{N}$ form a linearly independent set.

With this, define the following real quantities

$$H_l(y) = \begin{cases} \operatorname{Re} h_l(y), & l = 1,2,...,N, \\ \operatorname{Im} h_{l-N}(y), & l = N+1,...,2N, \end{cases} \tag{27}$$

$$B_l = \begin{cases} \operatorname{Re} b_l, & l = 1,2,...,N, \\ \operatorname{Im} b_{l-N}, & l = N+1,...,2N, \end{cases} \tag{28}$$

and the elements e_{ij} of a square $2N$ matrix

$$e_{ij} = \begin{cases} \delta_{(i+N)j}, & i = 1,2,...,N, \\ -\delta_{i(j+N)}, & i = N+1,...,2N, \end{cases} \tag{29}$$

where δ_{ij} is the Kronecker delta.

The system of N linear functional equations (23) now can be reduced to the following set of $2N$ real linear functional equations

$$(H_i, \varphi) = \sum_{j=1}^{2N} e_{ij}\,(H_j, \psi) + B_i, \quad i = 1,...,2N, \tag{30}$$

where $(u,v) = \int_D u(y)\,v(y)\,dy$.

Let $\{\Theta_m\}_{m=1}^{2N}$ be a suitably chosen basis of real continuous functions on D, vanishing on the boundary of D, and for which $\det\{a_{ij}\} \neq 0$, where

$$a_{ij} = (H_j, \Theta_j). \tag{31}$$

The solution of systems (30) may be found in the form

$$\varphi(x) = \sum_{j=1}^{2N} c_j \Theta_j(x) + \varphi^{\perp}(x),$$

where $\varphi^{\perp}(x)$ is orthogonal to span $\{H_l\}_{l=1}^{2N}$. The coefficients c_i are then obtained from the resulting algebraic system

$$\sum_{j=1}^{2N} a_{ij} c_j = B_i + \sum_{j=1}^{2N} e_{ij}(H_j, \psi).$$

If $\{\tilde{a}_{ij}\}$ denotes the inverse matrix of $\{a_{ij}\}$,

$$\sum_{k=1}^{2N} a_{ik}\,\tilde{a}_{kj} = \delta_{ij},$$

the general solution $\varphi(x)$ has the explicit form

$$\varphi(x) = \Phi(x) + \mathbb{K}\psi + \varphi^{\perp}(x), \tag{32}$$

where

$$\Phi(x) = \sum_{j=1}^{2N} \xi_j(x) B_j, \tag{33}$$

$$K\psi = \sum_{j,k=1}^{2N} \xi_j(x)\,e_{jk}(H_k, \psi), \tag{34}$$

$$\xi_i(x) = \sum_{j=1}^{2N} \tilde{a}_{ji} \Theta_j(x), \quad i = 1,2,...,N. \tag{35}$$

As is apparent, the solution for $\varphi(x)$ is not unique since $\varphi^{\perp}$ is undetermined. Hence Eq. (32) represents a family of solutions. An additional condition needs to be employed to select out of this family of solutions one particular solution. A common approach is to select the smoothest solution in some appropriate sense. Here the solution with the minimum $\mathscr{L}_2(D)$ norm will be selected. Apart from being the smoothest in this norm, it will be shown later that this condition has other important ramifications. The minimum norm solution $\varphi_M(x)$ is obtained by directly requiring that each element in the basis $\{\Theta_j\}_{j=1}^{2N}$ be a linear combination of $\{H_j\}_{j=1}^{2N}$. Then $\Phi(x)$ and $\mathbb{K}\psi$ will lie in the space spanned by $\{H_j\}_{j=1}^{2N}$, and the minimum $\mathscr{L}_2(D)$ norm is obtained by setting $\varphi^{\perp}(x) \equiv 0$.

The results are summarized as follows.

Lemma 1: If $\{H_j\}_{j=1}^{2N}$ form a linear independent set, then the minimum $\mathscr{L}_2(D)$ norm solution $\varphi_M(x)$ of the system of linear functions (23) is given by

$$\varphi_M(x) = \Phi_M(x) + \mathbb{K}_M\psi, \tag{36}$$

where $\Phi_M(x)$ and $\mathbb{K}_M\psi$ are given by Eqs. (32), (34), and (35), with the restriction that the basis $\{\Theta_j\}_{j=1}^{2N}$ be chosen so that each $\Theta_j(x)$ is a linear combination of $\{H_j\}_{j=1}^{2N}$.

The additional result is obtained:L

Lemma 2: $\|K\|_2 = \|\mathbb{K}_M\|_2 = 1$.

Proof: From Eqs. (35) and (31) it follows that

$$(\xi_i, H_j) = \sum_{k=1}^{2N} \tilde{a}_{ki}(\Theta_k, H_j) = \delta_{ij},$$

hence the sets $\{\xi_j\}_{j=1}^{2N}$ and $\{H_j\}_{j=1}^{2N}$ form a biorthonormal system. The operators $\mathbb{K}$ and $\mathbb{K}_M$ have the property

$$\mathbb{K}\xi_n = \sum_{j=1}^{2N} e_{jn}\xi_j(x) = \begin{cases} -\xi_{n+N}, & n = 1,2,...,N, \\ \xi_{n-N}, & n = N+1,...,2N. \end{cases}$$

It can thus be shown that the compact operator $\mathbb{K}$ (or $\mathbb{K}_M$) of finite rank has two distinct nonzero eigenvalues $\pm i$, each with multiplicity N. The eigenvalue i has the N independent eigenfunctions $\xi_n + i\xi_{n+N}$, $n = 1,2,...,N$, and the eigenvalue $-i$ has the N independent eigenfunctions $\xi_n - i\xi_{n+N}$, $n = 1,2,...,N$. It thus follows that the operators $\mathbb{K}$ or $\mathbb{K}_M$ have $\mathscr{L}_2(D)$ unit norm.

When N approaches ∞, the functions $H_i(y)$ become close together in the sense of the $\mathscr{L}_2(D)$ norm and the matrix $\{a_{ij}\}$ becomes singular. This instability and the effect of errors in the data can be removed by using a regularization or approximate method.[3,4] It is best to solve Eq. (30) for φ in terms of ψ and $\{B_i\}$ separately. In order to keep the results $\|\mathbb{K}_M\psi\|_2 = 1$, a judicious choice of a finite set of $2M$ equations

$$(H_i, \varphi) = \sum_{j=1}^{2N} e_{ij}(H_j, \psi),$$

for $i = i_1, i_2, ..., i_M, N+i_1, N+i_2, ..., N+i_M$, should be solved as in the previous manner. Essentially, these equations are picked so that the distances between the H_i's involved, $\|H_i - H_j\|_2$, are not too small. To find the approximate solution $\Phi_M(x)$, set $\Phi_M(x) = \Sigma_{j=1}^{2N} c_j\Theta_j(x)$ for a suitable chosen basis Θ_j (which could be a linear combination of the set of H_i's used in the equation to find $\mathbb{K}_M$). Then find the least-squares solution of the equations

$$(H_i, \varphi) = B_i.$$

Thus the c_i's are obtained from

$$\min_{|c_i|} \left\{ \sum_{i=1}^{2N} \left(\sum_{j=1}^{2N} a_{ik}c_k - B_i \right)^2 + \alpha \sum_{k=1}^{2N} c_k^2 \right\}$$

for some suitable $\alpha > 0$.

SOLUTION OF THE QUADRATIC INTEGRAL EQUATION

Upon substitution of the expression $\varphi_M(x)$,

$$\varphi_M(x) = \Phi_M(x) + \mathbb{K}_M\psi$$

(obtained from solving the linear function equations), into Eq. (21), namely,

$$\psi = \psi(L_I\psi - L_R\varphi) + \varphi(L_R\psi + L_I\varphi),$$

one obtains a quadratic integral equation for $\psi(x)$, which will be represented by the general expression

$$\psi = S(\psi). \tag{37}$$

In order to investigate the solution of this equation, some properties of the operators L_R and L_I will be required. It should be recalled that the kernels of the operators L_R and L_I correspond to the real and imaginary parts of

$$k^2G(x,y,v_*)[u_*(x)\bar{u}_*(y)]^{-1},$$

respectively. Because of assumption (iii) on the properties of u_* together with the properties of the Green's function [see Eqs. (6) and (7)], the kernels will be weakly polar kernels,[5] hence the operators not only map $C(\bar{D})$ in $C(\bar{D})$ and $\mathscr{L}_2(D)$ into $\mathscr{L}_2(D)$, but also $\mathscr{L}_2(D)$ into $C(\bar{D})$. In this latter case the respective norm for L_R has the estimate

$$\|L_R\| \leqslant \max_{x\in\bar{D}} \int_D \{\mathrm{Re}[k^2G(x,y,v_*)\,u_*^{-1}(x)\,\bar{u}_*^{-1}(y)]\}^2\,dy$$

and a similar result holds for $\|L_I\|$.

Results of Rall[6] on Newton's method and related iterated schemes for quadratic equations like Eq. (37) can be employed, but first, Eq. (37) must be placed in the "normal" form. To accomplish this the following Taylor expansion for quadratic operators is used:

$$S(\psi) = S(\psi_0) + S'(\psi_0)(\psi - \psi_0) + A(\psi - \psi_0, \psi - \psi_0).$$

Here A is the bilinear operator

$$A(\chi,\chi) = \chi(L_I - L_R\mathbb{K}_M)\chi + (\mathbb{K}_M\chi)\cdot(L_R + L_I\mathbb{K}_M)\chi, \tag{38}$$

and $S'(\psi)$ is the Fréchet derivative of $S(\psi)$, given explicitly by

$$S'(\psi)h = a(x)h + \mathbb{M}h, \tag{39a}$$

with

$$a(x) = (L_I - L_R\mathbb{K}_M)\psi - L_R\Phi_M, \tag{39b}$$

and the linear operator $\mathbb{M}$ given by

$$\mathbb{M}h = \psi(L_I - L_R\mathbb{K}_M)h + (\mathbb{K}_M\psi + \Phi_M)\cdot(L_R + L_I\mathbb{K}_M)h + [(L_R + L_I\mathbb{K}_M)\psi + L_I\Phi_M]\cdot\mathbb{K}_M h. \tag{39c}$$

With this result Eq. (37) can be reduced to Rall's normal form

$$(\psi-\psi_0)=[I-S'(\psi_0)]^{-1}A(\psi-\psi_0,\psi-\psi_0)+[I-S'(\psi_0)]^{-1}[S(\psi_0)-\psi_0], \tag{40}$$

for which the following result may be stated.

Lemma 3: If

$$4\{\|[I-S'(\psi_0)]^{-1}\|_2\}^2\|A\|_2\|S(\psi_0)-\psi_0\|_2\leqslant 1, \tag{41}$$

both the Newton's and the modified Newton's iterative procedures applied to Eq. (40) (starting from the initial approximation ψ_0) converge in the $\mathscr{L}_2(D)$ sense to a unique solution ψ of Eq. (40), which lies in the ball

$$\|\psi-\psi_0\|_2\leqslant 2\|[I-S'(\psi_0)]^{-1}\|_2\|S(\psi_0)-\psi_0\|_2 .$$

Proof: With the substitution $\psi-\psi_0=\chi$, $[I-S'(\psi_0)]^{-1}[S(\psi_0)-\psi_0]=\zeta$, Eq. (40) reduces to the form

$$\chi=B(\chi,\chi)+\zeta,$$

where B is a bilinear operator. From Rall,[6] convergence follows if $4\|B\|_2\|\zeta\|_2\leqslant 1$. The solution χ satisfies the inequality $\|\chi\|_2\leqslant 2\|\zeta\|_2[1+(1-4\|B\|_2\|\zeta\|_2)^{1/2}]^{-1}$.

From this the following more useful (although not as sharp) estimates can be obtained for the important case when the initial approximation in the iterative schemes is $\psi_0\equiv 0$.

Lemma 4: If

$$\|\Phi_M\|_2\leqslant\tfrac{1}{2}\{\|L_I\|+\|L_R\|+[2\|L_I\|(\|L_I\|+\|L_R\|)]^{1/2}\}^{-1}, \tag{42}$$

then both Newton's and the modified method, with initial approximation $\psi_0\equiv 0$, applied to Eq. (40) converges to a unique solution ψ such that

$$\|\psi\|_2\leqslant\|\Phi_M\|_2[2(\|L_I\|+\|L_R\|)/\|L_I\|]^{-1/2} . \tag{43}$$

Proof: The basic problem is to reduce constraint (41) to a simpler form. At first the additional assumption

$$\|L_R\Phi_M\|_c<1$$

will be imposed (later on it will be shown to be automatically satisfied).

From Eq. (39a) it is seen that

$$[I-S'(0)]h=[1+L_R\Phi_M][h-\mathbb{M}_1 h],$$

where

$$\mathbb{M}_1 h=[1+L_R\Phi_M]^{-1}[\Phi_M(L_R+L_I\mathbb{K}_M)h+L_I\Phi_M\cdot\mathbb{K}_M h] . \tag{44}$$

Using the above assumption, it follows that

$$\|(I-S'(0))^{-1}\|_2\leqslant(1-\|L_R\Phi_M\|_c)^{-1}(1-\|\mathbb{M}_1\|_2)^{-1} .$$

From Eq. (44) and the result that $\|K_M\|_2=1$, it is seen that

$$(1-\|L_R\Phi_M\|_c)\|\mathbb{M}_1\|_2\leqslant\|\Phi_M\|_2(\|L_R\|+2\|L_I\|),$$

hence

$$\|(I-S'(0))^{-1}\|_2\leqslant[1-2\|\Phi_M\|_2(\|L_R\|+\|L_I\|)]^{-1} .$$

For the case $\psi_0\equiv 0$, it can be shown that

$$\|S(\psi_0)-\psi_0\|_2=\|\Phi_M L_I\Phi_M\|_2\leqslant\|L_I\Phi_M\|_c\|\Phi_M\|_2\leqslant\|L_I\|(\|\Phi_M\|_2)^2 .$$

Finally, an estimate for $\|A\|_2$ is needed. It can be shown

$$\|A(\chi,\chi)\|_2\leqslant\|\chi\|_2\|(L_I-L_RK_M)\chi\|_c+\|K_M\chi\|_2\|(L_R+L_IK_M)\chi\|_c ,$$

$$\|A\|_2\leqslant 2[\|L_I\|+\|L_R\|] .$$

Finally, on combining results, inequality (41) reduces to

$$8(\|\Phi_M\|_2)^2\|L_I\|(\|L_I\|+\|L_R\|)\leqslant[1-2\|\Phi_M\|_2(\|L_R\|+\|L_I\|)]^2 \tag{45}$$

Taking the square root of both sides and rearranging terms, one obtains the desired result:

$$\|\Phi_M\|_2\leqslant\tfrac{1}{2}\{\|L_I\|+\|L_R\|+[2\|L_I\|(\|L_I\|+\|L_R\|)]^{1/2}\}^{-1} .$$

It is immediately obvious that since $\|L_R\Phi_M\|\leqslant\|L_R\|\,\|\Phi_M\|_2$, the assumption $\|L_R\Phi_M\|_c<1$ that was introduced in the beginning of this proof is automatically satisfied if inequality (42) is satisfied.

From Lemma 3 the condition on the size of the solution ψ reduces to

$$\|\psi\|_2\leqslant 2(\|\Phi_M\|_2)^2\|L_I\|[1-2\|\Phi_M\|_2(\|L_I\|+\|L_R\|)]^{-1},$$

and on using Eq. (45) this simplifies to

$$\|\psi\|_2\leqslant\|\Phi_M\|_2[2(\|L_I\|+\|L_R\|)/\|L_I\|]^{-1/2} .$$

Note that it was also shown by Rall that the successive-approximation scheme applied to the normal form converges if in the inequality (41) the factor 4 is replaced by a factor 8. However, even if successive approximation was employed, one needs to invert an integral equation to go from the original equation $\psi=S(\psi)$ to the normal form. It is of interest to see if the method of successive approximations is valid as applied to the equation in the original form, especially if the initial approximation $\psi_0\equiv 0$ is used, since no inversion of an integral equation would then be required.

Lemma 4: If inequality (42) is satisfied for $\|\Phi_M\|_2$, then the method of successive approximation

$$\psi_0=0,$$

$$\psi_{n+1}=S(\psi_n), \quad n=0,1,2,\cdots,$$

converges in the $\mathscr{L}_2(D)$ norm to a solution ψ which lies in the ball given by inequality (43).

Proof: A brief sketch of the proof will be given. Since

$$\|\psi_0-S(\psi_0)\|_2\leqslant(\|\Phi_M\|_2)^2\|L_I\| ,$$

$$\|S'(\psi)\|_2\leqslant(\|L_I\|+\|L_R\|)(4\|\psi\|_2+2\|\Phi_M\|_2) ,$$

it can be shown (Vainberg[7]) that the equation $\psi=S(\psi)$ is majorized by the equation

$$2\tilde{l}t^2-t(1-2\tilde{l}\|\Phi_M\|_2)+\|L_I\|(\|\Phi_M\|_2)^2=0 ,$$

where $\tilde{l}=\|L_I\|+\|L_R\|$.

It can be shown that this has positive real roots if inequality (42) is satisfied. Hence the result follows.

It has been shown that the various iterative techniques converge to ψ in the $\mathscr{L}_2(D)$ sense. The results can be made stronger as is shown in the following.

Theorem 1: The solution of equation $\psi=S(\psi)$ obtained by the successive approximation scheme in Lemma 4, or by Newton's method starting for $\psi_0=0$ as given by Lemma 3, is a continuous function in D and vanishes on the boundary of D. Furthermore, inequality (22) is satisfied by the pair φ_m, ψ, where $\varphi_M=\Phi_M+\mathbb{K}_M\psi$.

Proof: By construction, the function $\varphi_M=\Phi_M+\mathbb{K}_M\psi$ is continuous on D and vanishes on the boundary of D. Equation $\psi=S'(\psi)$ can be written in the form

$$4J\int_S \frac{dX\wedge dY}{(1+X^2+Y^2)^2} = (2n+\frac{\nu}{2})\cdot\pi, \tag{4.6}$$

where $n = 0,1,2,\cdots$. With the aid of the stereographic projection

$$X = \cot(\vartheta/2)\cdot\cos\varphi,$$
$$Y = \cot(\vartheta/2)\cdot\sin\varphi, \tag{4.7}$$

(4.6) reads

$$J\int_S \sin\vartheta d\vartheta d\varphi = (2n+\nu/2)\pi. \tag{4.6'}$$

Equation (4.6) is just regarded as a quantization condition à la Bohr–Sommerfeld in the curved phase space S^2. However, in contrast to the usual (flat) phase space quantization condition, the integer n does not take arbitrary positive values because the integral (4.6) is just proportional to the area on a unit sphere encircled by the closed orbit and is bounded by 4π.

Finally we examine the formula (4.6) for the simplest Hamiltonian $\tilde{H} = -\hbar\omega J_z$, which was taken up in the previous section. The classical orbit (3.45) is written in angle variables as

$$\vartheta = \vartheta_0 = \text{const}, \quad \varphi = \omega t + \varphi_0, \tag{4.8}$$

which just describes a circle on a unit sphere. The energy is given by

$$E = -\omega\hbar J_z = -\omega\hbar J\cos\vartheta_0. \tag{4.9}$$

The index ν is given by 2 for the solution (4.8), since the singularity of the semiclassical propagator, as is seen from Eq. (3.49), appears twice per period, i.e., $\sin\omega t$ vanishes at two points $t = \pi/\omega, 2\pi/\omega$. Thus Eq. (4.6') yields

$$2\pi J(1-\cos\vartheta_0) = (2n+1)\pi. \tag{4.10}$$

Using the relation $J_z = J\cos\vartheta_0$, we get

$$J_z = J - (n+\tfrac{1}{2}).$$

From (4.11), J_z takes an integer or half-integer value between $-(J-1/2)$ and $J-1/2$. Hence we obtain an energy spectrum

$$E_m = -\omega\hbar m \quad [\, m = -(J-\tfrac{1}{2}), -(J-\tfrac{3}{2}),\cdots, J-\tfrac{1}{2})\,].$$

This spectrum clearly differs from the exact result of the quantum theory, i.e., the magnitude of the spin is reduced by $\hbar/2$. This discrepancy originates from the index $\nu = 2$. It suggests that in the semiclassical approximation one makes a replacement $J\to J+1/2$. This is correct in the large J limit, where the semiclassical picture becomes accurate.

5. CONCLUDING REMARKS

In this paper we have investigated a semiclassical analysis of the spin system and obtained a closed form of the semiclassical propagator. The formula obtained here is useful for the approximate calculation of the energy spectra of an asymmetric top and the many-body systems which can be described by the quasispin formulation, etc.[16]

The essential point of our treatment is that the semiclassical propagator in the curved phase space ($\simeq$two-dimensional sphere), which appears at first sight rather complicated, can be handled on the same footing as that in the usual flat phase space. In this way, the present method would provide us a promising tool for the semiclassical analysis of the path integral in a curved phase space with more complicated geometrical structure. As an immediate extension, we can consider the spin systems with many degrees of freedom, e.g., a spin chain.

The other problem is the extension to the path integral in the coherent states for the unitary group of higher dimension which has been proposed elsewhere[17] in connection with the quantization of nuclear collective motions. The present method would provide us with a useful basis for this subject. As for a path integral on the group manifold SU(n), Dowker found a compact formula for the propagator by applying the DeWitt formula for the Riemannian space as the case of SU(2) propagator,[18] whereas the path integral form in Ref. 17 is given by the functional integral on the homogeneous space $U(m+n)/U(m)\times U(n)$; hence the semiclassical propagator may result in a quite different form from the one for the group manifold SU(n).

ACKNOWLEDGMENTS

The authors are grateful to Dr. Tōru Suzuki for several discussions which are helpful for this work. Their thanks are also due to the members of the Nuclear Theory Group of Kyoto University. One of the authors (Y.M.) acknowledges the Japan Society for the Promotion of Science for financial support.

APPENDIX

In this appendix we evaluate the reduced propagator (3.14) by adopting the discretization procedure for functional Gaussian integral.[19] The Nth approximation for $\tilde{K}$ reads

$$\tilde{K}_N = (2\pi\hbar)^{-N}\int\prod_{k=1}^{N-1}dx_k\prod_{k=1}^{N}dy_k \times\exp\Big[(i/\hbar)\sum_{k=1}^{N}\{y_k(x_k-x_{k-1}) - \tfrac{1}{2}\epsilon(A_kx_k^2+2B_kx_ky_k+C_ky_k^2)\}\Big]. \tag{A1}$$

with $\epsilon\equiv(t''-t')/N$. By performing the integration over y variables, we get

$$\tilde{K}_N = (2\pi\hbar)^{-N}(-2\pi i\hbar)^{N/2}\prod_{k=1}^{N}(C_k\epsilon)^{-1/2} \int\prod_{k=1}^{N-1}dx_k\exp\Big[\frac{i}{\hbar}L_N^{(2)}\Big]. \tag{A2}$$

$L_N^{(2)}$ is the Nth approximation of the Lagrangian

$$L^{(2)} = \int_{t'}^{t''}\Big[\frac{1}{2C}\dot{x}^2 - \frac{B}{C}x\dot{x} + \tfrac{1}{2}\Big(\frac{B^2}{C}-A\Big)x^2\Big]dt. \tag{A3}$$

By using the integration by part and noting the boundary $x_0 = x_N = 0$, the Nth approximation of $L^{(2)}$ becomes

$$L_N^{(2)} = \sum_{k=1}^{N}\Big\{\frac{(x_k-x_{k-1})^2}{2C_k\epsilon} - \frac{1}{2}\Big[A - \frac{B^2}{C} - \Big(\frac{B}{C}\Big)'\Big]_k x_k^2\cdot\epsilon\Big\}, \tag{A4}$$

where $(\,)'$ denotes the differentiation with respect to t and $(\,)_k$ the value at the time $t_k = t' + k\epsilon$. Noting that the (A4) is

nonsingular, hence $[I - S'(\psi)]^{-1}$ is a bounded operator. Thus it follows that δv is related to errors in the data by a linear bounded operator of the form

$$\delta v = \sum_{j=1}^{2N} F_j \delta B_j .$$

Hence small errors in the data produced finite size errors in the solution δv.

APPLICATION TO THE LINEARIZED THEORY

The nonlinear results can be used to easily obtain higher order correction terms to the linearized theory.[9] The linearized approach is based upon (1) the perturbation of the problem by assuming that the unknown quantity $n(x)$ is a small perturbation of a known value $n_*(x)$, and (2) the subsequent linearization of the problem by taking as the perturbed solution $u(x)$ of Eq. (4), that given by the first two terms in the Neumann expansion

$$u(x) \sim u_*(x) + k^2 \int_D G(x, y; v_*)[v(y) - v_*(y)] u_*(y)\, dy .$$

Consequently, measurements made at points x_l, $l = 1, 2, ..., N$, in the near field yield the system of linear functional equations over the unknown perturbation $(v - v_*)$:

$$k^2 \int_D G(x_l, y; v_*)[v(y) - v_*(y)] u_*(y)\, dy = b_l , \qquad l = 1, ..., N ,$$

with a similar result corresponding to Eq. (11b) for measurements in the far field. In either case, the general form for the linear functional equation is

$$\int_D h_l(y) |u_*(y)|^2 [v(y) - v_*(y)]\, dy = b_l , \tag{51}$$

where $h_l(y)$ is given by Eq. (24) or (25).

Since the system of equations corresponding to the nonlinear approach, Eq. (23), reduce to the linearized set given by Eq. (51) if

$$\varphi = |u_*(x)|^2 [v(x) - v_*(x)] ,$$

and $\psi = 0$, the minimum $\mathscr{L}_2(D)$ norm solution of Eq. (51) will be given by

$$|u_*(x)|^2 [v(x) - v_*(x)] = \Phi_M(x) ,$$

where $\Phi_M(x)$ is given by Eq. (33). Thus the linearized approach yields the solution

$$v(x) \sim v_*(x) + \Phi_M(x)/|u_*(x)|^2 .$$

In comparison, the nonlinear approach yields various higher-order approximations by taking successive approximations of the equation $\psi = S(\psi)$ starting from $\psi_0 = 0$. Thus the zeroth-order solution will be given by

$$\varphi_0(x) = \Phi_M(x) , \quad \psi_0(x) = 0 ,$$

yielding from Eq. (26) the zeroth-order approximation

$$v_0(x) \sim v_*(x) + \Phi_M(x)/|u_*|^2 [1 + L_R \Phi_M] . \tag{52}$$

The first-order approximation is given by

$$\psi_1(x) = S(\psi_0) = \Phi_M L_I \Phi_M , \tag{53}$$

$$\varphi_1(x) = \Phi_M(x) + \mathbb{K}_M \psi_1 , \tag{54}$$

Yielding the value

$$v_1(x) \sim v_*(x) + \frac{\varphi_1}{|u_*|^2 [1 + L_R \varphi_1 - L_I \psi_1]} . \tag{55}$$

As an illustration, consider the particular case where the incident wave is a plane wave $u^i = \exp(ik^i \cdot x)$ and $n(x)$ is assumed to be a perturbation of $n_*(x) \equiv 1$. In this case

$$v_*(x) \equiv 0 , \quad |u_*(x)| = |u_*^i(x)| = 1 ,$$

$$L_R f = \frac{k^2}{4\pi} \int_D \frac{\cos[k|x - y| - ik^i \cdot (x - y)]\, f(y)\, dy}{|x - y|} ,$$

with a similar representation for L_I except that the cosine term is replaced by a sine.

If measurements are made in the far field in the scattered directions k_l^s, $l = 1,2,...,N$, then

$$h_l(x) = \frac{k^2}{4\pi} \eta(x) \exp(-ik_l^s \cdot x + ik^i \cdot x) .$$

The solution to the linearized problem in this case is given by

$$v(x) \sim \Phi_M(x) ,$$

which is equivalent to the Born approximation.[10] The nonlinear iterative approximations (52) and (53) then yield corrections to the Born approximation, i.e.,

$$v_0(x) \sim \Phi_M(x)/[1 + L_R \Phi_M(x)] .$$

COMMENTS

There still remains to be done a considerable amount of analysis for the single incident wave case, including the investigation of the number of solutions to the equation $\psi = S(\psi)$ and the means of selecting the initial approximation ψ_0 when inequality (50) fails. In addition, the results should be extended to the case where measurements are made on the scattered fields generated by more than one incident field. The equation $\psi = S(\psi)$ would then have to be replaced by a system of quadratic integral equations.

Other possible stategies for choosing the component ϕ^1 need to be investigated.

[1] V.H. Weston, "Non-Linear Approach to Inverse Scattering" J. Math. Phys. **20**, 53 (1979).
[2] R. Leis, "Vorlesungen uber Partielle Differential-Gleichungen Zweiter Ordung", Bibliographisches Institut, Mannheim, 1967.
[3] J.W. Hilgers, "Non-Iterative Methods for Solving Operator Equations of the First Kinds" M.R.C. Report 1413, 1974.
[4] M.Z. Nashed, "Generalized Inverses and Applications" (Academic, New York, 1976).
[5] V.S. Vladimirov, "Equations of Mathematical Physics" (Dekker, New Yew York, 1971).
[6] L.B. Rall, "Nonlinear Functional Analysis and Applications" (Academic, New York, 1971), pp. 383–94.
[7] M.M. Vainberg, "Variational Methods for the Study of Non-Linear Operators", (Holden-Day, San Fransisco, 1964).
[8] G. Strang and G. Fix, "An Analysis of the Finite Element Method", (Prentice-Hall, Englewood Cliffs, 1973).
[9] V.G. Romanov, "Integral Geometry and Inverse Problems for Hyperbolic Equations", (Springer-Verlag, Berlin, 1974).
[10] E. Wolf, "Three-Dimensional Structure Determination of Semi-Transport Objects from Holographic Data," Opt. Commun. **1** (4), 153–56 (1969).

Diffusion equation derived from the space–time transport equation in anisotropic random media

K. Furutsu
Radio Research Laboratories, Koganei-shi, Tokyo 184, Japan

(Received 5 June 1979; accepted for publication 26 September 1979)

The diffusion equation is derived from the ordinary space–time transport equation with an arbitrary scattering cross section $\sigma(\Omega|\Omega')\neq\sigma(\Omega'|\Omega)$. After various equations are Fourier transformed with respect to the space–time coordinates, the transport equation is shown to be reduced to an integral equation of the Fredholm type with respect to the angle of scattering, with the variables of Fourier transformation kept constant. The solution is then expanded in the eigenfunction series and, under the diffusion condition, its convergence of series is generally good enough to keep only the first term of the series except for the case where the waves are scattered mostly in forward direction, as in turbulent air. The boundary equations on the surface of medium discontinuity are then obtained, also based on the eigenfunction expansion of the quantities continuous across the boundary surface. A simple application is made to the pulse wave propagation in a space of semi-infinite scattering medium.

I. INTRODUCTION

It has been known[1] that the average intensity of waves in stationary random media can be obtained, to a very good approximation, by solving the ordinary transport equation of the form

$$[\Omega\cdot\partial/\partial\rho + c^{-1}\partial/\partial t + \gamma_t(\Omega)]I(\Omega,\bar{\rho}) = \int d\Omega'\sigma(\Omega|\Omega')I(\Omega',\bar{\rho}) + J_c(\Omega,\bar{\rho}). \quad (1.1)$$

Here, $\rho = (\rho_1,\rho_2,\rho_3)$ denotes the space coordinates, t the time, and $\bar{\rho} = (\rho,t)$ represents the space–time coordinates together; $\Omega = (\Omega_1,\Omega_2,\Omega_3)$ with $\Omega^2 = 1$ denotes the unit vector in space, and $I(\Omega,\bar{\rho})$ is the angular distribution function of the wave, expressing the intensity of wave propagating in the direction Ω at $\bar{\rho}$; $\sigma(\Omega|\Omega')$ is the scattering cross section per unit angle and per unit volume, giving rise to the scattering of the wave propagating in the direction Ω' into Ω, and $J_c(\Omega,\bar{\rho})$ is an external wave source; the term $\gamma_t(\Omega)$ is conveniently divided into two parts, accoring to

$$\gamma_t(\Omega) = \gamma(\Omega) + \gamma_a(\Omega), \quad \gamma(\Omega) = \int d\Omega'\sigma(\Omega'|\Omega), \quad (1.2)$$

where $\gamma(\Omega)$ gives the total scattering cross section for the wave propagating in the direction Ω, whereas $\gamma_a(\Omega)$ gives rise to an absorption of the wave. When the reciprocity relation is kept in the scattering process, it holds the relation $\sigma(\Omega|\Omega') = \sigma(-\Omega'|-\Omega)$ which is generally not equal to $\sigma(\Omega'|\Omega)$. This sort of anisotropy often happens in nature when asymmetrical scatterers are likely to be oriented in a particular direction in space, as in rain, snow, plasma, etc. The transport equation can be derived, to a good approximation, from the equation of the space–time coherence function of the wave (which has the form of the Bethe–Salpeter equation) in almost the same way as in the time-independent case, but discussion of the approximation involved in Eq. (1.1) is outside of the scope of the present paper.

Solving the space–time transport equation of the form (1.1) provides a useful means of investigating the pulse wave broadening, not only in media of discrete random scatterers[2,3], but also in turbulent media where the scattering is made essentially in the forward direction[4–7]. Here, in the latter case, the transport equation can be approximated by a simple equation by means of an approximate Fourier transformation with respect to Ω[6,7], but it is generally not an easy task to solve the transport equation even when $\sigma(\Omega|\Omega')$ is a function only of $\Omega\cdot\Omega'$ and the equation is time-independent; Case and other people developed an analytical method which utilizes an expansion of $I(\Omega,\bar{\rho})$ in terms of a set of singular eigenfunctions[8,9], but the most popular way seems to be the numerical methods of discrete ordinates by Sick and Chandrasekhar.

On the other hand, as the distances of wave propagation become sufficiently large beyond the coherence distance of the wave, the wave intensity tends to be given by an asymptotic expression which is known to obey a sort of diffusion equation. In a recent paper[3], a diffusion equation was derived from the equation of the two-frequency mutual coherence function and was given by a second order differential equation in time, while, starting from the Fokker–Planck equation, Titulaer[10] derived a diffusion equation of first order in time for a Brownian particle in the high-friction case and showed a systematic expansion prodedure in terms of the inverse friction coefficient. This is also the case of the diffusion equation for the concentration near the nonequilibrium steady state derived by Gantsevich *et al.*[11] Indeed, in the case of isotropic scattering where an exact solution is obtainable, the diffusion equation can be given only in the first order in time (Sec. II), and this is also the general case of anisotropic scattering (Sec. III).

In this paper, the diffusion equation will be derived from the space–time transport equation with an arbitrary $\sigma(\Omega|\Omega')\neq\sigma(\Omega'|\Omega)$ and also the boundary conditions to be satisfied on surface of medium discontinuity, including that of free space. The method is based on an eigenfunction expansion of various physical quantities and is essentially an application of the method that was previously used for the time-independent transport equation.[12]

 0022-2488/80/040765-13$1.00

II. PRELIMINARIES AND BASIC EQUATIONS

In terms of the angular distribution function $I(\mathbf{\Omega}, \bar{\rho})$ introduced in Eq. (1.1), the wave intensity $I(\bar{\rho})$ and the flux vector $\mathbf{I}(\bar{\rho}) = (I_1, I_2, I_3)(\bar{\rho})$ are given by

$$I(\bar{\rho}) = \int d\,\mathbf{\Omega} I(\mathbf{\Omega}, \bar{\rho}),$$
$$I_j(\bar{\rho}) = \int d\,\mathbf{\Omega}\, \Omega_j I(\mathbf{\Omega},\bar{\rho}), \quad j = 1,2,3. \tag{2.1}$$

Hence, the $\mathbf{\Omega}$ integration on both sides of Eq. (1.1) with the aid of Eq. (1.2) yields the equation of continuity

$$(\partial/\partial\rho)\cdot\mathbf{I}(\bar{\rho}) + (\gamma_a + c^{-1}\partial/\partial t) I(\bar{\rho}) = J_c(\bar{\rho}), \tag{2.2}$$

where

$$\gamma_a = \int d\,\mathbf{\Omega}\, \gamma_a(\mathbf{\Omega}) I(\mathbf{\Omega},\bar{\rho}) \quad \int d\,\mathbf{\Omega}\, I(\mathbf{\Omega},\bar{\rho}),$$
$$J_c(\bar{\rho}) = \int d\,\mathbf{\Omega}\, J_c(\mathbf{\Omega},\bar{\rho}). \tag{2.3}$$

Since the transport equation depends only on $\partial/\partial\bar{\rho}$ but not on $\bar{\rho}$ explicitly, it is most convenient to treat the equation in terms of its Fourier transform with respect to $\bar{\rho}$: With the notation $\bar{\lambda} = (\boldsymbol{\lambda}, \nu)$ and $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)$, the equation obeyed by

$$\tilde{I}(\mathbf{\Omega},\bar{\lambda}) = \int d\bar{\rho}\, I(\mathbf{\Omega},\bar{\rho}) \exp[i(\boldsymbol{\lambda}\cdot\boldsymbol{\rho} - \nu t)], \quad d\bar{\rho} = d\rho\, dt, \tag{2.4}$$

becomes

$$[i(\nu/c - \mathbf{\Omega}\cdot\boldsymbol{\lambda}) + \gamma_t(\mathbf{\Omega})]\tilde{I}(\mathbf{\Omega},\bar{\lambda}) = \int d\,\mathbf{\Omega}'\sigma(\mathbf{\Omega}|\mathbf{\Omega}')\tilde{I}(\mathbf{\Omega}',\bar{\lambda}) + \tilde{J}_c(\mathbf{\Omega},\bar{\lambda}), \tag{2.5}$$

where $\tilde{J}_c(\mathbf{\Omega},\bar{\lambda})$ is the corresponding Fourier transform of $J_c(\mathbf{\Omega},\bar{\rho})$.

To put Eq. (2.5) in the conventional form in terms of the coherent wave part $\tilde{I}_c(\mathbf{\Omega},\bar{\lambda})$ and incoherent wave part $\tilde{I}_s(\mathbf{\Omega},\bar{\lambda})$, we introduce the new quantity

$$\tilde{J}_s(\mathbf{\Omega},\bar{\lambda}) = \int d\,\mathbf{\Omega}'\sigma(\mathbf{\Omega}|\mathbf{\Omega}')\tilde{I}(\mathbf{\Omega}',\bar{\lambda}), \tag{2.6}$$

and multiply both sides of Eq. (2.5) with

$$\tilde{U}(\mathbf{\Omega},\bar{\lambda}) = [i(\nu/c - \mathbf{\Omega}\cdot\boldsymbol{\lambda}) + \gamma_t(\mathbf{\Omega})]^{-1}, \tag{2.7}$$

whence, as a particular solution of Eq. (2.5),

$$\tilde{I}(\mathbf{\Omega},\bar{\lambda}) = \tilde{I}_s(\mathbf{\Omega},\bar{\lambda}) + \tilde{I}_c(\mathbf{\Omega},\bar{\lambda}), \tag{2.8}$$
$$\tilde{I}_s(\mathbf{\Omega},\bar{\lambda}) = \tilde{U}\tilde{J}_s(\mathbf{\Omega},\bar{\lambda}), \quad \tilde{I}_c(\mathbf{\Omega},\bar{\lambda}) = \tilde{U}\tilde{J}_c(\mathbf{\Omega},\bar{\lambda}), \tag{2.9}$$

while the general solution is obtained by adding to the coherent part $\tilde{I}_c(\mathbf{\Omega},\bar{\lambda})$ a function, given in terms of the complex δ function by

$$\delta[i(\nu/c - \mathbf{\Omega}\cdot\boldsymbol{\lambda}) + \gamma_t(\mathbf{\Omega})] f(\mathbf{\Omega},\bar{\lambda}),$$

where $f(\mathbf{\Omega},\bar{\lambda})$ is an arbitrary function to be determined by the boundary condition.

Here, by Eq. (2.6), the equation of $\tilde{J}_s$ is found to be

$$\tilde{J}_s(\mathbf{\Omega},\bar{\lambda}) = \int d\,\mathbf{\Omega}'\sigma(\mathbf{\Omega}|\mathbf{\Omega}')\tilde{U}\tilde{J}_s(\mathbf{\Omega}',\bar{\lambda}) + \tilde{J}'_c(\mathbf{\Omega},\bar{\lambda}), \tag{2.10}$$

where

$$\tilde{J}'_c(\mathbf{\Omega},\bar{\lambda}) = \int d\,\mathbf{\Omega}'\sigma(\mathbf{\Omega}|\mathbf{\Omega}')\tilde{I}_c(\mathbf{\Omega}',\bar{\lambda}) = \int d\,\mathbf{\Omega}'\sigma(\mathbf{\Omega}|\mathbf{\Omega}')\tilde{U}\tilde{J}_c(\mathbf{\Omega}',\bar{\lambda}). \tag{2.11}$$

Equation (2.10) forms an integral equation of the Fredholm type with respect to $\mathbf{\Omega}$, where $\bar{\lambda}$ is kept to be constant and $\tilde{U}(\mathbf{\Omega},\bar{\lambda})$ is regarded as the weighting function, and is the basic equation in the following.

In the same way as in Eqs. (2.8) and (2.9), we obtain from Eq. (2.1) the expressions of total intensity and flux vector components, given in the form

$$\tilde{I}(\bar{\lambda}) = \int d\,\mathbf{\Omega}\,\tilde{I}(\mathbf{\Omega},\bar{\lambda}) = \tilde{I}_s(\bar{\lambda}) + \tilde{I}_c(\bar{\lambda}),$$
$$\tilde{I}_j(\bar{\lambda}) = \int d\,\mathbf{\Omega}\,\Omega_j\tilde{I}(\mathbf{\Omega},\bar{\lambda}) = \tilde{I}_{sj}(\bar{\lambda}) + \tilde{I}_{cj}(\bar{\lambda}), \quad j = 1,2,3, \tag{2.12}$$

where

$$\tilde{I}_s(\bar{\lambda}) = \int d\,\mathbf{\Omega}\tilde{U}\tilde{J}_s(\mathbf{\Omega},\bar{\lambda}),$$
$$\tilde{I}_{sj}(\bar{\lambda}) = \int d\,\mathbf{\Omega}\,\Omega_j\tilde{U}\tilde{J}_s(\mathbf{\Omega},\bar{\lambda}), \text{ etc.,} \tag{2.13}$$

in which, as in Eqs. (2.10) and (2.11), $\tilde{U}(\mathbf{\Omega},\bar{\lambda})$ plays the role of the weighting function when integrating with respect to $\mathbf{\Omega}$.

On the other hand, from the equation of continuity (2.2), we find

$$-i\boldsymbol{\lambda}\cdot\tilde{\mathbf{I}}(\bar{\lambda}) + (i\nu/c + \tilde{\gamma}_a) I(\bar{\lambda}) = \tilde{J}_c(\bar{\lambda}) \equiv \int d\,\mathbf{\Omega}\,\tilde{J}_c(\mathbf{\Omega},\bar{\lambda}), \tag{2.14}$$

where

$$\tilde{\gamma}_a(\bar{\lambda}) = \int d\,\mathbf{\Omega}\gamma_a(\mathbf{\Omega})\tilde{I}(\mathbf{\Omega},\bar{\lambda}) \Big/ \int d\,\mathbf{\Omega}\tilde{I}(\mathbf{\Omega},\bar{\lambda}), \tag{2.15}$$

while the corresponding equation of continuity for $\tilde{I}_s$ alone becomes

$$-i\boldsymbol{\lambda}\cdot\tilde{\mathbf{I}}_s(\bar{\lambda}) + (i\nu/c + \tilde{\gamma}_{as})\tilde{I}_s(\bar{\lambda}) = \tilde{J}'_c(\bar{\lambda}), \tag{2.16}$$

where

$$\tilde{J}'_c(\bar{\lambda}) = \int d\,\mathbf{\Omega}\tilde{J}'_c(\mathbf{\Omega},\bar{\lambda}) = \int d\,\mathbf{\Omega}\,\gamma(\mathbf{\Omega})\tilde{I}_c(\mathbf{\Omega},\bar{\lambda}), \tag{2.17}$$

and $\tilde{\gamma}_{as}$ is defined in the same way as in Eq. (2.15) with $\tilde{I}$ replaced by $\tilde{I}_s$. Equation (2.16) is obtained by multiplying both sides of the first equation in Eq. (2.9) by $\tilde{U}^{-1}$ and performing the $\mathbf{\Omega}$ integration over the entire solid angle.

A. Diffusion equation in case of isotropic scattering

To see how the diffusion equation is derived from the above equations, we consider the simplest case of isotropic scattering in which $\sigma(\mathbf{\Omega}|\mathbf{\Omega}') = (4\pi)^{-1}\gamma$ is independent of $\mathbf{\Omega}$ and $\mathbf{\Omega}'$, and γ_a is also constant, and in which an exact expression is obtainable for the diffusion term of the wave. From Eqs. (2.10) and (2.11), it is immediately found that $\tilde{J}_s(\mathbf{\Omega},\bar{\lambda})$ and $\tilde{J}'_c(\mathbf{\Omega},\bar{\lambda})$ are independent of $\mathbf{\Omega}$, yielding

$$\tilde{J}_s(\mathbf{\Omega},\bar{\lambda}) = (4\pi)^{-1}\tilde{J}_s(\bar{\lambda}), \quad \tilde{J}'_c(\mathbf{\Omega},\bar{\lambda}) = (4\pi)^{-1}J'_c(\bar{\lambda}), \tag{2.18}$$

and hence, by Eq. (2.13), also

$$\tilde{I}_s(\bar{\lambda}) = (4\pi)^{-1}\int d\,\mathbf{\Omega}\tilde{U}(\mathbf{\Omega},\bar{\lambda})\tilde{J}_s(\bar{\lambda}),$$

$$\tilde{I}_{sj}(\bar{\lambda}) = (4\pi)^{-1}\int d\,\mathbf{\Omega}\Omega_j\tilde{U}(\mathbf{\Omega},\bar{\lambda})\tilde{J}_s(\bar{\lambda}),\quad j=1,2,3, \tag{2.19}$$

whereas $\tilde{J}_c(\mathbf{\Omega},\bar{\lambda})$ is a given function of $\mathbf{\Omega}$ in contrast with the $\mathbf{\Omega}$ independent $\tilde{J}_s$ and $\tilde{J}_c'$.

Here, on substituting Eq. (2.18) into (2.10), we obtain the equation of $\tilde{J}_s(\bar{\lambda})$ of the form

$$[1-A(\bar{\lambda})]\tilde{J}_s(\bar{\lambda}) = \tilde{J}_c'(\bar{\lambda}), \tag{2.20}$$

with

$$A(\bar{\lambda}) = (4\pi)^{-1}\int d\,\mathbf{\Omega}\gamma\tilde{U}(\mathbf{\Omega},\bar{\lambda}), \tag{2.21}$$

which becomes, on employing Eq. (2.7),

$$A(\bar{\lambda}) = \frac{i\gamma}{2\lambda}\log\left[\frac{i(\nu/c-\lambda)+\gamma_t}{i(\nu/c+\lambda)+\gamma_t}\right],\quad \lambda=|\boldsymbol{\lambda}|, \tag{2.22}$$

where the function log[ ] tends to zero as $\lambda\to 0$.

On the other hand, by Eq. (2.21), $\tilde{I}_s(\bar{\lambda})$ in Eq. (2.19) can be exhibited by

$$\tilde{I}_s(\bar{\lambda}) = \gamma^{-1}A(\bar{\lambda})\tilde{J}_s(\bar{\lambda}), \tag{2.23}$$

which gives, in terms of a particular solution of Eq. (2.20),

$$\tilde{I}_s(\bar{\lambda}) = \gamma^{-1}[1-A(\bar{\lambda})]^{-1}A(\bar{\lambda})\tilde{J}_c'(\bar{\lambda}). \tag{2.24}$$

Hence, in space–time coordinates,

$$\tilde{I}_s(\bar{\rho}) = \gamma^{-1}(2\pi)^{-4}\int d\bar{\lambda}\,[1-A]^{-1}A(\bar{\lambda})\tilde{J}_c'(\bar{\lambda}) \times\exp[i(\nu t-\boldsymbol{\lambda}\cdot\boldsymbol{\rho})],\quad d\bar{\lambda}=d\,\boldsymbol{\lambda}\,d\nu, \tag{2.25}$$

which gives the solution in the infinite space.

In the integral (2.25), we assume that $\tilde{J}_c'(\bar{\rho})$ is distributed within a small domain of $|\gamma\bar{\rho}|\leqslant 1$ in space and time, so that $\tilde{J}_c'(\bar{\lambda})$ can be regarded to be nearly constant in the important range of integration of $|\bar{\lambda}/\gamma|\ll 1$ when $|\gamma\bar{\rho}|\gg 1$. Here, to perform the ν integration for given $\boldsymbol{\lambda}$, first, we note that the integrand is regular everywhere on the ν plane with a cut drawn between the two points at (Fig. 1)

$$\nu/c = \pm\lambda + i\gamma_t \tag{2.26}$$

on the upper half-plane, except at one pole, say of $\nu=\omega(\lambda)$, given by $1-A(\bar{\lambda})=0$. Here, by Eq. (2.22), the pole is exactly at

$$i\omega/c = -\gamma_t + \lambda\cot(\lambda/\gamma) = -\gamma_a - \gamma\left[\frac{1}{3}\left(\frac{\lambda}{\gamma}\right)^2 + \frac{1}{45}\left(\frac{\lambda}{\gamma}\right)^4 + \frac{2}{945}\left(\frac{\lambda}{\gamma}\right)^6+\cdots\right],\quad \mathrm{Im}(\omega)>0, \tag{2.27}$$

which tends to zero as $\lambda\to 0$ when $\gamma_a=0$. Also, the integrand tends to zero as $\nu\to\infty$ since $A(\bar{\lambda})\to 0$.

Thus, when $t>0$, the ν integral is found to be given in terms of the residue value at the pole and of the contour integral around the branch cut, while the contributions from singularities of $\tilde{J}_c'(\bar{\lambda})$ are negligible when $\gamma ct\gg 1$, since they are to be distributed in the range $|\nu/c\gamma|\gtrsim 1$ by the assumption. The last situation is the same also for the contour integral, having the factor $\exp(-\gamma_t ct)$, as is evident from the branch points given by Eq. (2.26). Thus, when $\gamma ct\gg 1$, the dominant contribution to the integral is found to be made by the pole and, in its neighborhood, $A(\bar{\lambda})$ is given by

$$1-A(\bar{\lambda}) \sim -\left.\frac{\partial A}{\partial\nu}\right|_{\nu=\omega}[\nu-\omega(\lambda)],\quad \nu\sim\omega, \tag{2.28}$$

$$\left.\frac{\partial A}{\partial\nu}\right|_{\nu=\omega} = -i(c\gamma)^{-1}w^{-1}(\lambda),$$

$$w(\lambda) = 1+\frac{1}{3}\left(\frac{\lambda}{\gamma}\right)^2 + O\left[\left(\frac{\lambda}{\gamma}\right)^4\right],$$

and hence, by Eq. (2.27),

$$1-A(\bar{\lambda}) \sim \left[\frac{1}{\gamma}\left(i\frac{\nu}{c}+\gamma_a\right)+\frac{1}{3}\left(\frac{\lambda}{\gamma}\right)^2 + \frac{1}{45}\left(\frac{\lambda}{\gamma}\right)^4+\cdots\right]w^{-1}(\lambda). \tag{2.29}$$

Thus, substituting the expression (2.29) into the integral (2.25) to the order of $(\lambda/\gamma)^2$, we obtain, after the ν integration, the $\boldsymbol{\lambda}$ integral given by

$$I_s(\bar{\rho}) \sim c\gamma(2\pi)^{-3}\int d\,\boldsymbol{\lambda}\,w(\lambda)\tilde{J}_c'(\boldsymbol{\lambda}) \times\exp[-(\gamma_a+(3\gamma)^{-1}\lambda^2)ct - i\boldsymbol{\lambda}\cdot\boldsymbol{\rho}],\quad \gamma ct\gg 1, \tag{2.30}$$

and $I_s(\bar{\rho})=0$ for $t<0$. Here, since $\gamma ct\gg 1$ and $\tilde{J}_c'(\boldsymbol{\lambda}) \equiv\tilde{J}_c'(\bar{\lambda})|_{\nu=\omega}$ is nearly constant over the range $|\boldsymbol{\lambda}/\gamma|\ll 1$, the integration of Eq. (2.30) can be effected by use of the saddle point approximation, yielding

$$I_s(\bar{\rho}) \sim c\gamma w(\boldsymbol{\lambda}_s)\tilde{J}_c'(\boldsymbol{\lambda}_s)[3\gamma/4\pi ct]^{3/2} \times\exp[-\gamma_a ct-(3\gamma/4ct)\boldsymbol{\rho}^2],\quad \boldsymbol{\lambda}_s = -i\gamma(3\boldsymbol{\rho}/2ct), \tag{2.31}$$

where $\boldsymbol{\lambda}_s$ is the saddle point. Here, since the assumption has been made that $|\nu/c\gamma|\sim 3^{-1}|\lambda/\gamma|^2\ll 1$ in the derivation of Eq.

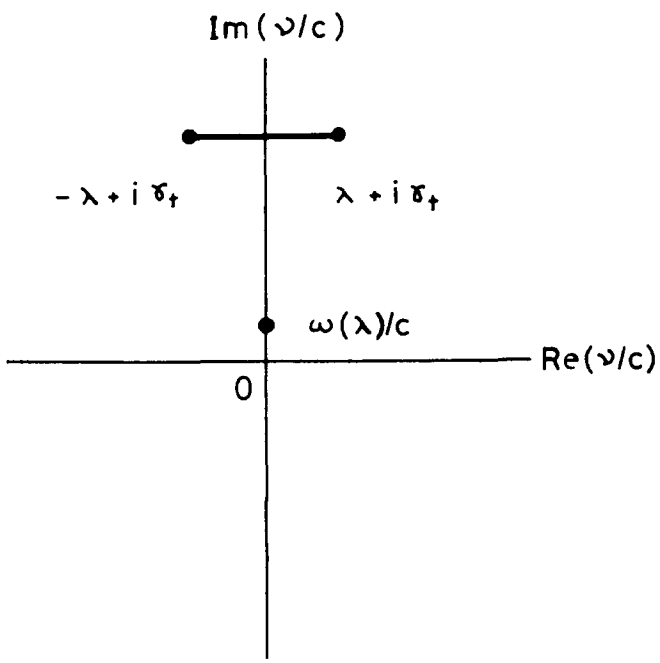


FIG. 1. Branch cut and pole for integral (2.25).

(2.30), the expression (2.31) is valid for $\gamma ct \gg 1$ and $|\lambda_s/\gamma|^2 = (3\rho/2ct)^2 \ll 1$.

Also, it is to be noted that, in the integrand of Eq. (2.30), the function $w(\lambda)$ is combined with $\tilde{J}_c'(\lambda)$ to form an effective source, given by $w\tilde{J}_c'(\lambda)$.

This is also the general case of anisotropic scattering to be treated in the following, and the effect of $w(\lambda) \sim 1$ will not be considered any further with the understanding that $\tilde{J}_c'(\bar{\rho})$ is to be replaced by the effective one when necessary.

So far we have seen that deriving the diffusion equation is reduced to finding the root of ν obeying $1 - A(\bar{\lambda}) = 0$ as a function of λ, $\omega(\lambda)$, which tends to zero as $\lambda \to 0$ when $\gamma_a = 0$, or, more simply, $\omega(\lambda)$ in power series of λ/γ to the first few orders. Therefore, $\omega(\lambda)$ is found systematically as follows: First, $\tilde{U}(\Omega,\bar{\lambda})$ in Eq. (2.7) is expanded in a power series of the form

$$\tilde{U}(\Omega,\bar{\lambda}) = \gamma^{-1}(\Omega)[1 + h(\Omega,\bar{\lambda}) + h^2(\Omega,\bar{\lambda}) + \cdots], \tag{2.32}$$

with

$$h(\Omega,\bar{\lambda}) = [i(\Omega\cdot\lambda - \nu/c) - \gamma_a(\Omega)]/\gamma(\Omega), \tag{2.33}$$

where h is of the order of $|\lambda/\gamma|$, $|\nu/c\gamma|$, and γ_a/γ, or of $|\bar{\lambda}/\gamma|$ for short.

Here, in the particular case of isotropic scattering, the substitution of Eq. (2.32) with (2.33) into Eq. (2.21) yields a power series of $\bar{\lambda}/\gamma$, given by

$$\begin{aligned} 1 - A(\bar{\lambda}) = {} & \frac{\gamma_a}{\gamma} + i\frac{\nu}{c\gamma}\left[1 - i\frac{\nu}{c\gamma} - \frac{1}{3}\left(\frac{\lambda}{\gamma}\right)^2\right] \\ & + \frac{1}{3}\left(\frac{\lambda}{\gamma}\right)^2 - i\frac{2}{3}\left(\frac{\lambda}{\gamma}\right)^2\left(\frac{\nu}{c\gamma}\right) \\ & - \frac{1}{5}\left(\frac{\lambda}{\gamma}\right)^4 - \left(\frac{\lambda}{\gamma}\right)^2\left(\frac{\nu}{c\gamma}\right)^2, \end{aligned} \tag{2.34}$$

to the order of $(\lambda/\gamma)^4$, $(\nu/c\gamma)^2$, and γ_a/γ. Therefore, the root $\omega(\lambda)$ is obtained in a power series of $(\lambda/\gamma)^2$ of the form

$$i\omega/c\gamma = \sum_{n=0}^{\infty} f_n(\lambda/\gamma)^{2n}, \tag{2.35}$$

where the unknown f_n is found by substituting $\nu = \omega$ into Eq. (2.34) for $1 - A = 0$ and equating the resulting coefficient of $(\lambda/\gamma)^{2n}$, $n = 0, 1, 2, \ldots$, to zero. Thus, to the order of $(\lambda/\gamma)^2$,

$$i\omega/c\gamma = -\gamma_a/\gamma - 3^{-1}(\lambda/\gamma)^2, \tag{2.36}$$

and the term of order $(\lambda/\gamma)^4$, to be added on the right-hand side, is obtained by the substitution of ω of Eq. (2.36) into the higher order terms on the right-hand side of Eq. (2.34), yielding $-45^{-1}(\lambda/\gamma)^4$ in accordance with that given by Eq. (2.27). Also, the factor $\partial A/\partial\nu$ in Eq. (2.28) is obtained from Eq. (2.34) to the order of $(\lambda/\gamma)^2$.

In the same way, on employing Eq. (2.19), the corresponding expressions of $\tilde{I}_s(\bar{\lambda})$ and $\tilde{I}_{sj}(\bar{\lambda})$ valid in neighborhood of the pole become

$$\begin{aligned} & \tilde{I}_s(\bar{\lambda}) = \gamma^{-1}A(\bar{\lambda})\tilde{J}_s(\bar{\lambda}) \sim \gamma^{-1}\tilde{J}_s(\bar{\lambda}), \\ & \tilde{I}_{sj}(\bar{\lambda}) \sim i(3\gamma^2)^{-1}\lambda_j[1 + 15^{-1}(\lambda/\gamma)^2 + O\{(\lambda/\gamma)^4\}]\tilde{J}_s(\bar{\lambda}), \end{aligned} \tag{2.37}$$

where ν appears only in $\tilde{J}_s(\bar{\lambda})$. Here, it will be noted that the substitution of Eq. (2.37) into the equation of continuity (2.16) reproduces Eq. (2.20) with $A(\bar{\lambda})$ given by Eq. (2.29), except for the factor $w^{-1}(\lambda) \sim 1$.

Thus, Eqs. (2.20), (2.29), and (2.37) are found to be consistent to each other and are expressed in the space-time coordinates by

$$\left[c^{-1}\frac{\partial}{\partial t} + \gamma_a - (3\gamma)^{-1}\left(\frac{\partial}{\partial\rho}\right)^2 + (45\gamma^3)^{-1}\left(\frac{\partial}{\partial\rho}\right)^4\right]I_s(\bar{\rho}) = \tilde{J}_c'(\bar{\rho}), \tag{2.38}$$

$$I_{sj}(\bar{\rho}) = -(3\gamma)^{-1}\frac{\partial}{\partial\rho_j}\left[1 - (15\gamma^2)^{-1}\left(\frac{\partial}{\partial\rho}\right)^2\right]I_s(\bar{\rho}), \quad j = 1, 2, 3, \tag{2.39}$$

giving the diffusion equation and the flux vector in terms of $I_s(\bar{\rho})$, respectively, to the order of $(\lambda/\gamma)^4$.

In the general case of anisotropic scattering, it will be shown in Sec. III that there exists a set of functions $A(\bar{\lambda})$, each of which obeys an equation of the form (2.20), but that, except for a special case, only one of the $A(\bar{\lambda})$'s gives the root $\nu = \omega(\lambda)$ of $1 - A(\bar{\lambda}) = 0$, tending to zero as $\lambda \to 0$ when $\gamma_a = 0$, and all the other $A(\bar{\lambda})$'s give those of tending to finite positive imaginary values in the same limit. This implies that, for the members of $A(\bar{\lambda})$ having roots of the latter class, the integrals corresponding to Eq. (2.25) have factors rapidly decreasing in time in the domain of diffusion and therefore are negligible as compared with that having the former $\omega(\lambda)$.

B. Characteristics of $I_c(\bar{\rho})$ and $\tilde{J}_c'(\bar{\rho})$

By Eq. (2.9), the intensity of coherent wave $I_c(\bar{\rho})$ can be expressed in terms of $U(\Omega,\bar{\rho})$, the Fourier inversion of $\tilde{U}(\Omega,\bar{\lambda})$, which, according to Eq. (2.7), is given by

$$\begin{aligned} U(\Omega,\bar{\rho}) &= (2\pi)^{-4}\int d\bar{\lambda}\tilde{U}(\Omega,\bar{\lambda})\exp[i(\nu t - \lambda\cdot\rho)] \\ &= c\exp(-\gamma_t\rho)\delta^3(\rho - \Omega ct), \quad t > 0, \\ &= 0, \quad t < 0, \quad \rho = |\rho|. \end{aligned} \tag{2.40}$$

Thus, with the aid of the relation

$$\int d\,\mathbf{\Omega}\delta^3(\boldsymbol{\rho}-\mathbf{\Omega}ct\,)f(\mathbf{\Omega})=\rho^{-2}\delta(\rho-ct\,)f(\,\boldsymbol{\rho}/\rho), \tag{2.41}$$

we obtain from Eq. (2.9)

$$\begin{aligned} I_c(\bar{\rho}) &= \int d\bar{\rho}'\int d\,\mathbf{\Omega}U(\mathbf{\Omega},\bar{\rho}-\bar{\rho}')J_c(\mathbf{\Omega},\boldsymbol{\rho}',t') \\ &= \int d\boldsymbol{\rho}'|\boldsymbol{\rho}-\boldsymbol{\rho}'|^{-2}\exp(-\gamma_t|\boldsymbol{\rho}-\boldsymbol{\rho}'|) \\ &\quad\times J_c[(\boldsymbol{\rho}-\boldsymbol{\rho}')/|\boldsymbol{\rho}-\boldsymbol{\rho}'|,\boldsymbol{\rho}',t-|\boldsymbol{\rho}-\boldsymbol{\rho}'|/c], \end{aligned} \tag{2.42}$$

which shows that $I_c(\bar{\rho})$ is different from zero only within the domain of distance of the order γ_t^{-1} from the source J_c.

Here, to see the difference between the original source $J_c(\mathbf{\Omega},\bar{\rho})$ in Eq. (1.1) and the secondary source $\tilde{J}_c'(\mathbf{\Omega},\bar{\rho})$ defined by Eq. (2.11), we consider a point source of the form

$$J_c(\mathbf{\Omega},\bar{\rho})=f(\mathbf{\Omega})\delta^4(\bar{\rho}), \tag{2.43}$$

and, on following the procedure of Eq. (2.42), use Eq. (2.17) to obtain

$$\begin{aligned} J_c'(\bar{\rho}) &= \int d\,\mathbf{\Omega}J_c'(\mathbf{\Omega},\bar{\rho}) \\ &= \gamma\rho^{-2}\exp(-\gamma_t\rho)f(\boldsymbol{\rho}/\rho)\delta(t-\rho/c), \end{aligned} \tag{2.44}$$

where γ and γ_t may be functions of $\mathbf{\Omega}$ set equal to $\boldsymbol{\rho}/\rho$.

Equation (2.44) shows that $J_c'(\bar{\rho})$ moves with the velocity c and is different from zero only in the domain of $\rho \lesssim \gamma_t^{-1}$ in space and of $t\lesssim(c\gamma_t)^{-1}$ in time. Therefore, when viewed from the space–time point of $\gamma_t\rho\gg 1$ and $\gamma_t ct\gg 1$, $J_c'(\bar{\rho})$ can be regarded also as a point source and has the relations

$$\int d\bar{\rho}J_c'(\bar{\rho})=\int d\,\mathbf{\Omega}(\gamma/\gamma_t)f(\mathbf{\Omega})\simeq\int d\bar{\rho}J_c(\bar{\rho}). \tag{2.45}$$

III. EIGENFUNCTION EXPANSIONS OF PHYSICAL QUANTITIES

The integral equation (2.10) is of Fredholm type with the weighting function $\tilde{U}$ and therefore the solutions can be expanded by the set of eigenfunctions $f_A(\mathbf{\Omega},\bar{\lambda})$ defined by the eigenvalue equation

$$\begin{aligned} &\int d\,\mathbf{\Omega}'\sigma(\mathbf{\Omega}|\mathbf{\Omega}')\tilde{U}(\mathbf{\Omega}',\bar{\lambda}\,)f_A(\mathbf{\Omega}',\bar{\lambda}\,)=A(\bar{\lambda}\,)f_A(\mathbf{\Omega},\bar{\lambda}\,), \\ &\int d\,\mathbf{\Omega}'\bar{f}_A(\mathbf{\Omega}',\bar{\lambda}\,)\tilde{U}(\mathbf{\Omega}',\bar{\lambda}\,)\sigma(\mathbf{\Omega}'|\mathbf{\Omega})=A(\bar{\lambda}\,)\bar{f}_A(\mathbf{\Omega},\bar{\lambda}\,). \end{aligned} \tag{3.1}$$

Here, $A(\bar{\lambda}\,)$ is the eigenvalue and $\bar{f}_A(\mathbf{\Omega},\bar{\lambda}\,)$ is the adjoint eigenfunction, normalized according to

$$\int d\,\mathbf{\Omega}\;\bar{f}_A\tilde{U}f_B(\mathbf{\Omega},\bar{\lambda}\,)=\delta_{AB}, \tag{3.2}$$

and hence by the completeness of the set of eigenfunctions, it holds the relation

$$\sum_A f_A(\mathbf{\Omega},\bar{\lambda}\,)\bar{f}_A(\mathbf{\Omega}',\bar{\lambda}\,)=\delta_U(\mathbf{\Omega}-\mathbf{\Omega}'), \tag{3.3}$$

where $\delta_U(\mathbf{\Omega}-\mathbf{\Omega}')=0$ for $\mathbf{\Omega}\neq\mathbf{\Omega}'$ and

$$\int d\,\mathbf{\Omega}\tilde{U}\delta_U(\mathbf{\Omega}-\mathbf{\Omega}')=1. \tag{3.4}$$

In terms of the eigenvalues and the eigenfunctions, $\sigma(\mathbf{\Omega}|\mathbf{\Omega}')$ can be exhibited by

$$\sigma(\mathbf{\Omega}|\mathbf{\Omega}')=\sum_A A(\bar{\lambda}\,)f_A(\mathbf{\Omega},\bar{\lambda}\,)\bar{f}_A(\mathbf{\Omega}',\bar{\lambda}\,), \tag{3.5}$$

and, in the same way, $\tilde{J}_s(\mathbf{\Omega},\bar{\lambda}\,)$ by

$$\tilde{J}_s(\mathbf{\Omega},\bar{\lambda}\,)=\sum_A C_A(\bar{\lambda}\,)f_A(\mathbf{\Omega},\bar{\lambda}\,). \tag{3.6}$$

Here, the equation of C_A is obtained by the substitution of Eq. (3.6) into Eq. (2.10) and the subsequent use of the orthogonality (3.2). Thus,

$$[1-A(\bar{\lambda}\,)]C_A(\bar{\lambda}\,)=\int d\,\mathbf{\Omega}\bar{f}_A\tilde{U}\tilde{J}_c'(\mathbf{\Omega},\bar{\lambda}\,). \tag{3.7}$$

Now, on performing the $\mathbf{\Omega}$ integration on both sides of Eq. (3.6), we obtain an eigenfunction series of the form

$$\tilde{J}_s(\bar{\lambda}\,)\equiv\int d\,\mathbf{\Omega}\;\tilde{J}_s(\mathbf{\Omega},\bar{\lambda}\,)=\sum_A\tilde{J}_A(\bar{\lambda}\,), \tag{3.8}$$

where

$$\tilde{J}_A(\bar{\lambda}\,)=C_A(\bar{\lambda}\,)\int d\,\mathbf{\Omega}\;f_A(\mathbf{\Omega},\bar{\lambda}\,) \tag{3.9}$$

obeys the equation

$$[1-A(\bar{\lambda}\,)]\tilde{J}_A(\bar{\lambda}\,)=\tilde{J}_{cA}'(\bar{\lambda}\,), \tag{3.10}$$

with a source term $\tilde{J}_{cA}'(\bar{\lambda}\,)$, defined by

$$\begin{aligned} \tilde{J}_{cA}'(\bar{\lambda}\,) &= \int d\,\mathbf{\Omega}\;f_A(\mathbf{\Omega},\bar{\lambda}\,)\int d\,\mathbf{\Omega}'\;\bar{f}_A\tilde{U}\tilde{J}_c'(\mathbf{\Omega}',\bar{\lambda}\,) \\ &= A(\bar{\lambda}\,)\int d\,\mathbf{\Omega}\;f_A(\mathbf{\Omega},\bar{\lambda}\,)\int d\,\mathbf{\Omega}'\;\bar{f}_A\tilde{U}\tilde{J}_c(\mathbf{\Omega}',\bar{\lambda}\,). \end{aligned} \tag{3.11}$$

Here, it will be noted that Eq. (3.10) has the same form as Eq. (2.20) for the case of isotropic scattering.

In the space–time coordinates $\bar{\rho}$, Eq. (3.8) is expressed by

$$J_s(\bar{\rho})\equiv\int d\,\mathbf{\Omega}\,J_s(\mathbf{\Omega},\bar{\rho})=\sum_A J_A(\bar{\rho}), \tag{3.12}$$

where, by Eq. (3.10), $J_A(\bar{\rho})$ obeys the equation

$$[1-A(i\partial/\partial\bar{\rho})]J_A(\bar{\rho})=\tilde{J}_{cA}'(\bar{\rho}), \tag{3.13}$$

in terms of the notation $\partial/\partial\bar{\rho}=(\partial/\partial\boldsymbol{\rho},\;-\partial/\partial t)$.

In order to obtain the corresponding eigenfunction series of wave intensity $I_s(\bar{\rho})$ and flux vector $\mathbf{I}_s(\bar{\rho})=(I_{s_1},I_{s_2},I_{s_3})(\bar{\rho})$, we introduce new quantities, defined by

$$\begin{aligned} &\beta_A(\bar{\lambda}\,)=\langle\tilde{U}(\mathbf{\Omega},\bar{\lambda}\,)\rangle_A, \\ &\alpha_{Aj}(\bar{\lambda}\,)=\langle\Omega_j\tilde{U}(\mathbf{\Omega},\bar{\lambda}\,)\rangle_A,\quad j=1,2,3, \end{aligned} \tag{3.14}$$

in terms of the notation

$$\langle Q(\mathbf{\Omega})\rangle_A=\int d\,\mathbf{\Omega}Q(\mathbf{\Omega})f_A(\mathbf{\Omega},\bar{\lambda}\,)\Big/\int d\,\mathbf{\Omega}f_A(\mathbf{\Omega},\bar{\lambda}\,). \tag{3.15}$$

Hence, on substituting Eq. (3.6) into (2.13), we obtain

$$\begin{aligned} \tilde{I}_s(\bar{\lambda}\,) &= \sum_A C_A(\bar{\lambda}\,)\int d\,\mathbf{\Omega}\tilde{U}f_A(\mathbf{\Omega},\bar{\lambda}\,) \\ &= \sum_A C_A(\bar{\lambda}\,)\beta_A(\bar{\lambda}\,)\int d\,\mathbf{\Omega}f_A(\mathbf{\Omega},\bar{\lambda}\,), \end{aligned}$$

which, by Eq. (3.9), can be expressed by a series of the form

$$\tilde{I}_s(\bar{\lambda}) = \sum_A \tilde{I}_A(\bar{\lambda}), \quad \tilde{I}_A(\bar{\lambda}) = \beta_A(\bar{\lambda})\tilde{J}_A(\bar{\lambda}) \tag{3.16}$$

in terms of $\tilde{J}_A(\bar{\lambda})$. In the same way,

$$\tilde{I}_{sj}(\bar{\lambda}) = \sum_A \tilde{I}_{Aj}(\bar{\lambda}), \quad \tilde{I}_{Aj}(\bar{\lambda}) = \alpha_{Aj}(\bar{\lambda})\tilde{J}_A(\bar{\lambda}). \tag{3.17}$$

Thus, in the space–time coordinates, we formally obtain

$$I_s(\bar{\rho}) = \sum_A \beta_A(i\partial/\partial\bar{\rho})J_A(\bar{\rho}),$$

$$\tilde{I}_{sj}(\bar{\rho}) = \sum_A \alpha_{Aj}(i\partial/\partial\bar{\rho})J_A(\bar{\rho}), \quad j = 1,2,3. \tag{3.18}$$

Here, $J_A(\bar{\rho})$ obeys the equation (3.13) and its general solution is given by

$$J_A(\bar{\rho}) = (2\pi)^{-4}\int d\bar{\lambda}\exp[i(\nu t - \boldsymbol{\lambda}\cdot\boldsymbol{\rho})] \times\{[1 - A(\bar{\lambda})]^{-1}\tilde{J}'_{cA}(\bar{\lambda}) + \delta[1 - A(\bar{\lambda})]g(\bar{\lambda})\}, \quad d\bar{\lambda} = d\boldsymbol{\lambda}\, d\nu, \tag{3.19}$$

where $g(\bar{\lambda})$ is an arbitrary function to be determined by boundary conditions.

The usefulness of the eigenfunction expansions (3.8), (3.16), and (3.17) depends on their convergence properties, of course, and also on whether the eigenfunction $f_A(\boldsymbol{\Omega},\bar{\lambda})$ and eigenvalue $A(\bar{\lambda})$ can be obtained to a sufficiently good approximation. Here, in the eigenvalue equation (3.1), the variable $\bar{\lambda}$ appears only through the function $\tilde{U}(\boldsymbol{\Omega},\bar{\lambda})$ of Eq. (2.7), which can be regarded as a function of $\bar{\lambda}/\gamma$ instead of $\bar{\lambda}$. Therefore, the eigenvalue $A(\bar{\lambda})$ can also be regarded as a function of $\bar{\lambda}/\gamma$. To investigate the convergence properties under the condition of the diffusion approximation in which, as we have seen in the previous example of isotropic scattering, various equations are treated in the range of $|\bar{\lambda}/\gamma| \ll 1$, we first observe that, as $\bar{\lambda}/\gamma \to 0$,

$$\tilde{U}(\boldsymbol{\Omega},\bar{\lambda}) \to \gamma_t^{-1}(\boldsymbol{\Omega}) \sim \gamma^{-1}(\boldsymbol{\Omega}), \quad \gamma \gg \gamma_a, \tag{3.20}$$

and, correspondingly,

$$A(\bar{\lambda}) \to a, \quad f_A(\boldsymbol{\Omega},\bar{\lambda}) \to f_a(\boldsymbol{\Omega}). \tag{3.21}$$

Therefore, if the above limiting values are used in the integrand of Eq. (3.19), and if $a \neq 1$, the integral simply becomes $\delta^4(\bar{\rho})$ except for a constant factor, provided that the source $J'_c(\bar{\rho})$ is distributed within a small domain of $|\gamma\bar{\rho}| \lesssim 1$, giving $\tilde{J}'_c(\bar{\lambda})$ and $\tilde{J}'_{cA}(\bar{\lambda})$ nearly constant within the range $|\bar{\lambda}/\gamma| \ll 1$. Therefore, it follows that the contributions of these terms to the series (3.12) become negligible in the domain of space and time where the distances from the source are large enough to be $|\gamma\bar{\rho}| \gg 1$. This result implies that, in the series (3.12), the terms with the limiting eigenvalues $a \neq 1$ are of short range and, like the coherent wave $I_c(\bar{\rho})$, are terms decreasing rapidly in space and time. However, as will be be seen in Sec. IV, there is at least one eigenfunction $f_1(\boldsymbol{\Omega})$ having the eigenvalue $a = 1$, and the corresponding $J_A(\bar{\rho})$ of Eq. (3.19) is a long range function, giving the leading term in the domain of diffusion.

Here, on performing the $\boldsymbol{\Omega}$ integration on both sides of Eq. (3.1), we obtain, by use of Eq. (1.2),

$$\int d\boldsymbol{\Omega}\gamma(\boldsymbol{\Omega})\tilde{U}f_A(\boldsymbol{\Omega},\bar{\lambda}) = A(\bar{\lambda})\int d\boldsymbol{\Omega} f_A(\boldsymbol{\Omega},\bar{\lambda}), \tag{3.22}$$

where, on the left-hand side, Eq. (2.7) gives

$$\gamma(\boldsymbol{\Omega})\tilde{U}(\boldsymbol{\Omega},\bar{\lambda}) = 1 + (i\boldsymbol{\Omega}\cdot\boldsymbol{\lambda} - i\nu/c - \gamma_a)\tilde{U}(\boldsymbol{\Omega},\bar{\lambda}). \tag{3.23}$$

Therefore, on using Eq. (3.23) in (3.22), we obtain an expression of $1 - A(\bar{\lambda})$, given by

$$1 - A(\bar{\lambda}) = \langle(-i\boldsymbol{\Omega}\cdot\boldsymbol{\lambda} + i\nu/c + \gamma_a)\tilde{U}(\boldsymbol{\Omega},\bar{\lambda})\rangle_A, \tag{3.24}$$

in terms of the notation (3.15). Here, the right-hand side of Eq. (3.24) formally tends to zero as $\bar{\lambda} \to 0$ when $\gamma_a = 0$, but this never means that, as $\bar{\lambda} \to 0$, *all* the $A(\bar{\lambda})$'s tend to unity, since, according to the definition (3.15), $\langle Q\rangle_A$ may have a finite value as $\bar{\lambda} \to 0$ even when $Q \to 0$ if $\int d\boldsymbol{\Omega} f_A \to 0$, simultaneously. In fact, this is the case for most of the f_A's, as will be seen in Sec. IV.

If $\int d\boldsymbol{\Omega} f_a \neq 0$, on the other hand, then Eq. (3.24) shows that $A \to 1$ as $\bar{\lambda} \to 0$ and, in terms of β_A and $\boldsymbol{\alpha}_A = (\alpha_{A_1}, \alpha_{A_2}, \alpha_{A_3})$, defined by Eq. (3.14), $1 - A(\bar{\lambda})$ is exhibited by

$$1 - A(\bar{\lambda}) = -i\boldsymbol{\lambda}\cdot\boldsymbol{\alpha}_A(\bar{\lambda}) + (i\nu/c + \gamma_{aA})\beta_A(\bar{\lambda}), \tag{3.25}$$

where the absorption coefficient γ_{aA} is defined by

$$\gamma_{aA} = \langle\gamma_a\tilde{U}\rangle_A/\langle\tilde{U}\rangle_A. \tag{3.26}$$

Thus, substituting the expression (3.25) into Eq. (3.10), the equation of $\tilde{I}_A(\bar{\lambda})$, defined in Eq. (3.16), is found to be given by

$$[-i\beta_A^{-1}\boldsymbol{\alpha}_A\cdot\boldsymbol{\lambda} + i\nu/c + \gamma_{aA}]\tilde{I}_A(\bar{\lambda}) = \tilde{J}'_{cA}(\bar{\lambda}), \tag{3.27}$$

and the corresponding flux vector $\tilde{\mathbf{I}}_A(\bar{\lambda}) = (\tilde{I}_{A_1}, \tilde{I}_{A_2}, \tilde{I}_{A_3})(\bar{\lambda})$ by

$$\tilde{I}_{Aj}(\bar{\lambda}) = \beta_A^{-1}\alpha_{Aj}(\bar{\lambda})\tilde{I}_A(\bar{\lambda}), \quad j = 1,2,3, \tag{3.28}$$

in terms of $\tilde{I}_A(\bar{\lambda})$. Here, between $\tilde{I}_A$ and $\tilde{\mathbf{I}}_A$, it holds by Eq. (3.27) that

$$-i\boldsymbol{\lambda}\cdot\tilde{\mathbf{I}}_A(\bar{\lambda}) + (i\nu/c + \gamma_{aA})\tilde{I}_A(\bar{\lambda}) = \tilde{J}'_{cA}(\bar{\lambda}), \tag{3.29}$$

which represents the equation of continuity of each eigenfunction component. Indeed, on using Eq. (3.29) with (3.11), we find the equation of continuity (2.16) reproduced in terms of the eigenfunction series, or, in the space–time coordinates,

$$\frac{\partial}{\partial\boldsymbol{\rho}}\cdot\mathbf{I}_s(\bar{\rho}) + \left(c^{-1}\frac{\partial}{\partial t} + \gamma_{as}\right)I_s(\bar{\rho}) = J'_c(\bar{\rho}), \tag{3.30}$$

with the aid of the relation

$$\int d\boldsymbol{\Omega}\gamma_a(\boldsymbol{\Omega})\tilde{I}_s(\boldsymbol{\Omega},\bar{\lambda}) = \sum_A C_A(\bar{\lambda})\int d\boldsymbol{\Omega}\gamma_a\tilde{U}f_A(\boldsymbol{\Omega},\bar{\lambda}) = \sum_A\langle\gamma_a\tilde{U}\rangle_A\tilde{J}_A(\bar{\lambda}) = \sum_A\gamma_{aA}\tilde{I}_A(\bar{\lambda}), \tag{3.31}$$

by virtue of the definition (3.26).

So far we have not made any approximation and have found that, when $\int d\boldsymbol{\Omega} f_a \neq 0, A(\bar{\lambda}) \to 1$ as $\bar{\lambda} \to 0$. When this is the case, $A(\bar{\lambda})$ can be conveneintly obtained according to Eq. (3.25) with $\boldsymbol{\alpha}_A$ and β_A given by Eq. (3.14) and, to first order of $\bar{\lambda}$, the latter quantities can be given in the form

$$\alpha_{Aj}(\bar{\lambda}) = \alpha_j + i\left[\sum_{k=1}^{3}A_{jk}\lambda_k - A_j\nu/c\right], \quad j = 1,2,3, \tag{3.32}$$

$$\beta_A(\bar{\lambda}) = \beta + i\left[\sum_{k=1}^{3} B_k\lambda_k - B\,\nu/c\right].$$

Here, the necessary subscript A has been dropped on the right-hand side, and $\alpha, \beta \sim O[\gamma^{-1}]$, whereas A_{jk}, A_j, B_j, $B \sim O[\gamma^{-2}]$, as will been seen later in Eq. (4.21) and (4.22). Thus, on substituting Eq. (3.32) into (3.25), we obtain an expression of $1 - A(\bar{\lambda})$ to the order of $(\bar{\lambda}/\gamma)^2$.

The next task is to find the root $\nu = \omega(\bar{\lambda})$ obeying $1 - A(\bar{\lambda}) = 0$ in a power series of λ of the general form

$$i\omega/c = f_0 + \sum_{j=1}^{3} f_j\lambda_j + \sum_{j,k=1}^{3} f_{jk}\lambda_j\lambda_k + \cdots, \tag{3.33}$$

by following the procedure of Eq. (2.34) and (2.35). However, to obtain the result to the order of λ^2, we can follow the method described after Eq. (2.36), i.e., to the first order of $\bar{\lambda}$, Eq. (3.25) gives

$$1 - A(\bar{\lambda}) = -i\boldsymbol{\lambda}\cdot\boldsymbol{\alpha} + (i\nu/c + \gamma_{aA})\beta, \tag{3.34}$$

$$\gamma_{aA} = \int d\,\boldsymbol{\Omega}\gamma_a(\boldsymbol{\Omega})\gamma^{-1}f_1(\boldsymbol{\Omega}) \quad \int d\,\boldsymbol{\Omega}\gamma^{-1}f_1(\boldsymbol{\Omega}),$$

with $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$, and therefore, by putting $\nu/c = \boldsymbol{\alpha}\cdot\boldsymbol{\lambda}/\beta$ in the right-hand side of Eq. (3.32) and then substituting the result in Eq. (3.25), the root $\omega(\boldsymbol{\lambda})$ is obtained to the order of λ^2 and γ_a.

Thus, at the pole of $\nu = \omega(\boldsymbol{\lambda})$, $\alpha_{Aj}(\bar{\lambda})$, and $\beta_A(\bar{\lambda})$ in Eq. (3.32) are given by

$$\begin{aligned}\alpha_{Aj}(\boldsymbol{\lambda}) &\equiv \alpha_{Aj}(\bar{\lambda})\big|_{\nu=\omega} \\ &= \alpha_j + i\sum_{k=1}^{3}(A_{jk} - \beta^{-1}A_j\alpha_k)\lambda_k, \\ \beta_A(\boldsymbol{\lambda}) &\equiv \beta_A(\bar{\lambda})\big|_{\nu=\omega} \\ &= \beta + i\sum_{k=1}^{3}(B_k - \beta^{-1}B\alpha_k)\lambda_k,\end{aligned} \tag{3.35}$$

and hence, in neighborhood of the pole,

$$1 - A(\bar{\lambda}) \sim [-i\boldsymbol{\lambda}\cdot\boldsymbol{\alpha}_A(\bar{\lambda}) + (i\nu/c + \gamma_{aA})\beta_A(\bar{\lambda})]w^{-1}(\boldsymbol{\lambda}), \tag{3.36}$$

where the function $w(\boldsymbol{\lambda})$, defined by

$$\begin{aligned}&\frac{\partial}{\partial\nu}A(\bar{\lambda})\bigg|_{\nu=\omega} = -i\,c^{-1}\beta_A(\bar{\lambda})w^{-1}(\boldsymbol{\lambda}), \\ &w(\boldsymbol{\lambda}) = 1 - i\sum_{j=1}^{3}\beta^{-1}\{A_j - \beta^{-1}B\alpha_j\}\lambda_j + O[(\boldsymbol{\lambda}/\gamma)^2],\end{aligned} \tag{3.37}$$

corresponds to $w(\lambda)$ in Eq. (2.28).

Here, the final result can be given in a more compact form by Eq. (3.27) rather than by Eq. (3.10). Thus, in neighborhood of the pole $\nu = \omega(\boldsymbol{\lambda})$, $\tilde{I}_A(\bar{\lambda}) \sim \beta_A(\boldsymbol{\lambda})\tilde{J}_A(\bar{\lambda})$ [cf. Eq. (2.37)] and

$$\beta_A^{-1}\alpha_{Aj}(\boldsymbol{\lambda}) = \beta^{-1}\alpha_j + i\sum_{k=1}^{3} D_{jk}\lambda_k, \quad j = 1,2,3, \tag{3.38}$$

where, by Eq. (3.35),

$$D_{jk} = \beta^{-1}[A_{jk} - \beta^{-1}(A_j\alpha_k + \alpha_jB_k) + \beta^{-2}B\alpha_j\alpha_k], \tag{3.39}$$

and $D_{jk} = D_{kj}$ in the symmetrical case of $\sigma(\boldsymbol{\Omega}|\boldsymbol{\Omega}') = \sigma(\boldsymbol{\Omega}'|\boldsymbol{\Omega})$ (Sec. IV). Hence, the use of Eq. (3.36) yields the equation of $\tilde{I}_A$, given by

$$\begin{aligned}&[i\nu/c + \gamma_{aA} - i\beta^{-1}\boldsymbol{\alpha}\cdot\boldsymbol{\lambda} + \boldsymbol{\lambda}\cdot\mathbf{D}\cdot\boldsymbol{\lambda}]\,\tilde{I}_A(\bar{\lambda}) \\ &\quad = w(\boldsymbol{\lambda})\,\tilde{J}'_{cA}(\bar{\lambda}),\end{aligned} \tag{3.40}$$

in terms of the 3×3 matrix $\mathbf{D}$ having the matrix elements D_{jk}, or, on replacing $w\tilde{J}'_{cA}$ by $\tilde{J}'_{cA}$, the space–time equation

$$\begin{aligned}&\left[c^{-1}\frac{\partial}{\partial t} + \gamma_{aA} + \beta^{-1}\boldsymbol{\alpha}\cdot\frac{\partial}{\partial\boldsymbol{\rho}} - \frac{\partial}{\partial\boldsymbol{\rho}}\cdot\mathbf{D}\cdot\frac{\partial}{\partial\boldsymbol{\rho}}\right]I_A(\bar{\rho}) \\ &\quad = J'_{cA}(\bar{\rho}).\end{aligned} \tag{3.41}$$

In the same way, from Eqs. (3.28) and (3.29) together with Eq. (3.38), we obtain

$$\mathbf{I}_A(\bar{\rho}) = \left(\beta^{-1}\boldsymbol{\alpha} - \mathbf{D}\cdot\frac{\partial}{\partial\boldsymbol{\rho}}\right)I_A(\bar{\rho}), \tag{3.42}$$

$$\frac{\partial}{\partial\boldsymbol{\rho}}\cdot\mathbf{I}_A(\bar{\rho}) + \left(\gamma_{aA} + c^{-1}\frac{\partial}{\partial t}\right)I_A(\bar{\rho}) = J'_{cA}(\bar{\rho}). \tag{3.43}$$

So far it has been supposed that $f_A(\boldsymbol{\Omega},\bar{\lambda})$ is the only eigenfunction having the eigenvalue tending to unity as $\bar{\lambda}\to0$. However, as will be seen in Sec. IV, several eigenfunctions have eigenvalues equally tending to values very close to unity as $\bar{\lambda}\to0$, when the scattering is made mostly in forward direction, as in turbulent air, so that $\sigma(\boldsymbol{\Omega}|\boldsymbol{\Omega}')$ is different from zero only in a narrow range of $\boldsymbol{\Omega}\sim\boldsymbol{\Omega}'$. Even when this is the case, the expansion (3.32) is still valid, but is quite limited within a range of very small values of $|\bar{\lambda}/\gamma| \ll 1$.

IV. EIGENFUNCTION $f_A(\boldsymbol{\Omega}, \bar{\lambda})$ AND EVALUATION OF $\boldsymbol{\alpha}, \beta$, and $\mathbf{D}$

In Sec. III, the wave intensity I_s and flux vector $\mathbf{I}_s$ have been expanded in the eigenfunction series (3.18) where $\beta_A(\bar{\lambda})$ and $\boldsymbol{\alpha}_A(\bar{\lambda})$ are defined by Eq. (3.14) and $J_A(\bar{\rho})$ obeys the equation (3.13). Here, in the domain of diffusion where $|\bar{\lambda}/\gamma| \ll 1$, only the term of the eigenfunction f_A with the eigenvalue A tending to unity as $\bar{\lambda}\to0$ becomes dominant and, for this eigenfunction, the linear approximation (3.32) of $\beta_A(\bar{\lambda})$ and $\boldsymbol{\alpha}_A(\bar{\lambda})$ is enough to obtain the diffusion equation to the second order in space.

Here, to obtain β_A and $\boldsymbol{\alpha}_A$ to the first order of $\bar{\lambda}$, we can employ a perturbative method and begin with the eigenvalue equation (3.1) for $\bar{\lambda} = 0$ and $\gamma_a = 0$:

$$\begin{aligned}&\int d\,\boldsymbol{\Omega}'\sigma(\boldsymbol{\Omega}|\boldsymbol{\Omega}')\gamma^{-1}(\boldsymbol{\Omega}')f_a(\boldsymbol{\Omega}') = a\,f_a(\boldsymbol{\Omega}), \\ &\int d\,\boldsymbol{\Omega}'\bar{f}_a(\boldsymbol{\Omega}')\gamma^{-1}(\boldsymbol{\Omega}')\sigma(\boldsymbol{\Omega}'|\boldsymbol{\Omega}) = a\,\bar{f}_a(\boldsymbol{\Omega}),\end{aligned} \tag{4.1}$$

with the normalization

$$\int d\Omega\,\bar{f}_a\gamma^{-1}f_b(\Omega) = \delta_{ab}, \tag{4.2}$$

where $\sigma(\boldsymbol{\Omega}|\boldsymbol{\Omega}') \neq \sigma(\boldsymbol{\Omega}'|\boldsymbol{\Omega})$ and $\gamma(\boldsymbol{\Omega})$ is defined by Eq. (1.2).

Here, it holds that $|a| \leqslant 1$ since, by Eq. (4.1),

$$\int d\,\boldsymbol{\Omega}'\sigma(\boldsymbol{\Omega}|\boldsymbol{\Omega}')\gamma^{-1}(\boldsymbol{\Omega}')|\,f_a(\boldsymbol{\Omega}')| \geqslant |a||\,f_a(\boldsymbol{\Omega})|, \tag{4.3}$$

and hence, by the $\boldsymbol{\Omega}$ integration on both sides,

$$(1 - |a|)\int d\,\boldsymbol{\Omega}|\,f_a(\boldsymbol{\Omega})| \geqslant 0, \text{ or } |a| \leqslant 1. \tag{4.4}$$

On performing the $\boldsymbol{\Omega}$ integration on both sides of the first equation of Eq. (4.1), we obtain

$$\int d\,\Omega\, f_a(\Omega) = a \int d\,\Omega\, f_a(\Omega), \tag{4.5}$$

which leads to the important results that

$$a = 1 \quad \text{if} \quad \int d\,\Omega\, f_a(\Omega) \neq 0, \tag{4.6}$$

whereas

$$\int d\,\Omega\, f_a(\Omega) = 0 \quad \text{if} \quad a \neq 1. \tag{4.7}$$

Also, we can directly show, by Eqs. (4.1) and (4.2), that

$$\bar{f}_1(\Omega) = (4\pi)^{-1}\gamma(\Omega), \qquad (4\pi)^{-1}\int d\,\Omega\, f_1(\Omega) = 1, \tag{4.8}$$

$$\int d\,\Omega\, \bar{f}_1 \gamma^{-1} f_a(\Omega) = (4\pi)^{-1}\int d\,\Omega\, f_a(\Omega) = 0, \quad a \neq 1, \tag{4.9}$$

where the last relation is another version of Eq. (4.7) and indicates that all the $f_a(\Omega)$'s with $a \neq 1$ have the vanishing average value.

To obtain $f_A(\Omega,\bar{\lambda})$ to the first order of $\bar{\lambda}$, we substitute, on using Eqs. (2.32) with (2.33),

$$\tilde{U}(\Omega,\bar{\lambda}) = \gamma^{-1}(\Omega)[1 + h(\Omega,\bar{\lambda})], \quad |h| \ll 1,$$
$$f_A = f_a + \sum_{b \neq a} f_b C_{bA}, \tag{4.10}$$

into the eigenvalue equation (3.1) and subsequently use the orthogonality (4.2) to obtain

$$(A - b)C_{bA} = b\sum_c \langle h \rangle_{bc} C_{cA}, \quad C_{aA} = 1, \tag{4.11}$$

in terms of the notation

$$\langle h \rangle_{bc} = \int d\,\Omega \gamma^{-1} \bar{f}_b h f_c(\Omega),$$
$$\langle h \rangle_{1c} = (4\pi)^{-1}\int d\,\Omega h f_c(\Omega). \tag{4.12}$$

Here, in Eq. (4.11), $A \sim a$ and $|C_{bA}| \ll 1$, $b \neq a$, and therefore, to the first order of h,

$$C_{bA} \sim (a - b)^{-1} b \langle h \rangle_{ba}, \quad b \neq a, \tag{4.13}$$

while, from Eq. (4.10),

$$(4\pi)^{-1}\int d\,\Omega\, f_A = \int d\,\Omega\, \bar{f}_1 \gamma^{-1} f_A = C_{1A}, \tag{4.14}$$

showing that the average value of f_A, $a \neq 1$, tends to zero as $\bar{\lambda} \to 0$ when $\gamma_a = 0$.

In the same way, on putting

$$\bar{f}_A = \bar{f}_a + \sum_{b \neq a} \bar{C}_{Ab} \bar{f}_b, \tag{4.15}$$

we obtain

$$\int d\,\Omega\, \bar{f}_A \gamma^{-1} f_A = 1 + \sum_{b \neq a} \bar{C}_{Ab} C_{bA},$$
$$\bar{C}_{Ab} C_{bA} \sim (a - b)^{-2} b^2 \langle h \rangle_{ab} \langle h \rangle_{ba}. \tag{4.16}$$

Here, the condition of applicablility of the present perturbative method is $|\bar{C}_{Ab} C_{bA}| \ll 1$ and therefore turns out to be

$$(a/b - 1)^{-2} |\bar{\lambda}/\gamma|^2 \ll 1, \qquad |h| \sim |\bar{\lambda}/\gamma|, \tag{4.17}$$

which becomes a very severe condition of $|\bar{\lambda}/\gamma|$ when $a \sim b$.

When $f_A \to f_1$ as $\bar{\lambda} \to 0$ and also $f_1(\Omega)$ is the only eigenfunction having the eigenvalue of unity, then it follows from Eq. (4.14) that

$$(4\pi)^{-1}\int d\,\Omega\, f_A(\Omega,\bar{\lambda}) = 1. \tag{4.18}$$

Therefore, according to the definition (3.14) with Eq. (4.10), we obtain, to the first order of $\bar{\lambda}$,

$$\alpha_{Aj}(\bar{\lambda}) = (4\pi)^{-1}\int d\,\Omega \left[\Omega_j \gamma^{-1}(1 + h)(f_1 + \sum_{b \neq 1} f_b C_{bA})\right], \tag{4.19}$$

which gives, in terms of the notation of Eq. (4.12),

$$\alpha_{Aj}(\bar{\lambda}) = \langle \Omega_j \gamma^{-1} \rangle_{11} + \langle \Omega_j \gamma^{-1} h \rangle_{11} + \sum_{b \neq 1} \langle \Omega_j \gamma^{-1} \rangle_{1b} C_{bA}, \tag{4.20}$$

where the C_{bA}'s are given by Eq. (4.13). Thus, according to the definition (3.32), we find that

$$\alpha_j = \langle \Omega_j \gamma^{-1} \rangle_{11},$$
$$A_{jk} = \langle \Omega_j \Omega_k \gamma^{-2} \rangle_{11} + \sum_{b \neq 1} b(1 - b)^{-1} \langle \Omega_j \gamma^{-1} \rangle_{1b} \langle \Omega_k \gamma^{-1} \rangle_{b1},$$
$$A_j = \langle \Omega_j \gamma^{-2} \rangle_{11} + \sum_{b \neq 1} b(1 - b)^{-1} \langle \Omega_j \gamma^{-1} \rangle_{1b} \langle \gamma^{-1} \rangle_{b1}. \tag{4.21}$$

In the same way,

$$\beta = \langle \gamma^{-1} \rangle_{11}, \quad B = \langle \gamma^{-2} \rangle_{11} + \sum_{b \neq 1} b(1 - b)^{-1} \langle \gamma^{-1} \rangle_{1b} \langle \gamma^{-1} \rangle_{b1},$$
$$B_j = \langle \Omega_j \gamma^{-2} \rangle_{11} + \sum_{b \neq 1} b(1 - b)^{-1} \langle \gamma^{-1} \rangle_{1b} \langle \Omega_j \gamma^{-1} \rangle_{b1}. \tag{4.22}$$

Here, A_{jk}, A_j, B, and B_j are of the order of γ^{-2} whereas α_j and β are of γ^{-1}.

In the case of symmetrical scattering where $\sigma(\Omega|\Omega') = \sigma(\Omega'|\Omega)$, it follows from Eq. (4.8) that

$$f_1(\Omega) = \gamma(\Omega)/\langle\gamma\rangle, \quad \langle\gamma\rangle = (4\pi)^{-1}\int d\,\Omega\gamma(\Omega),$$
$$\alpha_j = 0, \quad \beta = \langle\gamma\rangle^{-1}. \tag{4.23}$$

Also, the matrix $\mathbf{D}$ has the symmetrical matrix elements, given by $A_{jk} = A_{kj}$, and $A_j = B_j \neq 0$ when $\gamma(\Omega)$ depends on Ω.

As an example, it will be interesting to consider the case of the form $\sigma(\Omega|\Omega') = \sigma(\Omega\cdot\Omega')$ which has been widely assumed by previous authors. In this case, $\sigma(\cos\theta)$ can be expanded, in terms of the Legendre functions $P_n(x)$, $n = 0, 1, 2, \cdots$, by

$$\sigma(\cos\theta) = (4\pi)^{-1}\gamma\left[1 + \sum_{n=1}^{\infty} (2n + 1) P_n(\cos\theta) a_n\right], \tag{4.24}$$

where the extinction coefficient γ is constant and

$$a_n = \gamma^{-1}\int d\,\Omega P_n(\cos\theta)\,\sigma(\cos\theta), \quad d\,\Omega = 2\pi \sin\theta\, d\theta,$$
$$P_n(1) = 1, \quad a_0 = 1. \tag{4.25}$$

Here, by (4.24), we can directly show that the spherical harmonics

$$Y_n^m(\Omega) = Y_n^m(\theta,\varphi) = \left.\begin{matrix}\cos m\varphi \\ \sin m\varphi\end{matrix}\right\} P_n^m(\cos\theta), \quad n \geqslant m \geqslant 0,$$

$$\int d\Omega [Y_n^m(\Omega)]^2 = 4\pi(2n+1)^{-1}\frac{(n+m)!}{(n-m)!}\begin{cases}1, & m=0,\\ 1/2, & m\neq 0,\end{cases} \tag{4.26}$$

are the eigenfunctions of Eq. (4.1) and have the eigenvalues a_n's, i.e.,

$$\int d\Omega' \gamma^{-1}\sigma(\Omega\cdot\Omega') Y_n^m(\Omega') = a_n Y_n^m(\Omega). \tag{4.27}$$

On the other hand,

$$\bar{f}_1(\Omega) = (4\pi)^{-1}\gamma, \quad f_1(\Omega) = 1, \tag{4.28}$$

and hence

$$\langle \Omega_j \Omega_k \rangle_{1\,1} = (4\pi)^{-1}\int d\Omega \Omega_j \Omega_k = \frac{1}{3}\delta_{jk}, \tag{4.29}$$

while, if $\Omega_j = \cos\theta$, then $\langle \Omega_j \rangle_{1b}$ is different from zero only when $f_b(\Omega) = P_1(\cos\theta) = \cos\theta$, yielding $\langle \Omega_j \rangle_{1b} = 1/3$. Also, $\bar{f}_b(\Omega) = (3/4\pi)\gamma\cos\theta$ and hence $\langle \Omega_k \rangle_{b1} = \delta_{kj}$. Thus,

$$\langle \Omega_j \rangle_{1b} \langle \Omega_k \rangle_{b1} = \tfrac{1}{3}\delta_{jk}, \tag{4.30}$$

and by Eqs. (4.21), (4.22), and (3.34),

$$\alpha_j = 0, \quad D_{jk} = \gamma A_{jk} = 3^{-1}(1-a_1)^{-1}\gamma^{-1}\delta_{jk},$$

$$\beta = \gamma^{-1}, \quad \gamma_{aA} = (4\pi)^{-1}\int d\Omega \gamma_a(\Omega), \tag{4.31}$$

where a_1 is given by Eq. (4.25).

On the other hand, in Eq. (3.11), we can use the approximation $f_A \sim f_1$, $\tilde{U} \sim \gamma^{-1}$, $|\bar{\lambda}/\gamma| \ll 1$, and Eq. (4.28), to obtain $\tilde{J}'_{cA}(\bar{\rho}) \sim \tilde{J}'_c(\bar{\rho})$ and thus the diffusion equation (3.41) becomes

$$[c^{-1}\partial/\partial t + \gamma_{aA} - 3^{-1}(1-a_1)^{-1}\gamma^{-1}(\partial/\partial\rho)^2] I_s(\bar{\rho}) = \tilde{J}'_c(\bar{\rho}), \tag{4.32}$$

while the flux vector is given, according to Eq. (3.42), by

$$\mathbf{I}_s(\bar{\rho}) = -3^{-1}(1-a_1)^{-1}\gamma^{-1}(\partial/\partial\boldsymbol{\rho}) I_s(\bar{\rho}). \tag{4.33}$$

Here, when the scattering is made mostly in forward direction, then Eq. (4.25) shows that $a_n \sim 1$ over the first several orders of n, while the second term on the left-hand side of Eq. (4.32) tends to the infinite as $a_1 \to 1$. This reflects the fact that the present perturbative method is valid only under the condition (4.17), which is not satisfied when $a_n \sim a_0 = 1$, $n = 1, 2, \cdots$.

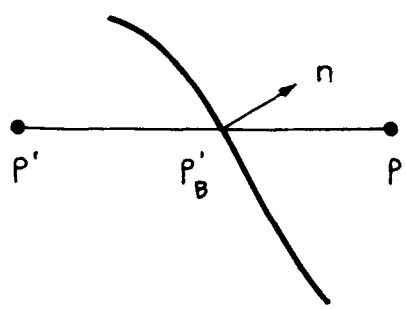


FIG. 2. Boundary of the medium discontinuity and notations for Eq. (5.2).

V. CONDITIONS AT BOUNDARY OF MEDIUM DISCONTINUITY

We first suppose that the space is divided into two parts by a surface, the one being free space and the other a homogeneous scattering medium (Fig. 2), and ask the boundary conditions to be satisfied by solutions of the diffusion equation (3.41). Here, in free space where no scatterer presents, $J_s = 0$ by the definition (2.6) and therefore J_s is not a quantity continuous across the boundary.

To find the latter quantities, we observe that $\tilde{U}^{-1}\tilde{I}_s = \tilde{J}_s$ by the first equation of Eq. (2.9) with $\tilde{U}$ given by Eq. (2.7) or, in the space–time coordinates,

$$[\Omega\cdot\partial/\partial\boldsymbol{\rho} + c^{-1}\partial/\partial t + \gamma_t] I_s(\Omega,\bar{\rho}) = J_s(\Omega,\bar{\rho}). \tag{5.1}$$

Thus, on following the usual procedure, we find the condition of continuity, given for arbitrary Ω by

$$(\mathbf{n}\cdot\Omega) I_s(\Omega,\bar{\rho})|_{\pm} \equiv (\mathbf{n}\cdot\Omega)[I_s(\Omega,\bar{\rho})|_{+} - I_s(\Omega,\bar{\rho})|_{-}] = 0, \tag{5.2}$$

where $\mathbf{n}$ is the unit vector normally directed into the scattering medium, say on the side $|_+$, from the free space side $|_-$, and the marks $|_{\pm}$ denote the quantities on the referred sides.

The quantities in Eq. (5.2) are functions of Ω and, in view of the right-hand side of Eq. (5.1) given by J_s, can be exhibited in terms of the previous eigenfunctions $f_a(\Omega')$'s. Thus, for each eigenfunction $f_a(\Omega)$, it holds

$$\int_{4\pi} d\Omega \gamma^{-1}\bar{f}_a(\Omega)(\mathbf{n}\cdot\Omega) I_s(\Omega,\bar{\rho})|_{\pm} = 0, \tag{5.3}$$

where the integration is performed on the entire range of solid angle.

Also, there should not be any contribution of flux from the free space side $|_-$, distributed within the range of $\mathbf{n}\cdot\Omega > 0$ and therefore, besides Eq. (5.3), we have the additional condition

$$\int_{\mathbf{n}\cdot\Omega>0} d\Omega \gamma^{-1}\bar{f}_a(\Omega)(\mathbf{n}\cdot\Omega) I_s(\Omega,\bar{\rho})|_{+} = 0. \tag{5.4}$$

Here, in the domain of diffusion, the components of the eigenfunction $\bar{f}_1(\Omega) = (4\pi)^{-1}\gamma(\Omega)$ are most important in Eqs. (5.3) and (5.4) and are given by

$$\int_{4\pi} d\Omega (\mathbf{n}\cdot\Omega) I_s(\Omega,\bar{\rho})|_{\pm} = \mathbf{n}\cdot\mathbf{I}_s(\bar{\rho})|_{\pm} = 0, \tag{5.5}$$

$$\int_{\mathbf{n}\cdot\Omega>0} d\Omega (\mathbf{n}\cdot\Omega) I_s(\Omega,\bar{\rho})|_{+} = 0, \tag{5.6}$$

whose expressions are independent of the natures of the scattering cross section. Here, the condition (5.5) is just the condition of continuity of total flux across the boundary surface and can also be derived directly from the equation of continuity (3.30). Since the diffusion equation (3.41) is the second order differential equation in space, the above two conditions are enough to determine I_s for a boundary surface of free space.

By Eq. (5.5) with Eq. (3.42), the normal component of total flux flowing away into the free space, say I_B, is given by

$$I_B = -\mathbf{n}\cdot\mathbf{I}_s|_{-} = -\mathbf{n}\cdot(\beta^{-1}\boldsymbol{\alpha} - \mathbf{D}\cdot\partial/\partial\boldsymbol{\rho}) I_s|_{+}, \tag{5.7}$$

while, on following the procedure leading to Eqs. (4.21) and (4.22) and also referring to Eqs. (3.38) and (3.39), we obtain

the condition (5.6) expressed, to the first order of $\gamma^{-1}\partial/\partial\rho$, in the form

$$[\beta^{-1}\alpha^{(n)} - \mathbf{D}^{(n)}\cdot\partial/\partial\rho]I_s(\bar{\rho})|_+ = 0. \tag{5.8}$$

Here, the vector $\mathbf{D}^{(n)} = (D_1^{(n)}, D_2^{(n)}, D_3^{(n)})$ is given by

$$D_j^{(n)} = \beta^{-1}[A_j^{(n)} - \beta^{-1}(A^{(n)}\alpha_j + \alpha^{(n)}B_j) + \beta^{-2}B\alpha^{(n)}\alpha_j], \tag{5.9}$$

where α_j, β, B_j, and B are given by Eqs. (4.21) and (4.22), and

$$\alpha^{(n)} = \langle \mathbf{n}\cdot\Omega\gamma^{-1}\rangle_{11}^{(n)},$$

$$A^{(n)} = \langle \mathbf{n}\cdot\Omega\gamma^{-2}\rangle_{11}^{(n)} + \sum_{b\neq 1} b(1-b)^{-1}\langle \mathbf{n}\cdot\Omega\gamma^{-1}\rangle_{1b}^{(n)}\langle\gamma^{-1}\rangle_{b1},$$

$$A_j^{(n)} = \langle \mathbf{n}\cdot\Omega\Omega_j\gamma^{-2}\rangle_{11}^{(n)}$$

$$+ \sum_{b\neq 1} b(1-b)^{-1}\langle \mathbf{n}\cdot\Omega\gamma^{-1}\rangle_{1b}^{(n)}\langle\Omega_j\gamma^{-1}\rangle_{b1}, \quad j = 1,2,3, \tag{5.10}$$

in terms of the notation similar to Eq. (4.12):

$$\langle Q\rangle_{ab}^{(n)} = \int_{\mathbf{n}\cdot\Omega>0} d\Omega\gamma^{-1}\bar{f}_a Q f_b(\Omega). \tag{5.11}$$

In case of the previous example where $\sigma(\Omega|\Omega') = \sigma(\Omega\cdot\Omega')$, we find, on choosing $f_b(\Omega) = \cos(\mathbf{n}\cdot\Omega)$, that

$$\langle \mathbf{n}\cdot\Omega\gamma^{-1}\rangle_{11}^{(n)} = (4\gamma)^{-1}, \quad \langle \mathbf{n}\cdot\Omega\Omega_j\gamma^{-2}\rangle_{11}^{(n)} = (6\gamma^2)^{-1}n_j,$$
$$\langle \mathbf{n}\cdot\Omega\gamma^{-1}\rangle_{1b}^{(n)} = (6\gamma)^{-1}, \quad \langle\Omega_j\gamma^{-1}\rangle_{b1} = \gamma^{-1}n_j, \tag{5.12}$$

and all the contributions from the other f_b's identically vanish. Thus, on using $\alpha_j = B_j = 0$, we obtain from Eqs. (5.8), (5.9), and (5.10)

$$[I_s - \tfrac{2}{3}\gamma^{-1}(1-a_1)^{-1}\mathbf{n}\cdot\partial/\partial\rho I_s]|_+ = 0, \tag{5.13}$$

which, when combined with Eq. (4.33), enables I_B in Eq. (5.7) to be expressed by

$$I_B = -\mathbf{n}\cdot\mathbf{I}_s = \tfrac{1}{2}I_s, \tag{5.14}$$

which means that one half of I_s on the boundary surface contributes to the normal component of flux into free space.

When the source J_c is in the free space side, the explicit expression of the source term J_c' in the diffusion equation (3.41) becomes, on using (2.17) and the method of deriving Eq. (2.42),

$$J_c'(\bar{\rho}) = \int d\Omega\gamma(\Omega)I_c(\Omega,\bar{\rho})$$
$$= \int d\rho'|\rho-\rho'|^{-2}\gamma(\Omega)\exp[-\gamma_t(\Omega)|\rho-\rho_B'|]$$
$$\times J_c(\Omega, \rho', t-|\rho-\rho'|/c),$$
$$\Omega = (\rho-\rho')/|\rho-\rho'|, \tag{5.15}$$

where ρ_B' is the point on boundary surface of free space, made by the line connecting the points ρ and ρ' (Fig. 2).

The above boundary conditions can be generalized to the case where the domain of the side $|_-$ is not free space but another scattering medium having properties different from those of $|_+$. In that case, the mark $|_+$ in Eq. (5.6) is replaced by $|_-^+ = |_+ - |_-$, and this is the case also for Eqs. (5.8) and (5.13).

In the particular case of $\sigma(\Omega|\Omega') = \sigma(\Omega\cdot\Omega')$, the new boundary condition yields, when combined with the condition (5.5), simply the conditions of continuity of I_s and $\mathbf{n}\cdot\mathbf{I}_s$, where $\mathbf{I}_s$ is given by Eq. (4.33) in terms of the gradiant of $I_s(\bar{\rho})$. In the general case of anisotropic scattering, however, the continuity of I_s at boundary surface is not always guaranteed by Eqs. (5.5) and (5.6) or Eq. (5.8) with $|_+$ replaced by $|_-^+$, in spite of the condition $I_s(\Omega,\bar{\rho})|_-^+ = 0$ which directly follows from Eq. (5.2). However, as far as the diffusion equation is given to the second order in space, the two conditions (5.5) and new equation (5.8) are necessary and sufficient at the expense of the I_s continuity.

VI. A SIMPLE EXAMPLE—PULSE WAVE PROPAGATION IN A SEMI-INFINITE SCATTERING MEDIUM

Here, the diffusion equation (4.32) is conveniently given in the form

$$[\gamma_a + c^{-1}\partial/\partial t - (\eta\gamma)^{-1}(\partial/\partial\rho)^2]I_s(\bar{\rho}) = J_c'(\bar{\rho}),$$
$$\eta = 3(1-a_1), \tag{6.1}$$

where J_c' is given by Eq. (5.15) and γ_a represents γ_{aA}. Hence, the equation of the Fourier transform

$$I_s(\rho,\nu) = \int_{-\infty}^{\infty} dt\, e^{-i\nu t}I_s(\rho,t), \tag{6.2}$$

with respect to the time t, becomes

$$[\gamma_a + i\nu/c - (\eta\gamma)^{-1}(\partial/\partial\rho)^2]I_s(\rho,\nu) = J_c'(\rho,\nu), \tag{6.3}$$

where $J_c'(\rho,\nu)$ is the corresponding Fourier transform of $J_c'(\rho,t)$. The solution of Eq. (6.3) is given, in terms of the Green's function $G_\nu(\rho|\rho')$ to be given later, by

$$I_s(\rho,\nu) = \int d\rho' G_\nu(\rho|\rho')J_c'(\rho',\nu). \tag{6.4}$$

Here, in the coordinate system $\rho = (\rho_1,\rho_2,\rho_3)$, we suppose that the domain of $\rho_3 > 0$ is filled with random scatterers whereas the domain of $\rho_3 < 0$ is free space, and ask the plane wave solution homogeneous over the $\rho_1\rho_2$ plane. Then, the Green's function $G_\nu(\rho_3|\rho_3')$ is given in the form

$$G_\nu(\rho_3|\rho_3') = \eta\gamma[I^\infty\partial I^0/\partial\rho_3 - I^0\partial I^\infty/\partial\rho_3]^{-1}$$
$$\times I^0(\rho_<,\nu)I^\infty(\rho_>,\nu). \tag{6.5}$$

Here, $\rho_<$ and $\rho_>$ denote the smaller and larger of ρ_3 and ρ_3', and $I^0(\rho_3,\nu)$ is the solution of the homogeneous equation

$$[\gamma_a + i\nu/c - (\eta\gamma)^{-1}(\partial/\partial\rho_3)^2]I(\rho_3,\nu) = 0, \tag{6.6}$$

satisfying the boundary condition of Eq. (5.13) at $\rho_3 = 0$, while $I^\infty(\rho_3,\nu)$ is the solution which tends to zero as $\rho_3 \to +\infty$. In order to prove Eq. (6.5), we only need, first, to note that the factor $[\ ]^{-1}$ is a constant, and then to check, after the substitution of $I_s = G_\nu$ in Eq. (6.3), the term resulting from the discontinuity at $\rho_3 = \rho_3'$, giving $\delta(\rho_3 - \rho_3')$ for J_c'.

Here, in terms of the variable

$$k = \{(\gamma_a + i\nu/c)\eta\gamma\}^{1/2}, \quad \mathrm{Re}[k] > 0, \tag{6.7}$$

Equation (6.6) gives

$$I^0(\rho_3,\nu) = e^{k\rho_3} + \frac{2k(\eta\gamma)^{-1} - 1}{2k(\eta\gamma)^{-1} + 1}e^{-k\rho_3}, \tag{6.8}$$

$$I^{\infty}(\rho_3,\nu) = e^{-k\rho_3},$$

and hence, according to Eq. (6.5),

$$G_\nu(\rho_3|\rho_3') = (2k)^{-1}\eta\gamma I^0(\rho_<,\nu) I^{\infty}(\rho_>,\nu). \quad (6.9)$$

On the other hand, to excite a plane wave pulse homogeneous on the $\rho_1\rho_2$ plane, the source function $J_c(\Omega, \rho, t)$ is assumed to have the form

$$J_c(\Omega,\rho,t) = I_0\delta(t-\rho_c/c)\delta(\rho_3-\rho_c)\delta(\Omega_1)\delta(\Omega_2), \quad (6.10)$$

where $\rho_c < 0$ is the coordinate of the source in the free space and I_0 is a constant. Hence, on referring to Eq. (5.15), we obtain

$$\begin{aligned} J_c'(\rho_3,t) &= I_0\gamma\delta(t-\rho_3/c)\exp(-\gamma_t\rho_3), \quad \rho_3>0, \\ &= 0, \quad \rho_3<0, \end{aligned} \quad (6.11)$$

whose Fourier transform with respect to time is

$$J_c'(\rho_3,\nu) = I_0\gamma\exp[-(\gamma_t + i\nu/c)\rho_3], \quad \rho_3>0. \quad (6.12)$$

Thus, on employing the expressions of Eqs. (6.9) and (6.12) in Eq. (6.4), we obtain, when $\rho_3 = 0$,

$$\begin{aligned} I_s(0,\nu) &= \int_0^{\infty} d\rho G_\nu(0|\rho) J_c'(\rho,\nu) \\ &= I_0(2k)^{-1}\eta\gamma^2(k+\gamma_t + i\nu/c)^{-1} I^0(0,\nu), \end{aligned} \quad (6.13)$$

and hence, by Eq. (6.8),

$$\begin{aligned} I_s(0,t) &= (2\pi)^{-1}\int_{-\infty}^{\infty} d\nu\, I_s(0,\nu)e^{i\nu t} \\ &= I_0\gamma^2\pi^{-1}\int_{-\infty}^{\infty} d\nu(2k/\eta+\gamma)^{-1} \\ &\quad \times(k+\gamma_t + i\nu/c)^{-1}e^{i\nu t}, \end{aligned} \quad (6.14)$$

where $\gamma_t = \gamma + \gamma_a$.

After changing the variable of integration from ν to k, the integral (6.14) is found to be expressed in terms of the function $W(x)$, defined by

$$\begin{aligned} W(x) &= (\pi i)^{-1}\int_{-\infty}^{\infty} du\,(u-ix)^{-1}e^{-u^2} \\ &= \frac{2}{\sqrt{\pi}} e^{x^2}\int_x^{\infty} dt\, e^{-t^2}, \quad |\arg(x)| < \pi/2, \end{aligned} \quad (6.15)$$

whence

$$\begin{aligned} I_s(0,t) &= 2c\gamma I_0\exp(-\gamma_a ct)(4/\eta-1)^{-1} \\ &\quad \times[2\,\mathrm{Re}[\xi W\{\xi(\eta\gamma ct)^{1/2}\}] - W\{\tfrac{1}{2}(\eta\gamma ct)^{1/2}\}], \end{aligned} \quad (6.16)$$

where Re[···] indicates the real part and ξ is a complex constant, given by

$$\xi = \tfrac{1}{2}[1 + i(4/\eta-1)^{1/2}], \quad \eta = 3(1-a_1) < 3. \quad (6.17)$$

The flux component I_B away from the boundary at $\rho_3 = 0$ into the free space of $\rho_3 < 0$ is given by $I_s/2$ according to Eq. (5.14) and therefore, utilizing the expansion[13]

$$\begin{aligned} W(x) &= \sum_{n=0}^{\infty}(-x)^n/\Gamma(1+n/2) \\ &\sim(\pi^{1/2}x)^{-1}\left[1 + \sum_{m=1}^{\infty}(-)^m\, 1\cdot3\cdots(2m-1)(2x^2)^{-m}\right], \end{aligned}$$

$$|x| \gg 1, \quad |\arg(x)| < 3\pi/4, \quad (6.18)$$

Eq. (6.16) gives

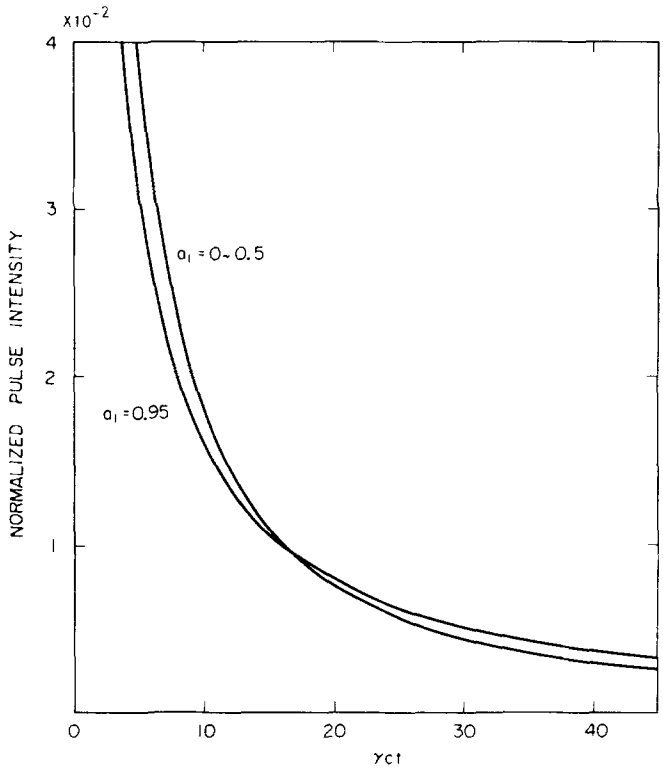


FIG. 3. Intensity of a plane wave pulse backscattered from a boundary of a homogeneous scattering medium. The normalized intensity $I_B/[c\gamma\exp(-\gamma_a ct)]$ is shown as a function of γct for several values of a_1.

$$\begin{aligned} &I_B/[c\gamma I_0\exp(-\gamma_a ct)] \\ &\quad \sim\pi^{-1/2}(\eta\gamma ct)^{1/2}, \quad \eta\gamma ct \ll 1 \\ &\quad \sim(2\pi^{1/2})^{-1}\eta(2+\eta)(\eta\gamma ct)^{-3/2}, \quad \eta\gamma ct \gg 1, \end{aligned} \quad (6.19)$$

which indicates that, except for the factor $\exp(-\gamma_a ct)$, I_B decreases with the time by the factor $t^{-3/2}$ for $t\sim\infty$. Curves of $I_B/[c\gamma I_0\exp(-\gamma_a ct)]$ are shown in Fig. 3 as a function of γct for several values of a_1.

On the other hand, when a source is embedded in the scattering medium at $\rho_3 = \rho > 0$, I_s on the boundary surface is given, to a good approximation, by the Green's function $G_t(0|\rho)$ in view of Eq. (2.45), and

$$\begin{aligned} G_t(0|\rho) &= (2\pi)^{-1}\int_{-\infty}^{\infty} d\nu\, G_\nu(0|\rho)e^{i\nu t} \\ &= \tfrac{1}{2}c\eta\gamma\exp[-\gamma_a ct - \eta\gamma\rho^2/(4ct)] \\ &\quad \times[2\pi^{-1/2}(\eta\gamma ct)^{-1/2} - W(z)], \end{aligned} \quad (6.20)$$

$$z = \tfrac{1}{2}(\eta\gamma ct)^{1/2}(1+\rho/ct), \quad (6.21)$$

and, when $\eta\gamma ct \gg 1$ and $(\rho/ct)^2 \ll 1$,

$$\begin{aligned} G_t(0|\rho) &\sim\pi^{-1/2}c\eta\gamma(\eta\gamma ct)^{-3/2}(\eta\gamma\rho+2) \\ &\quad \times\exp[-\eta\gamma\rho^2/(4ct) - \gamma_a ct], \end{aligned} \quad (6.22)$$

which again decreases with the time by the factor $t^{-3/2}\exp(-\gamma_a ct)$.

In Fig. 4, curves of $(c\eta\gamma)^{-1}\exp(\gamma_a ct)G_t(0|\rho)$ are shown as a function of $\eta\gamma ct$ for several values of $\eta\gamma\rho$.

VII. SUMMARY AND DISCUSSION

In order to derive the diffusion equation, it is most sim-

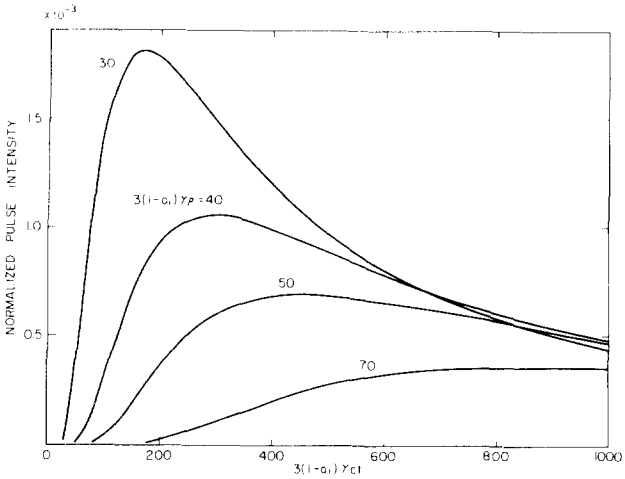


FIG. 4. Intensity of a plane wave pulse propagated into free space away from a point source embedded in a semi-infinite scattering medium. The normalized intensity $(c\eta\gamma)^{-1}\exp(\gamma_a ct)G_1(0|\rho)$, given by Eq. (6.20), is shown as a function of $\eta\gamma ct$ for several values of the numerical distance $\eta\gamma\rho$ from boundary to source.

ple to treat various equations in terms of their Fourier transforms with respect to the space–time coordinates $\bar{\rho}$ since the transport equation (1.1) does not depend on the coordinates explicitly, although it does on $\partial/\partial\bar{\rho}$. This is also closely connected to the fact that the Fourier transform of the transport equation can be reduced to an integral equation of the Fredholm type with respect to the angle of scattering, with the variable of Fourier transformation $\bar{\lambda}$ kept to be a constant parameter (Sec. II). This enables the solutions of the integral equation to be expanded in the eigenfunction series (3.8), where each term of the series obeys Eq.(3.10) and is represented by the integral (3.19) in space and time. When integrating with respect to ν in the latter integral for given λ, there exists one eigenfunction with the eigenvalue $A(\bar{\lambda})$ which yields the pole of $[1-A(\bar{\lambda})]^{-1}$, $\omega(\lambda)$, tending to zero as $\lambda\to 0$ when $\gamma_a = 0$. Here, in the domain of diffusion where the space–time distances $\bar{\rho}$ from the wave source are large enough to be $|\gamma\bar{\rho}|\gg 1$ or $|\bar{\lambda}/\gamma|\ll 1$ in the important range of integration of Eq. (3.19), the dominant contribution to the integral is given by this pole in terms of the residue value and therefore obeys an equation, given by Eq. (3.10) or (3.13), with $1-A(\bar{\lambda})\sim -[\nu-\omega(\lambda)]\partial A/\partial\nu|_{\nu=\omega}$. Therefore, the diffusion equation thus obtained necessarily becomes the first order differential equation in time, and $\omega(\lambda)$, which is a function of λ/γ with $\omega(0) = i\gamma_a$, can be expanded in a power series of λ/γ to arbitrary order, giving a corrected diffusion equation when the order exceeds two. Here, the contributions from the terms of other eigenfunctions are negligible since, when $|\bar{\lambda}/\gamma|\ll 1$, $1-A(\bar{\lambda})\sim 1-a$, $|a|<1$, giving by the Fourier inversion in Eq. (3.19) a function proportional to $\delta^4(\bar{\rho})$ or a short range function negligible in the range $|\gamma\bar{\rho}|\gg 1$. In the present case, the important physical quantities are the total wave intensity $I_s(\bar{\rho})$ and flux vector $\mathbf{I}_s(\bar{\rho})$ and their Fourier transforms are given by the corresponding eigenfunction series (3.16) and (3.17), respectively. Here, when performing their Fourier inversions, the dominant terms are again given by the residue values at the pole $\nu = \omega(\lambda)$ with β_A and α_A given by Eq. (3.35) to the first order of λ. In this connection, the expression (3.25) for $A(\bar{\lambda})$, given in terms of $\beta_A(\bar{\lambda})$ and $\alpha_A(\bar{\lambda})$, is very useful for evaluating $A(\bar{\lambda})$ and is also directly connected to the equation of continuity (3.29) for each eigenfunction component. The exceptional case is when the scattering is made mostly in the forward direction, as in turbulent air, and then the convergence of series becomes poor and also the present perturbative method with respect to $\bar{\lambda}/\gamma$ is not possible (Sec. IV).

In the general case of anisotropic scattering where $\sigma(\Omega|\Omega')\neq\sigma(\Omega'|\Omega)$, the diffusion equation is given by Eq. (3.41) and, in the infinite space, its solution for a space–time point source shows that the waves diffuse in an elliptical form and its center moves with the velocity $c\alpha/\beta$ in time. Here, it will be noted that, starting from the Boltzmann or Fokker–Planck equation, the diffusion equations of the form (3.41) were previously obtained for the concentration near the nonequilibrium steady state[11] and also, in a more general form, for a Brownian particle in the high-friction case.[10] Comparing the method in Ref. 10 with the present one, particularly, one finds that φ_n and $P_{|n|}$ are similar to f_a and f_A in the present paper, respectively, but are different in that, while φ_n is the eigenfunction of the operator C according to the usual definition and $P_{|n|}$ is not defined by an eigenvalue equation, both f_a and f_A are the eigenfunctions explicitly defined by the eigenvalue equations (4.1) and (3.1), where $\gamma^{-1}(\Omega)$ and $\tilde{U}(\Omega,\bar{\lambda})$ are regarded as the weighting functions in the respective equations. Here, when solving the transport equation (1.1), the latter method seems to be essential since the scattering cross section $\sigma(\Omega|\Omega')$ can be exhibited in terms of the eigenfunctions and eigenvalues by Eq. (3.5) or by the corresponding expression in terms of the f_a and a's, where the effect of $\tilde{U}(\Omega,\bar{\lambda})$ or the extinction coefficient $\gamma(\Omega)$ is inevitably included and needs not to be taken into account separately.

The boundary equations on the surface of the medium discontinuity can also be exhibited in terms of the eigenfunction f_a's introduced in Sec. IV and, in the case of the boundary of free space, the equations are given by Eqs. (5.3) and (5.4). In the diffusion approximation, these conditions are reduced to the two conditions (5.5) and (5.6), which belong to the eigenfunction dominant in the domain of diffusion. Here, the first condition coincides with that given by the equation of continuity (3.30) for the total flux of the incoherent wave. Here, if the diffusion equation was given to the fourth order in space, as in Eq. (2.38), then two more boundary conditions would become necessary and be given by Eq. (5.3) and (5.4) for the f_a having the eigenvalue a nearest to unity. In the special case of time-independent isotropic scatterers where $a_1 = 0$, the solution for a semi-infinite slab is known as the Milne solution, and it was pointed out in Ref. 3 that the condition (5.13) is a good approximation to the exact condition of this case.

[1] A through review paper was given by Yu.N. Barabanenkov, Yu.A. Kravtsov, S.M. Rytov, and V.I. Tatarskii, Sov. Phys. Usp. **13**, 551 (1971). See also R. L. Fante, J. Opt. Soc. Am. **64**, 592 (1974); I.M. Besieris, J. Math. Phys. **17**, 1707 (1976).

[2] I. Sreenivashiah, A. Ishimaru, and S.T. Hong, Radio Sci. **11**, 775 (1976).

[3] A. Ishimaru, J. Opt. Soc. Am **68**, 1045 (1978). The diffusion equation, corresponding to Eq. (6.1), contains a second order term in time, and it

was concluded that, in the diffusion approximation, the wave has a characteristic speed of $c/\sqrt{3}$. Correspondingly, the boundary condition also differs from Eq. (5.13), except for the special case of time-independent isotropic scattering.

[4]V.I. Shishov, Sov. Astron. **17**, 598 (1974).
[5]L.C. Lee and J.R. Jokipii, Astrophys. J. **201**, 532 (1975).
[6]L. Dolin, Radiophys. Quantum Electron. (USSR) **7**, 380 (1964).
[7]K. Furutsu, J. Math. Phys. **20**, 617 (1979).
[8]K.M. Case, Ann. Phys. (N.Y.) **9**, 1 (1960).
[9]N.J. McCormick and I. Kuščer, J. Math. Phys. **7**, 2036 (1966).
[10]U.M. Titulaer, Physica (Utrecht) A **91**, 321 (1978). The Fokker–Planck equation for the distribution function $P(u,x,t)$ of the velocity and position of a particle with mass m in an external potential $\Phi(x)$, and in a medium with temperature $(\kappa\beta)^{-1}$ and friction coefficient γ, was given in the form

$$\frac{\partial}{\partial t}P(u,x,t)=\left[\gamma C-u\frac{\partial}{\partial x}+\frac{1}{m}\frac{\partial\Phi}{\partial x}\frac{\partial}{\partial u}\right]P(u,x,t),$$

where C is an operator defined by

$$C=\frac{1}{m\beta}\left(\frac{\partial}{\partial u}\right)^2+\frac{\partial}{\partial u}u,$$

and the dominant term on the right-hand side is given by γC when γ is sufficiently large. To obtain its solutions, the eigenfunction $\varphi_n(u)$ of C was defined by the eigenvalue equation

$$C\varphi_n(u)=-n\varphi_n(u),\quad n=0,1,2,\cdots,$$

and special solutions were constructed in the form

$$P_{[n]}(u,x,t)=C_{[n]}(x,t)\varphi_n(u)=\gamma^{1}P^{(1)}_{[n]}(u,x,t)+\gamma^{2}P^{(2)}_{[n]}(u,x,t)+\cdots,$$

together with the time derivative of the form

$$\frac{\partial}{\partial t}C_{[n]}(x,t)=[-n\gamma+\partial^{(0)}_{[n]}+\gamma^{1}\partial^{(1)}_{[n]}+\gamma^{2}\partial^{(2)}_{[n]}+\cdots]C_{[n]}(x,t),$$

where the $\partial^{(l)}_{[n]}$ are linear differential operators to be determined by the integrability condition in the course of calculation. Thus, the equation of $P_{[n]}$ was shown to be systematically obtained to any order of γ^{-1} by several examples.
[11]S.V. Gantsevich, V.L. Gurevich, and R. Katilius, Phys. Cond. Matter **18**, 165 (1974).
[12]K. Furutsu, Radio Sci. **10**, 29 (1975).
[13]M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1964), p. 297.

Subsystems in physics described by bilinear maps between the corresponding vector spaces

Dirk Aerts[a)]
Theoretische Natuurkunde, Vrije Universiteit Brussel, Pleinlaan 2, B 1050 Brussel, Belgium

(Received 26 June 1979; accepted for publication 19 October 1979)

We show that whenever a physical system is composed of two subsystems, there exists a (σ_1,σ_2)-linear map between their generalized Hilbert spaces which describes this composition. As a consequence, subsystems of a physical system described by a generalized Hilbert space over a division ring K are always described by a generalized Hilbert space over a subdivision ring of K.

1. INTRODUCTION

Since it seems to be rather difficult to give a physical description of the whole universe, we are always obliged in physics to give a description of a subsystem of a bigger physical system.

Taking into account the results of two preceding papers[1,2] it seems to be appropriate to give the following definition of a subsystem: Given two physical systems S_1 and S, then S_1 is a subsystem of S iff there exists a third physical system S_2 such that S is composed out of S_1 and S_2. We will say that S is composed of S_1 and S_2 if the three following conditions are satisfied:

(1) The structure of S_1 and S_2 is preserved being considered as parts of S.

(2) A measurement on one of the system S_1 or S_2 does not disturb the other one.

(3) When we know the state of S_1 and of S_2 then we know also the state of S.

We want to look at this definition on a level as general as possible, therefore we will describe a physical system by means of the collection of all its properties, which seems to be the most general language at hand at this moment. We'll follow the interpretation given in Ref. 3, where from some basic physical requirements it is shown that the collection of all the properties of a physical system has to have the structure of a P lattice.[4] With this description it is indeed very easy to express mathematically the three requirements we asked for a physical system to be a subsystem. In Ref. 2 we were able to prove that for clasical systems, if S_1 and S_2 are described by the phase spaces Γ_1 and Γ_2, then (1), (2), and (3) force S to be described by the Cartesian product $\Gamma_1\times\Gamma_2$ of the two phase spaces. On the other hand for quantum systems if S_1 and S_2 are described by two complex Hilbert spaces, then (1), (2), and (3) force S to be described by the tensor product $\mathscr{H}_1\otimes\mathscr{H}_2$ of the two Hilbert spaces. This gives us two motivations for our definition. First of all the requirements (1), (2), and (3) seem to be physically intuitive and secondly they give the expected answers in the special cases of systems described by phase spaces and systems described by complex Hibert spaces. Since every classical system can be described by a phase space, the interpretation of (1), (2), and (3) for three classical systems was made in Ref. 2. For pure quantum systems however we restricted ourselves in Ref. 2 and 1 to the study of those quantum systems described by a complex Hibert space and all our results depend essentially on this restriction. In general a pure quantum system is described by a generalized Hilbert space (see Sec. 2) which is a vector space over an arbitrary division ring with involution. We'll try to see in this paper what the meaning of (1), (2), and (3) will be for general pure quantum systems.

We will prove that if a physical system is composed of two subsystems there exists a (σ_1, σ_2)-linear map (for a definition see Sec. 2) between their generalized Hilbert spaces which describes this composition in a way that will be made precise in the following. In the same way as the theorem of Wigner[5,3] is used to describe mathematically the symmetries of a physical system by means of σ-linear maps, we can use our theorem to describe subsystems of a physical system.

In a forthcoming paper we will also use the results of this paper to study the possible subsystems of an ordinary physical quantum system described by a complex Hilbert space.

Another reason why this general approach seems to be interesting is that in a forthcoming paper we will try to solve the problem of the joint system of two physical systems. Indeed we can ask ourselves if it is possible, having two physical system S_1 and S_2 described by P lattices $\mathscr{L}_1$ and $\mathscr{L}_2$, to find the P lattice which describes the joint system S. If S_1 and S_2 are classical systems described by phase spaces Γ_1 and Γ_2 this problem is solved in Ref. 2 and the joint system is described by $\Gamma_1\times\Gamma_2$. If S_1 and S_2 are pure quantum systems described by complex Hilbert spaces $\mathscr{H}_1$, and $\mathscr{H}_2$ the problem is also solved in Ref. 2 and the joint system is described by $\mathscr{H}_1\otimes\mathscr{H}_2$ or $\mathscr{H}_1\otimes\mathscr{H}_2^*$. For general systems we need the results of this paper, and for example it will be interesting to see which will be the P lattice describing the joint system of a quantum system and a classical system, for this would enable us to describe the measuring apparatus together with the system during the measurement.

2. DEFINITIONS, NOTATION, AND PRESENTATION OF THE PROBLEM

If S is a physical system we will denote by $\mathscr{L}$ the collection of all the properties (yes–no experiments) of S. If S is a classical system described in phase space Γ, then $\mathscr{L}$ is the

[a)]Scientific collaborator at the Interuniversitair Institut voor kerniweteschappeu (as part of Program 21 EN).

 0022-2488/80/040778-11$1.00

collection of all the subsets of phasespace. We'll denote this collection $\mathscr{P}(\Gamma)$. If S is a quantum system described in the complex Hilbert space $\mathscr{H}$, then $\mathscr{L}$ is the collection of all the closed subspaces of $\mathscr{H}$. We'll denote this set likewise $\mathscr{P}(\mathscr{H})$. In general $\mathscr{L}$ has to be a weakly modular, orthocomplemented, complete, atomic lattice satisfying the covering law. We'll call such a lattice a Piron lattice (P lattice).[6]

2.1. *Definition*: A set $\mathscr{L}$ is a P lattice iff:

(1) $\mathscr{L}$, $<$ is a partial ordered set,

(2) $\mathscr{L}$, $<$ is a complete lattice. This means that for every family $(b_i)_{i\in I}$ of elements in $\mathscr{L}$ there exists a greatest lower bound $\wedge_{i\in I} b_i$ and a least upper bound $\vee_{i\in I} b_i$.

(3) $\mathscr{L}$, $<$ is an orthocomplemented lattice, with orthocomplementation $'$. This means that $'$ is a bijection, mapping $\mathscr{L}$ to $\mathscr{L}$ such that for $b, c\in\mathscr{L}$

(i) $(b')' = b$,

(ii) $b < c \Rightarrow c' < b'$,

(iii) $b \vee b' = I$, $b \wedge b' = 0$, where $I = \vee_{b\in\mathscr{L}} b$ and $0 = \wedge_{b\in\mathscr{L}} b$.

(4) $\mathscr{L}$ is weakly modular. This means that whenever $b < c$, the sublattice generated by $\{b, b', c, c'\}$ is distributive. A lattice $\mathscr{L}$ is said to be distributive whenever $b \vee (c \wedge d) = (b \vee c) \wedge (b \vee d)$ if $b, c, d\in\mathscr{L}$

(5) $\mathscr{L}$ is atomic. This means that for every element $b\in\mathscr{L}$ there exists an atom $p\in\mathscr{L}$ such that $p < b$ where we mean by atom an element $p\in\mathscr{L}$ such that whenever $x\in\mathscr{L}$ and $0 < x < p$, then $x = 0$ or $x = p$.

(6) $\mathscr{L}$ satisfies the covering law. This means that if p is an atom of $\mathscr{L}$ and $a, x\in\mathscr{L}$ such that $a \wedge p = 0$ and $a < x < a \vee p$, then $x = a$ or $x = a \vee p$.

For a physical justification of these requirements we refer to Ref. 3.

2.2. *Definition*: Two properties $b, c\in\mathscr{L}$ are said to be compatible iff the sublattice generated by $\{b, b', c, c'\}$ is distributive. We shall denote this by $b \leftrightarrow c$

2.3. The following results from lattice theory will be used frequently in the calculations:

(1) In a weakly modular, complete orthocomplemented lattice $\mathscr{L}$ if $b\in\mathscr{L}$ and $a_i\in\mathscr{L}$ $\forall i\in I$ and $b \leftrightarrow a_i$ $\forall i\in I$, then

$$\vee_{i\in I}(b \wedge a_i) = b \wedge (\vee_{i\in I} a_i)$$

(see Ref. 3, Theorem 2.21).

(2) In a weakly modular, orthocomplemented lattice $\mathscr{L}$ we have two criteria which enable us to see whether two elements are compatible:

$$\text{If } b, c\in\mathscr{L}: b \leftrightarrow c \Longleftrightarrow (b \wedge c) \vee (b' \wedge c) = c \Longleftrightarrow (b' \vee c') \wedge c < b'$$

(see Ref. 3, §2.2).

(3) In a weakly modular, orthocomplemented lattice $\mathscr{L}$, the triplet (b, c, d) is distributive whenever one of the three elements is compatible with each of the other two (see Ref. 3, Theorem 2.25).

For a classical system described in a phase space Γ, $\mathscr{P}(\Gamma)$ is a P lattice and $<$ is the inclusion of subsets and the orthocomplement of a subset is his complement. For a quantum system described in a complex Hilbert space $\mathscr{H}$, $\mathscr{P}(\mathscr{H})$ is also a P lattice and $<$ is also the inclusion of subspaces and the orthocomplement of a subspace is the orthogonal subspace.

2.4. *Definition*: A physical system where every element is compatible with every other is said to be a classical system. A physical system where the only elements which are compatible with every other are 0 and I is said to be a pure quantum system.

In general a system is intermediate between those two extremes. It is in fact a quantum system with superselection rules and can be considered as a combination of pure quantum systems (see Ref. 3, chapter 2).

$\mathscr{P}(\Gamma)$ is a classical system and $\mathscr{P}(\mathscr{H})$ is a pure quantum system. The following representation theorem shows us that a classical system is always a $\mathscr{P}(\Gamma)$.

2.5. *Theorem*: If $\mathscr{L}$ is the P lattice of a classical system, then $\mathscr{L}$ is isomorphic to $\mathscr{P}(\Gamma)$, where Γ is the set of all the atoms of $\mathscr{L}$. (For the proof see Ref. 3, p. 9.)

For the pure quantum systems the situation is more complicated. $\mathscr{P}(\mathscr{H})$ is not the most general pure quantum system. Still there exists a representation theorem for pure quantum systems. Let us first give some definitions.

2.6 *Definition*: Let K be a division ring. A mapping α: $K \to K$ is called an involutive antiautomorphism if:

(1) $\alpha(k + l) = \alpha(k) + \alpha(l)$,

(2) $\alpha(k \cdot l) = \alpha(l) \cdot \alpha(k)$,

(3) $\alpha(\alpha(k)) = k$ for $k, l\in K$.

2.7 *Definition*: Let (V, K) be a vector space and α an involutive antiautomorphism over K. A mapping

$$\varphi: V \times V \to K,$$
$$(x, y) \mapsto \varphi(x, y),$$

is called a definite Hermitian form if it satisfies

(1) $\varphi(x + ky, z) = \varphi(x, z) + k\varphi(y, z)$,

(2) $\alpha(\varphi(x, y)) = \varphi(y, x)$,

(3) $\varphi(x, x) = 0 \Rightarrow x = 0$.

If (V, K) is a vectorspace equipped with a definite Hermitian form φ and $x, y\in V$ such that $\varphi(x, y) = 0$, we say that x is orthogonal to y. If M is a subset of V we define the orthogonal complement of M,

$$M^{\perp} = \{x | x\in V \text{ and } \varphi(x, y) = 0 \ \forall y\in M\}.$$

This is a subspace of V. It is easy to see that always $M \subset M^{\perp\perp}$. If $M = M^{\perp\perp}$ we will say that M is a closed subspace. If V is finite dimensional all the subspaces of V are closed. If $M \subset G$ are subsets of V, then $G^{\perp} \subset M^{\perp}$. Every subspace of the form $M^{\perp}$ is closed because $M^{\perp} = M^{\perp\perp\perp}$. In fact we can see that $M^{\perp\perp}$ is the smallest closed subspace containing M. Therefore we call it the closure of M. For an arbitrary family of subspaces $(M_i)_{i\in I}$ we have that

$$(\text{lin}(M_i))^{\perp} = \cap_{i\in I} M_i^{\perp}, \tag{2.2}$$

where $\text{lin}(M_i)$ is the subspace generated by all the M_i's. From (2.2) follows that the intersection of an arbitrary family of closed subspaces is closed. The set of closed subspaces is therefore a complete lattice which we will denote by $\mathscr{P}(V)$.

It is an orthocomplemented lattice because

$$': \mathscr{P}(V) \to \mathscr{P}(V),$$
$$M \mapsto M^{\perp},$$

is an orthocomplementation (Definition 2.1). It is atomic, the one dimensional subspaces being the atoms, and it satisfies the covering law (see Ref. 3, p. 55). So if M and G are closed subspaces of V we will write $M < G$ instead of $M \subset G$ and M' instead of $M^{\perp}$ and if $(M_i)_{i\in I}$ is a family of closed subspaces $\cap_i M_i = \wedge_i M_i$ and $(\cup_i M_i)^{\perp\perp} = \vee_i M_i$.

2.8 *Definition*: Let (V, K) be a vector space equipped with a definite Hermitian form. We'll say that (V, K) is a generalized Hilbert space iff

$$M^{\perp} + M^{\perp\perp} = V,$$

for every subset M of V.

We can see that if (V, K) is a generalized Hilbert space, $\mathscr{P}(V)$ is a weakly modular lattice (see Ref. 3, p. 55), so $\mathscr{P}(V)$ is a P lattice.

Thus $\mathscr{P}(V)$ is a more general example of a P lattice then the one used in ordinary quantum mechanics. A representation theorem of P lattices describing pure quantum systems proves that it is the most general.

2.9. *Theorem*: If $\mathscr{L}$ is the P lattice of a pure quantum system and $\mathscr{L}$ has more than three orthogonal atoms, then $\mathscr{L}$ is isomorphic to a set $\mathscr{P}(V)$ where (V, K) is a generalized Hilbert space (for the proof see Ref. 3).

2.10 *Definition*: Let (V, K) be a vector space equipped with a definite Hermitian form. A subset M of V is "dense" iff $M^{\perp\perp} = V$.

If one takes the field in this representation theorem to be $\mathbb{C}$ and the involutive antiautomorphism to be the conjugation, one can prove that the V becomes a complex Hilbert space (see Ref. 3, p. 56). This justifies the name "generalized Hilbert space."

2.11 *Definition*: Let (V_1, K_1) and (V, K) be two vector spaces and let σ be a morphism of K_1 into K. A map U of V_1 into V will be called a σ-linear map if:

(i) $U(x_1 + y_1) = U(x_1) + U(y_1), \quad \forall x_1, y_1 \in V_1$,

(ii) $U(k_1 x_1) = \sigma(k_1) U(x_1), \quad \forall k_1 \in k_1$.

It is called a σ-linear isomorphism if σ is an isomorphism and if U is bijective.

2.12. *Definition*: Let $(V_i, K_i)_{i=1,\ldots,n}$ be n vector spaces and (V, K) also a vector space and let $\sigma_i : K_i \to K$ be morphisms. A map B of $V_1 \times \cdots \times V_n$ into V will be called a $(\sigma_1,\ldots,\sigma_n)$-linear map if

$$\forall_i = 1\ldots, n$$
$$B(x_1,\ldots,x_i + y_i,\ldots,x_n) = B(x_1,\ldots,x_i,\ldots,x_n) + B(x_1,\ldots,y_i,\ldots,x_n),$$
$$\forall i = 1,\ldots n, \quad B(x_1,\ldots,k_i x_i,\ldots,x_n) = \sigma_i(k_i) B(x_1,\ldots,x_i,\ldots,x_n).$$

2.13. *Definition*: Let (V_1, K_1) and (V, K) be two vector spaces with definite Hermitian forms φ_1 and φ, a σ-linear map U will be called a σ-linear isometry, if there exists a $d \in K$ such that

$$\varphi(U(x_1), U(y_1)) = \sigma(\varphi_1(x_1, y_1))d, \quad \forall x_1, y_1 \in V_1.$$

2.14. *Definition*: Let $(V_i, K_i)_{i=1,\ldots,n}$ be n vector spaces equipped with definite Hermitian forms $\varphi_1,\ldots,\varphi_n$ and (V, K) a vector space equipped with a definite Hermitian form φ. A $(\sigma_1,\ldots,\sigma_n)$-linear map B will be called a $(\sigma_1,\ldots,\sigma_n)$-linear isometry whenever $\exists d \in K$ such that

$$\varphi(B(x_1,\ldots,x_n), B(y_1,\ldots,y_n)) = \prod_{i=1}^{n} \sigma_i(\varphi_i(x_i, y_i))d, \quad \forall x_i, y_i \in V_i, \quad d \in K.$$

2.15. *Definition*: Let (V_1, K_1) and (V, K) be two vector spaces equipped with a definite Hermitian form. A σ-linear map U is said to be continuous if the following condition is satisfied:

If M_1 is a subset of V_1, then $U(M_1^{\perp\perp}) \subset U(M_1)^{\perp\perp}$.

2.16. *Definition*: Let (V_i, K_i) be n vector spaces equipped with definite Hermitian forms $\varphi_1,\ldots,\varphi_n$ and (V, K) a vector space equipped with a definite Hermitian form φ, a $(\sigma_1,\ldots,\sigma_n)$-linear map B is said to be continuous if the following condition is satisfied:

If $M_1,\ldots,M_n$ are subsets of $V_1,\ldots,V_n$, then $B(M_1^{\perp\perp},\ldots,M_n^{\perp\perp}) \subset B(M_1,\ldots,M_n)^{\perp\perp}$.

2.17. *Definition*: Let (V_1, K_1) and (V, K) be two vector spaces equipped with a definite Hermitian form. A σ-linear map U is said to be "dense" if it maps every dense subset of V_1 to a dense subset of V.

2.18. *Definition*: Let (V_i, K_i) be n vector spaces equipped with definite Hermitian forms $\varphi_1,\ldots,\varphi_n$ and (V, K) a vector space equipped with a definite Hermitian form φ, a $(\sigma_1,\ldots,\sigma_n)$-linear map B is said to be dense if every family $(M_i)_{i=1,\ldots,n}$ of dense subsets of $(V_i)_{i=1,\ldots,n}$ is mapped to a dense subset of V.

2.19. Let us now interpret the three conditions to be fulfilled for a system to be a subsystem of a bigger system. Suppose we have three systems S_1, S_2, and S with their corresponding P lattices $\mathscr{L}_1$, $\mathscr{L}_2$, and $\mathscr{L}$.

[1] The structure of S_1 and S_2 has to be conserved when they are considered as parts of S.

It is clear that every property of S_1 and S_2 has to be a property of S. So there must exist maps

$$h_1 : \mathscr{L}_1 \to \mathscr{L}, \quad h_2 : \mathscr{L}_2 \to \mathscr{L}.$$

h_1 and h_2 have to conserve the structure of $\mathscr{L}_1$ and $\mathscr{L}_2$. It is sufficient to ask the following:

(i) for $a_1, b_1 \in \mathscr{L}_1$ and $a_1 < b_1'$ we have $h_1(a_1) < h_1(b_1)'$, (2.3)

(ii) for $(a_i)_{i\in I}$, $a_i \in \mathscr{L}_1$ we have $h_1(\wedge_{i\in I} a_i) = \wedge_{i\in I} h_1(a_i)$. (2.4)

From these requirements follows immediately that

(iii) $h_1(0_1) = 0$,

(iv) $h_1(a_1') = h_1(a_1)' \wedge h_1(I_1)$,

(v) $h_1(\vee_{i\in I} a_i) = \vee_{i\in I} h_1(a_i)$.

So (i) and (ii) make h_1 to conserve the structure of $\mathscr{L}_1$. We ask the same properties for h_2.

Definition: A map h_1 from a P lattice $\mathscr{L}_1$ to a P lattice $\mathscr{L}$ which satisfies (i) and (ii) will be called a c-morphism.

$h_1(I_1)$ is a property which the system has whenever S_1 exists. So it is a property that the system always has. (2.5)

Therefore $h_1(I_1) = I$ and for the same reason $h_2(I_2) = I$.

Definition: A c-morphism between two P lattices such that the identity is mapped on the identity is called a unitary c-morphism.

[2] No measurement on S_1 may disturb S_2 and vice versa. The mathematical translation of this requirement is the following: If $a_1 \in \mathscr{L}_1$ and $a_2 \in \mathscr{L}_2$, then $h_1(a_1) \leftrightarrow h_2(a_2)$. This follows directly from the interpretation of the compatibility relation (see Ref. 3).

[3] When we know the state of S_1 and S_2 then we must know the state of S. The state of a physical system is the set of all the properties which are actually true for the system. Usually this state is represented by the infimum of all these properties which has to be an atom of the P lattice if the state represents maximal information (see Ref. 2, 1.2). So the atoms of the P lattice represent the states of the corresponding system.

The mathematical translation of our requirement then becomes: If p_1 is an atom of $\mathscr{L}_1$ and p_2 is an atom of $\mathscr{L}_2$, then $h_1(p_1) \wedge h_2(p_2)$ is an atom of $\mathscr{L}$.

2.20. *Conclusion*: If S_1, S_2, and S are three physical systems described by P lattices $\mathscr{L}_1$, $\mathscr{L}_2$, and $\mathscr{L}$, then S_1 and S_2 will be subsystems of S whenever

(1) There exists two unitary c-morphism

$$h_1 : \mathscr{L}_1 \to \mathscr{L}\,, \quad h_2 : \mathscr{L}_2 \to \mathscr{L}\,.$$

(2) $h_1(a_1) \leftrightarrow h_2(a_2)$ for every $a_1 \in \mathscr{L}_1$ and $a_2 \in \mathscr{L}_2$

(3) If p_1 is an atom of $\mathscr{L}_1$ and p_2 an atom of $\mathscr{L}_2$, then $h_1(p) \wedge h_2(p_2)$ is an atom of $\mathscr{L}$.

From now on we will always describe a physical system which has more than three orthogonal atoms by its generalized Hilbert space. If x is a vector of a generalized Hibert space (V, K), then $\bar{x}$ will represent the one-dimensional subspace generated by x.

We will now prove the following two theorems in the next section.

2.21. *Theorem*: If we have three physical systems described by P lattices $\mathscr{P}(V_1)$, $\mathscr{P}(V_2)$, and $\mathscr{P}(V)$, where (V_1, K_1), (V_2, K_2), and (V_2, K_2) are generalized Hilbert spaces and

$$B : V_1 \times V_2 \to V \quad \text{is a dense } (\sigma_1, \sigma_2)\text{-isometry}$$

and we define

$$h_1 : \mathscr{P}(V_1) \to \mathscr{P}(V)\,,$$
$$a_1 \to B(a_1, V_2)^{\perp\perp}\,,$$
$$h_2 : \mathscr{P}(V_2) \to \mathscr{P}(V)\,,$$
$$a_2 \to B(V_1, a_2)^{\perp\perp}\,.$$

Then

(1) h_1 and h_2 are unitary c-morphisms,

(2) $h_1(a_1) \leftrightarrow h_2(a_2)$ for $a_1 \in \mathscr{P}(V_1)$ and $a_2 \in \mathscr{P}(V_2)$,

(3) $h_1(p_1) \wedge h_2(p_2)$ is an atom if p_1 is an atom of $\mathscr{P}(V_1)$ and p_2 is an atom of $\mathscr{P}(V_2)$.

2.22. *Theorem*: If we have two physical systems with each more than two orthogonal atoms and described by the P lattices $\mathscr{P}(V_1)$ and $\mathscr{P}(V_2)$, where (V_1, K_1) and (V_2, K_2) are generalized Hilbert spaces, and a third physical system described by the P lattice $\mathscr{P}(V)$, where (V, K) is a generalized Hilbert space, and two maps

$$h_1 : \mathscr{P}(V_1) \to \mathscr{P}(V)\,, \quad h_2 : \mathscr{P}(V_2) \to \mathscr{P}(V),$$

such that (1), (2), and (3) of Theorem 2.21 are fulfilled, then there exists for every $z \in h_1(z_1) \wedge h_2(z_2)$ one unique dense (σ_1, σ_2)-linear isometry $B : V_1 \times V_2 \to V$ such that $B(z_1, z_2) = z$ and such that

$$h_1(a_1) = B(a_1, V_2)^{\perp\perp}\,,$$
$$h_2(a_2) = B(V_1, a_2)^{\perp\perp}\,,$$
$$h_1(a_1) \wedge h_2(a_2) = B(a_1, a_2)^{\perp\perp}\,.$$

These two theorems will prove that subsystems are described by dense (σ_1, σ_2)-linear isometries between the corresponding generalized Hilbert spaces.

3. PROOF OF THEOREMS 2.21 AND 2.22

We will prove Theorems 2.21 and 2.22 in different steps.

3.1. *Lemma*: If (V_1, K_1), (V_2, K_2), and (V, K) are three generalized Hilbert spaces with Hermitian form φ_1, φ_2 and φ, and

$B : V_1 \times V_2 \to V$ is a dense(σ_1, σ_2)-isometry

then

$$\text{(i) } B(a_1^{\perp}, V_2)^{\perp\perp} = B(a_1, V_2)^{\perp} \text{ if } a_1 \text{ is a subset of } V_1\,, \quad (3.1)$$

$$\text{(ii) } B(V_1, a_2^{\perp})^{\perp\perp} = B(V_1, a_2)^{\perp} \text{ if } a_2 \text{ is a subset of } V_2\,, \quad (3.2)$$

$$\text{(iii) } B(a_1, a_2)^{\perp\perp} = B(a_1, V_2)^{\perp\perp} \wedge B(V_1, a_2)^{\perp\perp} \text{ if } a_1 \text{ is a subset of } V_1 \text{ and } a_2 \text{ a subset of } V_2 \quad (3.3)$$

Proof: (i) Take a subset a_1 of V_1.

$$\begin{aligned} B(a_1^{\perp}, V_2)^{\perp\perp} &= \{B(x_1, x_2) | \varphi_1(x_1, y_1) = 0 \text{ for } y_1 \in a_1 \\ &\quad \text{and } x_2 \in V_2\}^{\perp\perp} \\ &= \{B(x_1, x_2) | \sigma_1(\varphi_1(x_1, y_1)) = 0 \text{ for } y_1 \in a_1 \\ &\quad \text{and } x_2 \in V_2\}^{\perp\perp} \\ &= \{B(x_1, x_2) | \varphi(B(x_1, x_2), B(y_1, x_2)) = 0\,, \\ &\quad y_1 \in a_1,\ x_2 \in V_2\}^{\perp\perp} \\ &= [B(V_1, V_2) \cap B(a_1, V_2)^{\perp}]^{\perp\perp}\,. \end{aligned}$$

Now $B(V_1, V_2) \cap B(a_1, V_2)^{\perp} \subset B(a_1, V_2)^{\perp}$. So $[B(V_1, V_2) \cap B(a_1, V_2)^{\perp}]^{\perp\perp} < B(a_1, V_2)^{\perp}$. Therefore $B(a_1^{\perp}, V_2)^{\perp\perp} < B(a_1, V_2)^{\perp}$. Since $\mathscr{P}(V)$ is weakly modular, $B(a_1^{\perp}, V_2)^{\perp\perp}$ and $B(a_1, V_2)^{\perp}$ are compatible. So from Sec. 2.3 (2) it follows that

$$B(a_1, V_2)^{\perp} = B(a_1^{\perp}, V_2)^{\perp\perp} \vee [B(a_1, V_2)^{\perp} \wedge B(a_1^{\perp}, V_2)^{\perp}]\,.$$

Suppose $z \in B(a_1, V_2)^{\perp} \wedge B(a_1^{\perp}, V_2)^{\perp}$, then

$$\varphi(z, B(y_1, x_2)) = 0 \ \forall y_1 \in a_1^{\perp}\,, \ \forall x_2 \in V_2\,,$$
$$\varphi(z, B(y_1, x_2)) = 0 \ \forall y_1 \in a_1\,, \ \forall x_2 \in V_2\,,$$

so

$$\varphi(z, B(y_1, x_2)) = 0 \ \forall y_1 \in a_1 \cup a_1^{\perp}\,, \ \forall x_2 \in V_2\,.$$

Now $a_1 \cup a_1^{\perp}$ is a dense subset of V_1 because

$$(a_1 \cup a_1^{\perp})^{\perp} = a_1^{\perp} \wedge a_1^{\perp\perp} = 0\,,$$

so $B(a_1 \cup a_1^{\perp}, V_2)$ is a dense subset of V. But then $B(a_1 \cup a_1^{\perp}, V_2)^{\perp} = 0$. So $z = 0$ which proves that $B(a_1, V_2)^{\perp} \wedge B(a_1^{\perp}, V_2)^{\perp} = 0$ and $B(a_1, V_2)^{\perp} = B(a_1^{\perp}, V_2)^{\perp\perp}$.

(ii) In an analogous way we prove that

$$B(V_1, a_2)^{\perp} = B(V_1, a_2^{\perp})^{\perp\perp}.$$

(iii) $B(a_1, a_2) = B(a_1, V_2) \cap B(V_1, a_2)$,

so

$$B(a_1, a_2)^{\perp\perp} = [B(a_1, V_2) \cap B(V_1, a_2)]^{\perp\perp}$$
$$= B(a_1, V_2)^{\perp\perp} \wedge B(V_1, a_2)^{\perp\perp}.$$

■

3.2. *Lemma*: If (V_1, K_1), (V_2, K_2), and (V, K) and are three generalized Hilbert spaces with Hermitian forms φ_1, φ_2, φ, and if $B : V_1 \times V_2 \to V$ is a (σ_1, σ_2)-isometry, then B is dense if and only if B is continuous and $B(V_1, V_2)^{\perp\perp} = 0$.

Proof: If B is continuous and $B(V_1, V_2)^{\perp\perp} = V$ and $M_1^{\perp\perp} = V_1$ and $M_2^{\perp\perp} = V_2$, then $V = B(M_1^{\perp\perp}, M_2^{\perp\perp})^{\perp\perp} \subset B(M_1, M_2)^{\perp\perp}$. So $V = B(M_1, M_2)^{\perp\perp}$ which proves that B is dense.

Suppose that B is dense then we see immediately that $B(V_1, V_2)^{\perp\perp} = V$. Take a_1 a subset of V_1 and a_2 a subset of V_2. Using (3.1), (3.2), and (3.3) we have:

$$B(a_1^{\perp\perp}, a_2^{\perp\perp})^{\perp\perp} = B(a^{\perp\perp}, V_2)^{\perp\perp} \wedge B(V_1, a_2^{\perp\perp})^{\perp\perp}$$
$$= B(a_1^{\perp}, V_2)^{\perp} \wedge B(V_1, a_2^{\perp})^{\perp}$$
$$= B(a_1, V_2)^{\perp\perp} \wedge B(V_1, a_2)^{\perp\perp}$$
$$= B(a_1, a_2)^{\perp\perp},$$

so $B(a_1^{\perp\perp}, a_2^{\perp\perp}) \subset B(a_1, a_2)^{\perp\perp}$,

which proves that B is continuous.

■

These two lemmas are sufficient to prove Theorem 2.21.

3.3. *Theorem*: If (V_1, K_1), (V_2, K_2), and (V, K) are three generalized Hilbert spaces with Hermitian forms φ_1, φ_2, and φ, and $B : V_1 \times V_2 \to V$ is a dense (σ_1, σ_2)-isometry and we define maps

$$h_1 : \mathscr{P}(V_1) \to \mathscr{P}(V),$$
$$a_1 \to B(a_1, V_2)^{\perp\perp},$$
$$h_2 : \mathscr{P}(V_2) \to \mathscr{P}(V),$$
$$a_2 \to B(V_1, a_2)^{\perp\perp},$$

then (1) h_1 and h_2 are unitary c-morphisms,

(2) $h_1(a_1) \leftrightarrow h_2(a_2)$ for $a_1 \in \mathscr{P}(V_1)$ and $a_2 \in \mathscr{P}(V_2)$,

(3) $h_1(p_1) \wedge h_2(p_2)$ is an atom whenever p_1 is an atom of $\mathscr{P}(V_1)$ and p_2 is an atom of $\mathscr{P}(V_2)$.

Proof: (1) If $a_1 \in \mathscr{P}(V_1)$, then using Lemma 3.1 we have

$$h_1(a_1^{\perp}) = B(a_1^{\perp}, V_2)^{\perp\perp} = B(a_1^{\perp\perp}, V_2)^{\perp}$$
$$= B(a_1, V_2)^{\perp} = h_1(a_1)^{\perp}.$$

Take a family $(a_i)_{i \in I}$ with $a_i \in \mathscr{P}(V_1)$ then using again lemma 3.1 we have

$$\bigwedge_{i \in I} h_1(a_i) = \bigwedge_{i \in I} B(a_i, V_2)^{\perp\perp}$$
$$= \bigwedge_{i \in I} B(a_i^{\perp}, V_2)^{\perp} \quad \text{Put } a_i^{\perp} = b_i$$
$$= \bigcap_{i \in I} B(b_i, V_2)^{\perp}$$
$$= B\left(\bigcup_{i \in I} b_i, V_2\right)^{\perp}$$
$$= B\left(\left(\bigcap_{i \in I} b_i\right)^{\perp}, V_2\right)^{\perp\perp}$$
$$= B\left(\bigcap_{i \in I} b_i^{\perp}, V_2\right)^{\perp\perp}$$
$$= B\left(\bigwedge_{i \in I} a_i, V_2\right)^{\perp\perp} = h_1\left(\bigwedge_{i \in I} a_i\right),$$

$h_1(V_1) = B(V_1, V_2)^{\perp\perp} = V$ (see Lemma 3.2).

So h_1 is a map that fulfills (2.3), (2.4), and (2.5) and is therefore a unitary c-morphism.

In an analogous way we prove that h_2 is a unitary c-morphism.

(2) Take $a_1 \in \mathscr{P}(V_1)$ and $a_2 \in \mathscr{P}(V_2)$. We want to see that $h_1(a_1) \leftrightarrow h_2(a_2)$. Since $\mathscr{P}(V)$ is a P-lattice we can use the condition given by Sec. 2.3 (2),

$$h_1(a_1) \leftrightarrow h_2(a_2)$$
$$\Longleftrightarrow [h_1(a_1)^{\perp} \vee h_2(a_2)^{\perp}] \wedge h_2(a_2) < h_1(a_1)^{\perp}.$$

We have

$$h_1(a_1)^{\perp} \vee h_2(a_2)^{\perp} = [h_1(a_1) \wedge h_2(a_2)]^{\perp} = B(a_1, a_2)^{\perp}.$$

Now take $z \in [h_1(a_1)^{\perp} \vee h_2(a_2)^{\perp}] \wedge h_2(a_2)$, then

$$\varphi(z, B(x_1, x_2)) = 0 \ \forall x_1 \in a_1, \ \forall x_2 \in a_2,$$
$$\varphi(z, B(y_1, y_2)) = 0 \ \forall y_1 \in V_1, \ \forall y_2 \in a_2.$$

Take now $x_2 \in V_2$. Since V_2 is a generalized Hilbert space we have

$$a_2 + a_2^{\perp} = V_2.$$

So there exist $t_2 \in a_2$ and $z_2 \in a_2^{\perp}$ such that $x_2 = t_2 + z_2$. If $x_1 \in a_1$ we have

$$\varphi(z, B(x_1, x_2)) = \varphi(z, B(x_1, t_2) + B(x_1, t_2))$$
$$= \varphi(z, B(x_1, t_2)) + \varphi(z, B(x_1, z_2))$$
$$= 0.$$

So $z \in B(a_1, V_2)^{\perp} = h_1(a_1)^{\perp}$. This proves that $h_1(a_1) \leftrightarrow h_2(a_2)$.

(3) If $\bar{x}_1$ is an atom of $\mathscr{P}(V_1)$ and $\bar{x}_2$ an atom of $\mathscr{P}(V_2)$ we have from Lemma 3.1 that

$$h_1(\bar{x}_1) \wedge h_2(\bar{x}_2)$$
$$= B(\bar{x}_1, V_2)^{\perp\perp} \wedge B(V_1, \bar{x}_2)^{\perp\perp}$$
$$= B(\bar{x}_1, \bar{x}_2)^{\perp\perp}$$
$$= \overline{B(x_1, x_2)} \text{ and this is an atom of } \mathscr{P}(V).$$

■

This concludes the proof of Theorem 2.21. We will show next that every two subsystems are described in this way by a dense (σ_1, σ_2)-linear isometry. For this we need to find a linear structure underlying the c-morphisms h_1 and h_2. Let us note a result from lattice theory which says that any c-morphisms between P lattices representing pure quantum system is injective (See 3, pp. 31, 33).

3.4. *Theorem*: Suppose we have three P lattices $\mathscr{L}_1$, $\mathscr{L}_2$, and $\mathscr{L}$ and two maps

$$h_1 : \mathscr{L}_1 \to \mathscr{L}, \quad h_2 : \mathscr{L}_2 \to \mathscr{L},$$

which satisfy the coupling conditions:

(1) h_1 and h_2 are unitary c-morphisms, (3.4)

(2) $h_1(a_1) \leftrightarrow h_2(a_2) \quad \forall a_1 \in \mathscr{L}_1, \quad \forall a_2 \in \mathscr{L}_2,$

(3) $h_1(p_1) \wedge h_2(p_2)$ is an atom whenever p_1 and p_2 are atoms.

For every atom p_1 of $\mathscr{L}_1$ and for every atom p_2 of $\mathscr{L}_2$ consider following maps:

$$u_{p_2} : \mathscr{L}_1 \to [0, h_2(p_2)] = \{a_2 | a_2 \in \mathscr{L}_2 \text{ and } a_2 < h_2(p_2)\},$$
$$a_1 \to h_1(a_1) \wedge h_2(p_2),$$
$$v_{p_1} : \mathscr{L}_2 \to [0, h_1(p_1)],$$
$$a_2 \to h_2(a_2) \wedge h_1(p_1),$$

then u_{p_1} and v_{p_1} are c-morphisms which map atoms to atoms.

Proof: Since h_2 is injective we have $h_2(p_2) \neq 0$, and $[0, h_2(p_2)]$ is a P lattice with relative orthocomplementation $a^r = a' \wedge h_2(p_2)$ (see Ref. 3, p. 30).

$$\begin{aligned} u_{p_2}(\vee_i a_i) &= h_1(\vee_i a_i) \wedge h_2(p_2) \\ &= [\vee_i h_1(a_i)] \wedge h_2(p_2) \quad \text{using Sec. 2.3 (1)} \\ &\quad \text{and (3,4) we have:} \\ &= \vee_i (h_1(a_i) \wedge h_2(p_2)) \\ &= \vee_i u_{p_2}(a_i), \end{aligned}$$

$$\begin{aligned} u_{p_2}(a') &= h_1(a') \wedge h_2(p_2) \quad \text{using Sec. 2.3 (2) we have:} \\ &= [h_1(a') \vee h_2(p_2)'] \wedge h_2(p_2) \\ &= [h_1(a) \wedge h_2(p_2)]' \wedge h_2(p_2) \\ &= u_{p_2}(a)' \wedge h_2(p_2) \\ &= u_{p_2}(a)^r. \end{aligned}$$

$u_{p_2}(p_1) = h_1(p_1) \wedge h_2(p_2)$ and this is an atom if p_1 and p_2 are atoms.

$$u_{p_2}(I_1) = h_1(I_1) \wedge h_2(p_2) = h_2(p_2).$$

This proves that u_{p_2} is a unitary c-morphism that maps atoms to atoms. In an analagous way we prove that v_{p_1} is a unitary c-morphism that maps atoms to atoms. ■

Using the notation of the preceding theorem we can see that $v_{p_1}(\mathscr{L}_2)$ and $u_{p_2}(\mathscr{L}_1)$ are P lattices with atoms which are also atoms of $\mathscr{L}$; we will denote these P lattices $\mathscr{L}_{p_1}$ and $\mathscr{L}_{p_2}$.

3.5. *Lemma*: If we consider the following map

$$f_{q_1,p_1} : \mathscr{L}_{p_1} \to \mathscr{L}_{q_1}$$
$$a \to [a \vee h_1(r_1)] \wedge h_1(q_1),$$

where p_1, q_1, r_1 are different atoms of $\mathscr{L}_1$ such that $r_1 < p_1 \vee q_1$, then $f_{q_1,p_1}(a) = v_{q_1} v_{p_1}^{-1}(a)$ and so is independent of r_1.

Proof: If $a \in \mathscr{L}_{p_1}$, then $a = v_{p_1}(a_2) = h_2(a_2) \wedge h_1(p_1)$. This gives us:

$$f_{q_1,p_1}(a) = [(h_2(a_2) \wedge h_1(p_1)) \vee h_1(r_1)] \wedge h_1(q_1).$$

Since $h_1(r_1) \leftrightarrow h_2(a_2)$ and $h_1(p_1) \leftrightarrow h_2(a_2)$ we can use Sec. 2.3 (3) and get:

$$\begin{aligned} f_{q_1,p_1}(a) &= [h_2(a_2) \vee h_1(r_1)] \wedge [h_1(p_1) \vee h_1(r_1)] \wedge h_1(q_1) \\ &= [h_2(a_2) \vee h_1(r_1)] \wedge h_1((p_1 \vee r_1) \wedge q_1) \\ &= [h_2(a_2) \vee h_1(r_1)] \wedge h_1(q_1). \end{aligned}$$

Since $h_1(q_1) \leftrightarrow h_2(a_2)$ and $h_1(r_1) \leftrightarrow h_2(a_2)$ we can again use Sec. 2.3 (3) and get:

$$\begin{aligned} f_{q_1,p_1}(a) &= [h_2(a_2) \wedge h_1(q_1)] \vee [h_1(r_1) \wedge h_1(q_1)] \\ &= h_2(a_2) \wedge h_1(q_1) \\ &= v_{q_1}(a_2) \\ &= v_{q_1} v_{p_1}^{-1}(a). \end{aligned}$$

■

3.6. *Lemma*: Suppose that from the three P lattices $\mathscr{L}_1$, $\mathscr{L}_2$, and $\mathscr{L}$ of Theorem 3.4, $\mathscr{L}_1$ and $\mathscr{L}$ are of the form $\mathscr{P}(V_1)$ and $\mathscr{P}(V)$, where (V_1, K_1) and (V, K) are generalized Hilbert spaces, and let h_1, h_2, u_{p_2}, $v_{\bar{x}_2}$ be as in Theorem 3.4, where $x_1 \in V_1$, and let $f_{\bar{y}_1,\bar{x}_1}$ be as in lemma 3.5 where $y_1, x_1 \in V_1$. If $x \in V$ such that $\bar{x} \in \mathscr{L}_{\bar{x}_1}$, then

$$x \in h_1(\bar{y}_1) + h_1(x_1 - y_1) \ \forall y_1 \in V_1.$$

Proof: If $\bar{x} \in \mathscr{L}_{\bar{x}_1}$, then we can consider $f_{\bar{y}_1,\bar{x}_1}(\bar{x})$ and $f_{\overline{x_1 - y_1},\bar{x}_1}(\bar{x})$ which are atoms of $\mathscr{L}_{\bar{y}_1}$ and $\mathscr{L}_{\overline{x_1 - y_1}}$, $f_{\bar{y}_1,\bar{x}_1}(\bar{x}) \vee f_{\overline{x_1 - y_1},\bar{x}_1}(\bar{x})$

$$\begin{aligned} &= v_{\bar{y}_1} v_{\bar{x}_1}^{-1}(\bar{x}) \vee v_{\overline{x_1 - y_1}} v_{\bar{x}_1}^{-1}(\bar{x}) \\ &= [h_2(v_{\bar{x}_1}^{-1}(\bar{x})) \wedge h_1(\bar{y}_1)] \vee [h_2(v_{\bar{x}_1}^{-1}(\bar{x})) \wedge h_1(\overline{x_1 - y_1})] \\ &= h_2(v_{\bar{x}_1}^{-1}(\bar{x})) \wedge [h_1(\bar{y}_1) \vee h_1(\overline{x_1 - y_1})] \end{aligned}$$

And since $\bar{x} = h_2(v_{\bar{x}_1}^{-1}(\bar{x})) \wedge h_1(\bar{x}_1)$ we have that $\bar{x} < f_{\bar{y}_1,\bar{x}_1}(\bar{x}) \vee f_{\overline{x_1 - y_1},\bar{x}_1}(\bar{x})$. So there exists a vector $w \in f_{\overline{y_1 - x_1}}(\bar{x})$ and a vector $t \in f_{\overline{x_1 - y_1},\bar{x}_1}(\bar{x})$ such that $x = w + t$ and we also have $w \in h_1(\bar{y}_1)$ and $t \in h_1(\overline{x_1 - y_1})$, thus $x \in h_1(\bar{y}_1) + h_1(\overline{x_1 - y_1})$. ■

3.7. The vectors w and t which we constructed for $\bar{x} \in \mathscr{L}_{\bar{x}_1}$ and for every $y_1 \in V_1$ in the preceding lemma will be denoted

$$w = F_{y_1,x_1}(x), \quad t = F_{x_1 - y_1,x_1}(x).$$

So we have $x = F_{y_1,x_1}(x) + F_{x_1 - y_1,x_1}(x)$ where in a certain sense $F_{y_1,x_1}(x)$ and $F_{x_1 - y_1,x_1}(x)$ are the projections of x on $h_1(\bar{y}_1)$ and on $h_1(x_1 - y_1)$. Note that for an arbitrary vector of $h_1(\bar{x}_1)$ these projections do not necessarily exist since $h_1(\bar{x}_1)$ is not necessarily a part of $h_1(\bar{y}_1) + h_1(\overline{x_1 - y_1})$ although it is contained in $h_1(\bar{y}_1) \vee h_1(x_1 - y_1)$. But when they exist they are unique. Indeed suppose $x = r + s$, with $r \in h_1(\bar{y}_1)$ and $s \in h_1(\overline{x_1 - y_1})$, then

$$F_{y_1,x_1}(x) + F_{x_1 - y_1,x_1}(x) = r + s.$$

So

$$F_{y_1,x_1}(x) - r = s - F_{x_1 - y_1,x_1}(x) = z$$

and $z \in h_1(\bar{y}_1) \wedge h_1(\overline{x_1 - y_1})$, so $z = 0$. This gives us

$$r = F_{y_1,x_1}(x) \quad \text{and} \quad s = F_{x_1 - y_1,x_1}(x).$$

So we can consider the set of maps

$$F_{y_1,x_1} : h_1(\bar{x}_1) \to h_1(\bar{y}_1) \quad \text{for} \quad x_1 \neq 0 \neq y_1 \in V_1,$$

the domain of F_{y_1,x_1} being all the vectors $x \in V$ such that $\bar{x} \in \mathscr{L}_{\bar{x}_1}$.

We will now study the structure of this set.

3.8. *Lemma*: Let $\mathscr{P}(V_1)$, $\mathscr{L}_2$, $\mathscr{P}(V)$, h_1 and h_2 be as in Lemma 3.6 and suppose further that $\mathscr{P}(V_1)$ has more than two orthogonal atoms. Let $\{F_{y_1,x_1} : y_1, x_1 \in V_1, x_1 \neq 0 \neq y_1\}$ be the set of maps constructed in 3.7. This set has the following properties:

(1) If x is in the domain of F_{y_1,x_1} and $k \in K$, then kx is in

the domain of F_{y_1,x_1}, if y is in the domain of F_{y_1,x_1} such that $x+y$ is in the domain of F_{y_1,x_1}, then we have:

$$F_{y_1,x_1}(kx) = kF_{y_1,x_1}(x),$$

$$F_{y_1,x_1}(x+y) = F_{y_1,x_1}(x) + F_{y_1,x_1}(y).$$

(2) The domain of F_{x_1,y_1} is the image of F_{y_1,x_1} and $F_{y_1,x_1} = F^{-1}_{x_1,y_1}$. We have the following chain rule:

(3) $F_{z_1,y_1} \circ F_{y_1,x_1} = F_{z_1,x_1}$ for every $y_1 \in V_1$.

(4) $F_{x_1,x_1}(x) = x$ for $\bar{x} \in \mathscr{L}_{\bar{x}_1}$.

(5) $F_{k_1y_1,k_1x_1} = F_{y_1,x_1}$ for every $k_1 \in K_1$,

Proof: If x is in the domain of F_{y_1,x_1}, then $\bar{x} \in \mathscr{L}_{\bar{x}_1}$ and so kx is in the domain of F_{y_1,x_1} for $k \in K$

$$\begin{aligned} kx &= F_{y_1,x_1}(kx) + F_{x_1-y_1,x_1}(kx) \\ &= kF_{y_1,x_1}(x) + kF_{x_1-y_1,x_1}(x), \end{aligned}$$

and since the decomposition is unique we must have

$$F_{y_1,x_1}(kx) = kF_{y_1,x_1}(x).$$

In an analogous way we have

$$F_{y_1,x_1}(x+y) = F_{y_1,x_1}(x) + F_{y_1,x_1}(y).$$

This proves (1).

Consider $F_{y_1,x_1} : h_1(\bar{y}_1) \to h_1(\bar{x}_1)$. If $\bar{y} \in \mathscr{L}_{\bar{y}_1}$ we have

$$y = F_{x_1,y_1}(y) + F_{y_1-x_1,y_1}(y)$$

and $\overline{F_{x_1,y_1}(y)} \in \mathscr{L}_{\bar{x}_1}$ so we have

$$F_{x_1,y_1}(y) = F_{y_1,x_1}F_{x_1,y_1}(y) + F_{x_1-y_1,x_1}F_{x_1,y_1}(y).$$

So

$$\begin{aligned} y &= F_{y_1,x_1}F_{x_1,y_1}(y) + F_{x_1-y_1,x_1}F_{x_1,y_1}(y) \\ &\quad + F_{y_1-x_1,y_1}(y) \end{aligned}$$

This gives us

$$\begin{aligned} t &= y - F_{y_1,x_1}F_{x_1,y_1}(y) \\ &= F_{x_1-y_1,x_1}F_{x_1,y_1}(y) + F_{y_1-x_1,y_1}(y). \end{aligned}$$

This proves that $t \in h_1(\bar{y}_1) \wedge h_1(\overline{x_1-y_1})$, so $t=0$ which proves that

$$y = F_{y_1,x_1}F_{x_1,y_1}(y).$$

In an analogous way we prove that $F_{x_1,y_1}F_{y_1,x_1}(x) = x$ for $x \in h_1(\bar{x}_1)$ such that $\bar{x} \in \mathscr{L}_{\bar{x}_1}$. This proves (2).

Now take $x_1, y_1, z_1 \in V_1$ such that x_1, y_1, and z_1 are linearly independent. Since $\mathscr{P}(V_1)$ has more than two orthogonal atoms there always exist three linearly independent vectors in V_1,

$$\overline{y_1-z_1} < \bar{y}_1 \vee \bar{z}_1 \quad \text{and} \quad \overline{y_1-z_1} < \overline{y_1-x_1} \vee \overline{x_1-z_1}.$$

So $\overline{y_1-z_1} < (\bar{y}_1 \vee \bar{z}_1) \wedge (\overline{y_1-x_1} \vee \overline{x_1-z_1})$ and since y_1, z_1, x_1 are linearly independent we have

$$\overline{y_1-z_1} = (\bar{y}_1 \vee \bar{z}_1) \wedge (\overline{y_1-x_1} \vee \overline{x_1-z_1}).$$

So

$$\begin{aligned} &h_1(\overline{y_1-z_1}) \\ &\quad = (h_1(\bar{y}_1) \vee h_1(\bar{z}_1)) \wedge (h_1(\overline{y_1-x_1}) \vee h_1(\overline{x_1-y_1})). \end{aligned}$$

Now take $\bar{x} \in \mathscr{L}_{\bar{x}_1}$. Then we have

$$\begin{aligned} x &= F_{y_1,x_1}(x) + F_{y_1-x_1,x_1}(x) \\ &= F_{z_1,x_1}(x) + F_{z_1-x_1,x_1}(x). \end{aligned}$$

So

$$\begin{aligned} &F_{y_1,x_1}(x) - F_{z_1,x_1}(x) \\ &\quad = F_{z_1-x_1,x_1}(x) - F_{y_1-x_1,x_1}(x) = t. \end{aligned}$$

Then $t \in h_1(\overline{y_1-z_1})$, so we have

$$F_{y_1,x_1}(x) = F_{z_1,x_1}(x) + t.$$

But

$$F_{y_1,x_1}(x) = F_{z_1,y_1}F_{y_1,x_1}(x) + F_{z_1-y_1,y_1}F_{y_1,x_1}(x).$$

So since the decomposition is unique we must have

$$F_{z_1,x_1}(x) = F_{z_1,y_1}F_{y_1,x_1}(x).$$

If x_1, y_1 are linearly dependent, when $x_1 \in \bar{y}_1$ it is now possible for us to define F_{y_1,x_1}. Indeed take a $t_1 \notin \bar{x}_1$ and define

$$F_{y_1,x_1} = F_{y_1,t_1} \circ F_{t_1,x_1}.$$

It is clear that this definition does not depend on the choice of t_1 because of the just proven chain rule. It is also easy to see that (2) and (3) hold even when the vectors are not linearly independent because we have

$$\begin{aligned} F_{x_1,y_1} &= F_{x_1,t_1} \circ F_{t_1,y_1} = F^{-1}_{t_1,x_1} \circ F^{-1}_{y_1,t_1} \\ &= (F_{y_1,t_1} \circ F_{t_1,x_1})^{-1} = F_{y_1,x_1}, \end{aligned}$$

$$F_{x_1,y_1} \circ F_{y_1,z_1} = F_{x_1,t_1} \circ F_{t_1,y_1} \circ F_{y_1,z_1} = F_{x_1,z_1}.$$

It is clear that $F_{x_1,x_1}(x) = x$ if $x \in h_1(\bar{x})$ and $\bar{x} \in \mathscr{L}_{\bar{x}_1}$. Since we have $\overline{k_1y_1} = \bar{y}_1$ and $\overline{k_1x_1} = \bar{x}_1$ and $\overline{k_1(x_1-y_1)} = \overline{x_1-y_1}$, (5) is a direct consequence of the definition of F_{y_1,x_1}.

Remark: For this proof we essentially need the fact that V_1 is a vectors space of at least dimension three. If $\dim V_1 = 2$ we can still construct F_{y_1,x_1} but it is impossible to prove the chain rule.

3.9. *Lemma*: Let $\mathscr{P}(V_1)$, $\mathscr{L}_2$, $\mathscr{P}(V)$, h_1, h_2, and $\{F_{y_1,x_1} : x_1, y_1 \in V_1, x_1 \neq 0 \neq y_1\}$ be as in Lemma 3.8 and take $x_1, y_1, z_1 \in V_1$ such that $y_1 + z_1 \neq 0$. If $x \in V$ such that $\bar{x} \in \mathscr{L}_{\bar{x}_1}$, we have:

$$F_{y_1,x_1}(x) + F_{z_1,x_1}(x) \in h_1(\overline{y_1+z_1}),$$

$$F_{y_1,x_1} + F_{z_1,x_1} = F_{y_1+z_1,x_1}.$$

Proof: Take x_1, y_1, z_1 to be linearly independent. Then we have

$$\overline{y_1+z_1} = (\bar{y}_1 \vee \bar{z}_1) \wedge (\bar{x}_1 \vee \overline{x_1-y_1-z_1}),$$

$$\overline{x_1-y_1-z_1} = (\overline{x_1-y_1} \vee \bar{z}_1) \wedge (\overline{x_1-z_1} \vee \bar{y}_1).$$

Take

$$\bar{x} \in \mathscr{L}_{\bar{x}_1}, \quad x - F_{y_1,x_1}(x) \in h_1(\overline{x_1-y_1}),$$

$$x - F_{z_1,x_1}(x) \in h_1(\overline{x_1-z_1}).$$

So we have

$$\begin{aligned} &x - F_{y_1,x_1}(x) - F_{z_1,x_1}(x) \\ &\quad \in (h_1(\overline{x_1-y_1}) \vee h(\bar{z}_1)) \wedge (h(\overline{x_1-z_1}) \vee h_1(\bar{y}_1)) \\ &\quad = h_1(\overline{x_1-y_1-z_1}). \end{aligned}$$

So $F_{y_1,x_1}(x) + F_{z_1,x_1}(x) \in h_1(\bar{x}_1) \vee h_1(\overline{x_1-y_1-z_1})$. This gives us

$$F_{y_1,x_1}(x) + F_{z_1,x_1}(x)$$

$$\in[h_1(\bar{y}_1)\vee h_1(\bar{z}_1)]\wedge\left[h_1(\bar{x}_1)\vee h_1(\overline{x_1-y_1-z_1})\right]$$
$$=h_1(\overline{y_1+z_1})\,.$$

We also have

$$x=F_{y_1,x_1}(x)+F_{z_1,x_1}(x)+x-F_{y_1,x_1}(x)-F_{z_1,x_1}(x)\,,$$
$$x=F_{y_1+z_1,x_1}(x)+F_{x_1-y_1-z_1,x_1}(x)\,.$$

So

$$F_{y_1,x_1}(x)+F_{z_1,x_1}(x)-F_{y_1+z_1,x_1}(x)$$
$$=F_{y_1,z_1}(x)+F_{z_1,x_1}(x)-x+F_{x_1-y_1-z_1,x_1}(x)=t\,.$$

Thus we have $t\in h_1(\overline{y_1+z_1})\wedge h_1(\overline{x_1-y_1-z_1})$ and so $t=0$ which proves that

$$F_{y_1+z_1,x_1}(x)=F_{y_1,x_1}(x)+F_{z_1,x_1}(x).$$

Suppose now that $y_1\in\bar{z}_1$ such that $y_1+z_1\neq 0$ and suppose $x_1\notin\bar{z}_1$. Then there always exists a t_1 such that $\{x_1,t_1,y_1+z_1\}$ and $\{x_1,t_1+y_1,z_1\}$ are linearly independent sets. We have

$$\begin{aligned}F_{t_1+y_1+z_1,x_1}&=F_{t_1+y_1,x_1}+F_{z_1,x_1}\\&=F_{t_1,x_1}+F_{y_1,x_1}+F_{z_1,x_1}\\&=F_{t_1,x_1}+F_{y_1+z_1,x_1}\end{aligned}$$

and this proves that $F_{y_1+z_1,x_1}=F_{y_1,x_1}+F_{z_1,x_1}$.

There is still one case to examine. Suppose $x_1\in\bar{y}_1\vee\bar{z}_1$ and $y_1+z_1\neq 0$ (y_1 and z_1 can be linearly dependent). We choose a t_1 such that $t_1\notin\bar{y}_1\vee\bar{z}_1$ and have

$$\begin{aligned}F_{y_1+z_1,x_1}&=F_{y_1+z_1,t_1}F_{t_1,x_1}=F_{y_1,t_1}F_{t_1,x_1}+F_{z_1,t_1}F_{t_1,x_1}\\&=F_{y_1,x_1}+F_{z_1,x_1}\,.\end{aligned}$$

This proves Lemma 3.9.

3.10. *Lemma*: Let $\mathscr{P}(V_1)$, $\mathscr{L}_2$, $\mathscr{P}(V)$, h_1, h_2, and $\{F_{y_1,x_1}:x_1,y_1\in V_1,x_1\neq 0\neq y_1\}$ be as in Lemma 3.8 and take $k_1\in K_1$ and $\bar{x}\in\mathscr{L}_{\bar{x}_1}$. We then have

$$F_{k_1y_1,x_1}(x)=\sigma_1(k_1,x,x_1)F_{y_1,x_1}(x)\,,$$

where $\sigma_1(\cdot,x,x_1):K_1\to K$ is a morphism.

Proof: If $\bar{x}\in\mathscr{L}_{\bar{x}_1}$ and $k_1\in K_1$, we know that (see Lemma 3.5)

$$\overline{F_{k_1y_1,x_1}(x)}=f_{\overline{k_1y_1},\bar{x}_1}(\bar{x})=f_{\bar{y}_1,\bar{x}_1}(\bar{x})=\overline{F_{y_1,x_1}(x)}\,.$$

So

$$F_{k_1y_1,x_1}(x)=\sigma_1(k_1,y_1,x_1,x)F_{y_1,x_1}(x)\,,$$
$$\sigma_1(k_1,y_1,x_1,x)\in K\,.$$

Now take $y_1,z_1\in V_1$ and $\bar{y}_1\neq\bar{z}_1$,

$$\begin{aligned}&F_{k_1(y_1+z_1),x_1}(x)\\&\quad=\sigma_1(k_1,y_1+z_1,x_1,x)F_{y_1+z_1,x_1}(x)\\&\quad=\sigma_1(k_1,y_1+z_1,x_1,x)(F_{y_1,x_1}(x)+F_{z_1,x_1}(x)).\end{aligned}$$

But we have also

$$\begin{aligned}&F_{k_1(y_1+z_1),x_1}(x)\\&=F_{k_1y_1,x_1}(x)+F_{k_1z_1,x_1}(x)\\&=\sigma_1(k_1,y_1,x_1,x)F_{y_1,x_1}(x)+\sigma_1(k_1,z_1,x_1,x)F_{z_1,x_1}(x)\,.\end{aligned}$$

Since $\bar{y}\neq\bar{z}$ we must have

$$\sigma_1(k_1,y_1,x_1,x)=\sigma_1(k_1,y_1+z_1,x_1,x)=\sigma_1(k_1,z_1,x_1,x)\,.$$

This proves that σ_1 is independent of y_1. σ_1 will certainly depend on x_1 and x. To somewhat alleviate the notation we will often not explicitly write this dependence since most of the time we will work with a fixed x_1 and a fixed x.

$$\begin{aligned}\sigma_1(k_1+l_1)F_{y_1,x_1}(x)&=F_{(k_1+l_1)y_1,x_1}(x)\quad k_1,l_1\in K_1\\&=F_{k_1y_1,x_1}(x)+F_{l_1y_1,x_1}(x)\\&=(\sigma_1(k_1)+\sigma_1(l_1))F_{y_1,x_1}(x)\,.\end{aligned}$$

So we have

$$\sigma_1(k_1+l_1)=\sigma_1(k_1)+\sigma_1(l_1)\,,$$
$$\begin{aligned}\sigma_1(k_1\cdot l_1)F_{y_1,x_1}(x)&=F_{k_1\cdot l_1y_1,x_1}(x)\\&=\sigma_1(k_1)F_{l_1y_1,x_1}(x)\\&=\sigma_1(k_1)\cdot\sigma_1(l_1)F_{y_1,x_1}(x).\end{aligned}$$

This proves that: $\sigma_1(k_1\cdot l_1)=\sigma_1(k_1)\cdot\sigma_1(l_1)$.

□

3.11. Suppose that we have $\mathscr{P}(V_1)$, $\mathscr{P}(V_2)$, $\mathscr{P}(V)$, h_1, h_2 as in Theorem 2.22 and that we construct for h_1 the set of maps $\{F_{y_1,x_1}:y_1,x_1\in V_1,x_1\neq 0\neq y_1\}$ and for h_2 the set maps $\{G_{y_2,x_2}:y_2,x_2\in V_2,x_2\neq 0\neq y_2\}$ defined in 3.7. It is evident so that the set $\{G_{y_2,x_2};y_2,x_2\in V_2,x_2\neq 0\neq y_2\}$ also fulfills the properties proved in Lemmas 3.8, 3.9, and 3.10 where we change every assertion on $\mathscr{P}(V_1)$ and h_1 in an analogous assertion on $\mathscr{P}(V_2)$ and h_2. So we have the following properties:

(1) If x is in the domain of G_{y_2,x_2} and $k\in K$, then kx is in the domain of G_{y_2,x_2}, if y is in the domain of G_{y_2,x_2} such that $x+y$ is in the domain of G_{y_2,x_2}, then we have

$$G_{y_2,x_2}(kx)=kG_{y_2,x_2}(x)\,,$$
$$G_{y_2,x_2}(x+y)=G_{y_2,x_2}(x)+G_{y_2,x_2}(y)\,.$$

(2) The domain of G_{x_2,y_2} is the image of G_{y_2,x_2}, and $G_{y_2,x_2}=G^{-1}_{x_2,y_2}$.

(3) $G_{z_2,y_2}\circ G_{y_2,x_2}=G_{z_2,x_2}$ for every $y_2\in V_2$.

(4) $G_{x_2,x_2}(x)=x$ for $\bar{x}\in\mathscr{L}_{\bar{x}_2}$.

(5) $G_{k_2y_2,k_2x_2}=G_{y_2,x_2}$ for every $k_2\in K_2$.

(6) $G_{y_2,x_2}+G_{z_2,x_2}=G_{y_2+z_2,x_2}$ if $y_2,x_2,z_2\in V_2$ such that $y_2+z_2\neq 0$.

(7) $G_{k_2y_2,x_2}(x)=\sigma_2(k_2,x,x_2)G_{y_2,x_2}(x)$ if $k_2\in K_2$, y_2, $x_2\in V_2$, $\bar{x}\in\mathscr{L}_{\bar{x}_2}$,

where $\sigma_2(\cdot,x,x_2):K_2\to K$ is a morphism.

But we can prove more.

3.12. *Lemma*: Let $\mathscr{P}(V_1)$, $\mathscr{P}(V_2)$, $\mathscr{P}(V)$, h_1, h_2, $\{F_{y_1,x_1}\}$, $\{G_{y_2,x_2}\}$, σ_1, and σ_2 be as in 3.11 and suppose $\bar{z}=h_1(\bar{z}_1)\wedge h_2\bar{z}_2)$, where $z\in V$, $z_1\in V_1$, and $z_2\in V_2$, then we have:

(1) $\overline{G_{y_2,z_2}(x)}\in\mathscr{L}_{\bar{z}_2}$. and $\overline{F_{y_1,z_1}(z)}\in\mathscr{L}_{\bar{z}_1}$.

(2) $F_{y_1,z_1}G_{y_2,z_2}(z)=G_{y_2,z_2}F_{y_1,z_1}(z)$.

(3) For $\forall x_1\in V_2$, $\forall x_2\in V_2$, $\forall k_1\in K_1$, $\forall k_2\in K_2$, and $\forall k\in K$ we have:

(i) $\sigma_1(k_1,F_{x_1,z_1}(z),x_1)=\sigma_1(k_1,z,z_1)$
$$=\sigma_1(k_1,G_{x_2,z_2}(z),z_1)\,,$$

$$\sigma_2(k_2, G_{x_2,z_2}(z), x_2) = \sigma_2(k_2, z, z_2)$$
$$= \sigma_2(k_2, F_{x_1,z_1}(z), z_2),$$
(ii) $\sigma_1(k_1, kz, z_1) = k\sigma_1(k_1, z, z_1)k^{-1}$,
$$\sigma_2(k_2, kz, z_2) = k\sigma_2(k_2, z, z_2)k^{-1},$$
(iii) $\sigma_1(k_1, z, z_1)\cdot\sigma_2(k_2, z, z_2)$
$$= \sigma_2(k_2, z, z_2)\cdot\sigma_1(k_1, z, z_1).$$

Proof: $\overline{G_{y_2,x_2}(x)}$
$= u_{\bar{y}_2}u_{\bar{z}_2}^{-1}(\bar{z})$ and $\bar{z} = v_{\bar{z}_1}(\bar{z}_2) = u_{\bar{z}_2}(\bar{z}_1)$
$= u_{\bar{y}_2}(\bar{z}_1)$
$= v_{\bar{z}_1}(\bar{y}_2) \in \mathscr{L}_{\bar{z}_1}$.

In an analogous way we prove that $\overline{F_{y_1,x_1}(z)} \in \mathscr{L}_{\bar{z}_2}$. This proves (1) and so this proves that $F_{y_1,z_1}G_{y_2,z_2}(z)$ and $G_{y_2,z_2}F_{y_1,z_1}(z)$ are well defined.

We have $z = G_{y_2,z_2}(z) + G_{z_2-y_2,z_2}(z)$, so from Lemma 3.8 (1) we have

$$F_{y_1,z_1}(z) = F_{y_1,z_1}G_{y_2,z_2}(z) + F_{y_1,z_1}G_{z_2-y_2,z_2}(z).$$

But we also have

$$F_{y_1,z_1}(z) = G_{y_2,z_2}F_{y_1,z_1}(z) + G_{z_2-y_2,z_2}F_{y_1,z_1}(z),$$

so

$$F_{y_1,z_1}G_{y_2,z_2}(z) - G_{y_2,z_2}F_{y_1,z_1}(z)$$
$$= G_{z_2-y_2,z_2}F_{y_1,z_1}(z) - F_{y_1,z_1}G_{z_2-y_2,z_2}(z) = t.$$

This proves that $t \in h_1(\bar{y}_1) \wedge h_2(\bar{y}_2)$ and $t \in h_1(\bar{y}_1) \wedge h_2(\overline{z_2 - y_2})$, and so $t = 0$ which gives us

$$F_{y_1,z_1}G_{y_2,z_2}(z) = G_{y_2,z_2}F_{y_1,z_1}(z).$$

This proves (2).

We have

$$F_{k_1y_1,x_1}F_{x_1,z_1}(z) = \sigma_1(k_1, F_{x_1,z_1}(z), x_1)F_{y_1,z_1}(z)$$
$$= F_{k_1y_1,z_1}(z)$$
$$= \sigma_1(k_1, z, z_1)F_{y_1,z_1}(z),$$

so

$$\sigma_1(k_1, F_{x_1,z_1}(z), x_1) = \sigma_1(k_1, z, z_1),$$

We also have

$$F_{k_1z_1,z_1}G_{x_2,z_2}(z) = \sigma_1(k_1, G_{x_2,z_2}(z), z_1)G_{x_2,z_2}(z)$$
$$= G_{x_2,z_2}F_{k_1z_1,z_1}(z)$$
$$= G_{x_2,z_2}(\sigma_1(k_1, z, z_1)x)$$
$$= \sigma_1(k_1, z, z_1)G_{x_2,z_2}(x).$$

So we have

$$\sigma_1(k_1, G_{x_2,z_2}(z), z_1) = \sigma_1(k_1, z, z_1).$$

This completes the proof of (3) (i).

Next consider

$$F_{k_1x_1,z_1}(kz) = kF_{k_1x_1,z_1}(z) = k\sigma_1(k_1, z, z_1)F_{x_1,z_1}(z)$$
$$= \sigma_1(k_1, kz, z_1)F_{x_1,z_1}(kz)$$
$$= \sigma_1(k_1, kz, z_1)kF_{x_1,z_1}(z).$$

So we have

$$\sigma_1(k_1, kz, z_1) = k\sigma_1(k_1, z, z_1)k^{-1}.$$

This proves (3) (ii).

Now consider

$$F_{k_1x_1,z_1}G_{k_2x_2,z_2}(z)$$
$$= \sigma_1(k_1, G_{k_2x_2,z_2}(z), z_1)F_{x_1,z_1}G_{k_2x_2,z_2}(z)$$
$$= \sigma_1(k_1, z, z_1)F_{x_1,z_1}(\sigma_2(k_2, z, z_2)G_{x_2,z_2}(z))$$
$$= \sigma_1(k_1, z, z_1)\sigma_2(k_2, z, z_2)F_{x_1,z_1}G_{x_2,z_2}(z)$$
$$= G_{k_2x_2,z_2}F_{k_1x_1,z_1}(z)$$
$$= \sigma_2(k_2, z, z_2)\sigma_1(k_1, z, z_1)G_{x_2,z_2}F_{x_1,z_1}(z).$$

And since $F_{x_1,z_1}G_{x_2,z_2}(z) = G_{x_2,z_2}F_{x_1,z_1}(z)$ we conclude

$$\sigma_1(k_1, z, z_1)\cdot\sigma_2(k_2, z, z_2) = \sigma_2(k_2, z, z_2)\cdot\sigma_1(k_1, z, z_1).$$

We are now in the position to construct the (σ_1, σ_2)-linear map which we need to prove Theorem 2.22.

3.13. *Theorem*: Let $\mathscr{P}(V_1)$, $\mathscr{P}(V_2)$, $\mathscr{P}(V)$, h_1, h_2, $\{F_{y_1,x_1}\}$, $\{G_{y_2,x_2}\}$, σ_1, and σ_2 be as in 3.11 and fix three vectors $z_1 \in V_2$, $z_2 \in V_2$, and $z \in V$ such that $z \in h_1(\bar{z}_1) \wedge h_2(\bar{z}_2)$. Let us define the following map:

$$B(z, z_1, z_2, \cdot, \cdot) : V_1 \times V_2 \to V$$
$$(x_1, x_2) \to F_{x_1,z_1}G_{x_2,z_2}(z) = B(z, z_1, z_2, x_1, x_2)$$

This map has the following properties.

(1) $B(z, z_1, z_2, \cdot, \cdot)$ is a $(\sigma_1(\cdot, z, z_1), \sigma_2(\cdot, z, z_2))$-linear map.

(2) It is an isometry and

$$\varphi(B(y_1, y_2), B(x_1, x_2)) = \sigma_1(\varphi_1(y_1, x_1))\sigma_2(\varphi_2(y_2, x_2))d,$$

where $d = \varphi(z, z)\sigma_1(\varphi_1(z_1, z_1))^{-1}(\sigma_2(z_2, z_2))^{-1}$.

(3) It is a dense (σ_1, σ_2)-linear isometry.

(4) It generates h_1 and h_2 in the following way:

$$h_1(a_1) = B(a_1, V_2)^{\perp\perp} \quad \text{for } a_1 \in \mathscr{P}(V_1),$$
$$h_2(a_2) = B(V_1, a_2)^{\perp\perp} \quad \text{for } a_2 \in \mathscr{P}(V_2).$$

(5) We have for $(x_1, x_2) \in V_1 \times V_2$,

(i) $B(F_{x_1,z_1}(z), x_1, z_2, \cdot, \cdot) = B(z, z_1, z_2, \cdot, \cdot)$
$$= B(G_{x_2,z_2}(z), z_1, x_2, \cdot, \cdot),$$
(ii) $B(kz, z_1, z_2, \cdot, \cdot) = kB(z, z_1, z_2, \cdot, \cdot)$, for $k \in K$.

(6) If $W : V_1 \times V_2 \to V$ is another (θ_1, θ_2)-linear map that generates h_1 and h_2, then there exists a $k \in K$ such that

$$W = kB(z, z_1, z_2, \cdot, \cdot)$$

and

$$\theta_1 = k\sigma_1(\cdot, z, z_1)k^{-1}, \quad \theta_2 = k\sigma_2(\cdot, z, z_2)k^{-1}.$$

Proof: In the case where z, z_1, and z_2 are fixed we will not write the dependence of B, σ_1, and σ_2 on these vectors. Take $x_1, y_1 \in V_1$, $x_2, y_2 \in V_2$, $k_1 \in K_1$, and $k_2 \in K_2$.

$$B(x_1 + y_1, x_2) = F_{x_1+y_1,z_1}G_{x_2,z_2}(z)$$
$$= F_{x_1,z_1}G_{x_2,z_2}(z) + F_{y_1,z_1}G_{x_2,z_2}(z)$$
$$= B(x_1, x_2) + B(y_1, x_2).$$

In an analogous way we prove that

$$B(x_1, x_2 + y_2) = B(x_1, x_2) + B(x_1, y_2).$$

We also have

$$B(x_1, k_2x_2) = F_{x_1,z_1}G_{k_2x_2,z_2}(z)$$
$$= \sigma_2(k_2)B(x_1, x_2)$$

and

$$B(k_1x_1, x_2) = \sigma_1(k_1)B(x_1, x_2).$$

This proves (1).

Let us put $t = F_{x_1,z_1} G_{x_2,z_2}(z)$.

$$\varphi(B(y_1,y_2), B(x_1,x_2)) = \varphi(F_{y_1,z_1} G_{y_2,z_2}(z), F_{x_1,z_1} G_{x_2,z_2}(z)) = \varphi(F_{y_1,x_1} G_{y_2,x_2}(t), t),$$

where $t = F_{x_1,z_1} G_{x_2,z_2}(z)$.

We know that $t \in h_1(\bar{x}_1) \wedge h_2(\bar{x}_2)$ (see Lemma 3.12). Put

$$t_1 = \varphi_1(x_1,x_1)\cdot\varphi_1(y_1,x_1)^{-1} y_1,$$
$$t_2 = \varphi_2(x_2,x_2)\cdot\varphi_2(y_2,x_2)^{-1} y_2.$$

Then we have $t_1 - x_1 \perp x_1$ which implies $\overline{h_1(t_1 - x_1)} \perp h_1(\bar{x}_1)$ and also $t_2 - x_2 \perp x_2$ which implies $h_2(t_2 - x_2) \perp h_2(\bar{x}_2)$

We also can write

$$t = F_{t_1,x_1}(t) + F_{t_1-x_1,x_1}(t) = G_{t_2,x_2}(t) + G_{t_2-x_2,x_2}(t).$$

So

$$t = F_{t_1,x_1} G_{t_2,x_2}(t) + F_{t_1,x_1} G_{t_2-x_2,x_2}(t) + F_{t_1-x_1,x_1} G_{t_2,x_2}(t) + F_{t_1-x_1,x_1} G_{t_2-x_2,x_2}(t). \quad (3.5)$$

Now

$$F_{t_1,x_1} G_{t_2-x_2,x_2}(t) \in h_1(\bar{t}_1) \wedge \overline{h_2(t_2 - x_2)}$$

and since $h_2(t_2 - x_2) \perp h_2(\bar{x}_2)$ we have also $F_{t_1,x_1} G_{t_2-x_2,x_2}(t) \perp t$. It is easy to see that also

$$F_{t_1-x_1,x_1} G_{t_2,x_2}(t) \perp t \quad \text{and} \quad F_{t_1-x_1,x_1} G_{t_2-x_2,x_2}(t) \perp t. \quad (3.6)$$

From (3.5) and (3.6) it now easily follows that

$$\varphi(t,t) = \varphi(F_{t_1,x_1} G_{t_2,x_2}(t), t).$$

So we also have

$$\varphi(F_{y_1,x_1} G_{y_2,x_2}(t), t) = \varphi(F_{\varphi_1(y_1,x_1)\cdot\varphi_1(x_1,x_1)^{-1} t_1, x_1} G_{\varphi_2(y_2,x_2)\cdot\varphi_2(x_2,x_2)^{-1} t_2, x_2}(t), t)$$
$$= \sigma_1(\varphi_1(y_1,x_1)\cdot\varphi_1(x_1,x_1)^{-1})\cdot\sigma_2(\varphi_2(y_2,x_2)\cdot\varphi_2(x_2,x_2)^{-1})\varphi(t,t)$$
$$= \sigma_1(\varphi_1(y_1,x_1))\sigma_2(\varphi_2(y_2,x_2))\cdot\sigma_1(\varphi_1(x_1,x_1)^{-1})\sigma_2(\varphi_2(x_2,x_2)^{-1}) \cdot\varphi(B(x_1,x_2), B(x_1,x_2)).$$

So we have

$$\varphi(B(y_1,y_2), B(x_1,x_2)) = \sigma_1(\varphi_1(y_1,x_1))\sigma_2(\varphi_2(y_2,x_2))g(x_1,x_2),$$

where

$$g(x_1,x_2) = \sigma_1(\varphi_1(x_1,x_1)^{-1}) \cdot\sigma_2(\varphi_2(x_2,x_2)^{-1})\varphi(B(x_1,x_2), B(x_1,x_2)).$$

Now take $x_1, z_1, y_1 \in V_1$ such that $\bar{x}_1 \neq \bar{z}_1$, and take $x_2, y_2 \in V_2$,

$$\varphi(B(y_1,y_2), B(x_1+z_1,x_2)) = \sigma_1(\varphi_1(y_1,x_1+z_1))\cdot\sigma_2(\varphi_2(y_2,z_2))g(x_1+z_1,x_2)$$
$$= \sigma_1(\varphi_1(y_1,x_1))\sigma_2(\varphi_2(y_2,z_2))g(x_1+z_1,x_2) + \sigma_1(\varphi_1(y_1,x_1))\sigma_2(\varphi_2(y_2,z_2))g(x_1+z_1,x_2). \quad (3.7)$$

But we also have

$$\varphi(B(y_1,y_2), B(x_1+z_1,x_2)) = \varphi(B(y_1,y_2), B(x_1,x_2)) + \varphi(B(y_1,y_2), B(z_1,x_2))$$
$$= \sigma_1(\varphi_1(y_1,x_1))\sigma_2(\varphi_2(y_2,x_2))g(x_1,x_2) + \sigma_1(\varphi_1(y_1,z_1))\sigma_2(\varphi_2(y_2,z_2))g(z_1,z_2). \quad (3.8)$$

From (3.7) and (3.8) we have for $\forall y_1 \in V_1$ and $\forall y_2 \in V_2$

$$0 = \sigma_1(\varphi_1(y_1,x_1))\sigma_2(\varphi_2(y_2,x_2))[g(x_1+z_1,x_2) - g(x_1,x_2)] + \sigma_1(\varphi_1(y_1,z_1))\sigma_2(\varphi_2(y_2,z_2)) \times [g(x_1+z_1,x_2) - g(z_1,x_2)]. \quad (3.9)$$

(i) Take $y_1 \perp z_1$ and $y_2 \not\perp z_2$. Since $\bar{z}_1 \neq \bar{x}_1$ we have $y_1 \not\perp x_1$, this implies that $\sigma_1(\varphi_1(y_1,z_1)) = 0$ and $\sigma_1(\varphi_1(y_1,x_1)) \neq 0$. From (3.9) we then get

$$g(x_1,x_2) = g(x_1+z_1,x_2), \quad \forall x_2 \in V_2.$$

(ii) Take $y_1 \perp x_1$ and $y_2 \not\perp z_2$. Again we have $y_1 \not\perp z_1$ and so from (3.9) we have

$$g(z_1,x_2) = g(x_1+z_1,x_2), \quad \forall x_2 \in V_2.$$

Combining (i) and (ii) we have proven that

$$g(x_1,x_2) = g(z_1,x_2), \quad \forall x_2 \in V_2, \ \forall x_1, z_1 \in V_1.$$

If $\bar{x}_1 = \bar{z}_1$, we can always take a $t_1 \in V_1$ such that $\bar{t}_1 \neq \bar{x}_1$. We then conclude using the foregoing result, that

$$g(x_1,x_2) = g(t_1,x_2) = g(z_1,x_2), \quad \forall x_2 \in V_2.$$

In the same way we prove that

$$g(x_1,x_2) = g(x_1,z_2), \quad \forall x_1 \in V_1 \text{ and } \forall x_2, z_2 \in V_2.$$

So g has to be a constant element of K. Let us call this element d. We then have

$$\varphi(B(y_1,y_2), B(x_1,x_2)) = \sigma_1(\varphi_1(y_1,x_1))\sigma_2(\varphi_2(y_2,x_2))d.$$

And, since $z = B(z_1,z_2)$ we easily see that

$$d = \varphi(z,z)\cdot\sigma_1(\varphi_1(z_1,z_1)^{-1})\sigma_2(\varphi_2(z_2,z_2)^{-1}).$$

This concludes the proof of (2).

Now take M_1 to be a dense subset of V_1 and M_2 to be a dense subset of V_2, then we have

$$B(M_1,M_2)^{\perp\perp} = \bigvee_{x_1 \in M_1,\, x_2 \in M_2} \overline{B(x_1,x_2)} = \bigvee_{x_1 \in M_1,\, x_2 \in M_2} h_1(\bar{x}_1) \wedge h_2(\bar{x}_2)$$

using Sec. 2.3 (1) we have:

$$= \bigvee_{x_1 \in M_1} \left[h_1(\bar{x}_1) \wedge \left(\bigvee_{x_2 \in M_2} h_2(\bar{x}_2) \right) \right] = \bigvee_{x_1 \in M_1} \left[h_1(\bar{x}_1) \wedge h_2\left(\bigvee_{x_2 \in M_2} \bar{x}_2 \right) \right]$$

using again Sec. 2.3 (1) we get:

$$= h_1\left(\bigvee_{x_1 \in M_1} \bar{x}_1 \right) \wedge h_2(M_2^{\perp\perp}) = h_1(V_1) \wedge h_2(V_2) = V.$$

This proves (3)

Now take $a_1 \in \mathscr{P}(V_1)$. It is always so that $a_1 = \bigvee_{x_1 \in a_1} \bar{x}_1$; thus

$$h_1(a_1) = \bigvee_{x_1 \in a_1} h_1(\bar{x}_1) = \left[\bigvee_{x_1 \in a_1} h_1(\bar{x}_1) \right] \wedge \left[\bigvee_{x_2 \in V_2} h_2(\bar{x}_2) \right]$$

and using twice Sec. 2.3 (1) we have:

$$h_1(a_1) = \bigvee_{x_1 \in a_1,\, x_2 \in V_2} (h_1(\bar{x}_1) \wedge h_2(\bar{x}_2))$$
$$= \bigvee_{x_1 \in a_1,\, x_2 \in V_2} \overline{B(x_1, x_2)}$$
$$= B(a_1, V_2)^{\perp\perp} .$$

In the same way we prove that $h_2(a_2) = B(V_1, a_2)^{\perp\perp}$ which concludes the proof of (4).

Let us now prove (5)

$$B(F_{x_1,z_1}(z), x_1, z_2, y_1, y_2) = F_{y_1,x_1} G_{y_2,z_2} F_{x_1,z_1}(z)$$
$$= F_{y_1,z_1} G_{y_2,z_2}(z)$$
$$= B(z, z_1, z_2, y_1, y_2)$$
$$= F_{y_1,z_1} G_{y_2,x_2} G_{x_2,z_2}(z)$$
$$= B(G_{x_2,z_2}(z), z_1, x_2, y_1, y_2) ,$$

$$B(kz, z_1, z_2, y_1, y_2) = F_{y_1,z_1} G_{y_2,z_2}(kz)$$
$$= kB(z, z_1, z_2, y_1, y_2) .$$

Now take $W: V_1 \times V_2 \to V$ to be another (θ_1, θ_2)-linear map that generates h_1 and h_2. We must have $\overline{W(x_1, x_2)} = h_1(\bar{x}_1) \wedge h_2(\bar{x}_2) = \overline{B(x_1, x_2)}$ and thus we can find an element $g(x_1, x_2) \in K$ such that

$$W(x_1, x_2) = g(x_1, x_2) B(x_1, x_2) .$$

Take $x_1, y_1 \in V_1$ such that $x_1 \notin \bar{y}_1$,

$$W(x_1 + y_1, x_2) = g(x_1 + y_1, x_2) B(x_1 + y_1, x_2)$$
$$= g(x_1 + y_1, x_2) B(x_1, x_2)$$
$$+ g(x_1 + y_1, x_2) B(y_1, x_2) .$$

But we also have

$$W(x_1 + y_1, x_2) = W(x_1, x_2) + W(y_1, x_2)$$
$$= g(x_1, x_2) B(x_1, x_2) + g(y_1, x_2) B(y_1, x_2) .$$

Since $\bar{x}_1 \neq \bar{y}_1$ we have proven that

$$g(x_1, x_2) = g(x_1 + y_1, x_2) = g(y_1, x_2) \quad \forall x_2 \in V_2 .$$

In the same way we can prove that

$$g(x_1, x_2) = g(x_1, y_2) , \quad y_2, x_2 \in V_2 \text{ and } \forall x_1 \in V_1 ,$$

so $g(x_1, x_2)$ has to be independent of x_1 and x_2. Putting $W(z_1, z_2) = t$, then we just proved that

$$W = B(t, z_1, z_2, \cdot, \cdot) , \quad \theta_1 = \sigma_1(\cdot, t, z_1) ,$$
$$\theta_2 = \sigma_2(\cdot, t, z_2) .$$

And since there always exists a $k \in K$ such that $t = kz$ we conclude using (5) (ii) and Lemma 3.12 (3) (ii) that

$$W = kB(z, z_1, z_2, \cdot, \cdot) , \quad \theta_1 = k\sigma_1(\cdot, z, z_1)k^{-1} ,$$
$$\theta_2 = k\sigma_2(\cdot, z, z_2)k^{-1} .$$

■

This concludes the proof of Theorem 2.22.

[1] D. Aerts and I. Daubechies, "A characterization of subsystems in physics," Lett. Math. Phys. 3, 11–17 (1979).

[2] D. Aerts and I. Daubechies, "Physical justification for using the tensor product to describe two quantum systems as one joint system," Helv. Phys. Acta 51, 661 (1978).

[3] C. Piron, *Foundation of Quantum Physics* (Benjamin, New York, 1976).

[4] For the definition of a P lattice see Sec. 2.

[5] E.P. Wingner, *Group Theory* (Academic, New York, 1959).

[6] P lattices are often called "propositional system," but this term interferes with the word "physical system," which is often used in the text.

A probabilistic formulation of quantum theory

Alexander Bach
Institut für Theoretische Physik I, Universität Münster, D-4400 Münster, Germany

(Received 18 July 1979; accepted for publication 27 November 1979)

By means of integration in the Hilbert space H which is associated to a quantum mechanical system we demonstrate that the expectation value of quantum mechanics, $E_W(A) = \mathrm{tr}(WA)$, can be expressed as $E_W(A) = \int_H d\mu_W(\phi)\,(1/2)\langle\phi|A|\phi\rangle$. Here μ_W denotes a probability measure in H which is determined by the statistical operator. We discuss this hidden variables representation of quantum mechanics.

I. INTRODUCTION

It is widely accepted opinion that there is no hidden variables theory for quantum mechanics. On the other hand it is this aspect of quantum mechanics which was recognized but never satisfactorily understood. There is a never ending series of attempts to obtain a hidden variables theory for quantum mechanics.[1] In general these approaches do not seem convincing.

In this paper we present a hidden variables theory of quantum mechanics which we believe satisfactory regarding the following result: The hidden variables are the wavefunctions of quantum mechanics themselves and it is well known that these have no empirical meaning.

As the hidden variables now are to be thought of as elements of an infinite dimensional Hilbert space the probability measure describing a quantum mechanical expectation value has to be defined on an infinite dimensional space.

We demonstrate that the characteristic measures on Hilbert space are determined by the statistical operator of the system and that the measurable functions which allow one to evaluate the expectation values are determined by the operators which represent the observables. Nevertheless, these never admit dispersion-free ensembles.

We remark that we are not concerned with Feynman path integrals nor with functional integrals used in quantum field theory but with integrals which are set up by means of a measure in an infinite dimensional space.[2]

Our paper is organized as follows. In Sec. II we establish an isometry between real and complex Hilbert spaces which allows us to give a formulation of quantum theory based on the field of reals. We demonstrate that this isometry leaves invariant the characteristic properties of operators which are used in quantum theory.

In Sec. III we review Borel measures on real Hilbert spaces. It is observed that the covariance operator of such a measure and the statistical operator of quantum mechanics have the same properties.

This fact is used in Sec. IV to set up a measure in the real Hilbert space by means of the real representation of the statistical operator. We then prove the formula for the expectation value of quantum mechanics. An application of the inverse isometry enables us to set up this probabilistic formulation on the complex Hilbert space associated with a quantum mechanical system.

In Sec. V we discuss the probabilistic formulation of quantum theory which we have obtained. We investigate the fact that quantum mechanics does not admit dispersion-free ensembles and elucidate the difference between classical and quantum mechanics.

Section VI is devoted to a general discussion concerning the restrictions and possible extensions of probabilistic quantum theory.

II. THE CANONICAL ISOMETRY BETWEEN A COMPLEX AND A REAL HILBERT SPACE

Let $(H, \langle\cdot|\cdot\rangle)$ be a complex separable Hilbert space and $(H_0, \langle\cdot|\cdot\rangle_0)$ a real separable Hilbert space such that $2\dim(H) = \dim(H_0)$ if $\dim(H) < \infty$ and $\dim(H_0) = \infty$ if H is infinite dimensional. Denote by $\{\phi_n\}_{n=1}^{\infty}$, $\{e_n\}_{n=1}^{\infty}$ a complete orthonormal system for H, H_0, respectively.

We define the canonical isometry $U, U: H \to H_0$ as follows. Let be $|\psi\rangle \in H$, $z_n \in \mathbb{C}$, $z_n = a_n + ib_n$;

$$|\psi\rangle = \sum_n z_n|\phi_n\rangle = \sum_n (a_n + ib_n)|\phi_n\rangle, \tag{1}$$

and set

$$U|\psi\rangle = \sum_n (a_n|e_{2n-1}\rangle + b_n|e_{2n}\rangle). \tag{2}$$

U is an $\mathbb{R}$-linear operator with domain H and image H_0. Furthermore U is an isometry as we have

$$\langle U\psi|U\psi\rangle_0 = \sum_n (a_n^2 + b_n^2) = \sum_n z_n^+ z_n = \langle\psi|\psi\rangle. \tag{3}$$

On the basis $\{|\phi_n\rangle, n \in \mathbb{N}\}$ the isometry acts as

$$U|\phi_n\rangle = |e_{2n-1}\rangle, \quad Ui|\phi_n\rangle = |e_{2n}\rangle, \tag{4}$$

from which follows that $\{U|\phi_n\rangle, n \in \mathbb{N}\}$ is a complete orthonormal system for a subspace of H_0, the orthogonal complement of which is generated by $\{Ui|\phi_n\rangle, n \in \mathbb{N}\}$.

We define the linear operator $U^{-1}: H_0 \to H$ by

$$|x\rangle \in H_0, \quad |x\rangle = \sum_n c_n|e_n\rangle, \quad c_n \in \mathbb{R}, \tag{5}$$

$$U^{-1}|x\rangle = \sum_n (c_{2n-1} + ic_{2n})|\phi_n\rangle. \tag{6}$$

On the basis $\{|e_n\rangle, n \in \mathbb{N}\}$ the operator U^{-1} acts as

 0022-2488/80/040789-05$1.00

$$U^{-1}|e_n\rangle = |\phi_{(n+1)/2}\rangle, \quad n \text{ odd},$$
$$U^{-1}|e_n\rangle = i|\phi_{n/2}\rangle, \quad n \text{ even}, \tag{7}$$

such that $\{U^{-1}|e_n\rangle, n\in\mathbb{N}\}$ is overcomplete in H. Furthermore we obtain

$$UU^{-1} = \mathbb{1}_{H_0}, \quad U^{-1}U = \mathbb{1}_H, \tag{8}$$

such that U^{-1} is an isometry as well. Remark that U is not unitary as $\mathbb{R}\ni\langle U\phi\,|\,U\psi\rangle_0 \neq \langle\phi\,|\,\psi\rangle\in\mathbb{C}$. Denote by $\mathscr{U}(H)$ the space of linear bounded self-adjoint operators on H, by $\mathscr{T}(H)$ the Banach space of operators of trace class defined by H and by $\mathscr{S}(H)$ the convex space of positive, self-adjoint operators of trace class which satisfy $\mathrm{tr}_H(\cdot) = 1$. Let us denote the spaces of operators subject to the same conditions on H_0 by $\mathscr{U}(H_0)$, $\mathscr{T}(H_0)$, and $\mathscr{S}(H_0)$, respectively, with the slight difference that for elements of $\mathscr{S}(H_0)$ we require $\mathrm{tr}_{H_0}(\cdot) = 2$.

We now define an operator T on the Banach space $\mathscr{L}(H)$ of linear bounded operators on H with values in the Banach space $\mathscr{L}(H_0)$ by

$$TB = UBU^{-1} =: B_0, \quad B\in\mathscr{L}(H). \tag{9}$$

T is an $\mathbb{R}$-linear operator. B_0 is isometrically equivalent to B as we have $\|TB\|_0 = \|B\|$, hence T defines an isometry on the space of operators.

By means of this isometry we obtain a one-to-one relation between the spaces $\mathscr{U}(H)$, $\mathscr{T}(H)$, $\mathscr{S}(H)$ and $\mathscr{U}(H_0)$, $\mathscr{T}(H_0)$, $\mathscr{S}(H_0)$, respectively. For compact operators it is easy to show by means of the matrix representation that TB has the same properties with respect to the classifications given above as B. We remark that a matrix element $B_{nm} = \langle|\phi_n\,|B\phi_m\rangle$ transforms as

$$B_{nm} \overset{T}{\to} \begin{pmatrix} \mathrm{Re}(B_{nm}) & -\mathrm{Im}(B_{nm}) \\ \mathrm{Im}(B_{nm}) & \mathrm{Re}(B_{nm}) \end{pmatrix}. \tag{10}$$

This demonstrates that operators belonging to $\mathscr{S}(H)$ are represented in H_0 by unitarily equivalent ones which satisfy $\mathrm{tr}_{H_0}(\cdot) = 2$.

For elements of $\mathscr{U}(H)$ we conclude as follows. We set up the spectral decomposition of A

$$A = \int_{\mathbb{R}} \lambda\, dP_\lambda, \tag{11}$$

where $\{P_\lambda, \lambda\in\mathbb{R}\}$ is a bounded projection-valued measure and obtain

$$A_0 = TA = \int_{\mathbb{R}} \lambda\, d(TP_\lambda). \tag{12}$$

Clearly $\{(UP_\lambda U^{-1}), \lambda\in\mathbb{R}\}$ is a bounded projection-valued measure with values in $\mathscr{L}(H_0)$.

Finally let us remark that for $A\in\mathscr{U}(H_0)$, $A_0 = TA \in\mathscr{U}(H_0)$ the identity

$$\langle x|A_0|x\rangle_0 = \langle U^{-1}x|A\,|U^{-1}x\rangle, \tag{13}$$

can be derived by means of the spectral decomposition of A and regarding that U is not linear.

III. BOREL MEASURE IN HILBERT SPACES

Let $(H_0,\langle\cdot|\cdot\rangle_0)$ be a real separable Hilbert space and denote by $\mathscr{B}(H_0)$ the Borel field of H_0. This is the σ field generated by the open sets of H_0. A measure defined on $(H_0,\mathscr{B}(H_0))$ is called a Borel measure. It is a probability measure if $\mu(H_0) = 1$. For further information on Borel measures in real Hilbert spaces see Ref. 2.

Most important for the following considerations is the covariance operator C_μ of a measure μ defined on H_0. It is defined by $C_\mu: H_0\to H_0$, $|y\rangle, |z\rangle\in H_0$,

$$\langle y|C_\mu z\rangle_0 := \int_{H_0} d\mu(x)\, \langle y|x\rangle_0\, \langle x|z\rangle_0. \tag{14}$$

If C_μ exists it is positive, self-adjoint and if

$$\int_{H_0} d\mu(x)\, |x|_0^2 < \infty, \quad |x|_0^2 = \langle x|x\rangle_0, \tag{15}$$

C_μ is of trace class.[2] Then it is easy to show that

$$\mathrm{tr}_{H_0}(C_\mu) = \int_{H_0} d\mu(x)\, |x|_0^2. \tag{16}$$

As an example let us consider the Dirac measure δ_{x_0} on $(H_0,\mathscr{B}(H_0))$ which is defined by

$$\delta_{x_0}(B) = \begin{cases} 1, & x_0\in B, \\ 0, & x_0\notin B, \end{cases} \quad B\in\mathscr{B}(H_0). \tag{17}$$

As in finite-dimensional spaces we have

$$\int_{H_0} d\delta_{x_0}(x) f(x) = f(x_0), \tag{18}$$

from which we obtain the covariance operator

$$C_{\delta_{x_0}} = |x_0\rangle\,\langle x_0|. \tag{19}$$

The measure μ to which a given covariance operator C_μ corresponds is not uniquely determined. The most important exceptions are—as in finite-dimensional spaces—Gaussian measures which are uniquely determined by the covariance operator and the mean $|m_\mu\rangle\in H_0$ which is defined by

$$\langle y|m_\mu\rangle_0 = \int_{H_0} d\mu(x)\, \langle y|x\rangle_0. \tag{20}$$

A Gaussian measure is defined by the relation

$$\mu(\{\, y\in H_0, \langle x|\,y\rangle_0 \leqslant a\,\}) = \int_{-\infty}^{a} [2\pi\sigma(x)]^{-1/2} \exp\left(-\frac{[t-m(x)]^2}{2\sigma(x)}\right) dt, \tag{21}$$

where

$$m(x) = \langle m_\mu|x\rangle_0, \quad \sigma(x) = \langle x|C_\mu\,|x\rangle_0. \tag{22}$$

We now consider measures on the complex measure space $(H,\mathscr{B}(H))$. As the isometry U^{-1} is linear bounded, hence continuous, it is $(H_0,\mathscr{B}(H_0)) - (H,\mathscr{B}(H))$ measurable. By means of U^{-1} we define the image measure of μ_0, μ_0 measure on $(H_0,\mathscr{B}(H_0))$, by

$$\mu(B) = \mu_0(U(B)), \quad B\in\mathscr{B}(H). \tag{23}$$

Let $f, f: H\to\mathbb{R}$, be a measurable function, then

$$\int_H d\mu(\phi) f(\phi) = \int_{H_0} d\mu_0(x) f(U^{-1}(x)). \tag{24}$$

In the following we confine ourselves to probability measures μ_0 on H_0. In this case μ defined by (23) is a probability measure too.

IV. QUANTUM MECHANICS AND INTEGRATION IN HILBERT SPACES

By means of the mathematical tools of the Secs. II and III we now are going to demonstrate how to express quantum mechanical expectation values by measure integrals which are defined on the Hilbert space which by familiarity is associated with a quantum mechanical system.

We denote the quantum mechanical expectation value of an observable which is represented by the operator $A\in\mathscr{U}(H)$ by $E_W(A)$ if the state of the system is given by the statistical operator $W\in\mathscr{S}(H)$. $E_W(A)$ is evaluated as

$$E_W(A) = \mathrm{tr}_H(WA). \tag{25}$$

Let us now apply the isometry between H and the real Hilbert space H_0 which we defined in Sec. II. We present the operators A and W by the unitarily equivalent ones and denote them by A_0, W_0. As we know $A_0\in\mathscr{U}(H_0)$ and $W_0\in\mathscr{S}(H_0)$. This yields

$$E_W(A) = \tfrac{1}{2}\mathrm{tr}_{H_0}(W_0 A_0). \tag{26}$$

As $W_0\in\mathscr{S}(H_0)$ we can interpret W_0 as the covariance operator of a probability measure on $(H_0,\mathscr{B}(H_0))$ which we denote by μ_{W_0}. Note that this does not determine the measure uniquely.

By means of this measure we establish the following identity where $W_0\in\mathscr{S}(H_0)$ and $A_0\in\mathscr{U}(H_0)$:

$$\mathrm{tr}_{H_0}(W_0 A_0) = \int_{H_0} d\mu_{W_0}(x)\,\langle x|A_0|x\rangle_0. \tag{27}$$

For the proof we need the spectral decomposition of A_0

$$A_0 = \int_{H_0} \lambda\, dP_\lambda. \tag{28}$$

We proceed as follows:

$$\begin{aligned}
&\int_{H_0} d\mu_{W_0}(x)\,\langle x|A_0|x\rangle_0 \\
&\quad= \int_{H_0} d\mu_{W_0}(x)\left\langle x\left|\left(\int_{\mathbb{R}}\lambda\, dP_\lambda\right)x\right.\right\rangle_0 \\
&\quad= \int_{H_0} d\mu_{W_0}(x)\sum_{n,m}\langle x|e_n\rangle_0\left\langle e_n\left|\left(\int_{\mathbb{R}}\lambda\, dP_\lambda\right)e_m\right.\right\rangle_0 \\
&\qquad\times\langle|e_m|x\rangle_0 \qquad (29)\\
&\quad= \int_{H_0} d\mu_{W_0}(x)\sum_{n,m}\langle e_m|x\rangle_0\,\langle x|e_n\rangle_0 \\
&\qquad\times\left\{\int_{\mathbb{R}}\lambda\, d(\langle e_n|P_\lambda\, e_m\rangle_0)\right\}.
\end{aligned}$$

By (25), Lebesgue's dominated convergence theorem, and the theorem of Fubini

$$\begin{aligned}
&\int_{H_0} d\mu_{W_0}(x)\,\langle x|A_0|x\rangle_0 \\
&\quad= \sum_{n,m}\int_{H_0} d\mu_{W_0}(x)\left\{\int_{\mathbb{R}}\lambda\, d(\langle e_n|P_\lambda\, e_m\rangle_0)\right\} \\
&\qquad\times\langle e_m|x\rangle_0\,\langle x|e_n\rangle_0 \qquad (30)\\
&\quad= \sum_{n,m}\left\{\int_{\mathbb{R}}\lambda\, d(\langle e_n|P_\lambda\, e_m\rangle_0)\right\} \\
&\qquad\times\int_{H_0} d\mu_{W_0}(x)\,\langle e_m|x\rangle_0\,\langle x|e_n\rangle_0.
\end{aligned}$$

By the definition (14) of the covariance operator we obtain

$$\begin{aligned}
&\int_{H_0} d\mu_{W_0}(x)\,\langle x|A_0|x\rangle_0 \\
&\quad= \sum_{n,m}\langle e_m|W_0|e_n\rangle_0\,\langle e_n|\left(\int_{\mathbb{R}}\lambda\, dP_\lambda\right)|e_m\rangle_0 \qquad (31)\\
&\quad= \sum_m\langle e_m|W_0 A_0|e_m\rangle_0 = \mathrm{tr}_{H_0}(W_0 A_0).
\end{aligned}$$

For convenience let us express the result (26) by means of the image measure μ_W [cf. Eq. (23)] on H

$$\begin{aligned}
E_W(A) &= \tfrac{1}{2}\mathrm{tr}_{H_0}(A_0 W_0) \\
&= \int_H d\mu_W(\phi)\,\tfrac{1}{2}\langle U\phi\,|UAU^{-1}|U\phi\rangle_0. \qquad (32)
\end{aligned}$$

Using Eq. (13) this can be expressed by the scalar product on H as

$$E_W(A) = \mathrm{tr}_H(WA) = \int_H d\mu_W(\phi)\,\tfrac{1}{2}\langle\phi\,|A\,|\phi\rangle. \tag{33}$$

Note that the factor 1/2 in the expectation functional

$$f_A(\phi) = \tfrac{1}{2}\langle\phi\,|A\,|\phi\rangle. \tag{34}$$

is due to $\mathrm{tr}_{H_0}(W_0) = 2$.

V. A PROBABILISTIC FORMULATION OF QUANTUM THEORY

The central result of this paper is contained in

$$E_W(A) = \int_H d\mu_W(\phi)\,f_A(\phi), \tag{35}$$

where

$$\begin{aligned}
&f_A(\phi) = \tfrac{1}{2}\langle\phi\,|A\,|\phi\rangle, \quad \mu_W(B) = \mu_{(TW)}(U(B)), \\
&B\in\mathscr{B}(H). \qquad (36)
\end{aligned}$$

The mappings

f: $\mathscr{U}(H)\to$ [measurable functions on $(H,\mathscr{B}(H))$],

μ: $\mathscr{S}(H)\to$ [probability measures on $(H,\mathscr{B}(H))$],

show indeed those properties characteristic for a hidden variables theory. For a review of hidden variables theories cf. Chap. 7 of Ref. 1. Note that these criteria apply to the real representation as well.

The space of hidden variables is the Hilbert space of the system. That the elements of H are hidden variables can be understood from the fact that quantum theory is a partially interpreted theory. This means that the formalism of quantum theory contains an element, a "theoretic term" without empirical meaning given by the wavefunctions. Thus it is not so surprising that these quantities without empirical meaning constitute the space of hidden variables.

In Ref. 3 Wiener and Siegel have established a hidden variables theory of quantum mechanics which resembles that one we present here insofar as the space of hidden variables in Ref. 3 is defined by the underlying Hilbert space too. But their definition of the mappings (36) is quite different from ours. The most important difference results from the fact that f in Ref. 3 implicitly depends on the statistical operator W of the system. Thus according to the definition of a

hidden variables theory Wiener's and Siegel's construction does not define such a theory.

The possibility demonstrated in (36) to set up a hidden variables theory of quantum mechanics seems to contradict v. Neumann's famous "impossibility proof." It was proved by v. Neumann[4] and later in a more refined manner by Gleason[5] that a probability measure π defined on the complete orthocomplementary lattice of closed subspaces of a Hilbert space H of dimension $\dim(H)\geqslant 3$ uniquely is defined by a positive, self-adjoint operator of trace class Π such that $\mathrm{tr}(\Pi)=1$ and

$$\pi(U)=\mathrm{tr}(\Pi P_U), \tag{37}$$

where P_U denotes the orthogonal projection on the subspace U.

This result justifies Eq. (25) which in turn was used by v. Neumann to demonstrate that there are no dispersion-free states in quantum theory. A dispersion-free state W is characterized by the fact that for all $A\in\mathcal{U}(H)$

$$E_W(A^2)-[E_W(A)]^2=0. \tag{38}$$

The fact that there are no disperson-free states has been regarded as the proof for the impossibility of hidden variables representations. Indeed, a formulation of quantum theory by means of classical probability theory admits the Dirac measure which is dispersion-free in each measure space. What we have done, however, is no more than to express the familiar term $\mathrm{tr}(WA)$ by means of a measure integral such that v. Neumann's argument still holds.

To elucidate the situation let us rewrite the left-hand side of Eq. (38) using the probabilistic formulation we have established

$$\int_H d\mu_W(\phi)f_{(A^2)}(\phi)-\left[\int_H d\mu_W(\phi)f_A(\phi)\right]^2. \tag{39}$$

Obviously this is different from the expression

$$\int_H d\mu_W(\phi)\,[f_A(\phi)]^2-\left[\int_H d\mu_W(\phi)f_A(\phi)\right]^2, \tag{40}$$

as in general

$$(f_A)^2\neq f_{(A^2)}. \tag{41}$$

The fact that (40) is used to define a dispersion-free random variable f_A on $(H,\mathcal{B}(H))$ and that this expression is different from (39) indicates that with respect to the quantum mechanical meaning of dispersion free the functional f_A cannot be regarded as an ordinary random variable on $(H,\mathcal{B}(H))$.

This becomes more evident if we consider a pure state which can, e.g., be represented by means of a Dirac measure in $(H,\mathcal{B}(H))$. Whereas expression (40) vanishes identically for a Dirac measure, due to Eq. (41) this is not the case for expression (39). This shows that the quantum mechanical meaning of dispersion free is different from that one of classical probability theory, and for that very reason different from classical physics. This fact allows the representation of quantum theory by means of integration in Hilbert space.

Equation (41) which is crucial for our reasoning is connected with the transformation properties of the functional f_A with respect to operator valued functions $k(A)\in\mathcal{U}(H)$. Substituting $k(A)$ for A, the mapping f_A has to be replaced by $f_{(k(A))}$ which cannot be effected by a substitution of the variable $\phi\in H$.

These transformation properties are different from those which hold in classical mechanics and show explicitly that the nonexistence of dispersion-free states depends on the mapping f and not on the measure. The formula $E_W(k(A))=\mathrm{tr}(Wk(A))$, however, corresponds to the classical analog.

We finally remark that these arguments do not depend on $\dim(H)$.

VI. DISCUSSION AND CONCLUSION

The most severe restrictions to the probabilistic version of quantum mechanics we present here are due to the fact that we confine our treatment to linear bounded self-adjoint operators. We do so in order to avoid singularities of the mapping f_A if we deal with unbounded operators A which can only be defined on a dense linear subset of H.

In this situation the concept of rigged Hilbert spaces has successfully been used in quantum mechanics. As there exists a theory of integration for one of the spaces of the Gel'fand triplet[6] it might be expected that this construction is apt to overcome these difficulties.

The above probabilistic formulation of quantum theory fulfills the formal aspects of a hidden variables theory. But as the hidden variables are the wavefunctions themselves the indeterminism of quantum mechanics still remains.

As we have seen this depends on the properties of the mapping f which associates to each observable an integrable function. Structurally f sets up the expectation value of an observable A with respect to the state $|\phi\rangle\langle\phi|$. The quantity thus obtained, $\langle\phi|A|\phi\rangle$, already contains the relevant quantum mechanical information and clearly demonstrates the difference compared to classical mechanics.

In contrast to the expression $\mathrm{tr}(WA)$ where all matrix elements of A with respect to a complete orthonormal system are needed, the probabilistic formulation gets along with the diagonal elements. But these constitute an uncountable set and the "summation over all states" is carried out with respect to this overcomplete set. These aspects of our theory resemble the well-known properties of the coherent state representation.

The bilinear structure of the functional f_A implies that the measures in Hilbert space are arbitrary up to the covariance operator as the other moments are empirically irrelevant. In the description of quantum theory by means of classical probability theory this indicates a degree of indeterminateness which discerns classical statistical mechanics and quantum mechanics.

These properties have an impact on quantum dynamics. Given a statistical operator $W(t)$, $t\in[0,T]$, a probability measure may be associated with it at each instant of time. This gives a stochastic process with state space $H^{[0,T]}$. But remember that it is possible to associate the measure at different times in a quite different manner. From this we conclude that there is no specific quantum mechanical stochastic process which in addition would contain more information than empirically relevant.

We remark that quite in analogy to the fact that Feyn-

man's path integral can be applied to problems of statistical mechanics, this is the fact for our formalism too. The canonical distribution of quantum statistical mechanics, $\rho = \exp(-\beta\mathscr{H})$, has just the properties which are needed for a covariance operator of a measure in H. This permits the representation of the partition function Z by means of the formula

$$Z = \mathrm{tr}(\rho) = \tfrac{1}{2}\int_H d\mu_\rho(\phi). \tag{42}$$

To summarize, we have given a probabilistic formulation of quantum theory by means of measure integrals in Hilbert space. Mathematically our formalism is identical to the well-known representations of quantum theory such that no new questions arise with respect to the interpretation of quantum mechanics. Concerning the question of hidden variables we have arrived at what might be called a vicious circle.

The present formalism is not able to deal with unbounded operator so far. As we have shown integration in function spaces applied to quantum mechanics is sensitive to the domain of an operator. This may be helpful for a better understanding of the role of the various topologies we are confronted with in infinite dimensional spaces—e.g., Gel'fand triplets—the empirical meaning of which is not evident.

ACKNOWLEDGMENTS

I thank Prof. J. Kamphusmann for clarifying discussions and Dr.D. Dürr, now at Rutgers College, for years of cooperation.

[1]M. Jammer, *The Philosophy of Quantum Mechanics* (Wiley, New York 1974).
[2]H.-H. Kuo, *Gaussian Measures in Banach Spaces* (Springer, Berlin, 1975).
[3]4N. Wiener and A. Siegel, Nuovo Cimento Suppl. **2**, 982 (1957).
[4]J. v. Neumann, *Die Mathematischen Grundlagen der Quantenmechanik* (Springer, Berlin, 1968).
[5]A.M. Gleason, J. Math. Mech. **6**, 885 (1957).
[6]I.M., Gelfand and N.J. Wilenkin, *Verallgemeinerte Funktionen, Band IV* (Deutscher Verlag, Berlin, 1964).

Positive quantum joint distributions

Leon Cohen[a] and Y. I. Zaparovanny[b]
Hunter College of the City University, New York, New York 10021

(Received 19 April 1979; accepted for publication 18 May 1979)

We demonstrate the existence of positive phase space density functions which yield the quantum mechanical marginal distributions of position and momentum.

I. INTRODUCTION

The possibility of formulating quantum mechanics in the phase space of position and momentum was first investigated by Wigner in 1932.[1] He studied a function $F(q,p)$ which yields the proper quantum mechanical marginal distributions when either one of the variables is integrated over,

$$\int F(q,p)\,dp = |\psi(q)|^2, \tag{1.1}$$

$$\int F(q,p)\,dq = |\phi(p)|^2, \tag{1.2}$$

where ψ and ϕ are the configuration and momentum wave functions respectively.

Since Wigner's original paper other phase space distributions have been found which also satisfy the above.[2] Cohen[3] has given an explicit formula for the set of all such functions. The description of quantum mechanics in terms of these functions is often called the phase space formulation of quantum mechanics. It has been applied to a wide variety of fields such as to quantum statistical mechanics, coherence properties of light and to questions relating to the foundations of quantum mechanics.

As Wigner pointed out in his first paper the function he presented cannot be considered a proper probability distribution because it may take on negative values. The other distributions which have been investigated also take on negative or even imaginary values. For this reason these functions are sometimes called quasiprobabilities. In Wigner's 1932 paper he stated that positive distributions which are bilinear in the wave function do not exist and he has recently published a proof demonstrating that fact.[4] Numerous authors have blanketly stated that positive distributions do not exist and have either quoted Wigner (although Wigner was quite clear that his proof applied only to bilinear forms) or have given other arguments. The aim of this paper is to show that positive phase functions do exist. They are, of course, not bilinear in the wave function.

We point out that positive distributions which do not satisfy Eqs. (1.1) and (1.2) exactly have been given by Bopp[5] and Kuryshkin.[6]

In Sec. II we present some of the basic ideas and prove some theorems relating to products of probability functions which will be subsequently used. The main results and the demonstration of positiveness is presented in Sec. III.

II. DENSITY FUNCTIONS AND CHARACTERISTIC FUNCTIONS

The set of all possible functions of q and p which are consistent with (1.1) and (1.2) is

$$F(q,p) = \frac{1}{4\pi^2}\int \exp(-i\theta q - i\tau p + i\theta u) f(\theta,\tau) \times \psi^*(u - \tfrac{1}{2}\tau\hbar)\psi(u + \tfrac{1}{2}\tau\hbar)\,d\theta\,d\tau\,du, \tag{2.1}$$

where $f(\theta,\tau)$ is any function satisfying

$$f(\theta,0) = f(0,\tau) = 1. \tag{2.2}$$

The Wigner distribution is obtained by taking $f = 1$. The characteristic function of a distribution and its marginals are defined as follows,

$$M(\theta,\tau) = \int\int F(q,p) e^{i\theta q + i\tau p}\,dq\,dp, \tag{2.3}$$

$$M_q(\theta) = \int |\psi(q)|^2 e^{i\theta q}\,dq, \tag{2.4}$$

$$M_p(\tau) = \int |\phi(p)|^2 e^{i\tau p}\,dp, \tag{2.5}$$

and furthermore satisfy

$$M_q(\theta) = M(\theta,0), \tag{2.6}$$

$$M_p(\tau) = M(0,\tau), \tag{2.7}$$

$$M_q(0) = M_p(0) = M(0,0) = 1. \tag{2.8}$$

Specifically for F as given by Eq. (2.1) the characteristic function is

$$M(\theta,\tau) = f(\theta,\tau)\int e^{i\theta u}\psi^*(u - \tfrac{1}{2}\tau\hbar)\psi(u + \tfrac{1}{2}\tau\hbar)\,du. \tag{2.9}$$

We now prove the following theorem relating probability distributions. Let $h(v,w)$ be a joint probability density of the random variables v, w defined over the range (0,1) and let $k(q)$ and $g(p)$ be any probability density distributions of the variables q and p which range over all space. Then, $H(q,p)$ defined by

$$H(q,p) = k(q)\,g(p)h(v,w), \tag{2.10}$$

is a proper probability function where v and w are taken to be

$$v(q) = \int_{-\infty}^{q} k(q')\,dq', \tag{2.11}$$

$$w(p) = \int_{-\infty}^{p} g(p')\,dp'. \tag{2.12}$$

Also, the marginals of $H(q,p)$ are given by

$$H_1(q) = k(q)h_1(v), \tag{2.13}$$

$$H_2(p) = g(p)h_2(w), \tag{2.14}$$

[a]Work done during tenure of a grant from The City University Faculty Research Award Program.

[b]Permanent address: Peoples' Friendship University, Moscow, U.S.S.R.

 0022-2488/80/040794-03$1.00

where h_1 and h_2 are the marginal distributions of $H(q,p)$.

We first note that since k, g, and h are positive H is also. To show that H is properly normalized consider

$$\int\int H(q,p)\,dq\,dp = \int\int_{-\infty}^{\infty} k(q)g(p)h(v(q),w(p))\,dq\,dp. \tag{2.15}$$

Making the change of variables

$$dq = \frac{dv}{k}, \tag{2.16}$$

$$dp = \frac{dw}{g}, \tag{2.17}$$

Eq. (2.15) becomes

$$\int_0^1\int_0^1 h(v,w)\,dv\,dw = 1, \tag{2.18}$$

by definition and hence H is a well-defined probability. To show the relation between the marginal as given by Eq. (2.13) we have

$$\begin{aligned} H_1(q) &= \int H(q,p)\,dp \\ &= \int k(q)\,g(p)h(v,w)\,dp \\ &= \int_0^1 k(q)h(v,w)\,dw \\ &= k(q)h_1(v(q)). \end{aligned} \tag{2.19}$$

Similarly for relation (2.14).

We define the characteristic functions of $H(q,p)$, $H_1(q)$, $H_2(p)$ by

$$N(\theta,\tau) = \int\int H(q,p)\exp(i\theta q + i\tau p)\,dq\,dp, \tag{2.20}$$

$$N_q(\theta) = \int H_1(q)e^{i\theta q}\,dq, \tag{2.21}$$

$$N_p(\tau) = \int H_2(p)e^{i\tau p}\,dp. \tag{2.22}$$

In the subsequent we specialize to the quantum case by taking

$$k(q) = |\psi(q)|^2, \tag{2.23}$$

$$g(p) = |\phi(p)|^2. \tag{2.24}$$

III. POSITIVE QUANTUM DISTRIBUTIONS

The class of distribution functions we shall show are positive are obtained by choosing

$$\begin{aligned} f(\theta,\tau) = M(\theta)M(\tau)\Bigg[1 + c\Bigg(&\frac{N(\theta,\tau)}{M(\theta)M(\tau)} \\ &- \frac{N(\theta)}{M(\theta)} - \frac{N(\tau)}{M(\tau)} + 1\Bigg)\Bigg] \\ &\times\left(\int \psi^*(u - \tfrac{1}{2}\tau\hbar)e^{i\theta u}\psi(u + \tfrac{1}{2}\tau\hbar)\,du\right)^{-1}, \end{aligned} \tag{3.1}$$

where c is a constant appropriately chosen as discussed below and the N's are obtained from any probability function $h(v,w)$ defined on the unit square, as outlined in Sec. II.

Putting $f(\theta,\tau)$ as given by Eq. (3.1) into Eq. (2.1) yields

$$F(q,p) = |\psi(q)|^2\,|\phi(p)|^2\,[1 + c\rho(q,p)], \tag{3.2}$$

where we have defined

$$\rho(q,p) = h(v(q),w(p)) - h_1(v(q)) - h_2(w(p)) + 1. \tag{3.3}$$

Special cases of distribution functions of this form have been studied in other contexts.[7-9] Gumbas,[8] in particular, has studied the case when $\rho(v,w) = (2v-1)(2w-1)$ which is obtained from our formulation by taking the probability distribution $h(v,w) = 4vw$. We note that the advantage of formulating F in terms $h(v,w)$ is that the marginals will always be satisfied for any h. Hence F's can be generated at will.

We now show that indeed the marginal distributions are satisfied and that c can be chosen to make F positive.

A. Marginals

Taking $\tau = 0$ in Eq. (3.1)

$$\begin{aligned} f(\theta,0) = M(\theta)&\left[1 + c\left(\frac{N(\theta,0)}{M(\theta)} - \frac{N(\theta)}{M(\theta)} - 1 + 1\right)\right] \\ &\times\left(\int |\psi(u)|^2 e^{i\theta u}\,du\right)^{-1} = 1. \end{aligned} \tag{3.4}$$

which follows by noting that $N(\theta,0) = N(\theta)$ and that the denominator is $M(\theta)$. Similarly for $f(0,\tau)$. Hence the F given by Eq. (3.2) satisfies Eqs. (1.1) and (1.2).

It is of interest to show directly from (3.2) and (3.3) that the marginal distributions are satisfied

$$\begin{aligned} \int F(q,p)\,dp &= \int |\psi(q)|^2\,|\phi(p)|^2\,(1 + c\rho(v,w))\,dp \\ &= |\psi(q)|^2 + c\int \rho(v,w)\,|\phi(p)|^2\,|\psi(q)|^2\,dp. \end{aligned} \tag{3.5}$$

But

$$\begin{aligned} \int \rho(v,w)|\phi(p)|^2\,|\psi(q)|^2\,dp &= \int [H(q,p) - H_1(q)|\phi(p)|^2 - H_2(p)|\psi(q)|^2 \\ &\quad + |\psi(q)|^2\,|\phi(p)|^2]\,dp \\ &= H_1(q) - H_1(q) - |\psi(q)|^2 + |\psi(q)|^2 = 0 \end{aligned} \tag{3.6}$$

and hence (1.1). Similarly for (1.2)

B. Positiveness

We show that for a given $\rho(v,w)$ a range of c's can always be found which makes F positive. From Eq. (3.3) we have

$$\begin{aligned} \int_0^1\int_0^1 \rho(v,w)\,dv\,dw &= \int [h(v,w) - h_1(v) - h_2(w) + 1]\,dv\,dw \\ &= 0, \end{aligned} \tag{3.7}$$

and therefore $\rho(v,w)$ must take on positive and negative values on the unit square. We follow an argument similar to that of Farlie.[9] Define the absolute minimum and maximum of

$\rho(v,w)$ by -1_1 and 1_2 respectively. Suppose c is positive then from (3.5) we must choose c such that

$$1 - cl_1 \geqslant 0, \tag{3.8}$$

or

$$c \leqslant 1/l_1. \tag{3.9}$$

Nevative c's force

$$|c| \geqslant 1/l_2, \tag{3.10}$$

and hence any c chosen such that

$$-1/l_2 \leqslant c \leqslant 1/l_1, \tag{3.11}$$

will make F positive.

We note that $\rho(v,w)$ may go to either plus or minus infinity but not both, otherwise we would be forced to choose c equal to zero.

We have shown the existence of an infinite class of positive joint quantum distributions which satisfy the proper quantum mechanical marginal functions. They can be generated at will: Choose any probability distribution defined on the unit square and form ρ as in (3.3). Find the absolute minimum and maximum of and choose any c in the range given by 3.10. The F defined by (3.2) will be positive and satisfy the quantum mechanical marginals given by Eqs. (1.1) and (1.2). We point out that consistent with Wigner's proof the F's given by Eq. (3.2) are not bilinear in ψ. Also, it is probably the case that there are positive F's of different functional form from those considered here.

The uncertainty principle is very often appealed to argue that positive joint distributions with correlations cannot exist. It has been previously pointed out that such arguments are not valid and in the light of having found positive distributions it may be worthwhile to point this out again.[10,11] The uncertainty principle is a statement regarding the root mean square deviations of position and momentum,

$$(\Delta q)^2 = \int (q - \langle q \rangle)^2 \, |\psi(q)|^2 \, dq, \tag{3.12}$$

$$(\Delta p)^2 = \int (p - \langle p \rangle)^2 |\phi(p)|^2 \, dp. \tag{3.13}$$

It is clear that Δq and Δp depend *only* on the marginal distributions and hence any proper joint distribution which yields the proper quantum mechanical distribution of position and momentum will yield the uncertainty relations. The well defined positive distributions given in this paper are such functions.

[1] E. Wigner, Phys. Rev. **40**, 749 (1932).

[2] Some of the early papers studying particular distributions are: Ya.P. Terletsky, JETP, 1290 (1937); D.I. Blokhintsev, J. Phys. (Moscow) **2**, 71 (1940); H.J. Groenewold, Physica **12**, 405 (1946); J.E. Moyal, Proc. Cambridge Philos. Soc. **45**, 99 (1949); T. Takabayasi, Prog. Theor. Phys. **11**, 341 (1959); G.A. Baker, Jr., Phys. Rev. **69**, 2198 (1959); H. Margenau and R.N. Hill, Progr. Theor. Phys. **26**, 722 (1961); C.L. Mehta, J. Math. Phys. **5**, 677 (1964).

[3] L. Cohen, J. Math. Phys. **7**, 781 (1966).

[4] E. Wigner, "Quantum-Mechanical Distribution Functions Revisited" in *Perspectives in Quantum Theory*, edited by W. Yourgrau and A. van der Merwe (M.I.T., Cambridge, 1971).

[5] F. Bopp, Ann. L'Inst. H. Poincaré **15**, 81 (1956).

[6] V.V. Kuryskin, Izv. Vusov Phys. **11**, 103 (1971).

[7] D. Margenstern, Mitt. Math. Stat. **8**, 234 (1956).

[8] E.J. Gumbel, Rev. Fac. Ci. Univ. Lisboa Ser. 2A CI. Mat. **7**, 179 (1959).

[9] D.J.G. Farlie, Biometrica **50**, 499 (1963).

[10] L. Cohen, Philos. Sci. **33**, 317 (1966).

[11] V.V. Kuryshkin, "Uncertainty Principle and the Problems of Joint Coordinate-Momentum Probability Density in Quantum Mechanics," in *The Uncertainty Principle and Foundations of Quantum Mechanics* (Wiley, London, 1977).

A Lie superalgebraic interpretation of the para-Bose statistics

A. Ch. Ganchev and T. D. Palev
Institute for Nuclear Research and Nulear Energy, Boul. Lenin 72, 1113 Sofia, Bulgaria

(Received 21 February 1979; accepted for publication 20 April 1979)

We show that n pairs of para-Bose operators generate the classical simple orthosymplectic Lie superalgebra $\mathrm{osp}(1,2n)\equiv B(0,n)$. The creation and annihilation operators are negative and positive root vectors, respectively, and they span a basis in the odd part of $B(0,n)$.

It is well known that generalizations of the ordinary quantum statistics are possible if one abandons the requirement for the commutator or the anticommutator of two fields to be a c-number.[1,2] The most popular one is the para-statistics proposed by Green.[1] In the case of spinor fields the anticommuting relations between the Fermi operators are replaced by double commutation relations for the para-Fermions. Ryan and Sudarshan have shown[3] that the para-Fermi statistics have well defined Lie-algebraical meaning: n pairs of para-Fermi creation and annihilation operators generate the simple Lie algebra B_n of the orthogonal group $\mathrm{SO}(n+1)$. This property simplifies considerably the investigation of the para-Fermi statistics. It has essentially been used in some recent new physical ideas based on the para-Fermi statistics.[4-7]

The fact that the para-Fermi operators generate a simple Lie algebra is important not only because it makes it possible to translate most of the problems of the para-Fermi statistics in the language of the simple Lie algebras, it also provides a basis for further generalizations. Indeed, the Fermi and para-Fermi quantization is actually a quantization according to certain representations of the odd-orthogonal algebra. Therefore, it is natural to ask whether one can quantize according to some representations of other classical Lie algebras. This question has been investigated.[2] It was shown that of every classical simple Lie algebra there corresponds a quantization which is logically compatible with the second quantization axioms.

The Lie algebraical properties of the para-Fermi operators are closely connected with the fact that their defining relations are expressed only in terms of commutators. This is not the case for para-Bose statistics. The para-Bose creation ($\xi = +$) and annihilation ($\xi = -$) operators a_i^ξ satisfy the relations

$$[\{a_i^\xi, a_j^\eta\}, a_k^\epsilon] = (\epsilon - \xi)\delta_{ik} a_j^\eta + (\epsilon - \eta)\delta_{kj} a_i^\xi. \tag{1}$$

(Throughout the paper $\xi, \eta, \epsilon = \pm$ or ± 1, $[x,y] = xy - yx$, and $\{x,y\} = xy + yx$.) Therefore, the para-Bose statistics do not allow generalizations along a Lie algebraical line. It is known that the linear envelope of all anticommutators, i.e, the set

$$L_0 = \mathrm{lin.env.}\{\{a_i^\xi, a_j^\eta\} | i,j = 1,\ldots,n; \quad \xi,\eta = \pm\}, \tag{2}$$

is isomorphic to the classical simple Lie algebra of the symplectic group $\mathrm{Sp}(2n)$.[8] The operators are not, however, elements of this algebra and cannot be elements of any other Lie algebra in such a way that they generate through commutations the whole algebra.

The relations (1) indicate however that the para-Bose operators can be considered as elements of a Lie superalgebra, generating the whole algebra. To show this, denote by L_1 the linear envelope of the para-Bose creation ($\xi = +$) and annihilation ($\xi = -$) operators,

$$L_1 = \mathrm{lin.env.}\{a_i^\xi | i = 1,\ldots,n; \quad \xi = \pm\}. \tag{3}$$

Let pB_n be the direct sum of the spaces L_0 and L_1,

$$pB_n = L_0 + L_1. \tag{4}$$

From now on the lower case index $\alpha = 0,1$ indicates that the corresponding element belongs to L_α, i.e., $a_\alpha \in L_\alpha$.

Define a multiplication

$$[a,b] \in pB_n, \quad a,b \in pB_n. \tag{5}$$

In the space pB_n by the relations

$$\begin{aligned} &[a_1,b_1] = \{a_1,b_1\} \quad a_1,b_1 \in L_1, \\ &[a_0,b_\alpha] = [a_0,b_\alpha] \quad a_0 \in L_0, b_\alpha \in L_\alpha. \end{aligned} \tag{6}$$

To obtain the multiplication for arbitrary elements one has to extend the relations (6) by linearity.

Let Z_2 be a ring with two elements (0,1) and multiplication

$$0+0 = 0, \quad 0+1 = 1, \quad 1+1 = 0. \tag{7}$$

From relations (2) one easily obtains

$$[a_\alpha, b_\beta] \in L_{\alpha+\beta}, \quad \alpha,\beta \in Z_2. \tag{8}$$

Therefore, pB_n is a Z_2-graded algebra. Moreover, from 6 it follows

$$[a_\alpha, b_\beta] = -(-1)^{\alpha\beta}[b_\beta, a_\alpha], \quad \alpha,\beta,\gamma \in Z_2, \tag{9}$$

$$[a_\alpha, [b_\beta, c_\gamma]] = [[a_\alpha, b_\beta], c_\gamma] + (-1)^{\alpha\beta}[b_\beta, [a_\alpha, c_\gamma]]. \tag{10}$$

The relations (8–10) are defining for a Lie superalgebra. For the abstract definition of a Lie superalgebra see Ref. 9. The subspaces L_0 and L_1 are the even and odd parts of the algebra. We collect the result.

Lemma 1: With respect to the multiplication (6) pB_n is a Lie superalgebra. The para-Bose operators $a_1^\pm, \ldots, a_n^\pm$ define a basis in the odd part $L_1 \subset pB_n$ and generate through multiplication the whole algebra pB_n.

We call the Lie superalgebra pB_n para-Bose algebra. If $A,B \subset pB_n$ we denote

$$[A,B] = \{[a,b] | a \in A, b \in B\}. \tag{11}$$

in particular,

$$\{A,B\} = \{\{a,b\} | a \in A, b \in B, A,B \subset L_1\}, \tag{12}$$

 0022-2488/80/040797-03$1.00

$$[A,B] = \{[a,b] \,|\, a \in A \subset L_0,\ b \in B \subset pB_n\}. \tag{13}$$

Lemma 2: The para-Bose algebra pB_n is simple.

Proof: We have to show that pB_n has no nontrivial ideals. Let $I \neq 0$ is an ideal in pB_n. Consider three cases.

(a) Suppose $I \cap L_0 \neq 0$ and let

$$0 \neq x \in I \cap L_0. \tag{14}$$

Since L_0 is isomorphic to the simple Lie algebra C_n,

$$[x, L_0] = L_0; \tag{15}$$

moreover,

$$[L_0, L_1] = L_1. \tag{16}$$

Therefore, L_0 and L_1 are contained in I. Hence $I = pB_n$.

(b) Suppose $0 \neq x \in I \cap L_1$. The para-Bose operators constitute a basis in L_1.

$$x = \sum_{j,\eta} \alpha_\eta^j \, a_j^\eta. \tag{17}$$

Let $\alpha_\xi^i \neq 0$. Using the relation

$$[\{a_i^{-\xi}, a_i^{-\xi}\}, a_j^\eta] = 2(\xi + \eta)\delta_{ij} \, a_i^{-\xi}, \tag{18}$$

we obtain

$$[\{a_i^{-\xi}, a_i^{-\xi}\}, x] = 4\xi \alpha_\xi^i \, a_i^{-\xi} \in I.$$

Therefore, $a_i^{-\xi} \in I$. From the relation

$$[\{a_k^{-\xi}, a_i^\xi\}, a_i^{-\xi}] = -2\xi \, a_k^{-\xi},$$

we conclude that

$$a_k^{-\xi} \in I, \quad k = 1,2,...,n. \tag{19}$$

Hence

$$[\{a_k^\xi, a_k^\xi\}, a_k^{-\xi}] = -4\xi \, a_k^\xi \in I. \tag{20}$$

The relations (19) and (20) indicates that $L_1 \subset I$. Using (16) once more we conclude that $I = pB_n$.

(c) It remains to consider the general case

$$0 \neq x = x_0 + x_1 \in I, \tag{21}$$

where $x_\alpha \in L_\alpha$ are nonzero elements.

Suppose $[x_0, x_1] \neq 0$; then

$$[x_0 + x_1, x_0] = [x_1, x_0] \in L_1 \cap I. \tag{22}$$

Remark that if

$$0 \neq x_1 \in L_1 \Rightarrow \{x_1, x_1\} \neq 0. \tag{23}$$

therefore if $[x_1, x_0] = 0$, then

$$0 \neq [x_0 + x_1, x_1] = \{x_1, x_1\} \in L_0 \cap I. \tag{24}$$

So this case reduces to (a) and (b). Therefore, the Lie superalgebra pB_n contains no ideals different from 0 and pB_n, i.e., it is simple.

Recently, the simple Lie superalgebras have been fully classified.[9] As in the Lie algebraical case the simple Lie superalgebra is completely determined by its root system. The roots are vectors from the Cartan subalgebra $\mathscr{H}$ which by definition is the Cartan subalgebra of the even part. If A is a lienar operator in $L = L_0 + L_1$ denote by

$$\mathrm{tr} A \,|_{L_\alpha}, \tag{25}$$

the trace of the operator A in the subspace L_α, $\alpha = 0,1$. Then the Killing form on $\dot{L}$ is defined as

$$(a,b) = \mathrm{tr}\ \mathrm{ad}\ a\ \mathrm{ad}\ b\,|_{L_0} - \mathrm{tr}\ \mathrm{ad}\ a\ \mathrm{ad}\ b\,|_{L_1} \quad a,b \in L. \tag{26}$$

Here ad a is an operator from the adjoint representation,

$$(\mathrm{ad}\ a)z = [a,z] \in L, \quad z \in L. \tag{27}$$

The Killing form of a simple Lie superalgebra is not degenerate. The basis

$$h_1,...,h_n \quad e_{\omega_1},...,e_{\omega p}, \tag{28}$$

in L can always be chosen such that for every $h \in \mathscr{H}$

$$[h, e_{\omega i}] = (h, \omega i) e_{\omega i} \quad i = 1,...,p. \tag{29}$$

Here,

$$h_1, h_2, ..., h_p \tag{30}$$

is an arbitrary basis in the Cartan subalgebra $\mathscr{H}$.

Returning to the para-Bose algebra we choose as a basis in pB_n the vectors

$$a_i^\xi, \qquad i = 1,2,...,n; \quad \xi = \pm, \quad \text{basis in } L_1 \tag{31a}$$

$$h_i = -\tfrac{1}{2}\{a_i^+, a_i^-\}, \quad i = 1,...,n, \tag{31b}$$

$$\{a_j^+, a_k^-\}, \qquad j \neq k = 1,...,n, \tag{31c}$$

$$\{a_p^\xi, a_q^\xi\}, \qquad p \leqslant q = 1,...,n; \quad \xi = \pm, \quad \text{basis in } L_0 \tag{31d}$$

The vectors $h_1,...,h_n$ constitute a basis in the Cartan subalgebra $\mathscr{H}$; the linear envelope of the vectors (31b) and (31c), gives the Lie subalgebra $\mathrm{gl}(n+1) \subset L_0 = c_n$.

From (26) and the structure relations (1) one easily obtains

$$(h_i, h_j) = \delta_{ij}(4n + 2), \tag{32}$$

so that (31b) is an orthogonal basis in $\mathscr{H}$. Let

$$h^1, h^2, ..., h^n \tag{33}$$

be the dual basis to (31b),

$$(h_i, h^j) = \delta_i^j. \tag{34}$$

A simple calculation gives ($h \in \mathscr{H}$):

$$\begin{aligned} &[h, a_i^\xi] = (h, -\xi h^i) a_i^\xi, \\ &[h, \{a_p^\xi, a_q^\eta\}] = (h, -\xi h^p - \eta h^q)\{a_p^\xi, a_q^\eta\}, \end{aligned} \tag{35}$$

therefore the root system Σ of pB_n is

$$\Sigma = \{\xi h^i, \xi h^i + \eta h^j \,|\, i,j = 1,...,n;\ \xi,\eta = \pm\}. \tag{36}$$

Hence the vectors (31a), (31c), (31d), are the root vectors of the Lie superalgebra. The correspondence with their roots is

$$a_i^\xi \leftrightarrow -\xi h^i, \tag{37}$$

$$\{a_i^\xi, a_j^\eta\} \leftrightarrow -\xi h^i - \eta h^j, \tag{38}$$

so that the para-Bose creation ($\xi = +$) and annihilation ($\xi = -$) operators are in the basis (33) negative and positive root vectors, respectively.

The only simple Lie superalgebra with root system (36) is the algebra OSP(1,2n) or in the Kac notation[9] the algebra $B(0,n)$. We collect the results obtained so far in a theorem.

Theorem: The para-Bose algebra pB_n of n pairs of creation and annihilation operators $a_1^\pm,...,a_n^\pm$ is isomorphic to

the classical simple Lie superalgebra $B(0,n)$. The para-Bose operators constitute a basis in the even part of the algebra and generate the whole algebra. The basis in the Cartan subalgebra of pB_n can always be chosen such that the creation and annihilation operators are negative and positive root vectors, respectively.

Corollary 1: The Bose creation and annihilation operators $a_1^{\pm},...,a_n^{\pm}$ considered as odd elements of a Lie superalgebra generate one particular infinite-dimensional representation of the algebra $B(0,n)$.

In other words the usual Fock representation of n pairs of Bose operators is an infinite-dimensional irreducible representation of the Lie superalgebra. Since $B(0,n)$ is generated by the para-Bose operators, from the above Theorem we immediately conclude

Corollary 2: There exists a one to one correspondence between the (irreducible) representations of a given number n of para-Bose operator and the (irreducible) representations of the Lie superalgebra $B(0,n)$.

Thus, the problem of finding all representations of the para-Bose operators is equivalent to the determination of all representations of the orthosymplectic Lie superalgebra. Inverting the problem, we can say that a class of representations of the orthosymplectic algebra labelled by one positive integer p, the so called order of the parastatistics, has been found by Green [1] long before the concept of Lie superalgebra was introduced.

Since the transition to an infinite set of creation and annihilation operators does not change the algebraical structure, one can view the ordinary second quantization of, say, scalar field as a quantization according to an irreducible representation of the infinite dimensional orthosymplectic Lie superalgebra.

[1]H.S. Green, Phys. Rev. **90**, 270 (1953).
[2]T. Palev, Thesis, Institute for Nuclear Research and Nuclear Energy, Sofia (1976). Part of the results were published in Compt. Rend. Acad. Bulg. Sci. **30**, 993 (1977); preprint JINR E2-10258 (1977); Proc. XVth ICCR, Povdiv **11**, 546; (1977); See also preprint JINR E2-11904 and E2-11905 (1978).
[3]C. Ryan and E.C.G. Sudarshan, Nucl. Phys. **47**, 207 (1963).
[4]A.B. Govorkov, JETP **54**, 1785 (1968).
[5]A.B. Govorkov, Int. J. Theor. Phys. **7**, 49 (1973).
[6]A.B. Govorkov, preprint JINR, E2-7485 (1973).
[7]A.J. Bracken and H.S. Green, J. Math. Phys. **14**, 1784 (1973).
[8]S. Kamefuchi and Y. Takahashi, Nucl. Phys. **36**, 177 (1960).
[9]V.G. Kac, Adv. Math. **26**, 8 (1977).

On the extremum problem of the energy functional on the set of Slater determinants

Giacomo Fonte
Istituto di Fisica Teorica dell'Università, Catania, Italy

(Received 16 October 1979; accepted for publication 30 November 1979)

Differential properties of the energy functional on the set of Slater determinants are defined according to a new differential calculus on complex Banach spaces previously considered. Such properties permit us to derive, with a new and rigorous procedure, the Hartree–Fock equations and to define an iterative method of the gradient type to minimize this functional on the above set. For a large class of one-body and two-body potentials, including Coulombic potentials, we show that our gradient method yields a decreasing sequence of values of the energy functional and a sequence of approximate solutions of the Hartree–Fock equations. The convergence of the latter sequence to a solution of these equations is proved in any finite-dimensional subspace.

I. INTRODUCTION

The idea of determining approximate stationary states of a many-fermion system by restricting the wave function to a Slater (S) determinant is old.[1] The corresponding precise mathematical problem consists in the minimum problem for the energy functional

$$E(\Psi) = \frac{\langle \Psi | H\Psi \rangle}{|\Psi|^2}$$

on the set of S determinants. Up to now this problem presents formidable difficulties, both in proving the existence of the minimum and in determining a reliable approximation to it. In fact, only in the last few years[2] has the existence of the absolute minimum been proved for Coulombic interactions, while the research of numerical methods, which are more efficient and reliable than the solution of the Hartree–Fock (HF) equations by the usual diagonalization procedure, is still in progress.[3]

Moreover, we note that most atomic and nuclear physicists do not use the correct mathematical language to formulate our extremum problem. This language, as we show in Ref. 4, is the semidifferential calculus on complex Banach spaces. Therefore we can say that until today the extremum problem of $E(\Psi)$ on the set of S determinants is not even rigorously formulated.

As an unpleasant consequence, neither the theoretical results of the modern calculus of variations[5] nor the numerical methods of the optimization theory[6] can be applied to it. For these reasons we have thought it useful to formulate this extremum problem according to the language of the semidifferential calculus. The results of the present paper can be summarized as follows. We establish (Secs. III and IV) some semidifferentiability properties of $E(\Psi)$; we can then avoid the unpleasant facts previously mentioned. Thus we can derive in Sec. III the HF equations by a rigorous procedure, and we can extend in Sec. V the well-known iterative method of optimization theory, the gradient method, to solve approximately our extremum problem. In fact, for a large class of potentials, including Coulombic potentials, we construct by our gradient method a decreasing sequence of values of $E(\Psi)$ and a sequence of approximate solutions of the HF equations. Convergence to a solution is proved only in the subspace case when the space is restricted to a finite-dimensional one.

Already other authors[7,8] have considered the minimization of $E(\Psi)$ on the set of S determinants by means of a gradient method.[9] However, as we shall discuss widely in the concluding remarks, such a method is not, strictly speaking, correct, nor is it expected to be the most efficient method of the gradient type to solve approximately our extremum problem. Moreover, these authors do not give any convergence proof to a solution of the HF equations. For an easier reading of the paper we have summarized the main notations in the next section.

II. NOTATION

$\mathsf{B}(\mathsf{L}^2(R^3;C^2))$: Banach space of all bounded operators on $\mathsf{L}^2(R^3;C^2)$ to $\mathsf{L}^2(R^3;C^2)$.
$\mathsf{B}(\mathsf{L}^2(R^3;C^2)\otimes\mathsf{L}^2(R^3;C^2))$: Banach space of all bounded operators on $\mathsf{L}^2(R^3;C^2)\otimes\mathsf{L}^2(R^3;C^2)$ to $\mathsf{L}^2(R^3;C^2)\otimes\mathsf{L}^2(R^3;C^2)$.
C: complex numbers.
$\mathsf{C}(\mathsf{L}^2(R^3;C^2))$: set of all closed operators from $\mathsf{L}^2(R^3;C^2)$ to $\mathsf{L}^2(R^3;C^2)$.
$\mathsf{D}(A)$: domain of definition of A, where A can be an operator or a functional.
$\mathsf{D}_{\Psi\Xi}[k,l]$: minor of the first rank assigned to the k row and l column of matrix $\{\langle\psi^r|\chi^s\rangle\}$, where ψ^r and χ^s, $r,s = 1,\ldots N$, are the one-particle functions associated with the S determinants Ψ and Ξ.
$\mathsf{D}_{\Psi\Xi}[k_1k_2,l_1l_2]$: minor of the second rank of the matrix $\{\langle\psi^r|\chi^s\rangle\}$ assigned to the k_1,k_2 rows and l_1,l_2 columns and antisymmetrized in these indices.

$$E(\Psi_n) = E_n = \langle\Psi_n|H\Psi_n\rangle/|\Psi_n|^2, \quad n = 0,1,2,\cdots.$$

$$E'_{c\psi^i_n} = r_c\left(h_n^{\mathrm{HF}}\psi^i_n - \sum_{r=1}^{N}\langle h_n^{\mathrm{HF}}\psi^i_n|\psi^r_n\rangle\psi^r_n\right);$$

strong partial gradient of $E(\Psi)$ with respect to ψ^i_n, $i = 1,2,\ldots N$, at $(\psi^1_n,\ldots,\psi^N_n)$, with $\langle\psi^i_n|\psi^j_n\rangle = \delta_{ij}$, $i,j = 1,\ldots,N$, $n = 0,1,2\cdots$.
$h = t + v$: one-particle part of one-body Hamiltonian operator.

 0022-2488/80/040800-07$1.00

$h^{HF} = h + \Sigma_{s=1}^{N} \langle \cdot \psi^s | \tilde{w} \psi^s \cdot \rangle$: Hartree–Fock Hamiltonian.
$h_n^{HF} = h + \Sigma_{s=1}^{N} \langle \cdot \psi_n^s | \tilde{w} \psi_n^s \cdot \rangle$: Hartree–Fock Hamiltonian at the nth step.
$H = \Sigma_{i=1}^{N} h^i + \Sigma_{i<j=1}^{N} w^{ij}$: Hamiltonian operator.
$L^2(R^3;C^2)$: space of square-integrable functions on R^3 with values in C^2, i.e., "one-particle functions of space and spin."
$L_a^2(R^{3N};C^{2N})$: space of antisymmetrized square-integrable functions on R^{3N} with values in C^{2N}, i.e., "N-fermion functions of space and spin."
N: number of identical fermions.
R: real numbers.
t: one-particle kinetic energy operator.
v: one-body potential.
w: two-body potential.
$\tilde{w}$: antisymmetrized w.
$\Psi = (N!)^{-1/2} \det\{\psi^1, \psi^2, \ldots \psi^N\}$: Slater determinant.
$\Psi_n = (N!)^{-1/2} \det\{\psi_n^1, \psi_n^2, \ldots, \psi_n^N\}$: Slater determinant at the nth step.
$\Psi[\cdot] = (N!)^{-1/2} \det\{\psi^1, \ldots \psi^{i-1} \cdot, \psi^{i+1}, \ldots, \psi^N\}$: Slater determinant where ψ^i is replaced with another vector of $L^2(R^3;C^2)$.
$\|\cdot\|_{L^2}$: norm of operators of $B(L^2(R^2;C^3))$, or of $B(L^2(R^3;C^2) \otimes L^2(R^3;C^2))$.
$\|\cdot\|_{z_c}$: norm of bounded operators on z_c to $L^2(R^3;C^2)$.

III. WEAK SEMIDIFFERENTIABILITY OF $E(\Psi)$ AND HARTREE-FOCK EQUATIONS

First of all we recall the definitions of the semidifferential calculus used in this paper so that it is self-contained; for more details we refer the reader to Ref. 4 and to references quoted therein. Let $X^1, \ldots, X^N$ be complex Hilbert spaces whose scalar products (linear in the first arguments) and norms we denote by $\langle \cdot | \cdot \rangle$ and $|\cdot|$, and A be a functional on $X = X^1 \oplus \cdots \oplus X^N$.[10]

Definition 1: *If for a fixed vector* $(\psi^1, \ldots, \psi^N) \in D(A)$ *there exist two vectors of* X^i, ${}^lA'_{\psi^i}$ *and* ${}^sA'_{\psi^i}$, *such that for every* $\phi^i \in X'^i$ (*linear manifold dense in* X^i)

$$\lim_{\alpha \to o} \frac{A(\psi^1, \ldots, \psi^i + \alpha\phi^i, \ldots, \psi^N) - A(\psi^1, \ldots, \psi^N)}{\alpha} = \langle \phi^i | {}^lA'_{\psi^i} \rangle + \langle {}^sA'_{\psi^i} | \phi^i \rangle, \tag{1}$$

then A *is said to be weakly partially semidifferentiable at* $(\psi^1, \ldots, \psi^N)$ *and* ${}^lA'_{\psi^i}$ ${}^sA'_{\psi^i}$ *are called respectively the weak partial linear gradient and the weak partial semilinear gradient of* A *at* $(\psi^1, \ldots, \psi^N)$ *with respect to* ψ^i.

We note that α is a real variable and the limit (1) is taken in the usual sense of norm convergence in X^i. However, this convergence may not be uniform with respect to ϕ^i.

Definition 2: *If the convergence in* (1) *is uniform with respect to* $\phi^i \in X'^i$ *on a ball* $|\phi^i| = k$, *then* A *is said to be strongly partially semidifferentiable at* $(\psi^1, \ldots, \psi^N)$ *and* ${}^lA'_{\psi^i}$ *and* ${}^sA'_{\psi^i}$ *are respectively called the strong partial linear gradient and the strong partial semilinear gradient of* A *at* $(\psi^1, \ldots, \psi^N)$ *with respect to* ψ^i. The total weak and strong semidifferentiability of A are defined analogously.[4]

Concerning the definition of semidifferential properties of $E(\Psi)$ on the set of S determinants, we note that such a set is not a linear manifold, therefore our semidifferential calculus cannot be directly applied.

However, we can overcome this difficulty by regarding $E(\Psi)$ not as a functional on $L_a^2(R^{3N};C^{2N})$,[11] considered at any S determinant $\Psi \neq 0 \in D(H)$, but as a functional on the space $L = L^2(R_1^3;C_1^2) \oplus \cdots \oplus L^2(R_N^3;C_N^2)$. In this space each S determinant $\Psi \neq 0$, built up from the linearly independent one-particle functions $\psi^1, \ldots, \psi^N$, is represented by the vector $(\psi^1, \ldots, \psi^N)$ and the domain of $E(\Psi)$ by some set $D(E)$ of these vectors. From now on our functional will always be regarded as a functional on L, but we shall continue to denote it as $E(\Psi)$.

In this section we define weak semidifferential properties only, so we need no particular hypothesis on H except the very general one that $D(H)$ is a set dense in $L_a^2(R^{3N};C^{2N})$.

Proposition 1: *The functional* $E(\Psi)$ *is, at any* $(\psi^1, \ldots, \psi^N) \in D(E)$, *weakly partially semidifferentiable with respect to* ψ^i, $i = 1,2,\ldots N$.

Proof: Let $(\psi^1, \ldots, \psi^N)$ be fixed. For any ϕ^i belonging to a set dense in $L^2(R^3;C^2)$ we have

$$\lim_{\alpha \to 0} \frac{E(\Psi[\psi^i + \alpha\phi^i]) - E(\Psi)}{\alpha} = \frac{1}{|\Psi|^2} \langle (H - E(\Psi))\Psi | \Psi[\phi^i] \rangle + \frac{1}{|\Psi|^2} \langle \Psi[\phi^i] | (H - E(\Psi))\Psi \rangle + \lim_{\alpha \to 0} \alpha \frac{\langle \Psi[\phi^i] | H\Psi[\phi^i] \rangle - E(\Psi) |\Psi[\phi^i]|^2}{|\Psi|^2 + \alpha\langle \Psi | \Psi[\phi^i] \rangle + \alpha\langle \Psi[\phi^i] | \Psi \rangle + \alpha^2 |\Psi[\phi^i]|^2} \tag{2}$$

Since the first and second term on the right-hand side can be written as scalar products in $L^2(R^3;C^2)$, as we shall see below, and

$$\lim_{\alpha \to 0} \alpha \frac{\langle \Psi[\phi^i] | H\Psi[\phi^i] \rangle - E(\Psi) |\Psi[\phi^i]|^2}{|\Psi|^2 + \alpha\langle \Psi | \Psi[\phi^i] \rangle + \alpha\langle \Psi[\phi^i] | \Psi \rangle + \alpha^2 |\Psi[\phi^i]|^2} = 0. \tag{3}$$

the proposition is proved on account of Definition 1.

In order to write $(1/|\Psi|^2)\langle (H - E(\Psi))\Psi | \Psi[\phi^i] \rangle$ and $(1/|\Psi|^2)\langle \Psi[\phi^i] | (H - E(\Psi))\Psi \rangle$ as scalar products in $L^2(R^3;C^2)$, we use the well-known formula for the matrix elements of an operator with respect to two S determinants.[12] In the case of the operator H this formula is written (for notation see Sec. II)

$$\langle \Xi | H\Psi \rangle = \sum_{kl} \langle \chi^k | h\psi^l \rangle D_{\Xi\Psi}[k,l] + \tfrac{1}{2} \sum_{\substack{k_1 k_2 \\ l_1 l_2}} \langle \chi^{k_1} \chi^{k_2} | w\psi^{l_1} \psi^{l_2} \rangle D_{\Xi\Psi}[k_1 k_2, l_1 l_2]. \tag{4}$$

For the sake of brevity we give the expression of our scalar products only in the case $\langle\psi^i|\psi^j\rangle = \delta_{ij}$, $i,j = 1,...N$. So for $(1/|\Psi|^2)\langle(H - E(\Psi))\,\Psi\,|\Psi\,[\phi^i]\rangle$ we have

$$\langle h^{\mathrm{HF}}\psi^i|\phi^i\rangle - \sum_{r=1}^{N}\langle h^{\mathrm{HF}}\psi^i|\psi^r\rangle\,\langle\psi^r|\phi^i\rangle\,, \tag{5}$$

and for $(1/|\Psi|^2)\langle\Psi\,[\phi^i]|(H - E(\Psi))\,\Psi\rangle$,

$$\langle\phi^i|h^{\mathrm{HF}}\psi^i\rangle - \sum_{r=1}^{N}\langle\psi^r|h^{\mathrm{HF}}\psi^i\rangle\,\langle\phi^i|\psi^r\rangle\,. \tag{6}$$

Therefore the vector of $\mathsf{L}^2(R^3;C^2)$,

$$E'_{\psi^i} = h^{\mathrm{HF}}\psi^i - \sum_{r=1}^{N}\langle h^{\mathrm{HF}}\psi^i|\psi^r\rangle\psi^r\,, \tag{7}$$

is by Definition 1 both the weak partial linear gradient and the weak partial semilinear gradient of $E(\Psi)$ at $(\psi^1,...,\psi^N)$ with respect to ψ^i, with $\langle\psi^i|\psi^j\rangle = \delta_{ij}$, $i,j = 1,...,N$. Because of the coincidence of these weak gradients,[13] from now on we shall speak only of weak partial gradients of $E(\Psi)$. Exactly as above, we could show that $E(\Psi)$ is weakly totally semidifferentiable in $\mathsf{D}(E)$ and the total weak gradient of it at $(\psi^1,...,\psi^N)\in\mathsf{D}(E)$, with $\langle\psi^i|\psi^j\rangle = \delta_{ij}$, is the vector of L

$$E' = (E'_{\psi^1}, E'_{\psi^2},..., E'_{\psi^N})\,. \tag{8}$$

Proposition 2: *The necessary condition of extremum for* $E(\Psi)$ *at* $(\psi^1,...,\psi^N)\in\mathsf{D}(E)$ *with* $\langle\psi^i|\psi^j\rangle = \delta_{ij}$, $i,j = 1,...,N$, *gives the* HF *equations.*

Proof: By virtue of Proposition 1, for (7) and (8) this extremum condition is given[4] by the equations (Euler equations)

$$h^{\mathrm{HF}}\psi^i - \sum_{r=1}^{N}\langle h^{\mathrm{HF}}\psi^i|\psi^r\rangle\psi^r = 0, \quad i = 1,...,N\,. \tag{9}$$

These are not yet the HF equations because of the off-diagonal matrix elements $\langle h^{\mathrm{HF}}\psi^i|\psi^r\rangle$ are, in general, different from zero. However, if we carry out a unitary transformation

$$\psi'^k = \sum_{i=1}^{N} c_i^k\,\psi^i\,, \quad k = 1,...,N\,, \tag{10}$$

the extremum condition (9) becomes

$$h'^{\mathrm{HF}}\psi'^i - \sum_{r=1}^{N}\langle h'^{\mathrm{HF}}\psi'^i|\psi'^r\rangle\psi'^r = 0\,,$$

where

$$h'^{\mathrm{HF}} = h + \sum_{s=1}^{N}\langle\cdot\psi'^s|\tilde{w}\psi'^s\cdot\rangle\,.$$

Since

$$\langle h'^{\mathrm{HF}}\psi'^i|\psi'^r\rangle = \sum_{q,p=1}^{N} c_q^i\,\langle h^{\mathrm{HF}}\psi^q|\psi^p\rangle c_p\,,$$

it is sufficient to choose a unitary transformation (10) diagonalizing the $N\times N$ matrix $\langle h^{\mathrm{HF}}\psi^q|\psi^p\rangle$ so that the HF equations

$$E'_{\psi'^i} = h'^{\mathrm{HF}}\psi'^i - \langle h'^{\mathrm{HF}}\psi'^i|\psi'^i\rangle\psi'^i = 0\,, \quad i = 1,...,N\,, \tag{11}$$

are obtained.

We note that the equations (11) are derived in quantum mechanics mostly by means of the Lagrange-multiplier rule. However, this rule, as we noted in Ref. 4, is applied disregarding the appropriate mathematical formalism. Here, on the contrary, as was suggested in Ref. 14, the HF equations have been derived with a rigorous, and at the same time simple procedure. In fact, we have not imposed the orthonormalization conditions $\langle\psi^i|\psi^j\rangle = \delta_{ij}$, as constraints, but we have merely written, in the appropriate mathematical formalism, the necessary condition of free extremum for $E(\Psi)$ on L at the special vector $(\psi^1,...,\psi^N)$ built up of orthonormal one-particle functions.

IV. STRONG SEMIDIFFERENTIABILITY OF $E(\Psi)$

Proposition 3: *In* L *the functional* $E(\Psi)$ *is not strongly partially semidifferentiable with respect to* ψ^i, $i = 1,...,N$, *at any* $(\psi^1,...,\psi^N)\in\mathsf{D}(E)$.

Proof: On account of Definition 2, it will be sufficient to show that the relationship of convergence (3) is not uniform with respect to all ϕ^i on a ball $|\phi^i| = k$. In fact, for any sufficiently small $\epsilon > 0$ we have

$$\left|\alpha\,\frac{\langle\Psi\,[\phi^i]|H\,\Psi\,[\phi^i]\rangle - E(\Psi)|\Psi\,[\phi^i]|^2}{|\Psi|^2 + \alpha\langle\Psi|\Psi\,[\phi^i]\rangle + \alpha\langle\Psi\,[\phi^i]|\Psi\rangle + \alpha^2|\Psi\,[\phi^i]|^2}\right| < \epsilon$$

for $|\alpha| < \delta(\phi^i)$, with

$$\delta(\phi^i) = \frac{\epsilon|\Psi|^2}{|\langle\Psi\,[\phi^i]|H\,[\phi^i]\rangle - E(\Psi)|\Psi\,[\phi^i]|^2|}\,;$$

but, owing to the unboundedness of H, in general

$$\operatorname*{Inf}_{|\phi^i|=k}\delta(\phi^i) = 0$$

for any k.

This proposition, together with Proposition 1, states that $E(\Psi)$ has only the weak semidifferentiability property in L. Such a property, though useful for deriving rigorously necessary conditions for extrema, is nevertheless inadequate to extend iterative methods of the optimization theory to our extremum problem. For this purpose the strong semidifferentiability will turn out to be more useful.

In this section we shall obtain for some classes of one-body and two-body potentials such a semidifferentiability property by introducing a new topology. The classes of potential which we shall consider are:

(a) $v\in\mathsf{B}(\mathsf{L}^2(R^3;C^2))$,

(b) v t-bounded[15] with t-bound smaller than 1,

(c) $h = t + v\in\mathsf{C}(\mathsf{L}^2(R^3;C^2))$ for one-body potentials, and

(a') $w\in\mathsf{B}(\mathsf{L}^2(R^3;C^2)\otimes\mathsf{L}^2(R^3;C^2))$,

(b') w $(t+t)$-bounded with $(t+t)$-bound smaller than 1,

(c') w $(h+h)$-bounded with $(h+h)$-bound smaller than 1 for two-body potentials. In each class the operators are supposed to be symmetric.

We think that the above classes are sufficiently general to include a lot of potentials currently used in nuclear or atomic physics. For instance, if v is a local potential defined by a bounded function, or a non local one satisfying the inequalities $\int|\langle\mathbf{r}|v\,\mathbf{r}'\rangle|d\mathbf{r}'\leqslant l$ and $\int|\langle\mathbf{r}|v\,\mathbf{r}'\rangle|d\mathbf{r}\leqslant l$ (l is a constant), then[16] it belongs to the class a); an analogous consideration holds for w. If v or w are defined by locally square-integrable functions bounded at infinity (Coulombic poten-

tials or potentials of the form r^{-m} with $m < \frac{3}{2}$), then[17-19] they belong respectively to the class b) and b'). For more potentials included in our classes we refer the reader to Ref. 19. However, we remark that for all these potentials every $(\psi^1,...,\psi^N) \in \mathsf{D}(E)$ is built up of linearly independent one-particle functions $\psi^i \in \mathsf{D}(h)$, $i = 1,...,N$. Therefore $\mathsf{D}(E) \subset \mathsf{D}(h^1) \oplus ... \oplus \mathsf{D}(h^N)$. The following proposition establishes how to choose a new topology on $\mathsf{D}(h^1) \oplus ... \oplus \mathsf{D}(h^N)$.

Proposition 4: *Let c be an operator belonging to* $\mathsf{C}(\mathsf{L}^2(R^3;C^2))$ with $\mathsf{D}(c) = \mathsf{D}(h)$. *Set,* $\forall \phi, \psi \in \mathsf{D}(h)$ *and* $0 < \eta < \infty$,

$$(\phi|\psi)_c = \langle \phi|\psi\rangle + \eta^{-2}\langle c\phi|c\psi\rangle , \tag{12}$$

$$\|\psi\|_c^2 = (\psi|\psi)_c , \tag{13}$$

then the expression (12) *defines a new scalar product on* $\mathsf{D}(h)$ *and* $\mathsf{D}(h)$, *in the topology of the new norm* (13), *becomes a Hilbert space.*

From now on we shall denote this space by z_c and the space $z_c^1 \oplus ... \oplus z_c^N$ by Z_c.

We omit the proof of this proposition because the proof of Proposition 3.3 of Ref. 4 holds equally well here, provided that the operators H and C, the space $\mathsf{L}^2(R^{3N})$, and the N-particle wavefunctions Φ and Ψ are replaced respectively by h, c, $\mathsf{L}^2(R^3;C^2)$, ϕ, and ψ. For the same reason we omit the proof of two next propositions.

Let us consider the operator $(I + \eta^{-2} c^*c)$, where c^* is the adjoint operator to c and I the identity operator in $\mathsf{L}^2(R^3;C^2)$, then we have[4,17]

Proposition 5: *The operator* (*metric operator*)

$$r_c = (I + \eta^{-2} c^*c)^{-1} \tag{14}$$

exists, is defined on the whole of $\mathsf{L}^2(R^3;C^2)$, *has as range* $\mathsf{D}(c^*c) \subset \mathsf{D}(h)$, *belongs to* $\mathsf{B}(\mathsf{L}^2(R^3;C^2))$ *with* $\|r_c\|_{\mathsf{L}^2} \leqslant 1$, *and is the unique operator with the property*

$$(r_c\psi|\phi)_c = \langle \psi|\phi\rangle, \forall\, \psi \in \mathsf{L}^2(R^3;C^2), \forall \phi \in \mathsf{D}(h) . \tag{15}$$

Proposition 6: *For any choice of the operator c and the number* η *in Proposition* 4, *every operator* $A \in \mathsf{C}(\mathsf{L}^2(R^3;C^2))$, *with* $\mathsf{D}(A) = z_c$, *is bounded if regarded as an operator from* z_c *to* $\mathsf{L}^2(R^3;C^2)$. *In other words,*

$$\operatorname*{Sup}_{0 \neq \phi \in z_c} \frac{|A\phi|}{\|\phi\|_c} = \|A\|_{z_c} < \infty .$$

For the self-adjointness of t,[2,19] and the properties of our potentials, some choices of the operator c satisfying the condition of Proposition 4 are: $c = t$, h, h^{HF} for potentials of the classes a), a'), and b), a')[20]; $c = t$ for potentials of the classes b), b'); $c = h$ for potentials of the classes c), a'), and c), c'), etc. On account of Propositions 4 and 6, we have, at least for the above choices of c,

Proposition 7: *In* Z_c *the functional* $E(\Psi)$ *is strongly partially semidifferentiable with respect to* ψ^i, $i = 1,...,N$, *at any* $(\psi^1,...,\psi^N) \in \mathsf{D}(E)$.

Proof: $E(\Psi)$ regarded as a functional on z_c is defined, as it is easily seen, in a neighborhood of ψ^i $I = \{\forall \phi^i : \|\phi^i\|_c \leqslant k$, with k sufficiently small$\}$. Moreover, as we shall show in Appendix A, since it turns out that

$$\operatorname*{Inf}_{\|\phi^i\|_c = k} \delta(\phi^i) \geqslant \frac{\epsilon|\Psi|^2}{k(p^i + |E(\Psi)|q^i)} ,$$

p^i and q^i constants,

the convergence relationship (3) is uniform with respect to all ϕ^i with $\|\phi^i\|_c = k$. So the proposition is proved.

By Definitions 1 and 2 we get the strong partial linear and the strong partial semilinear gradient of $E(\Psi)$ by writing the first and the second term of the right-hand side of (2) as scalar products in z_c. In the case of $\langle \psi^i|\psi^j\rangle = \delta_{ij}$, $i,j = 1,...,N$, these scalar products, on account of the expressions (5) and (6) and the property (15) of r_c, are written respectively

$$(r_c h^{HF}\psi^i|\phi^i)_c - \sum_{r=1}^{N} \langle h^{HF}\psi^i|\psi^r\rangle (r_c\psi^r|\phi^i)_c ,$$

and

$$(\phi^i|r_c h^{HF}\psi^i)_c - \sum_{r=1}^{N} \langle \psi^r|h^{HF}\psi^i\rangle (\phi^i|r_c\psi^r)_c .$$

The vector of z_c,

$$E'_{c\psi^i} = r_c\left(h^{HF}\psi^i - \sum_{r=1}^{N} \langle h^{HF}\psi^i|\psi^r\rangle \psi^r\right), \tag{16}$$

is therefore both the strong partial linear and the strong partial semilinear gradient of $E(\Psi)$ with respect to ψ^i, $i = 1,...,N$, at $(\psi^1,...,\psi^N)$, with $\langle \psi^i|\psi^j\rangle = \delta_{ij}$, $i,j = 1,...,N$. Because of the coincidence of these strong gradients[21], from now on we shall speak only of strong partial gradients of $E(\Psi)$.

Proposition 8: *The mapping* $(\psi^1,...,\psi^N) \to$ *strong partial gradient of* $E(\Psi)$, *from* Z_c *to* z_c, *is strongly continuous in* Z_c, *at any* $(\psi^1,...,\psi^N) \in \mathsf{D}(E)$.

Proof: The expression of $E'_{c\psi^i}$ at any $(\psi^1,...,\psi^N) \in \mathsf{D}(E)$ is given by a linear combination of vectors of z_c of the following kinds[22]:

d) $r_c\langle \psi^{k_1}|h\psi^{l_1}\rangle \psi^r$,
e) $r_c\langle \psi^{k_1}\psi^{k_2}|w\,\psi^{l_1}\psi^{l_2}\rangle \psi$,
f) $r_c\langle \cdot\psi^{k_2}|w\,\psi^{l_1}\psi^{l_2}\rangle$,
g) $r_c h\psi^r$,
h) $E(\Psi) r_c\psi^r$,

being $k_1,k_2 = 1,...i-1,i+1,...,N$, $r,l_1,l_2 = 1,...,N$, and $r \neq l_1, r \neq l_2$. This can be seen by taking into account the expression of the first and second terms on the right-hand side of (2), together with (4) and (15).

Now, by using the Schwarz inequality, the properties of our potentials (boundedness, relative boundedness, etc.), Proposition 6, the inequality $\|\cdot\|_c \geqslant |\cdot|$, and $\|r_c\|_{\mathsf{L}^2} \leqslant 1$, it can be easily shown that the mapping from Z_c to any vector of the kinds d), e), f), g), and h) is strongly continuous in Z_c at any $(\psi^1,...,\psi^N) \in \mathsf{D}(E)$.

V. GRADIENT METHOD

The gradient method is well known both as a minimization[23] technique and as a technique to solve nonlinear equations.[24] In this section we consider a gradient method that generalizes the usual one because it is defined by means of the strong gradient of the previous sections and thus depends on the metric operator r_c. Here we use this method to lower at each step the value of $E(\Psi)$ and at the same time solve

iteratively the HF equations.[25] The procedure of its application is the following:

(1) Carry out a choice for the operator c and the number η of Proposition 4.

(2) Compute at an initial guess $(\psi_0^1,...,\psi_0^N)\in D(E)$, with $\langle\psi_0^i|\psi_0^j\rangle=\delta_{ij}$, i.e., at the corresponding S determinant Ψ_0, $E(\Psi_0)=E_0$ and the strong partial gradients (see notation) $E'_{c\psi_0^i}$, $i=1,...,N$.

(3) Determine a new vector $(\tilde\psi_1^1,...,\tilde\psi_1^N)$ by means of the iterative formula

$$\tilde\psi_{n+1}^i=\psi_n^i-\alpha_n^i E'_{c\psi_n^i},\quad i=1,...,N,\ n=0,1,2,\cdots, \tag{17}$$

where $\alpha_n^1,...\alpha_n^N$ are real numbers to be specified below.

(4) Orthonormalize the one-particle functions $\tilde\psi_1^i$ and let ψ_1^i, $i=1,...,N$, be the orthonormalized ones.

(5) Compute at $(\psi_1^1,...,\psi_1^N)$, i.e., at the corresponding S determinant Ψ_1, $E(\Psi_1)=E_1$ and $E'_{c\psi_1^i}$, $i=1,...,N$, and so on. Concerning a suitable choice of the numbers $\alpha_n^1,\alpha_n^2,...,\alpha_n^N$, we have

Theorem: *Values* α_n^{-i}, $i=1,...,N$, *belonging to the domain* $\mathcal{D}=\{\alpha_n^i:|\alpha_n^i|\leqslant k^i, k^i$ *arbitrary constants*, $i=1,...,N\}$ *exist such that*

(1) $E_n>E_{n+1}$,

(2) $\lim_{n\to\infty}\|E'_{c\psi_n^i}\|_c=0$, $\quad i=1,...,N$.

Proof: Since the total weak gradient of $E(\Psi)$ exists [see (8)] at any $(\psi^1,...,\psi^N)\in D(E)$, we can write, by choosing in (17) $\alpha_n^1=\alpha_n^2=...\alpha_n^N=\alpha$,

$$E_{n+1}-E_n=-\alpha\sum_{i=1}^N\Big\langle E'_{c\psi_n^i}\Big|\Big(h_n^{\mathrm{HF}}\psi_n^i-\sum_{r=1}^N\langle\psi_n^i|h_n^{\mathrm{HF}}\psi_n^r\rangle\psi_n^r\Big)+\text{c.c.}+\epsilon.$$

This equation, on account of the property (15) of the metric operator r_c and of the expression for the strong partial gradient $E'_{c\psi_n^i}$ (see notation), becomes

$$E_{n+1}-E_n=-2\alpha\sum_{i=1}^N\|E'_{c\psi_n^i}\|_c^2+\epsilon,$$

where

$$\lim_{\alpha\to 0}\frac{\epsilon}{\alpha}=0.$$

Therefore, if α is a sufficiently small positive number,

$$E_{n+1}-E_n<0$$

and point 1) of the theorem holds for the particular choice $\alpha_n^1=\alpha_n^2\cdots\alpha_n^N=\alpha$.

However, better values of these variables can be obtained by a reasonable approximation of

$$\operatorname*{Inf}_{(\alpha_n^1,...,\alpha_n^N)\in\mathcal{D}}(E_{n+1}-E_n)$$

by means of a suitable minimization algorithm.[26] Let us now consider the value $E(\Psi_n[\psi_{n+1}^i])=E_{n+1}^i$ and the function of α_n^i

$$F(\alpha_n^i)=E_{n+1}^i-E_n=\big[(\alpha_n^i)^2(E(\Psi_n[E'_{c\psi_n^i}])-E_n)|\Psi_n[E'_{c\psi_n^i}]|^2-2\alpha_n^i\|E'_{c\psi_n^i}\|_c^2\big]/[|\Psi[\psi_{n+1}^i]|^2],$$

where ψ_{n+1}^i is determined according to (17), with α_n^i satisfying the inequalities

$$\delta\leqslant\alpha_n^i\leqslant\min\{k^i,|\sigma_n^i|-\delta\}, \tag{18}$$

being $\delta<k^i$, $0<\delta<\frac{1}{2}|\sigma_n^i|$ and

$$|\sigma_n^i|=\frac{2\|E'_{c\psi_n^i}\|_c^2}{|\langle\Psi[E'_{c\psi_n^i}]|H\Psi[E'_{c\psi_n^i}]\rangle-E_n|\Psi[E'_{c\psi_n^i}]|^2|}. \tag{19}$$

As shown in Ref. 27, $F(\alpha_n^i)$ is negative in the interval defined by (18) and hence

$$E_n>E_{n+1}^i. \tag{20}$$

Now there exist value $\bar\alpha_n^i$ of the variables α_n^i, $i=1,...,N$, belonging to $\mathcal{D}$ such that

$$E_{n+1}^i>E_{n+1}. \tag{21}$$

In fact, this inequality is surely satisfied if $\alpha_n^1,\alpha_n^2,...,\alpha_n^N$ are taken to minimize exactly on $\mathcal{D}$, $E_{n+1}-E_n$, when the

$$\operatorname*{Inf}_{(\alpha_n^1,...\alpha_n^N)\in\mathcal{D}}(E_{n+1}-E_n)$$

is attained on $\mathcal{D}$ (steepest-descent method), or otherwise, if they are taken to give the best possible approximation to

$$\operatorname*{Inf}_{(\alpha_n^1,...\alpha_n^N)\in\mathcal{D}}(E_{n+1}-E_n).$$

In any case, such values, $\bar\alpha_n^1,\bar\alpha_n^2,...,\bar\alpha_n^N$ can be identified in practice with a good approximation to

$$\operatorname*{Inf}_{(\alpha_n^1,...\alpha_n^N)\in\mathcal{D}}(E_{n+1}-E_n).$$

For these values of the variables α_n^i, both point 1) of the theorem, as is immediately seen by (20) and (21), and point 2), as we show now, hold.

In fact, the sequences $E_0, E_1^i, E_1,...,E_n, E_{n+1}^i, E_{n+1},...,$ $i=1,...,N$, on account of (20) and (21), are decreasing; moreover, owing to the boundedness from below of H, they are bounded from below and are therefore convergent. Thus

$$\lim_{n\to\infty}E_{n+1}^i-E_n=\lim_{n\to\infty}F(\alpha_n^i)=0,\quad i=1,...,N. \tag{22}$$

Now since $|\sigma_n^i|\geqslant d>0$, as we shall prove in Appendix B, we can take in (18) $0<\delta<\frac{1}{2}d$. With this choice of δ and by using the inequality $|\Psi_n[E'_{c,\psi_n^i}]|^2\leqslant N\|E'_{c\psi_n^i}\|_c^2$ (see Appendix B), it can be shown, exactly as in Ref. 27, that

$$F(\alpha_n^i)>\frac{\delta^2}{k^i+\delta}\frac{\|E'_{c\psi_n^i}\|_c^2}{1+k^{i2}N\|E'_{c\psi_n^i}\|_c^2}.$$

This inequality gives by (22)

$$\lim_{n\to\infty}\|E'_{c\psi_n^i}\|_c^2=0,\quad i=1,2,...,N.$$

Thus the theorem is proved.

We remark that the procedure of application of the gradient method which we have considered so far is not the only one by which the theorem can be proved. Another procedure which has this property, although the proof is more complicated, is that of splitting up a step of gradient methods into N

"sub-steps," where one single-particle wave function at a time is changed according to (17), with α_n^i satisfying (18) and $0 < \delta < \frac{1}{2}d$.[28]

The equations $E'_{\psi_n^i} = 0$ and $E'_{c\psi_n^i} = 0$, $i = 1,...,N$, have the same solutions because the metric operator r_c has the inverse $(I + \eta^{-2} c^* c)$. Therefore, in virtue of point 2) of the theorem, we can regard $(\psi_n^1,...,\psi_n^N)$ or $(\psi_n'^1,...,\psi_n'^N)$ [see (10)], if the off-diagonal elements $\langle h^{\mathrm{HF}}\psi_n^i | \psi_n^r \rangle$ are different from zero, as an approximate solution of the HF equations. The error for this approximation could be estimated by methods described in Ref. 29 (for instance, see p. 192). We note that point 2) of the theorem does not imply the convergence of the sequence $\{\psi_n^1,...,\psi_n^N\}$, $n = 0,1,2,...$. Actually, we have not been able to give sufficient conditions for this convergence. However, we have

Proposition 9: *If* $\lim_{n\to\infty} (\psi_n^1,...,\psi_n^N) = (\bar{\psi}^1,...,\bar{\psi}^N)$ *(strong limit in* Z_c*), then* $(\bar{\psi}^1,...,\bar{\psi}^N)$ *or* $(\bar{\psi}'^1,...,\bar{\psi}'^N)$ *[see* (10)*] solves the* HF *equations.*

Proof: By Proposition 8, the existence of the inverse of the metric operator r_c, and point 2) of the theorem, we have

$$\lim_{n\to\infty} E'_{c\psi_n^i} = E'_{c\bar{\psi}^i} = E'_{\bar{\psi}^i} = 0 .$$

Thus, if the off-diagonal elements $\langle h^{\mathrm{HF}}\bar{\psi}^i | \bar{\psi}^r \rangle$ are zero, $(\bar{\psi}^1,...,\bar{\psi}^N)$ solves the HF equations [see Ref. 11]; otherwise, a solution $(\bar{\psi}'^1,...,\bar{\psi}'^N)$ is obtained by means of the unitary transformation (10), which diagonalizes the $N \times N$ matrix $\langle h^{\mathrm{HF}}\bar{\psi}^q | \bar{\psi}^p \rangle$. So the proposition is proved. If we restrict the space $\mathsf{L}^2(R^3;C^2)$ to a finite-dimensional subspace M, we have, moreover, (call P_M the projection operator onto M)

Proposition 10: *The sequence* $\{(\mathsf{P}_\mathsf{M}\psi_n^1,...,\mathsf{P}_\mathsf{M}\psi_n^N)\}$, $n = 0,1,2,...$, *contains a subsequence* $\{(\mathsf{P}_\mathsf{M}\psi_{n'}^1,...,\mathsf{P}_\mathsf{M}\psi_{n'}^N)\}$, $n' = 0,1,2,...$, *such that* $\lim_{n'\to\infty} (\mathsf{P}_\mathsf{M}\psi_{n'}^1,...,\mathsf{P}_\mathsf{M}\psi_{n'}^N) = (\bar{\psi}_\mathsf{M}^1,...,\bar{\psi}_\mathsf{M}^N)$ *(strong limit in the topology of the norm of* Z_c*). The vector* $(\bar{\psi}_\mathsf{M}^1,...,\bar{\psi}_\mathsf{M}^N)$ *or the vector* $(\bar{\psi}_\mathsf{M}'^1,...,\bar{\psi}_\mathsf{M}'^N)$ *[see* (10)*] solves the* HF *equations restricted to the sub-space* $\mathsf{M}^1 \oplus \mathsf{M}^2 \oplus \cdots \mathsf{M}^N$.

Proof: Since at any n, $\langle \psi_n^i | \psi_n^j \rangle = \delta_{ij}$, $i,j = 1,...,N$, the sequence $\{(\mathsf{P}_\mathsf{M}\psi_n^1,...,\mathsf{P}_\mathsf{M}\psi_n^N)\}$, $n = 0,1,2,...$, is bounded in the topology of the norm of L, thus it contains[17] a subsequence $\{\mathsf{P}_\mathsf{M}\psi_{n'}^1,...,\mathsf{P}_\mathsf{M}\psi_{n'}^N)\}$, $n' = 0,1,...$, weakly convergent in $\mathsf{M}^1 \oplus \mathsf{M}^2 \cdots \oplus \mathsf{M}^N$. Now, in a finite dimensional space, weak convergence implies strong convergence and any two norms are equivalent.[17] Hence the sequence $\{(\mathsf{P}_\mathsf{M}\psi_{n'}^1,...,\mathsf{P}_\mathsf{M}\psi_{n'}^N)\}$, $n' = 0,1,2,...$, is strongly convergent both in the topology of the norm of L and in the topology of the norm of Z_c. Therefore, since Proposition 9 holds as well in $\mathsf{M}^1 \oplus \mathsf{M}^2 \oplus \cdots \oplus \mathsf{M}^N$, the proof is complete.

VI. CONCLUSIVE REMARKS

The gradient method presented in the previous section is actually a family of gradient methods depending on the choices of number η and operator c in the definition (12) of the scalar product. For the choice $c = I$, i.e., no change of the usual topology of $\mathsf{L}^2(R^3;C^2)$, $r_c \to I$, $E'_{c\psi_n^i} \to E'_{\psi_n^i}$ and thus (17) becomes

$$\tilde{\psi}_{n+1}^i = \psi_n^i - \alpha_n^i E'_{\psi_n^i}, \quad i = 1,...,N. \tag{23}$$

The gradient method defined by this formula is basically the method of Refs. 7 and 8, however, it has two deficiencies:

(1) Because the operator h has as range the whole space $\mathsf{L}^2(R^3;C^2)$, it may happen that even if ψ_n^i belongs to $\mathsf{D}(h)$, $\tilde{\psi}_{n+1}^i$ does not. In this case the sequence defined by (23) goes out of $\mathsf{D}(E)$ at the step n and the iterative process cannot work any more.

(2) In general, point 2) of the theorem does not hold.[30] We have these two unpleasant features because the method defined by (23) disregards the unboundedness of operators t, v, or w, and in practice can work only in finite-dimensional subspaces. On the contrary, our gradient method takes into account, by means of a metric operator $r_c \neq I$, the unboundedness of t and several kinds of unboundedness of v and w. Thus it does not share with the method defined by (23) the disadvantages 1) [r_c has as range $\mathsf{D}(c^*c) \subset \mathsf{D}(h)$] and 2), and it can work equally well in the whole space and in a finite-dimensional subspace.

However, as is widely discussed in Ref. 31, the pure and simple method (23), translated into the formalism of the density matrix and applied in finite dimensional subspaces to nuclear systems, is much faster and computationally simpler than the usual procedure for solving the HF equations. Now some applications[32] of our gradient method to exactly solvable problems, in the case of the total wave function Ψ not restricted to be a S determinant, clearly indicate that the metric operator greatly increases the convergence speed. Therefore the same advantageous effect of the operator r_c is also expected in the case of this paper where Ψ is restricted to be a S determinant.

So the introduction of a new topology, that is to say, of the metric operator r_c, has both theoretical and probably practical advantages.

The calculation of r_c, necessary for applying our gradient method, is in general a hard job, but for the choice $c = t$ this operator has the explicit form[33] (in the coordinate representation)

$$\langle \mathbf{y} | r_t | \mathbf{x} \rangle = \frac{2m\eta}{4\pi|\mathbf{x}-\mathbf{y}|} \exp\left(- \frac{\sqrt{2m\eta}}{2^{1/2}} |\mathbf{x}-\mathbf{y}| \right) \times \sin\left(\frac{\sqrt{2m\eta}}{2^{1/2}} |\mathbf{x}-\mathbf{y}| \right),$$

where m is the mass of the particle.

APPENDIX A

By the Schwarz inequality in $\mathsf{L}^2(R^3;C^2)$, the properties of our potentials, the inequality $\|\cdot\|_c \geq |\cdot|$, and Proposition 6, we get from expression (4):

$$|\langle \Psi[\phi^i] | H\Psi[\phi^i] \rangle| \leq \sum_{kl} p_{kl}^i \|\phi^i\|_c^2 + \tfrac{1}{2} \sum_{\substack{k_1 k_2 \\ l_1 l_2}} p_{k_1,k_2,l_1,l_2}^i \|\phi^i\|_c^2 . \tag{A1}$$

The coefficients p_{kl}^i and p_{k_1,k_2,l_1,l_2}^i depend on the potentials v and w, as well as on the one-particle functions. For instance, $p_{ii}^i = \|h\|_{z_c} |\psi^1|^2 \cdots |\psi^{i-1}|^2 |\psi^{i+1}|^2 \cdots |\psi^N|^2 (N-1)!$ if v belongs to the classes b) or c), and

$$p_{ji,ji}^i = \sqrt{2a}\, \|\psi^j\|_c^2 |\psi^1|^2 \cdots |\psi^{i-1}|^2 |\psi^{i+1}|^2 \cdots |\psi^{j-1}|^2 |\psi^{j+1}|^2 \cdots |\psi^N|^2 (N-2)!$$

(for the definition of the constant α see Ref. 15) if w belongs to the classes b') or c').
Similar expressions are found for the other coefficients.
Analogously, by the Schwarz inequality and $\|\cdot\|_c \geqslant |\cdot|$, we have

$$|\Psi[\phi^i]|^2 \leqslant \sum_k q_k^i \|\phi^i\|_c^2 , \tag{A.2}$$

where the coefficients q_k^i depend only on the one-particle functions. Now by (A1) and (A2) and by putting

$$p^i = \sum_{kl} p_{kl}^i + \tfrac{1}{2} \sum_{\substack{k_1,k_2 \\ l_1,l_2}} p_{k_1,k_2,l_1,l_2}^i \quad \text{and} \quad q^i = \sum_k q_k^i ,$$

we get

$$\begin{aligned}\delta(\phi^i) &= \frac{\epsilon|\Psi|^2}{|\langle\Psi[\phi^i]|H\Psi[\phi^i]\rangle - E(\Psi)|\Psi[\phi^i]|^2|} \\ &\geqslant \frac{\epsilon|\Psi|^2}{|\langle\Psi[\phi^i]|H\Psi[\phi^i]\rangle| + |E(\Psi)||\Psi[\phi^i]|^2} \\ &\geqslant \frac{\epsilon|\Psi|^2}{(p^i + |E(\Psi)|q^i)\|\phi^i\|_c^2}\end{aligned}$$

and hence

$$\inf_{\|\phi^i\|_c = k} \delta(\phi^i) \geqslant \frac{\epsilon|\Psi|^2}{(p^i + |E(\Psi)|q^i)k} .$$

APPENDIX B

By (A1) we obtain

$$\begin{aligned}&|\langle\Psi_n[E'_{c\psi_n^i}]|H\Psi_n[E'_{c\psi_n^i}]\rangle| \\ &\quad\leqslant \sum_{kl} \tilde{p}_{kl}\|E'_{c\psi_n^i}\|^2 + \tfrac{1}{2}\sum_{\substack{k_1,k_2 \\ l_1,l_2}} \tilde{p}_{k_1,k_2,l_1,l_2}\|E'_{c\psi_n^i}\|_c^2 .\end{aligned} \tag{B1}$$

Here, however since $\langle\psi_n^i|\psi_n^j\rangle = \delta_{ij}$, $i,j = 1,...,N$, $n = 0,1,2,...$, and our sequence $\{E_n\}$, $n = 0,1,2,...$, is decreasing, the coefficients $\tilde{p}_{kl}^i$ and $\tilde{p}_{k_1,k_2,l_1,l_2}^i$, result in being independent on the one-particle functions, i.e., on the index n. Moreover, by the Schwarz inequality and $\|\cdot\|_c \geqslant |\cdot|$, it can be easily shown that

$$|\Psi_n[E'_{c\psi_n^i}]|^2 \leqslant N\|E'_{c\psi_n^i}\|_c^2 . \tag{B2}$$

Now by (B1) and (B2), we get

$$\begin{aligned}|\sigma_n^i| &= \frac{2\|E'_{c\psi_n^i}\|_c^2}{|\langle\Psi[E'_{c\psi_n^i}]|H\Psi[E'_{c\psi_n^i}]\rangle - E_n|\Psi[E'_{c\psi_n^i}]|^2|} \\ &\geqslant \frac{2\|E'_{c\psi_n^i}\|_c^2}{(\tilde{p} + \lambda_0 N)\|E'_{c\psi_n^i}\|_c^2} = \frac{2}{\tilde{p} + \lambda_0 N} = d > 0 ,\end{aligned} \tag{B3}$$

where we have set

$$\tilde{p} = \sum_{kl} \tilde{p}_{kl} + \tfrac{1}{2}\sum_{\substack{k_1,k_2 \\ l_1,l_2}} \tilde{p}_{k_1,k_2,l_1,l_2}$$

and

$$\lambda_0 = \max\left\{\left|\inf_{\Psi\in D(E)} E(\Psi)\right|, |E_0|\right\} .$$

The value of $\tilde{p}$, if $v\in$ a) and $w\in$ a'), (for the sake of brevity we consider only this case) is

$$\begin{aligned}\tilde{p} &= (2N+1)\|t\|_{z_c} + (N^3 + 3N^2 + 3N)\|v\|_{L^2} \\ &\quad + (N^4 + 3N^2 + 7N^2 + 2N)\|w\|_{L^2} + (N^2 + N)\lambda_0 .\end{aligned}$$

ACKNOWLEDGMENTS

We are grateful to Professor G. Schiffrer for helpful suggestions we thank also Professor M. Di Toro for fruitful discussions.

[1] V. Fock, Z. Phys. **61**, 126 (1930); J.C. Slater, Phys. Rev. **35**, 210 (1930).
[2] E.H. Lieb and B. Simon, Commun. Math. Phys. **53**, 185 (1977).
[3] F. Catara, M. Di Toro, E. Pace, and G. Schiffrer, Nuovo Cimento **11A**, 733 (1972); G. Fonte, R. Mignani, and G. Schiffrer, Commun. Math. Phys. **33**, 293 (1973); M. Di Toro, G. Russo, and F. Duggan, *Alternative Methods for the Solution of Static Hartree-Fock and Related Problems* (University of Catania, Preprint 626).
[4] G. Fonte, Nuovo Cimento **49B**, 200 (1979).
[5] E.K. Blum, *Numerical Analysis and Computation* (Addison-Wesley, New York, 1972) Chap. 12; S.G. Miklin, *Mathematical Physics, an advanced course* (North-Holland, Amsterdam 1970) Chap. 3; M.M. Vainberg, *Variational Methods for the Study of Nonlinear Operators* (Holden-Day, San Francisco, 1964).
[6] E.S. Levitin and B.T. Poliak, Zh. Vychis. Mat. Mat. Fiz. **6**, 787 (1966); E.K. Blum, Ref. 5.
[7] A. Kelemen, J. Math. Phys. **16**, 1396 (1975).
[8] R. McWeeny, Proc. Soc. (London) A **235**, 496 (1975).
[9] In Ref. 8 the formalism of the density matrix is used.
[10] In the sequel of the work the space X will be identified with the space $L = L^2(R_1^3;C_1^2) \oplus \cdots \oplus L^2(R_N^3;C_N^2)$, the functional A with $E(\Psi)$, and $\langle\cdot|\cdot\rangle$ and $|\cdot|$ will denote the usual scalar product (linear in the first argument) and norm both of $L^2(R^3;C^2)$ and of $L_a^2(R^{3N};C^{2N})$.
[11] We consider fermions with two spin states. The general case of m spin states involves only notational changes.
[12] P.O. Löwdin, Phys. Rev. **97**, 1474 (1955).
[13] The two weak gradients coincide even if calculated at a vector $(\psi^1,...,\psi^N)$ with nonorthonormalized one-particle functions.
[14] G. Fonte and G. Schiffrer, Nucl. Phys. A **259**, 20 (1976).
[15] Let X be a Hilbert space with norm denoted by $|\cdot|$ and A and B two operators from X to X densely defined, then (see Refs. 17 and 19) A is said B-bounded if $D(A) \supset D(B)$ and if, for some constants a and b and all $\Psi\in D(B)$,

$$|A\Psi|^2 \leqslant a|\Psi|^2 + b|B\Psi|^2 .$$

The infimum of such b is called the B-bound of A.
[16] V.I Smirnov, *A Course of Higher Mathematics* (Pergamon Press, Oxford, 1964) Vol **5**, p. 509.
[17] T. Kato, *Perturbation Theory for Linear Operators* (Springer-Verlag, Berlin, 1966).
[18] T. Kato, Trans. Am. Math. Soc. **70**, 195 (1951).
[19] M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1975) Vol. II.
[20] We note that by Theorem 4.3, p. 287 of Ref. 17, the operators h and h^{III}, in this case, belong to $C(L^2(R^3;C^2))$.
[21] Reference 13 holds in this case of strong gradients as well.
[22] We recall that the expression (16) holds only when

$$\langle\psi^i|\psi^j\rangle = \delta_{ij} , \quad i,j = 1,...,N .$$

[23] See Ref. 6.
[24] M.Z. Nashed, J. Math. Mech., **13**, 765 (1964).
[25] We consider potentials v and w belonging to the classes a), b), c), a'), b'), and c').
[26] B. Pchénitchny and Y. Daniline, *Méthods Numériques dans les Problémes d'Extrémum* (Editions MIR-Moscou, 1977).
[27] G. Fonte and G. Schiffrer, Nuovo Cimento **37**, 63 (1977).
[28] In this case, however, at the first "sub-step" the strong partial gradient $E'_{c\psi_n^1}$ is calculated at $(\psi_0^1,\psi_0^2,...,\psi_0^N)$, at the second sub-step the strong partial gradient $E'_{c\psi_0^2}$ is calculated at $(\psi_1^1,\psi_0^2,...,\psi_0^N)$, and so on.
[29] M.A. Krasnosel'skiy, *Approximate solutions of operator equations* (Wolters-Noordhoff Publishing, Groningen, 1969).
[30] The inequality (B3), which is essential for the proof of point 2) of the theorem, does not hold in general in the gradient method (23).
[31] M. Di Toro, G. Russo, and F. Duggan, Ref. 3.
[32] M. Lappano, G. Schiffrer, and M. Sgroi, Lett. Nuovo Cimento **20**, 569 (1977).
[33] G. Schiffrer, private communication.

Dynamical symmetries of rotationally invariant, three-dimensional, Schrödinger equations

D. R. Truax[a)]
School of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455

(Received 19 July 1979; accepted for publication 30 November 1979)

All rotationally invariant, time-dependent potentials for the time-dependent, three-dimensional, Schrödinger equation are classified in terms of their dynamical symmetries. A comparison is made to their classification by kinematical algebras. The degeneracy and dynamical algebras are identified for each potential. The oscillator, constant, and Coulomb potentials are shown to be unique in the sense that their degeneracy algebras are all larger than o(3).

I. INTRODUCTION

Our explanations of a physical or chemical system at the microscopic level may require knowledge of the solutions and their properties of an appropriate Schrödinger equation. However, the intricacies of real systems forbid, in general, a complete description of all but the simplest systems. On the other hand, for many cases, the adoption of an efficacious model permits the simulation of a real system in many of its most significant features. For instance, one expects that both the model and modelled system will possess, if only approximately, the same conservation laws. Although any symmetry inherent in a model may be used to characterize states, determine transition probabilities, and in certain cases to solve the Schrödinger equation itself, more often direct methods are employed in the analysis with symmetry emerging after the fact to rationalize apparent degeneracies in the energy. Both the harmonic oscillator[1,2] and the Coulomb problem[1,2] may be cited as examples. The former has been applied to a number of fields including vibrational spectroscopy[2,3] and nuclear physics,[4] while the latter is important in the treatment of the nonrelativistic hydrogen atom,[1,2] with extensions to many electron atoms and molecules.[5]

In a recent work, Boyer[6] derived the maximal group of space–time transformations, called the kinematic group, for the time-dependent, n-dimensional Schrödinger equation with an arbitrary time-independent potential. The generators of these local space–time symmetries (rotations, translations, etc.) form a Lie symmetry algebra, called the kinematic algebra $\mathscr{G}$. A classification of model Hamiltonians was then obtained in terms of their kinematic algebras. Futhermore, Boyer[6] showed that these symmetry algebras for the constant potential, repulsive and attractive oscillators, and linear potential are isomorphic, previously demonstrated only for the constant and harmonic oscillator potentials.[7,8] The kinematic algebra for the constant potential,[6] $(o(3) \oplus sl(2,\mathbb{R}))\square w_3$ is called the Schrödinger algebra.[9]

More recently, Boyer, Sharp, and Winternitz[10] have systematically investigated the symmetry properties of the possibly nonlinear Schrödinger equation $u_{xx} + iu_t = F(x,t,u,\bar{u})$. They found all continuous subalgebras of the Schrödinger algebra and the most general interaction term $F(x,t,u,\bar{u})$ associated with each subalgebra. In each case, the symmetry subalgebra was used to obtain special solutions or separate variables for the resulting Schrödinger equation.

The characterization of the dynamical state of a quantum mechanical system requires the specification of a complete set of commuting observables.[1] If one of the commuting observables is the Hamiltonian, then the remaining compatible observables are constants of the motion which lead to the conservation laws for the system. A possible source of constants of the motion is the kinematic algebra. For rotationally invariant [o(3)-invariant] models such as the three-dimensional Coulomb system or oscillator, the generators of rotations, the components of the angular momentum, are constants of the motion which pertain to conservation of angular momentum for these cases.[1] Now, not all constants of the motion are compatible; the angular momentum operators do not commute with each other but form a basis for an o(3) Lie algebra.[6] Usually, the square of the total angular momentum, which is an invariant (Casimir) operator, and the z component of the angular momentum are chosen as the compatible variables.[11] These operators and the Hamiltonian characterize the dynamical states. The compatible observables form a set of simultaneously diagonizable operators, with their common eigenvectors determining the dynamical states. Hence, a multiplicity in the energy may occur. If $\bar{\mathscr{D}}$ denotes the algebra formed by the constants of the motion, then $\bar{\mathscr{D}}$ is called the degeneracy algebra, and the Hamiltonian is said to have a $\bar{\mathscr{D}}$ degeneracy.[1,11] $\bar{\mathscr{D}}$ is the smallest algebra whose representations completely span the states of individual degenerate levels of the system, and the generators of $\bar{\mathscr{D}}$ permit a laddering among the various states of a degenerate level. The primary problem, then, is the identification of all constants of the motion and hence the degeneracy algebra.

In certain instances, there exist constants of the motion which may not be found in the kinematic algebra. For certain rotationally invariant potentials, the energy multiplicity is greater than that predicted from rotational invariance alone. For instance, the energy degeneracy for the bound states of the Coulomb potentials[1,12–14] is n^2 rather than $2l+1$, the latter expected from conservation of angular momentum only.[15] The extra degeneracy occurs because of the existence of symmetries which are constants of the motion but are not kinematic in origin; they arise from the particular form of the force law and are termed dynamical symme-

[a)]Permanent address: Department of Chemistry, University of Calgary, Calgary, Alberta, Canada, T2N 1N4. Please address all correspondence to this address.

 0022-2488/80/040807-11$1.00

tries.[16] In addition to the generators of rotations, the Coulomb potential admits the three components of the Runge–Lenz vector as constants of the motion,[12-14,16] and for the bound states of the hydrogen atom, $\bar{\mathscr{D}}_c = o(4)$ which has o(3) as a subalgebra[12-14]; the degeneracy algebra $\bar{\mathscr{D}}_0$ for the three-dimensional oscillator is u(3) which also contains o(3) as a subalgebra[12-14]; the dynamical symmetries correspond to components of some quadrupole tensor.[16] The three-dimensional oscillator and Coulomb potentials are the only rotationally invariant potentials known to exhibit this unexpected degeneracy[13]; in this work we shall show that these two cases and the constant potential are the only ones with this property.

In addition to $\bar{\mathscr{D}}$, there may exist an algebra $\mathscr{D}$ which contains $\bar{\mathscr{D}}$ as a subalgebra and which has an irreducible representation which spans the entire set of dynamical states.[12,14] Such an algebra $\mathscr{D}$ is called a dynamical algebra.[12,14] It contains generators which permit a laddering between different degenerate levels and hence determine transition probabilities between states.[12,14] Since $\mathscr{D}$ has operators which do not commute with the Hamiltonian, it is a noninvariance algebra. The dynamical algebras for the three-dimensional oscillator[12] and the Coulomb potential[12,14] are $sp(6,\mathbb{R})\square w_3$ and o(4,2), respectively.

In this work we consider the following o(3)-invariant time-dependent Schrödinger equation

$$Q\Psi(\boldsymbol{\eta},t) = \left\{ \sum_{k=1}^{3} \partial_{kk} + 2i\,\partial_t - 2W(\eta) \right\} \Psi(\boldsymbol{\eta},t) = 0, \tag{1.1}$$

where $\hbar = 1$, $\partial_{kk} = \partial^2/\partial\eta_k^2$, and $\eta = (\eta_1^2 + \eta_2^2 + \eta_3^2)^{1/2}$. The operator Q is

$$Q = \sum_{k=1}^{3} \partial_{kk} + 2i\,\partial_t - 2W(\eta). \tag{1.2}$$

Equation (1.1) is equivalent to a one-body Schrödinger equation or a two-body equation with center-of mass removed.[17] For all systems of this type angular momentum is conserved, and $o(3) \oplus t_1$ is their minimal kinematic algebra. In Sec. II, Boyer's classification of rotationally invariant potentials in terms of their full kinematic algebras is briefly reviewed. In Sec. III, all potentials of the form $W(\eta)$ and their associated symmetry algebras realized as differential operators of order at most two are calculated. All geometric and dynamical invariances are identified for each case, and it is shown that the only rotationally invariant (two-body) potentials with degeneracy algebras greater than o(3) are the Coulomb, oscillator, and constant potentials. Furthermore, the Coulomb potential is proven to be unique in the sense that its dynamical symmetries, the components of the Runge–Lenz vector, cannot be written as symmetric products of elements of the kinematic algebra, as in the case for the oscillator and constant potentials. The structure of these second order algebras and their relationship to the boson representation is briefly outlined in Sec. IV.

II. KINEMATICAL ALGEBRAS FOR POTENTIALS OF THE FORM $W(\eta)$

Boyer[6] has given a complete classification of potentials $W(\eta)$ in terms of their kinematical algebra, and we briefly mention the procedure and results here. The form of the operators is important for the calculations in Secs. III and IV.

TABLE I. Potentials $W(\eta)$ and their associated kinematical algebras.

Name of potential	Potential $W(\eta)$	Algebra $\mathscr{G}$ [a]
Constant potential	κ	$(o(3) \oplus sl(2,\mathbb{R}))\square w_3$
Harmonic oscillator [b]	$\omega^2\eta^2/2 + \kappa$	$(o(3) \oplus sl(2,\mathbb{R}))\square w_3$
Centrifugal potential	λ/η^2	$o(3) \oplus sl(2,\mathbb{R})$
Radial oscillator	$\omega^2\eta^2/2 + \lambda/\eta^2$	$o(3) \oplus sl(2,\mathbb{R})$
Other	$W(\eta)$	$o(3) \oplus t_1$

[a] o(3) is the Lie algebra of the rotation group, $sl(2,\mathbb{R})$ is the Lie algebra of he special linear group, and w_3 is the seven dimensional Heisenberg–Weyl algebra.[19] t_1 is the one-dimensional algebra formed by the time displacement operator. $\oplus$ indicates a direct sum; $\square$ is the semidirect sum.[6]

[b] Both the attractive oscillator $\omega^2 > 0$ and the repulsive oscillator $\omega^2 < 0$ are included here. See Boyer.[6]

If we consider the set of local space–time transformations of the form

$$\Psi(\boldsymbol{\eta},t) \to T(g)\Psi(\boldsymbol{\eta},t) = \nu_g(g^{-1}(\boldsymbol{\eta},t))\Psi(g^{-1}(\boldsymbol{\eta},t)), \tag{2.1}$$

Where $g^{-1}(\boldsymbol{\eta},t)$ gives the local action of the transformation g on the space–time coordinates and ν_g is an arbitrary function called the multiplier; $T(g)$ is called a multiplier representation.[18] If we denote by $\mathscr{H}$ the space of solutions of Eq. (1.1), then we want the subset of transformations which leaves $\mathscr{H}$ invariant, i.e.,

$$QT(g)\Psi(\boldsymbol{\eta},t) = 0. \tag{2.2}$$

Such a subset forms a group. When Eq. (2.1) is expanded, the infinitesimal generators of first order in the derivatives associated with the transformation $T(g)$ have the form[6,18]

$$L = \sum_{k=1}^{3} A^k(\boldsymbol{\eta},t)\partial_k + A^t(\boldsymbol{\eta},t)\partial_t + A(\boldsymbol{\eta},t). \tag{2.3}$$

For L to be consistent with Eq. (2.2), it is required that

$$[L,Q] = R(\boldsymbol{\eta},t)Q, \tag{2.4}$$

where Q is given by Eq. (1.2), and $R(\boldsymbol{\eta},t)$ may be a complex-valued function of its arguments. The elements L form a Lie algebra[18] $\mathscr{G}$.

In Table I, the potentials and their respective kinematical algebras $\mathscr{G}$ are given. Generators of the o(3) subalgebra are

$$M_1 = i(\eta_3\partial_2 - \eta_2\partial_3), \quad M_2 = i(\eta_1\partial_3 - \eta_3\partial_1),$$
$$M_3 = i(\eta_2\partial_1 - \eta_1\partial_2), \tag{2.5}$$

the usual angular momentum operators. The generators of dilations and conformal transformations for (a) the constant and centrifugal potentials ($\kappa = 0$), are[18,20]

$$I_1 = \frac{i}{2}(\boldsymbol{\eta}\cdot\boldsymbol{\partial} + 2t\partial_t + 2i\kappa t + \tfrac{3}{2}),$$
$$I_2 = \frac{i}{2}\left(-t\boldsymbol{\eta}\cdot\boldsymbol{\partial} + (1-t^2)\partial_t + \frac{i\eta^2}{2} + i(1-t^2)\kappa - \tfrac{3}{2}t\right), \tag{2.6}$$
$$I_3 = \frac{i}{2}\left(t\boldsymbol{\eta}\cdot\boldsymbol{\partial} + (1+t^2)\partial_t - \frac{i\eta^2}{2} + i(1+t^2)\kappa + \tfrac{3}{2}t\right),$$

and (b) the harmonic and radial oscillator potentials ($\kappa = 0$) are

$$\begin{aligned} I_1 &= \frac{i}{2\omega}(\omega \cos 2\omega t \boldsymbol{\eta}\cdot\boldsymbol{\partial} + \sin 2\omega t\, \partial_t \\ &\quad + i(\omega^2\boldsymbol{\eta}^2 + \kappa)\sin 2\omega t + \tfrac{3}{2}\omega \cos 2\omega t), \\ I_2 &= \frac{i}{2\omega}(-\omega \sin 2\omega t \boldsymbol{\eta}\cdot\boldsymbol{\partial} + \cos 2\omega t\, \partial_t \\ &\quad + i(\omega^2\boldsymbol{\eta}^2 + \kappa)\cos 2\omega t - \tfrac{3}{2}\omega \sin 2\omega t), \\ I_3 &= \frac{i}{2\omega}(\partial_t + i\kappa). \end{aligned} \tag{2.7}$$

For the general potential $W(\eta)$, t_1 is a one-dimensional algebra consisting of the time translation operator $T = \partial_t$. Finally, the generators of the Heisenberg–Weyl algebra w_3 are for (a) the constant potential

$$C_i^{(k)} = i\partial_k, \quad C_2^{(k)} = i(t\partial_k - i\eta_k), \quad E = 1, \quad 1 \leqslant k \leqslant 3, \tag{2.8}$$

and (b) the harmonic oscillator potential

$$\left.\begin{aligned} C_1^{(k)} &= -i\omega^{-1/2}(\cos\omega t\, \partial_k + i\omega\eta_k \sin\omega t), \\ C_2^{(k)} &= -i\omega^{-1/2}(\sin\omega t\, \partial_k + i\omega\eta_k \cos\omega t), \end{aligned}\right\} \quad 1 \leqslant k \leqslant 3, \tag{2.9}$$

$$E = 1.$$

The commutation relations are

$$\begin{aligned} &[M_j, M_k] = i\epsilon_{jkl}M_l, \quad 1 \leqslant j,k,l \leqslant 3, \\ &[I_1, I_2] = -iI_3, \quad [I_2, I_3] = iI_1, \quad [I_3, I_1] = iI_2, \\ &[C_1^{(k)}, C_2^{(l)}] = i\delta_{kl}E, \quad 1 \leqslant k,l \leqslant 3, \\ &[C_1^{(k)}, C_1^{(l)}] = [C_2^{(k)}, C_2^{(l)}] = [C_1^{(k)}, E] = [C_2^{(k)}, E] = 0, \\ &[M_k, I_l] = 0, \quad 1 \leqslant k,l \leqslant 3, \\ &[M_1, I_l] = 0, \quad 1 \leqslant k,l \leqslant 3, \\ &[M_1, C_\sigma^{(k)}] = i(\delta_{2k}C_\sigma^{(3)} - \delta_{3k}C_\sigma^{(2)}, \\ &[M_2, C_\sigma^{(k)}] = i(\delta_{3k}C_\sigma^{(1)} - \delta_{1k}C_\sigma^{(3)}), \\ &[M_3, C_\sigma^{(k)}] = i(\delta_{1k}C_\sigma^{(2)} - \delta_{2k}C_\sigma^{(1)}), \quad 1 \leqslant k \leqslant 3, \quad 1 \leqslant \sigma \leqslant 2, \\ &[I_1, C_1^{(k)}] = \frac{-i}{2}C_1^{(k)}, \quad [I_2, C_1^{(k)}] = \frac{i}{2}C_2^{(k)}, \\ &[I_3, C_1^{(k)}] = \frac{-i}{2}C_2^{(k)}, \\ &\qquad\qquad 1 \leqslant k \leqslant 3, \\ &[I_1, C_2^{(k)}] = \frac{i}{2}C_2^{(k)}, \\ &[I_2, C_2^{(k)}] = \frac{i}{2}C_1^{(k)}, \quad [I_3, C_2^{(k)}] = \frac{i}{2}C_1^{(k)}. \end{aligned} \tag{2.10}$$

The isomorphism of the kinematical algebras for the constant potential and the oscillator are clear. Also note the isomorphism of the kinematical algebras for the radial oscillator and centrifugal potential. We shall exploit this isomorphism in the next section. The Coulomb potential is conspicuous by its absence from Table I; in terms of its kinematic algebra, it must be classified with the general form $W(\eta)$ as $o(3) \oplus t_1$. Its special status will become apparent in Sec. III.

III. SECOND ORDER SYMMETRY ALGEBRAS AND THEIR CLASSIFICATION

If we expand Eq. (2.1) to second order, the generators associated with the transformation $T(g)$ have the form

$$\begin{aligned} S = &\sum_{k=1}^{3} A^k(\boldsymbol{\eta},t)\partial_{kk} + \sum_{\substack{k,l=1 \\ k<l}}^{3} A^{kl}(\boldsymbol{\eta},t)\partial_{kl} \\ &+ \sum_{k=1}^{3} B^k(\boldsymbol{\eta},t)\partial_k + B^t(\boldsymbol{\eta},t)\partial_t + C(\boldsymbol{\eta},t). \end{aligned} \tag{3.1}$$

Also, we require that S satisfy

$$[S,Q] = UQ, \tag{3.2}$$

where U is a first order differential operator.[21] Now, if we denote by $q = \{RQ: R \text{ a complex-valued function}\}$, the set of all trivial symmetries, then for some R we have $S' = S + RQ$, where

$$\begin{aligned} S' = &\sum_{k=1}^{3} A^k(\boldsymbol{\eta},t)\partial_{kk} + \sum_{\substack{k,l=1 \\ k<l}}^{3} A^k(\boldsymbol{\eta},t)\partial_{kl} \\ &+ \sum_{k=1}^{3} B^k(\boldsymbol{\eta},t)\partial_k + C, \end{aligned} \tag{3.3}$$

with Eq. (3.2) reduced to

$$[S',Q] = 0. \tag{3.4}$$

Thus, every S is equivalent to an S' in which the time derivative has been removed, and S' commutes with Q. Denote the set of all such S' by $\mathcal{D}/q$.

Like members of the kinematic algebra $\mathcal{G}$, elements of $\mathcal{D}/q$ map solutions of Eq. (1.1) into solutions. Unlike $\mathcal{G}$, however, $\mathcal{D}/q$ is not, in general, a Lie algebra, as the commutator of two second order operators is usually a third order differential operator. Nevertheless, a subset $\mathcal{G}^*$ of $\mathcal{D}/q$ does form a Lie algebra, in particular for the kinematic algebra $\mathcal{G} \subseteq \mathcal{G}^* \subset \mathcal{D}/q$.

Now, we calculate a basis for $\mathcal{D}/q$ for Q given by Eq. (1.2) and determine the explicit form of $W(\eta)$ associated with each such $\mathcal{D}/q$. For all potentials, $W(\eta)$ we shall find that a minimal $\mathcal{D}/q$ exists which consists of the kinematic algebra $\mathcal{G}$ and the set $\{M_j^2, [M_k,M_l]_+ = M_kM_l + M_lM_k: 1 \leqslant j,\ k < l \leqslant 3\}$, where M_j, $1 \leqslant j \leqslant 3$ are the generators of rotations. The operators M_j^2 and $[M_k,M_l]_+$ are symmetric products of elements of o(3) and are said to be second order elements in the universal enveloping algebra of o(3). Denote these elements by $o^2(3)$ and call the minimal algebra $\mathcal{G} \cup o^2(3)$ (not a Lie algebra). With the single exception of the Coulomb potential, $\mathcal{D}/q$ will consist of $\mathcal{G}$ and any remaining elements of $\mathcal{D}/q$ may be written as symmetric products of elements of $\mathcal{G}$, i.e., second order elements of the universal enveloping algebra of $\mathcal{G}$. Potentials (Schrödinger equations) with this property will be labeled class I.[21] On the other hand, $\mathcal{D}/q$ for the Coulomb potential has elements not in $\mathcal{G}$ and which cannot be written as second order members of the universal enveloping algebra of $\mathcal{G}$; potentials like the Coulomb potential are termed class II.[21] Finally, we show that only for the oscillator, the constant, and Coulomb potentials is $\mathcal{G}^*$ strictly larger than o(3).

Substituting Eqs. (3.3) and (1.2) into Eq. (3.4), the fol-

lowing system of partial differential equations is obtained for the coefficients A^k, A^{kl}, B^k, and C:

$$\left.\begin{aligned}
&A^k_k = 0, \quad 1\leqslant k\leqslant 3,\\
&\left.\begin{aligned}A^l_k + A^{kl}_l &= 0,\\ A^k_l + A^{kl}_k &= 0,\end{aligned}\right\} \quad 1\leqslant k<l\leqslant 3,\\
&A^{23}_1 + A^{13}_2 + A^{12}_3 = 0,\\
&\sum_{k=1}^{3} A^l_{kk} + 2B^l_l + 2i A^l_t = 0, \quad 1\leqslant l\leqslant 3,\\
&\sum_{k=1}^{3} A^{lm}_{kk} + 2B^l_m + 2B^m_l + 2i A^{lm}_t = 0, \quad 1\leqslant l<m\leqslant 3,
\end{aligned}\right\} \tag{3.5a}$$

$$\sum_{k=1}^{3} B^l_{kk} + 2C_l + 2i B^l_t + 4A^l W_l + 2\sum_{\substack{k=1\\k\neq l}}^{3} A^{lk} W_k = 0, \quad 1\leqslant l\leqslant 3, \tag{3.5b}$$

$$\sum_{k=1}^{3} (C_{kk} + 2B^k W_k + 2A^k W_{kk}) + 2\sum_{\substack{k,l=1\\k<l}}^{3} A^{kl} W_{kl} + 2iC_t = 0. \tag{3.5c}$$

Equations (3.5a) may be integrated and the functions A^k, A^{kl}, and B^l obtained as polynomials in the coordinates (η_1,η_2,η_3). The coefficient of each term in the polynomial is either time independent (Greek letters α, β, γ, etc.) or time dependent (Latin letters, a, b, c, etc.). The resulting expressions for A^k, A^{kl}, and B^k are

$$\left.\begin{aligned}
A^1 &= \beta_{12}\eta_2^2 + \beta_{13}\eta_3^2 - \beta_1\eta_2\eta_3 + a_{12}\eta_2 + a_{13}\eta_3 + a_1,\\
A^2 &= \beta_{12}\eta_1^2 + \beta_{23}\eta_3^2 - \beta_2\eta_1\eta_3 + a_{21}\eta_1 + a_{23}\eta_3 + a_2,\\
A^3 &= \beta_{13}\eta_1^2 + \beta_{23}\eta_2^2 - \beta_3\eta_1\eta_3 + a_{31}\eta_1 + a_{32}\eta_2 + a_3,
\end{aligned}\right\} \tag{3.6a}$$

$$\left.\begin{aligned}
A^{12} &= -2\beta_{12}\eta_1\eta_2 + \beta_1\eta_1\eta_3 + \beta_2\eta_2\eta_3 - \beta_3\eta_3^2 - a_{12}\eta_1 - a_{21}\eta_2 + b_{123}\eta_3 + b_{12},\\
A^{13} &= \beta_1\eta_1\eta_2 - 2\beta_{13}\eta_1\eta_3 + \beta_3\eta_2\eta_3 - \beta_2\eta_2^2 - a_{13}\eta_1 - a_{31}\eta_3 + b_{132}\eta_2 + b_{13},\\
A^{23} &= \beta_2\eta_1\eta_2 + \beta_3\eta_1\eta_3 - 2\beta_{23}\eta_2\eta_3 - \beta_1\eta_1^2 - a_{23}\eta_2 - a_{32}\eta_3 - (b_{123} + b_{132})\eta_1 + b_{23},
\end{aligned}\right\} \tag{3.6b}$$

$$\begin{aligned}
B^1 &= i\dot{a}_{12}\eta_1\eta_2 - i\dot{a}_{13}\eta_1\eta_3 - (i\dot{a}_1 + \beta_{12} + \beta_{13})\eta_1 + i\dot{a}_{21}\eta_2^2 + i\dot{a}_{31}\eta_3^2 - i(\dot{b}_{123} + \dot{b}_{132})\eta_2\eta_3 + c_{12}\eta_2 + c_{13}\eta_3 + c_1,\\
B^2 &= -i\dot{a}_{21}\eta_1\eta_2 - i\dot{a}_{23}\eta_2\eta_3 - (i\dot{a}_2 + \beta_{12} + \beta_{23})\eta_2 + i\dot{a}_{12}\eta_1^2 + i\dot{a}_{32}\eta_3^2 + i\dot{b}_{132}\eta_1\eta_3\\
&\quad - (i\dot{b}_{12} + c_{12} - \beta_3)\eta_1 + c_{23}\eta_3 + c_2,\\
B^3 &= -i\dot{a}_{31}\eta_1\eta_3 - i\dot{a}_{32}\eta_2\eta_3 - (i\dot{a}_3 + \beta_{13} + \beta_{23})\eta_3 - i\dot{a}_{13}\eta_1^2 + i\dot{a}_{23}\eta_2^2 + i\dot{b}_{123}\eta_1\eta_2\\
&\quad - (i\dot{b}_{13} + c_{13} - \beta_2)\eta_1 - (i\dot{b}_{23} + c_{23} - \beta_1)\eta_2 + c_3,
\end{aligned} \tag{3.6c}$$

where $\dot{x} = dx/dt$. The remaining four conditions (3.5b) and (3.5c) will now be used to determine the explicit form for the potential W and the coefficients a, b, and c.

First, we can eliminate the term C_l from the three equations (3.5b) to obtain three expressions containing the coefficients A, B, their derivatives, and derivatives of W, where we have made use of the fact that $B^l_{kkm} = 0$, for all indices, k, l, m. Thus,

$$\begin{aligned}
&-i(B^l_{kt} - B^k_{lt})\\
&= -3A^k_l W_k + 3A^l_k W_l + 2(A^l - A^k)W_{kl}\\
&\quad + A^{kl}(W_{kk} - W_{ll})\\
&\quad + (A^{lm}_k - A^{km}_l)W_m + A^{lm}W_{km}\\
&\quad - A^{km}W_{lm}, \quad 1\leqslant k<l\neq m\leqslant 3.
\end{aligned} \tag{3.7}$$

If we substitute

$$\begin{aligned}
W_k &= (\eta_k/\eta)W',\\
W_{kl} &= (1/\eta)W'\delta_{kl} - (\eta_k\eta_l/\eta^3)G, \quad k\leqslant 1, l\leqslant 3,
\end{aligned} \tag{3.8}$$

along with the expressions (3.6) for the functions A^k, A^{kl}, B^k, and their derivatives into Eq. (3.7), we obtain, after some algebraic manipulation, three polynomials of the form

$$\sum_{\alpha} \eta_1^{\alpha_1}\eta_2^{\alpha_2}\eta_3^{\alpha_3} \sum_{\sigma} f^{(|\alpha|,k)}_{\alpha,\sigma}(\eta)\Gamma^{(|\alpha|,k)}_{\alpha,\sigma}(t) = 0, \quad 1\leqslant k\leqslant 3, \tag{3.9}$$

where α is a multi-index $|\alpha| = \alpha_1 + \alpha_2 + \alpha_3$, and

$$f^{(|\alpha|,k)}_{\alpha,\sigma}(\eta) = \sum_{n=1}^{2} g^s_n(\eta)W^{(n)} + g^s_0(\eta), \quad s = (|\alpha|,k,\alpha,\sigma). \tag{3.10}$$

The $\Gamma^{(|\alpha|,k)}_{\alpha,\sigma}$ are functions of the time-dependent coefficients a, b, c and their first or second derivatives.

The conditions (3.9), with explicit expressions for f and Γ, are given in Table II. For Eq. (3.10) to hold for all values of (η_1, η_2, η_3) either the f or Γ coefficient in each term of the double sum must vanish. This requirement in turn restricts the form of W and determines, in part, the time-dependent coefficients.

We proceed by enumerating the various cases which arise by inspection of Table II, stating the conditions on the potential $W(\eta)$ and the relationships among the coefficients a,b,c,d, and their derivatives. Once $W(\eta)$ has been obtained, Eqs. (3.5b) can be integrated for an explicit form of the function C.

Case 1: If we choose

$$b_{132}\neq 0, \quad b_{123}\neq 0, \tag{3.11a}$$

then we must have

$$W' = 0. \tag{3.12a}$$

Furthermore, together, Eqs. (III.12a) and (III.8) imply

TABLE II. Explicit expansions for polynomials (3.9).

			$k=1$		$k=2$		$k=3$	
$\lvert\alpha\rvert$	α	σ	f	Γ	f	Γ	f	Γ
0	0	1	0	0	0	0	0	0
1	1	1	0	0	0	0	0	0
2	(2,0,0)	1	$-G$	b_{12}	$-G$	b_{13}	η	$\ddot{b}_{23}-2i\ddot{c}_{23}$
		2	η	$\dot{b}_{12}-2i\dot{c}_{12}$	η	$\dot{b}_{13}-2i\dot{c}_{13}$	0	0
	(0,2,0)	1	G	b_{12}	η	$\dot{b}_{13}-2i\dot{c}_{13}$	$-G$	b_{23}
		2	η	$\ddot{b}_{12}-2i\ddot{c}_{12}$	0	0	η	$\dot{b}_{23}-2i\dot{c}_{23}$
	(0,0,2)	1	η	$\dot{b}_{12}-2i\dot{c}_{12}$	G	b_{13}	G	b_{23}
		2	0	0	0	$\dot{b}_{13}-2i\dot{c}_{13}$	η	$\dot{b}_{23}-2i\dot{c}_{23}$
	(1,1,0)	1	$2G$	a_1-a_2	$-G$	b_{23}	$-G$	b_{13}
	(1,0,1)	1	$-G$	b_{23}	$2G$	a_1-a_3	G	b_{12}
	(0,1,1)	1	G	b_{13}	G	b_{12}	$2G$	a_2-a_3
3	(3,0,0)	1	$G-3W'$	a_{12}	$G-3W'$	a_{13}	W'	$b_{132}-b_{123}$
		2	-3η	$\ddot{a}_{12}$	-3η	$\ddot{a}_{13}$	η	$\dot{b}_{132}-\dot{b}_{123}$
	(0,3,0)	1	$-(G-3W')$	a_{21}	$-W'$	b_{132}	$G-3W'$	a_{23}
		2	3η	$\ddot{a}_{21}$	$-2W'$	b_{123}	-3η	$\ddot{a}_{23}$
		3	0	0	$-\eta$	$2\dot{b}_{123}+\ddot{b}_{132}$	0	0
	(0,0,3)	1	$2(G-W')$	b_{132}	$-(G-3W')$	a_{31}	$-(G-3W')$	a_{32}
		2	$-W'$	b_{123}	3η	$\ddot{a}_{31}$	3η	$\ddot{a}_{32}$
		3	$-\eta$	$\ddot{b}_{123}+2\ddot{b}_{132}$	0	0	0	0
3	(2,1,0)	1	$-(G-3W')$	a_{21}	$-W'$	b_{132}	$-3W'$	a_{23}
		2	-3η	$\ddot{a}_{21}$	$G-2W'$	b_{123}	G	a_{13}
		3	0	0	$-\eta$	$2\dot{b}_{123}+\ddot{b}_{132}$	-3η	$\ddot{a}_{23}$
	(2,0,1)	1	$G-2W'$	b_{132}	$-(G-3W')$	a_{31}	$-G$	a_{12}
		2	$-W'$	b_{123}	3η	$\ddot{a}_{31}$	$3W'$	a_{32}
		3	$-\eta$	$\ddot{b}_{123}+2\ddot{b}_{132}$	0	0	3η	$\ddot{a}_{32}$
	(1,2,0)	1	$G-3W'$	a_{12}	$-3W'$	a_{13}	W'	b_{132}
		2	-3η	$\ddot{a}_{12}$	G	a_{23}	$G-W'$	b_{132}
		3	0	0	-3η	$\ddot{a}_{13}$	η	$\dot{b}_{132}-\dot{b}_{123}$
	(0,2,1)	1	$G-2W'$	b_{132}	$3W'$	a_{31}	$-(G-3W')$	a_{32}
		2	$G-W'$	b_{123}	$-G$	a_{21}	3η	$\ddot{a}_{32}$
		3	$-\eta$	$\ddot{b}_{123}+\ddot{b}_{132}$	3η	$\ddot{a}_{31}$	0	0
	(1,0,2)	1	G	a_{32}	$G-3W'$	a_{13}	$-(G-W')$	b_{132}
		2	$-3W'$	a_{12}	-3η	$\ddot{a}_{13}$	$-W'$	b_{123}
		3	-3η	$\ddot{a}_{12}$	0	0	η	$\dot{b}_{132}-\dot{b}_{123}$
	(0,1,2)	1	$-G$	a_{31}	$G-3W'$	a_{13}	$G-3W'$	a_{23}
		2	$3W'$	a_{21}	-3η	$\ddot{a}_{13}$	-3η	$\ddot{a}_{23}$
		3	-3η	$\ddot{a}_{13}$	0	0	0	0
	(1,1,1)	1	G	$a_{13}-a_{23}$	G	$a_{12}-a_{32}$	G	$a_{21}-a_{31}$

$$G=0. \tag{3.12b}$$

Thus, $W(\eta)$ is the constant potential

$$W(\eta)=\kappa. \tag{3.12c}$$

Consistent with Eqs. (3.11a), (3.12a), and (3.12b) are the following constraints on the a,b,c coefficients:

$$\left.\begin{aligned} &\ddot{b}_{132}=\ddot{b}_{123}=0,\\ &a_{ij}\neq 0,\\ &\ddot{a}_{ij}=0,\end{aligned}\right\}\quad 1\leqslant i\neq j\leqslant 3,$$

$$\left.\begin{aligned} &b_{ij}\neq 0,\\ &\ddot{b}_{ij}-2i\,\dot{c}_{ij}=0,\\ &a_i-a_j\neq 0\end{aligned}\right\}\quad 1\leqslant i<j\leqslant 3, \tag{3.11b}$$

The function C is given by

$$\begin{aligned} C=&-i(\dot{a}_{21}+\dot{a}_{31})\eta_1-i(\dot{a}_{12}+\dot{a}_{32})\eta_2\\ &-i(\dot{a}_{13}+\dot{a}_{23})\eta_3-\tfrac{1}{2}(\ddot{a}_1\eta_1^2+\ddot{a}_2\eta_2^2+\ddot{a}_3\eta_3^2)\\ &-i\dot{c}_{12}\eta_1\eta_2-i\dot{c}_{13}\eta_1\eta_3-i\dot{c}_{23}\eta_2\eta_3\\ &-i\dot{c}_1\eta_1-i\dot{c}_2\eta_2-i\dot{c}_3\eta_3+d, \end{aligned} \tag{3.13}$$

where d is a time-dependent function occurring in the integration of Eq. (3.5b).

Case 2: Now let

$$b_{ij}\neq 0,\quad 1\leqslant i<j\leqslant 3, \tag{3.14a}$$

Then we require

$$G=W'-\eta W''=0, \tag{3.15a}$$

whose solution is

$$W(\eta)=\omega^2\eta^2/2+\kappa, \tag{3.15b}$$

the harmonic (or repulsive) oscillator potential. Equations (3.15a) and (3.14a) together imply

$$\left.\begin{aligned} &\ddot{b}_{ij}-2i\,\dot{c}_{ij}=0,\\ &a_i-a_j\neq 0,\end{aligned}\right\}\quad 1\leqslant i<j\leqslant 3,$$

$$\ddot{a}_{ij}+\omega^2 a_{ij}=0,\quad 1\leqslant i\neq j\leqslant 3, \tag{3.14b}$$

$$\ddot{b}_{132}+\omega^2 b_{132}=0,$$

$$\ddot{b}_{123} + \omega^2 b_{123} = 0.$$

The expression for the C coefficient is

$$C = i(\dot{a}_{21} + \dot{a}_{31} + \dot{c}_1)\eta_1 + i(\dot{a}_{12} + \dot{a}_{32} + \dot{c}_2)\eta_2 + i(\dot{a}_{13} + \dot{a}_{23} + \dot{c}_3)\eta_3 + \tfrac{1}{2}\sum_k (\ddot{a}_k + 2\omega^2 a_k)\eta_k^2 + \sum_{i<j} (\ddot{b}_{ij} + \omega^2 b_{ij} - i\dot{c}_{ij})\eta_i\eta_j + d. \tag{3.16}$$

Case 3: When

$$a_{ij} \neq 0, \quad 1 \leqslant i \neq j \leqslant 3, \tag{3.17a}$$

we must have

$$G - 3W' = -2W' - \eta W'' = 0. \tag{3.18a}$$

This equation has the Coulomb potential

$$W(\eta) = -\lambda/\eta + \kappa \tag{3.18b}$$

as its solution. Consistency of Eqs. (3.17a), (3.18a), and (3.18b) with Table II require further that

$$\left.\begin{aligned} &\ddot{a}_{ij} = 0, \quad 1 \leqslant i \neq j \leqslant 3, \\ &b_{ij} = 0, \\ &\dot{c}_{ij} = 0, \end{aligned}\right\} \quad 1 \leqslant i < j \leqslant 3,$$
$$a_1 = a_2 = a_3 = a,$$
$$b_{132} = b_{123} = 0,$$
$$a_{km} - a_{lm} = 0, \quad a_{km} = a_{mk}, \quad 1 \leqslant k < l \neq m \leqslant 3. \tag{3.17b}$$

The expression for the C function is

$$C = -\ddot{a}\eta^2/2 - i(\dot{c}_1 + 2\dot{a}_{31})\eta_1 - i(\dot{c}_2 + 2\dot{a}_{12})\eta_2 - i(\dot{c}_3 + 2\dot{a}_{23})\eta_3 + a_{12}(\lambda\eta_2/\eta) + a_{31}(\lambda\eta_1/\eta) + a_{23}(\lambda\eta_3/\eta) + 2a(\lambda/\eta) + d. \tag{3.19}$$

Case 4: Now, if we let

$$\ddot{a}_{ij} = -\omega^2 a_{ij}, \quad 1 \leqslant i \neq j \leqslant 3, \tag{3.20a}$$

then the equation restricting the form of the potentials is

$$G - 3W' = -3\omega^2\eta. \tag{3.21a}$$

Its solution is the Coulombic oscillator potential

$$W(\eta) = \omega^2\eta^2/2 - \lambda/\eta + \kappa. \tag{3.21b}$$

Conditions (3.20a) and (3.21) together with Table II yield

$$\left.\begin{aligned} &b_{ij} = 0, \\ &\dot{c}_{ij} = 0 \end{aligned}\right\} \quad 1 \leqslant i < j \leqslant 3,$$
$$a_1 = a_2 = a_3 = a,$$
$$b_{132} = b_{123} = 0,$$
$$a_{km} - a_{lm} = 0, \quad a_{km} = a_{mk}, \quad 1 \leqslant k < l \neq m \leqslant 3. \tag{3.20b}$$

C has the form

$$C = -\tfrac{1}{2}(\ddot{a} + 2\omega^2 a)\eta^2 + a(2\lambda/\eta) - i(2\dot{a}_{31} + \dot{c}_1)\eta_1 - i(2\dot{a}_{12} + \dot{c}_2)\eta_2 - i(2\dot{a}_{23} + \dot{c}_3)\eta_3 + a_{12}(\lambda\eta_2/\eta) + a_{31}(\lambda\eta_1/\eta) + a_{23}(\lambda\eta_3/\eta) + d, \tag{3.21}$$

Case 5: Let

$$\left.\begin{aligned} &b_{ij} = 0, \quad 1 \leqslant i < j \leqslant 3, \\ &a_1 = a_2 = a_3 = a, \end{aligned}\right\}$$
$$\left.\begin{aligned} &a_{ij} = 0, \quad 1 \leqslant i \neq j \leqslant 3, \\ &b_{123} = b_{132} = 0. \end{aligned}\right\} \tag{3.22a}$$

Then $W(\eta)$ is an arbitrary function of its independent variable but is not of the type mentioned in the previous four cases. However, relaxing any of the constraints (3.22a) will achieve one of the four previous cases. Equations (3.22a) imply that

$$c_{ij} = 0, \quad 1 \leqslant i < j \leqslant 3, \tag{3.22b}$$

and

$$C = -\ddot{a}\eta^2/2 - 2aW - i\sum_{k=1}^{3} \dot{c}_k\eta_k + d. \tag{3.23}$$

For each of the above cases Eq. (3.5c) may now be applied to the A, B, and C functions to obtain conditions on the remaining, as yet undetermined, time-dependent coefficients a, b, c, and d. These may now be substituted back into (3.6) and the appropriate C function and S^1 determined for each case. An appropriate basis for each $\mathcal{Q}/q$ is chosen and expressed in terms of elements of a kinematical algebra wherever possible.

Case 1: Previously, it was mentioned that the kinematical algebras for the oscillator and constant potential are isomorphic. That their second order algebras $\mathcal{Q}/q$ are also isomorphic follows from the fact that both potentials are class I. With this in mind, we pass on to the calculation of the oscillator algebra.

Case 2: The condition (3.5c) for the harmonic oscillator potential is

$$-\tfrac{1}{2}\sum_{k=1}^{3} (\ddot{a}_k + 4\omega^2\dot{a}_k)\eta_k^2 + i\dot{d} - \tfrac{1}{2}\sum_{k=1}^{3} \ddot{a}_k + \sum_{k=1}^{3} (\ddot{c}_k + \omega^2 c_k)\eta_k - i\sum_{\substack{k,l=1 \\ k<l}}^{3} (\dddot{b}_{kl} + 2\omega^2\dot{b}_{kl} - i\ddot{c}_{kl})\eta_k\eta_l = 0. \tag{3.24}$$

For this to be true for all (η_1, η_2, η_3) we must have

$$\left.\begin{aligned} &\ddot{a}_k + 4\omega^2\dot{a}_k = 0, \quad 1 \leqslant k \leqslant 3, \\ &i\dot{d} - \tfrac{1}{2}\sum_{k=1}^{3} \ddot{a}_k = 0, \\ &\ddot{c}_k + \omega^2 c_k = 0, \quad 1 \leqslant k \leqslant 3, \\ &\dddot{b}_{kl} + 2\omega^2\dot{b}_{kl} - i\ddot{c}_{kl} = 0, \quad 1 \leqslant k < l \leqslant 3. \end{aligned}\right\} \tag{3.25}$$

Integrating the system of equations (3.14) and (3.25), we have

$$\left.\begin{aligned} &a_{kl} = (\alpha_{kl}^{(3)}/\sqrt{\omega})\sin\omega t + (\alpha_{kl}^{(2)}/\sqrt{\omega})\cos\omega t, \quad 1 \leqslant k \neq l \leqslant 3, \\ &a_k = (\alpha_k^{(3)}/2\omega)\sin 2\omega t - (\alpha_k^{(2)}/2\omega)\cos 2\omega t + \alpha_k^{(1)}, \quad 1 \leqslant k \leqslant 3, \\ &b_{kl} = (\gamma_l^{(2)}/2\omega)\sin 2\omega t - (\gamma_{kl}^{(1)}/2\omega)\cos 2\omega t + \epsilon_{kl}, \\ &b_{132} = \beta_7\cos\omega t + \beta_8\sin\omega t, \\ &b_{123} = \beta_9\cos\omega t + \beta_{10}\sin\omega t, \\ &c_{kl} = -\tfrac{1}{2}(\gamma_{kl}^{(2)}\cos 2\omega t + \gamma_{kl}^{(2)}\sin 2\omega t) + \gamma_{kl}^{(3)}, \quad 1 \leqslant k < l \leqslant 3, \\ &c_k = (\gamma_k^{(2)}/\sqrt{\omega})\cos\omega t + (\gamma_k^{(1)}/\sqrt{\omega})\sin\omega t, \quad 1 \leqslant k \leqslant 3, \\ &d = -\tfrac{1}{2}\sum_k (\alpha_k^{(3)}\cos 2\omega t + \alpha_k^{(2)}\sin 2\omega t) + \beta_{11}. \end{aligned}\right\} \tag{3.26}$$

A basis for $\mathcal{D}/q$ consists of

$$\left.\begin{aligned}
&E,\ C_1^{(k)},\ C_2^{(k)},\quad 1\leqslant k\leqslant 3,\\
&M_j^2,\ [M_k,M_l]_+,\quad 1<j,k<l\leqslant 3,\\
&[M_j,C_l^{(k)}]_+ \text{ subject to the constraint } \mathbf{M}\cdot\mathbf{C}_l=0,\\
&1\leqslant j,k\leqslant 3,\quad 1\leqslant l\leqslant 2,\\
&P_1^{(k)}=\tfrac{1}{4}[C_1^{(k)},C_2^{(k)}]_+,\quad P_2^{(k)}=\tfrac{1}{4}(C_1^{(k)2}-C_2^{(k)2}),\\
&P_3^{(k)}=\tfrac{1}{4}(C_1^{(k)2}+C_2^{(k)2}),\quad 1\leqslant k\leqslant 3,\\
&T_\pm^{(k,l)}=\tfrac{1}{2}\{[C_1^{(k)},C_2^{(l)}]_+\pm[C_1^{(l)},C_2^{(k)}]_+\},\\
&S_\pm^{(k,l)}=\tfrac{1}{2}\{[C_1^{(k)},C_1^{(l)}]_+\pm[C_2^{(k)},C_2^{(l)}]_+\},
\end{aligned}\right\}\quad 1\leqslant k<l\leqslant 3, \tag{3.27}$$

where

$$M_1=T_-^{(2,3)},\quad M_2=-T_-^{(1,3)},\quad M_3=T_-^{(1,2)}. \tag{3.28}$$

Since on the solution space $\mathcal{H}$ of Eq. (1.1) we have the identities

$$I_1=\sum_{k=1}^{3}P_1^{(k)},\quad I_2=\sum_{k=1}^{3}P_2^{(k)},\quad I_3=\sum_{k=1}^{3}P_3^{(k)}, \tag{3.29}$$

the harmonic oscillator is class I, where the $P_i^{(k)}$'s are defined in Eqs. (3.27).

Case 3: The remaining constraint (3.5c) for the Coulomb potential has the form

$$\begin{aligned}
&(i/2)\ddot{a}\eta^2+i\dot{a}(-\lambda/\eta)+\tfrac{3}{2}\ddot{a}-i\dot{d}+i\dot{a}_{12}(-\lambda\eta_2/\eta)\\
&\quad+i\dot{a}_{31}(-\lambda\eta_1/\eta)+i\dot{a}_{23}(-\lambda\eta_1/\eta)-(\ddot{c}_1\eta_1+\ddot{c}_2\eta_2\\
&\quad+\ddot{c}_3\eta_3)-(\lambda\eta_1/\eta^3)(c_1+a_{31})-(\lambda\eta_2/\eta_3)(c_2+a_{12})\\
&\quad-(\lambda\eta_3/\eta^3)(c_3+a_{23})=0,
\end{aligned} \tag{3.30}$$

which is valid only if

$$\left.\begin{aligned}
&\dot{a}=0,\\
&\dot{a}_{kl}=0,\quad 1\leqslant k<l\leqslant 3,\\
&c_1+a_{31}=c_2+a_{12}=c_3+a_{23}=0,\\
&\dot{d}=0.
\end{aligned}\right. \tag{3.31}$$

When the system of equations (3.17) and (3.31) are integrated, we get

$$\left.\begin{aligned}
&a=\alpha_1,\\
&\left.\begin{aligned}a_{kl}&=\alpha_{kl},\\ c_{kl}&=\gamma_{kl},\end{aligned}\right\}\quad 1\leqslant k<l\leqslant 3,\\
&c_1=-\alpha_{13},\quad c_2=-\alpha_{12},\quad c_3=-\alpha_{23},\\
&d=\alpha_2.
\end{aligned}\right. \tag{3.32}$$

A basis for $\mathcal{D}/q$ is provided by the symmetry generators

$$\left.\begin{aligned}
&M_k,E,\quad 1\leqslant k\leqslant 3,\\
&M_j^2,\ [M_k,M_l]_+,\quad 1\leqslant j,\ k<l\leqslant 3,\\
&T=\Delta_3+2\lambda/\eta,
\end{aligned}\right\} \tag{3.33a}$$

which can be identified on the solution space $\mathcal{H}$ as the time displacement operator $T'=-i\partial_t+\kappa$. Finally, we obtain the three components of the Runge–Lenz vector[13,14]

$$\mathbf{A}=\tfrac{1}{2}(\mathbf{M}\times\mathbf{C}_1-\mathbf{C}_1\times\mathbf{M})-\lambda\boldsymbol{\eta}/\eta, \tag{3.33b}$$

where $\mathbf{C}_1$ is given in Eq. (2.8) and $\mathbf{M}$ by Eq. (2.5).

Since the components of $\mathbf{A}$ cannot be written as symmetric products of members of the kinematic algebra $\mathcal{G}$, $\mathcal{D}/q$ for the Coulomb potential is class II.

Case 4: For the Coulombic oscillator, we have the following condition:

$$\begin{aligned}
&-i(\ddot{a}+4\omega^2a)\eta^2/2+\dot{a}(-i\lambda/\eta)+i\dot{d}-\tfrac{3}{2}\ddot{a}\\
&\quad+\sum_{k=1}^{3}(\ddot{c}_k+\omega^2c_k)\eta_k+(c_1+a_{31})(\lambda\eta_1/\eta^3)\\
&\quad+(c_2+a_{12})(\lambda\eta_2/\eta_3)+(c_3+a_{23})(\lambda\eta_3/\eta^3)\\
&\quad+i\dot{a}_{12}(\lambda\eta_2/\eta)+ia_{31}(\lambda\eta_1/\eta)+ia_{23}(\lambda\eta_3/\eta)=0,
\end{aligned}$$

which is consistent for all values (η_1,η_2,η_3) only if

$$\left.\begin{aligned}
&\dot{a}=0,\quad \ddot{a}+4\omega^2\dot{a}=0,\\
&\ddot{c}_k+\omega^2c_k=0,\quad 1\leqslant k\leqslant 3,\\
&c_1+a_{31}=c_2+a_{12}=c_3+a_{23}=0,\\
&\dot{a}_{12}=\dot{a}_{31}=\dot{a}_{23}=0,\\
&\dot{d}-\tfrac{3}{2}\ddot{a}=0.
\end{aligned}\right. \tag{3.34}$$

When Eqs. (3.20) and (3.34) are integrated, we find that

$$\left.\begin{aligned}
&a=\alpha_1,\\
&c_k=0,\quad 1\leqslant k\leqslant 3,\\
&a_{12}=a_{31}=a_{23}=0,\\
&d=\alpha_2,\\
&c_{kl}=\gamma_{kl},\quad 1\leqslant k<l\leqslant 3.
\end{aligned}\right. \tag{3.35}$$

The second order symmetry algebra is spanned by the basis

$$\left.\begin{aligned}
&M_k,E,\quad 1\leqslant k\leqslant 3,\\
&T=-\partial_t+\kappa,\\
&M_j^2,\ [M_k,M_l]_+,\quad 1\leqslant j\ k<l\leqslant 3.
\end{aligned}\right\} \tag{3.36}$$

This potential is class I, and $\mathcal{D}/q$ is not exceptional consisting only of $\mathcal{G}\cup o^2(3)$, the minimal $\mathcal{D}/q$ algebra.

Case 5: For the general potential $W(\eta)$, where $W(\eta)$ is none of cases 1, 2, 3, or 4, we have for Eq. (3.5c)

$$\begin{aligned}
&\frac{i}{2}\ddot{a}\eta^2+i\dot{a}(2W+\eta W')-i\dot{d}+\tfrac{3}{2}\ddot{a}\\
&\quad-\sum_{k=1}^{3}c_k(\eta_kW'/\eta)-\sum_{k=1}^{3}\ddot{c}_k\eta_k=0.
\end{aligned} \tag{3.37}$$

Three special situations can be identified:

$$\text{(a) } \dot{a}\neq 0. \tag{3.38a}$$

Therefore,

$$2W+\eta W'=0, \tag{3.39a}$$

whose solution is

$$W(\eta)=\lambda/\eta^2, \tag{3.39b}$$

the centrifugal potential. The remaining conditions on the coefficients a, b, c, and d are

$$\left.\begin{aligned}
&\ddot{a}=0,\\
&-i\dot{d}+\tfrac{3}{2}\ddot{a}=0,\\
&c_k=0,\quad 1\leqslant k\leqslant 3.
\end{aligned}\right\} \tag{3.38b}$$

Integrating the system of equations (3.22) and (3.38), we obtain

$$\left.\begin{aligned} &a = \alpha_3 t^2 + \alpha_2 t + \alpha_1, \\ &c_{kl} = \gamma_{kl}, \quad 1 \leqslant k < l \leqslant 3, \\ &d = -3i\alpha_3 t + \alpha_4. \end{aligned}\right\} \tag{3.40}$$

A basis for $\mathcal{D}/q$ is provided by the elements

$$\left.\begin{aligned} &M_j, E, \\ &M_j^2, [M_k, M_l]_+, \end{aligned}\right\} \quad 1 \leqslant j,\ k < l \leqslant 3, \tag{3.41a}$$

$$\left.\begin{aligned} &S_1 = \Delta_3 - 2\lambda/\eta^2, \\ &S_2 = t\Delta_3 - i\boldsymbol{\eta}\cdot\boldsymbol{\partial} - 2t\lambda/\eta^2, \\ &S_3 = t^2\Delta_3 - 2it\,\boldsymbol{\eta}\cdot\boldsymbol{\partial} - 3it - \eta^2 - 2t^2\lambda/\eta^2. \end{aligned}\right\} \tag{3.41b}$$

On the solution space $\mathscr{H}$ we may write Eq. (3.41b) in terms of the constant potential, Heisenberg–Weyl algebra (2.8):

$$\begin{aligned} I_1 &= -\tfrac{1}{2}(S_2 - \tfrac{3}{2}iE) = \tfrac{1}{4}\sum_{k=1}^{3}[C_1^{(k)}, C_2^{(k)}]_+ + t\lambda/\eta^2, \\ I_2 &= -\tfrac{1}{4}(S_1 - S_3) = \tfrac{1}{4}\sum_{k=1}^{3}(C_1^{(k)2} - C_2^{(k)2}) \\ &\quad + (1 - t^2)(\tfrac{1}{2}\lambda/\eta^2), \\ I_3 &= -\tfrac{1}{4}(S_1 + S_3) = \tfrac{1}{4}\sum_{k=1}^{3}(C_1^{(k)2} + C_2^{(k)2}) \\ &\quad + (1 + t^2)(\tfrac{1}{2}\lambda/\eta^2). \end{aligned} \tag{3.42}$$

Therefore, the centrifugal potential is class I.

(b) If

$$\ddot{a} = -4\omega^2\dot{a}, \quad \dot{a} \neq 0, \tag{3.43a}$$

then the potential $W(\eta)$ must satisfy

$$\eta W' + 2W = 2\omega^2\eta^2, \tag{3.44a}$$

which has as its solution the radial oscillator potential

$$W(\eta) = \omega^2\eta^2/2 + \lambda/\eta^2. \tag{3.44b}$$

Now Eq. (3.37) implies that

$$\left.\begin{aligned} &-i\dot{d} + \tfrac{1}{2}\ddot{a} = 0, \\ &c_k = 0, \quad 1 \leqslant k \leqslant 3. \end{aligned}\right\} \tag{3.43b}$$

Integrating the system (3.22) and (3.43), one obtains

$$\left.\begin{aligned} &a = (\alpha_1/2\omega)\sin 2\omega t - (\alpha_2/2\omega)\cos 2\omega t + (\alpha_3/2\omega), \\ &c_{kl} = \gamma_{kl} \quad 1 \leqslant k < l \leqslant 3, \\ &d = -\tfrac{1}{2}i(\alpha_1 \cos 2\omega t + \alpha_2 \sin 2\omega t) + \alpha_4. \end{aligned}\right\} \tag{3.45}$$

A basis for $\mathcal{D}/q$ is

$$\left.\begin{aligned} &M_j, E, \\ &M_j^2, [M_k, M_l]_+, \end{aligned}\right\} \quad 1 \leqslant j,\ k < l \leqslant 3,$$

$$\begin{aligned} S_1 &= \frac{1}{2\omega}(-2i\omega\cos 2\omega t\,\boldsymbol{\eta}\cdot\boldsymbol{\partial} + \sin 2\omega t\,\Delta_3 \\ &\quad + (\omega^2\eta^2 - 2\lambda/\eta^2)\sin 2\omega t - 3i\omega\cos 2\omega t), \\ S_2 &= -\frac{1}{2\omega}(2i\omega\sin 2\omega t\,\boldsymbol{\eta}\cdot\boldsymbol{\partial} + \cos 2\omega t\,\Delta_3 \\ &\quad + (\omega^2\eta^2 - 2\lambda/\eta^2)\cos 2\omega t + 3i\omega\sin 2\omega t), \\ S_3 &= \frac{1}{2\omega}(\Delta_3 - \omega^2\eta^2 - 2\lambda/\eta^2). \end{aligned} \tag{3.46}$$

On the solution space $\mathscr{H}$, S_1, S_2, S_3 may be interpreted as

$$\begin{aligned} I_1 &= -\tfrac{1}{2}S_1 = \tfrac{1}{4}\Big[\sum_{k=1}^{3}[C_1^{(k)}, C_2^{(k)}]_+ + (2\omega^{-1}\lambda/\eta^2)\Big], \\ I_2 &= \tfrac{1}{2}S_2 = \tfrac{1}{4}\Big[\sum_{k=1}^{3}(C_1^{(k)2} - C_2^{(k)2}) - (2\omega^{-1}\lambda/\eta^2)\Big], \\ I_3 &= -\tfrac{1}{2}S_3 = \tfrac{1}{4}\Big[\sum_{k=1}^{3}(C_1^{(k)2} + C_2^{(k)2}) + (2\omega^{-1}\lambda/\eta^2)\Big], \end{aligned} \tag{3.47}$$

where $C_1^{(k)}$, $C_2^{(k)}$ are elements of the Heisenberg–Weyl algebra, [Eq. (2.9)], for the harmonic oscillator. The radial oscillator is class I.

(c) If

$$\dot{a} = 0, \tag{3.48a}$$

then $W(\eta)$ is arbitrary [except for cases 1, 2, 3, 5(a), and 5(b)]. Furthermore,

$$\left.\begin{aligned} &c_k = 0, \quad 1 \leqslant k \leqslant 3, \\ &\dot{d} = 0. \end{aligned}\right\} \tag{3.48b}$$

Therefore,

$$\left.\begin{aligned} &a = \alpha_1, \\ &c_{kl} = \gamma_{kl}, \quad 1 \leqslant k < l \leqslant 3, \\ &d = \alpha^2. \end{aligned}\right\} \tag{3.49}$$

A basis for $\mathcal{D}/q$ consists of

$$\left.\begin{aligned} &\left.\begin{aligned} &M_j, E, \\ &M_j^2, [M_k, M_l]_+, \end{aligned}\right\} \quad 1 \leqslant j,\ k < l \leqslant 3, \\ &T = \Delta_3 - 2W. \end{aligned}\right\} \tag{3.50}$$

T on the solution space $\mathscr{H}$ of Eq. (1.1) becomes the time displacement operator. Thus, potentials of this type are class I, and $\mathcal{D}/q$ is the minimal algebra.

The general features of the $\mathcal{D}/q$ algebras have been outlined, and they will be further analyzed, and the structure of $\mathcal{G}^*$ identified in Sec. IV.

IV. DYNAMICAL AND DEGENERACY ALGEBRAS

At the beginning of the previous section it was remarked that in general we have the chain of algebras $\mathcal{G} \subseteq \mathcal{G}^* \subseteq \mathcal{D}/q$, where $\mathcal{G}^*$ is the largest subset of $\mathcal{D}/q$ closing under the Lie product. For the centrifugal potential, radial oscillator, and general potential, we have shown that $\mathcal{G} = \mathcal{G}^*$ in Sec. III. Therefore, the maximal Lie symmetry algebra of second order differential operators in these instances is equivalent to the kinematic algebra. The algebra $\mathcal{G}^*$ is larger than $\mathcal{G}$ only for the harmonic oscillator, the constant potential, and the Coulomb potential. In the first two cases, $\mathcal{G}_0^* = \{P_1^{(j)}, P_2^{(j)}, P_3^{(j)}, T_{\pm}^{(k,l)}, S_{\pm}^{(k,l)}, C_1^{(j)}, C_2^{(j)}, E;\ 1 < j,\ k < l < 3\}$; in the last case, $\mathcal{G}_c^* = \{M_j, A_j;\ 1 \leqslant j \leqslant 3\}$.

The commutation relations satisfied by the elements of $\mathcal{G}_0^*$ are given in the Appendix, and those for $\mathcal{G}_c^*$ may be found in Miller[24] or Gilmore.[27] $\mathcal{G}_0^*$ has the structure $\mathrm{sp}(6,\mathbb{R})\square w_3$ and hence $\mathcal{G}_0^* = \mathcal{D}_0$, the dynamical algebra for the harmonic oscillator in three dimensions. To demonstrate this it is instructive to transform the differential operators of $\mathcal{G}_0^*$ into the corresponding operators in the boson representation. This is achieved by first going to the Heisenberg picture[22] by letting $t \to 0$ in Eq. (3.27) and replacing ∂_t by H,

TABLE III. Analysis of $\mathscr{D}/q$ algebras for o(3)-invariant potentials.

Potential $W(\eta)$	Class	$\mathscr{D}/q$	Degeneracy algebra $\mathscr{D}$	Dynamical algebra $\mathscr{D}$
Harmonic ocsillator	I	$(\mathrm{sp}(6,\mathbb{R})\square w_3)\cup \mathrm{o}^2(3)$ $\supset \mathrm{u}(3)\supset \mathrm{o}(3)$ $\supset \mathrm{sl}(2,\mathbb{R})$	u(3)	$\mathrm{sp}(6,\mathbb{R})\square w_3$
Constant κ	I	isomorphic to oscillator		
Coulombic $-\lambda/\eta$	II	$\mathrm{o}(3)\cup\{\mathbf{A}\}\cup\mathrm{o}^2(3)$ [a]	$\mathrm{so}(4)\supset\mathrm{so}(3)$ $(E<0)$	$\mathrm{so}(4,2)$ $\supset\mathrm{so}(2,1)$ $\supset\mathrm{so}(4)$
Radial oscillator $\omega^2\eta^2/2$ $+\lambda/\eta^2$	I	$(\mathrm{o}(3)\oplus \mathrm{sl}(2,\mathbb{R}))\cup\mathrm{o}^2(3)$	o(3)	$\mathrm{o}(3)\oplus\mathrm{sl}(2,\mathbb{R})$
Centrifugal λ/η^2	I	isomorphic to radial oscillator		
General $W(\eta)$, none of the above	I	$(\mathrm{o}(3)\oplus t_1)\cup\mathrm{o}^2(3)$	o(3)	

[a] $\{\mathbf{A}\}$ is the set of Runge–Lenz vectors [Eq. (3.33b)].

the Hamiltonian.[6,21] In this picture, the form of the operators in Eq. (3.28) is retained with $\mathscr{C}_1^{(k)}$, $\mathscr{C}_1^{(k)}$ replacing $C_1^{(k)}$, $C_2^{(k)}$, respectively, where $\mathscr{C}_1^{(k)} = -i\omega^{-1/2}\partial_k$ and $\mathscr{C}_2^{(k)} = -\omega^{1/2}\eta_k$. The commutation rules are preserved.[21,22] Now if we let $a_k = (-i\mathscr{C}_1^{(k)} + \mathscr{C}_2^{(k)})/\sqrt{2}$ and $a_k^* = (i\mathscr{C}_1^{(k)} + \mathscr{C}_2^{(k)})/\sqrt{2}$, we have $[a_k, a_l^*] = \delta_{kl}E$. Transforming $\mathscr{D}_0$ into the boson model we have the operators

$$\mathscr{P}_1^{(j)} = \frac{i}{4}(a_j a_j - a_j^* a_j^*), \quad \mathscr{P}_2^{(j)} = \frac{1}{4}(a_j a_j + a_j^* a_j^*),$$

$$\mathscr{P}_3^{(j)} = \frac{1}{4}(a_j a_j^* + a_j^* a_j),$$

$$\mathscr{T}_+^{(k,l)} = i(a_k a_l - a_k^* a_l^*), \quad \mathscr{T}_-^{(k,l)} = i(a_k a_l^* - a_l a_k^*),$$

$$\mathscr{A}_+^{(k,l)} = a_k a_l^* + a_l a_k^*, \quad \mathscr{A}_-^{(k,l)} = -(a_k a_l + a_k^* a_l^*), \qquad (4.1)$$

$$\mathscr{C}_1^{(k)} = i(a_k - a_k^*)/\sqrt{2}, \quad \mathscr{C}_2^{(k)} = (a_k + a_k^*)/\sqrt{2},$$

$$E = 1, \quad 1\leqslant j,k<l\leqslant 3.$$

The commutation rules obeyed by these symmetry operators of $\mathscr{G}_0^*$ are identical to those in the Appendix.[21,22] It is straightforward now to show[12] that $\mathscr{G}_0^*$ is isomorphic to $\mathrm{sp}(6,\mathbb{R})\square w_3$.

The kinematical subalgebra $\mathscr{G}_0$ may be readily identified. The Heisenberg–Weyl algebra $w_3 = \{\mathscr{C}_1^{(k)}, \mathscr{C}_2^{(k)}, E: 1\leqslant k\leqslant 3\}$; $\mathrm{o}(3) = \{\mathscr{T}_-^{(k,l)}: 1\leqslant k<l\leqslant 3\}$ where, by Eq. (3.28), $M_1 = \mathscr{T}_-^{(1,2)}$, $M_2 = -\mathscr{T}_-^{(1,3)}$, and $M_3 = \mathscr{T}_-^{(1,2)}$. The sl(2,ℝ) algebra is realized by the operators

$$\mathscr{I}_1 = \sum_{j=1}^{3}\mathscr{P}_1^{(j)}, \quad \mathscr{I}_2 = \sum_{j=1}^{3}\mathscr{P}_2^{(j)},$$

$$\mathscr{I}_3 = \sum_{j=1}^{3}\mathscr{P}_3^{(j)} = \frac{1}{2\omega}H, \qquad (4.2)$$

and is locally isomorphic to so(2,1) [$\sim$sp(2)$\sim$su(1.1)].[20] It is called the spectrum generating algebra of the oscillator.[12] Utilizing Eq. (4.2) and the Appendix, the nine symmetry generators $\overline{\mathscr{D}}_0 = \{\mathscr{P}_3^{(j)}, \mathscr{T}_-^{(k,l)}, \mathscr{A}_+^{(k,l)}: 1\leqslant j, k<l\leqslant 3\}$ can be shown to satisfy the Heisenberg equation[22] such that $dX/dt = 0$, for all $X\in\overline{\mathscr{D}}_0$, the elements of $\overline{\mathscr{D}}_0$ are constants of the motion, $\overline{\mathscr{D}}_0$ is the degeneracy algebra for the oscillator, and $\overline{\mathscr{D}}_0$ is isomorphic to u(3). As commutation relations between the Heisenberg and Schrödinger pictures are preserved,[22] the structure of the algebra $\overline{\mathscr{D}}_0$ in the latter is identical.

For bound states ($E<0$), the degeneracy algebra for the Coulomb problem is o(4).[23–25] The six constants of the motion $\overline{\mathscr{D}}_c = \{M_j, A_j: 1\leqslant j\leqslant 3\}$ provide a basis for o(4).[23–25] For the continuum states ($E<0$), the algebra $\overline{\mathscr{D}}_c$ is isomorphic to o(3,1).[24,25] Although $\overline{\mathscr{D}}_c$ is obtained directly from our calculations, the spectrum generating algebra so(2,1) may be realized by noting the relationship with the so(2,1) [$\sim$sl(2,ℝ)] algebra of the harmonic oscillator.[26] The full dynamical algebra for the Coulomb potential is[14] o(4,2); a boson representation of this algebra has been obtained.[27]

The dynamical group of the radial oscillator (and the centrifugal potential) is $\mathscr{D}_r = \mathscr{G}$. The spectrum generating algebra is the sl(2,ℝ) $\sim$so(2,1) subalgebra of $\mathscr{G}$. The degeneracy algebra $\overline{\mathscr{D}}_r$ is o(3). For a treatment of the radial oscillator see Moshinsky *et al.*[26]

$\overline{\mathscr{D}}_w$ for the general potential $W(\eta)$ is also o(3).

It is mentioned that the elements of $\mathscr{D}/q$ are important for the classification of separable coordinate systems for Eq. (1.4).[21]

The above discussion has been summarized in Table III. We have illustrated a general technique for systematically computing all kinematic and dynamical symmetries associated with, in principle, any Schrödinger equation. Specifically, for rotationally invariant, two-body Schrödinger equations, explicit forms of potentials, their degeneracy algebras, and in certain instances their dynamical algebras have been identified and for the oscillator and Coulomb problems agree with those found by other means. The algebraic approach has the advantage of providing a general framework within which quantum mechanical problems may be analyzed.

ACKNOWLEDGMENTS

This work was supported by a grant from the National Research Council of Canada. The author also wishes to express his appreciation for the kind hospitality of the School of Mathematics of the University of Minnesota and to Willard Miller, Jr. for several helpful discussions.

APPENDIX

Commulation relations for $\mathscr{D}_0$ are as follows:

$$[P_1^{(k)}, P_2^{(l)}] = -iP_3^{(k)}\delta_{kl}, \quad [P_2^{(k)}, P_3^{(l)}] = iP_1^{(k)}\delta_{kl}, \quad [P_3^{(k)}, P_1^{(l)}] = iP_2^{(k)}\delta_{kl}, \quad 1\leqslant k, l\leqslant 3,$$

$$[P_1^{(j)},T_{\pm}^{(k,l)}] = \mp \frac{i}{2}(\delta_{jl}T_{\mp}^{(j,k)} + \delta_{jk}T_{\mp}^{(j,l)}),$$

$$[P_2^{(j)},T_{\pm}^{(k,l)}] = \frac{i}{2}(\delta_{jl}S_{\pm}^{(j,k)} + \delta_{jk}S_{\pm}^{(j,l)}),$$

$$[P_3^{(j)},T_{\pm}^{(k,l)}] = \frac{i}{2}(\delta_{jl}S_{\mp}^{(j,k)} \pm \delta_{jk}S_{\mp}^{(j,l)}),$$

$$[P_1^{(j)},S_{\pm}^{(k,l)}] = -\frac{i}{2}(\delta_{jl}S_{\mp}^{(j,k)} + \delta_{jk}S_{\mp}^{(j,l)}),$$

$$[P_2^{(j)},S_{\pm}^{(k,l)}] = \pm \frac{i}{2}(\delta_{jl}T_{\pm}^{(j,k)} + \delta_{jk}T_{\pm}^{(j,l)}),$$

$$[P_3^{(j)},S_{\pm}^{(k,l)}] = \pm \frac{i}{2}(\delta_{jl}T_{\mp}^{(j,k)} + \delta_{jk}T_{\mp}^{(j,l)}), \quad 1\leqslant j, \quad k<l\leqslant 3,$$

$$[T_{\pm}^{(i,j)},T_{\pm}^{(k,l)}] = -i(\delta_{jk}T_{-}^{(i,l)} + \delta_{il}T_{-}^{(j,k)} \pm \delta_{jl}T_{-}^{(i,k)} \pm \delta_{ik}T_{-}^{(j,l)}),$$

$$[T_{+}^{(i,j)},T_{-}^{(k,l)}] = i(-\delta_{jk}T_{+}^{(i,l)} + \delta_{il}T_{+}^{(j,k)} + \delta_{jl}T_{+}^{(i,k)} - \delta_{ik}T_{+}^{(j,l)}),$$

$$[T_{+}^{(i,j)},T_{-}^{(i,j)}] = 4i(P_1^{(i)} - P_1^{(j)}),$$

$$[S_{\pm}^{(i,j)},S_{\pm}^{(k,l)}] = \pm i(\delta_{jk}T_{-}^{(i,l)} + \delta_{il}T_{-}^{(j,k)} + \delta_{jl}T_{-}^{(i,k)} + \delta_{ik}T_{-}^{(j,l)}),$$

$$[S_{+}^{(i,j)},S_{-}^{(k,l)}] = -i(\delta_{jk}T_{+}^{(i,l)} + \delta_{il}T_{+}^{(j,k)} + \delta_{jl}T_{+}^{(i,k)} + \delta_{ik}T_{-}^{(j,l)}),$$

$$[S_{+}^{(i,j)},S_{-}^{(i,j)}] = -4i(P_1^{(i)} + P_1^{(j)}),$$

$$[T_{+}^{(i,j)},S_{\pm}^{(k,l)}] = -i(\delta_{jk}S_{\mp}^{(i,l)} + \delta_{il}S_{\mp}^{(j,k)} + \delta_{jl}S_{\mp}^{(i,k)} + \delta_{ik}S_{\mp}^{(j,l)}),$$

$$[T_{+}^{i,j},S_{+}^{(i,j)}] = -4i(P_2^{(i)} + P_2^{(j)}), \quad [T_{+}^{(i,j)},S_{-}^{(i,j)}] = -4i(P_3^{(i)} + P_3^{(j)}),$$

$$[T_{-}^{(i,j)},S_{\pm}^{(k,l)}] = i(-\delta_{jk}S_{\pm}^{(i,l)} + \delta_{il}S_{\pm}^{(j,k)} - \delta_{jk}S_{\pm}^{(i,k)} + \delta_{ik}S_{\pm}^{(j,l)}),$$

$$[T_{-}^{(i,j)},S_{+}^{(i,j)}] = -4i(P_3^{(i)} - P_3^{(j)}), \quad [T_{-}^{(i,j)},S_{-}^{(i,j)}] = -4i(P_2^{(i)} - P_2^{(j)}), \quad 1\leqslant i<j\leqslant 3$$

$$[P_1^{(j)},C_1^{(k)}] = -\frac{i}{2}C_1^{(j)}\delta_{jk}, \quad [P_2^{(j)},C_1^{(k)}] = \frac{i}{2}C_2^{(j)}\delta_{jk}, \quad [P_3^{(j)},C_1^{(k)}] = -\frac{i}{2}C_2^{(j)}\delta_{jk},$$

$$[P_1^{(j)},C_2^{(k)}] = \frac{i}{2}C_2^{(j)}\delta_{jk}, \quad [P_2^{(j)},C_2^{(k)}] = \frac{i}{2}C_1^{(j)}\delta_{jk}, \quad [P_3^{(j)},C_2^{(k)}] = -\frac{i}{2}C_1^{(j)}\delta_{jk}, \quad 1\leqslant j,k\leqslant 3,$$

$$[T_{\pm}^{(i,j)},C_1^{(k)}] = \mp i(\pm\delta_{jk}C_1^{(i)} + \delta_{ik}C_1^{(j)}),$$

$$[T_{\pm}^{(i,j)},C_2^{(k)}] = i(\pm\delta_{jk}C_2^{(i)} + \delta_{ik}C_2^{(j)}),$$

$$[S_{\pm}^{(i,j)},C_1^{(k)}], = \mp i(\delta_{jk}C_2^{(i)} + \delta_{ik}C_2^{(j)}),$$

$$[S_{\pm}^{(i,j)},C_2^{(k)}] = i(\delta_{jk}C_1^{(i)} + \delta_{ik}C_1^{(j)}), \quad 1\leqslant k, \quad i<j\leqslant 3,$$

$$[C_1^{(k)},C_2^{(l)}] = iE\delta_{kl}, \quad [C_1^{(k)},C_1^{(l)}] = [C_2^{(k)},C_2^{(l)}] = 0, \quad 1\leqslant k,l\leqslant 3.$$

[1] A. Messiah, *Quantum Mechanics*, Vols. 1 and 2, 3rd edition (Wiley, New York, 1961).

[2] A.S. Davydov, *Quantum Mechanics* (Addison-Wesley, Reading, Mass., 1968).

[3] E.B. Wilson, Jr., J.C. Decius, and P.C. Cross, *Molecular Vibrations* (McGraw-Hill, New York, 1955).

[4] See, for example, M. Moshinsky, SIAM (Soc. Ind. Appl. Math.) J. Appl. Math. **25**, 193 (1973).

[5] (a) C.E. Wulfman and Y. Takahata, J. Chem. Phys. **47**, 488 (1967); (b) C.E. Wulfman, *Group Theory and Its Applications*, Vol. 2, edited by E.M. Loebl (Academic, New York, 1975); (c) M. Moshinsky, Phys. Rev. **126**, 1880 (1962).

[6] C.P. Boyer, Helv. Phys. Acta. **47**, 589 (1974).

[7] U. Niederer, Helv. Phys. Acta. **45**, 802 (1972).

[8] U. Niederer, Helv. Phys. Acta. **46**, 191 (1973).

[9] $\oplus$ denotes direct sum; $\square$ denotes semidirect sum; sl(2,$\mathbb{R}$), o(3), and w_3 are the algebras for the real special linear group, the orthogonal group, and the real Heisenberg–Weyl group in three-dimensions, respectively.

[10] C.P. Boyer, R.T. Sharp, and P. Winternitz, J. Math. Phys. **17**, 1439 (1976).

[11] In general, the Hamiltonian, the Casimir operators, and a maximal Abelian subalgebra of the constants of the motion are selected to specify the dynamical state.

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

[13] W. Miller, Jr., *Symmetry Groups and Their Applications* (Academic, New York, 1972).

[14] A.O. Barut, *Dynamical Groups and Generalized Symmetries in Quantum Theory* (University of Canterbury, Christchurch, 1972).

[15] The letters n and l are commonly used for the principal quantum number and the angular momentum quantum number, respectively, in the treatment of bound states of the hydrogen atom.

[16] L.I. Schiff, *Quantum Mechanics*, 3rd edition (McGraw-Hill, New York, 1968).

[17] (a) J.O. Hirschfelder and J.S. Dahler, Proc. Natl. Acad. Sci. U.S.A. **42**, 363 (1956); (b) D.W. Jepsen and J.O. Hirschfelder, Proc. Natl. Acad. Sci. U.S.A. **45**, 249 (1959).

[18] See Ref. 13, pp. 196 ff.

[19] For a description of these algebras see Ref. 6, or Ref. 13. R. Gilmore, *Lie Groups, Lie Algebras and Some of Their Applications* (Wiley, New York, 1974). The Heisenberg–Weyl algebra w_3 is an algebra isomorphic to one

formed by the three coordinate variables and their conjugate momenta, i.e., $[x_i, p_j] = i\hbar\delta_{ij}$, $[x_i, x_j] = [p_i, p_j] = 0$, $1 \leqslant i, j \leqslant 3$.

[20]The symbol $\boldsymbol{\eta}\cdot\boldsymbol{\partial} = \Sigma_{k=1}^{3} \eta_k \partial_k$.

[21](a) W. Miller, Jr., "Symmetry, Separation of Variables, and Special Functions," in *Theory and Application of Special Functions* (Academic Press, New York, 1975); (b) W. Miller, Jr., *Symmetry and Separation of Variables* (Addison-Wesley, Reading, Mass., 1977).

[22]See Ref. 1, pp. 314 ff.

[23]See Ref. 12, pp. 300 ff.

[24]See Ref. 13, pp. 376 ff.

[25]See R. Gilmore, p. 428 in Ref. 19.

[26]M. Moshinsky, T.H. Seligman, and K.B. Wolf, J. Math. Phys. **13**, 901 (1972).

[27]See Ref. 12, p. 307.

On some representations of the Poincaré group on phase space. II[a)]

S. Twareque Ali[b)]

Department of Mathematics and Computer Science, University of Prince Edward Island, Charlottetown, P. E. I., Canada C1A 4P3

(Received 18 July 1979; accepted for publication 11 October 1979)

Representations of the proper Poincaré group on spaces of functions of the phase space variables **q** and **p** have been studied, by starting with the rotation subgroup SO(3), and carrying out an inducing procedure. A complete decomposition theory, for the physically interesting representations, have been developed. Explicit computations of the forms of the generators and certain systems of imprimitivity have been carried out. The role of the Newton–Wigner position and momentum operators, in the setting of phase space representations, has been discussed. The work is a continuation of an earlier paper on phase space representations, inspired ultimately by the possibility of developing a consistent relativistic quantum mechanics for single particles on phase space.

1. INTRODUCTION

In a previous paper[1] (hereafter referred to as I) we constructed representations of the Poincaré group $\mathscr{P}^{\uparrow}_{+}$ on spaces of functions of the phase space variables **q** and **p**. The reducible representations corresponding to classical systems of particles having spin j and mass $m \neq 0$, and hence adequate for the purposes of classical statistical mechanics, were constructed by an inducing procedure, starting from the subgroup $T \otimes \mathrm{SO}(3)$ of time translations and spatial rotations. The irreducible quantum mechanical representations, again for particles with spin j and mass $m \neq 0$, were constructed by means of a unitary transformation, starting with the standard Wigner representations. These latter representations were then extended to unitary reducible representations on an enlarged Hilbert space, which is more convenient for the purposes of deriving equations of motion in the presence of electromagnetic fields (cf. Ref. 2).

The present paper is a continuation of I. We first construct the quantum mechanical representations described in I by means of an inducing procedure, starting this time from the subgroup SO(3). We obtain thereby both the reducible representations on the enlarged Hilbert space as well as their irreducible sectors on the projected spaces, getting thus a deeper understanding of the unitary map $\tilde{\psi} \mapsto \psi_e$ used in I (cf. Theorem 1 in that paper). We also obtain a decomposition theory for representations built on the relativistic phase space Ω (see Sec. 2 for definition). These representations become useful whenever one wishes to construct a field theory on phase space (cf. Ref. 3). We construct, in addition, the infinitesimal generators for our representations, and look at some systems of imprimitivity.

The material in the sequel is organized into four sections. Section 2 deals with the inducing procedure used to construct the reducible representation, of which the quantum mechanical representations developed in I, in a rather *ad hoc* manner, from irreducible components. We obtain first an induced representation starting from the subgroup SO(3) and then develop a decomposition theory for this representation. We identify in it the quantum mechanical representations found in I and get an alternative proof of the isometry $\tilde{\psi} \mapsto \psi_e$, first derived in Ref. 4. In Sec. 3 we construct the infinitesimal generators for some of the representations obtained in Sec. 2 and briefly discuss some systems of imprimitivity associated with these representations as well as some useful resolutions of the identity. We conclude in Sec. 4 with a brief comment on some of the physical implications of using phase space representations of quantum mechanics and of group contractions to go from our representations to representations of the Galilei group. Some possibilities of further work are also suggested.

2. REPRESENTATIONS INDUCED FROM SO(3)

Much of the notation followed in this paper is the same as that introduced in I. Results derived in that paper will be used with our further comment.

Consider the subgroup SO(3) of $\mathscr{P}^{\uparrow}_{+}$ and let

$$\mathscr{V} = \mathscr{P}^{\uparrow}_{+}/\mathrm{SO}(3) \tag{2.1}$$

be the associated left coset. For $R \in \mathrm{SO}(3)$, let $\mathscr{D}^{j}(R)$ be the usual $2j+1$ dimensional unitary spinor representation on the Hilbert space $\mathscr{K}^{j}$. The "velocity hyperboloid"

$$\mathscr{V}^{c} = \{u \mid u_0 = \sqrt{\mathbf{u}^2 + c^2} = \gamma c\}, \tag{2.2}$$

can be used to parametrize pure Lorentz transformations, and indeed we have the (topological) isomorphism

$$\mathscr{V} \simeq \mathbb{R}^4 \times \mathscr{V}^{c}. \tag{2.3}$$

Setting

$$\Omega = \mathbb{R}^4 \times \mathscr{V}^{c}, \tag{2.4}$$

we see that Ω carries the invariant measure

$$d\Omega = d^4q \frac{d^3\mathbf{u}}{u_0}. \tag{2.5}$$

Let

$$\mathscr{H} = L^2(\Omega, d\Omega) \otimes \mathscr{K}^{j}, \tag{2.6}$$

and denote elements in $\mathscr{H}$ by f.

The representation $\mathscr{U}^{j}$ of SO(3) can be seen (using the methods of I) to induce on $\mathscr{H}$ the following unitary representation of $\mathscr{P}$

a)Work supported in part by the NSERC.

b)Present address: Institut für Theoretische Physik der Technischen Universität Clausthal, 3392 Clausthal-Zellerfeld, Fed. Rep. of Germany.

 0022-2488/80/040818-12$1.00

$$(U(a,\Lambda)f)(q,u) = \mathscr{D}^j(\Lambda_u^{-1}\Lambda\Lambda_{\Lambda^{-1}u})f(\Lambda^{-1}(q-a),\Lambda^{-1}u), \tag{2.7}$$

where Λ_u is the same as Λ_v introduced in I. This representation is highly reducible, and we proceed immediately, therefore, to examine some of its irreducible components.

It will be useful for what follows to introduce the forward and backward momentum hyperboloids

$$\mathscr{V}_m^{\pm} \equiv \pm m\mathscr{V} = \{p | p_0 = \pm\sqrt{\mathbf{p}^2 + m^2c^2} = \pm m\gamma c\}, \tag{2.8}$$

for particles with mass m and having positive and negative energies, respectively. We shall also use the full velocity (two sheeted) hyperboloid

$$\mathscr{V}_m = \mathscr{V}_m^+ \cup \mathscr{V}_m^-. \tag{2.9}$$

Further, if

$$\Omega_m^{\pm} = \mathbb{R}^4 \times \mathscr{V}_m^{\pm} \tag{2.10}$$

represent the relativistic phase spaces, for positive and negative energies, and if

$$\Omega_m = \Omega_m^+ \cup \Omega_m^-, \tag{2.11}$$

then $\Omega_m^{\pm}$ carry the invariant measures

$$d\Omega_m^{\pm} = d^4q\,\frac{d^3\mathbf{p}}{\pm p_0}, \tag{2.12}$$

while Ω_m carries the invariant measure

$$d\Omega_m = d^4q\,d\nu_m, \tag{2.13}$$

where

$$d\nu_m = 2d^4p\,\delta(p^2 - m^2c^2). \tag{2.14}$$

The sets $\mathscr{V}_m^{\pm}$ themselves carry the invariant measures

$$d\nu_m^{\pm} = \frac{d^3\mathbf{p}}{\pm p_0}, \tag{2.15}$$

while $\mathscr{V}_m$ carries the invariant measure $d\nu_m$.

The unitary irreducible representations of $\mathscr{P}^{\uparrow}_+$ for mass m and spin j are implemented (for positive and negative energies) on the Hilbert spaces

$$\tilde{\mathscr{H}}_{m,j}^{\pm} = L^2(\mathscr{V}_m^{\pm}, d\nu_m^{\pm}) \otimes \mathscr{K}^j, \tag{2.16}$$

by means of the operators $\tilde{\mathscr{V}}_{m,j}^{\pm}(a,\Lambda)$:

$$(\tilde{\mathscr{V}}_{m,j}^{\pm}(a,\Lambda)\tilde{\phi}^{\pm})(k) = \mathscr{D}^j(\Lambda_k^{-1}\Lambda\Lambda_{\Lambda^{-1}k})e^{(i/\hbar)ka}\tilde{\phi}^{\pm}(\Lambda^{-1}k), \tag{2.17}$$

where $\tilde{\phi}^{\pm} \in \tilde{\mathscr{H}}_{m,j}^{\pm}$. If $\mathscr{F}_m$ is the isometric relativistic Fourier transform, we write

$$\tilde{\phi}_m^{\pm} = \mathscr{F}_m\phi_m^{\pm}, \tag{2.18}$$

$$\phi^{\pm}(x) = \frac{1}{h^{3/2}}\int_{\mathscr{V}_m} \tilde{\phi}^{\pm}(k)e^{-(i/\hbar)kx}d\nu_m^{\pm}, \tag{2.19}$$

where $\phi^{\pm}$ are elements of the Hilbert space $\mathscr{H}_{m,j}^{\pm}$. On $\mathscr{H}_{m,j}^{\pm}$ the scalar products are defined by

$$\langle\phi_1^{\pm}|\phi_2^{\pm}\rangle = \frac{i\hbar}{2}\int_{\mathbb{R}^3}\langle\phi_1^{\pm}(x)|\overleftrightarrow{\partial}_0|\phi_2^{\pm}(x)\rangle_j d^3\mathbf{x}, \tag{2.20}$$

where $\langle|\rangle_j$ denotes the scalar product in $\mathscr{K}^j$, and for any scalar functions χ_1 and χ_2 of x:

$$\chi_1(x)\overleftrightarrow{\partial}_0\chi_2(x) = \chi_1(x)\frac{\partial}{\partial x_0}\chi_2(x) - \left(\frac{\partial}{\partial x_0}\chi_1(x)\right)\chi_2(x). \tag{2.21}$$

The two representations $\tilde{V}_{m,j}^{\pm}$ can be combined to form a single (reducible) representation

$$\tilde{V}_{m,j} = (\tilde{V}_{m,j}^+, \tilde{V}_{m,j}^-), \tag{2.22}$$

on

$$\tilde{\mathscr{H}}_{m,j} = \tilde{\mathscr{H}}_{m,j}^+ \oplus \tilde{\mathscr{H}}_{m,j}^- = L^2(\mathscr{V}_m, d\nu_m) \otimes \mathscr{K}^j, \tag{2.23}$$

as

$$(\tilde{V}_{m,j}(a,\Lambda)\tilde{\phi})(k) = \mathscr{D}^j(\Lambda_k^{-1}\Lambda\Lambda_{\Lambda^{-1}k})e^{(i/\hbar)ka}\tilde{\phi}(\Lambda^{-1}k). \tag{2.24}$$

Consider now the spaces

$$\mathscr{H}(\Omega_m^{\pm}) = \mathscr{H}_{m,0}^{\pm} \otimes \tilde{\mathscr{H}}_{m,j}^{\pm}, \tag{2.25}$$

$$\mathscr{H}(\Omega_m) = \mathscr{H}(\Omega_m^+) \oplus \mathscr{H}(\Omega_m^-), \tag{2.26}$$

whose elements are $(2j+1)$ spinor valued functions of q and p. We denote vectors in $\mathscr{H}(\Omega_m^{\pm})$ and $\mathscr{H}(\Omega_m)$, respectively, by $f_m^{\pm}$ and f_m, and note that these Hilbert spaces carry unitary (reducible) representations $U_m^{\pm}$ and U_m, respectively, of $\mathscr{P}^{\uparrow}_+$:

$$(U_m^{\pm}(a,\Lambda)f_m^{\pm})(q,p) = \mathscr{D}^j(\Lambda_p^{-1}\Lambda\Lambda_{\Lambda^{-1}p})f_m^{\pm}(\Lambda^{-1}(q-a),\Lambda^{-1}p), \tag{2.27}$$

$$U_m = (U_m^+, U_m^-). \tag{2.28}$$

We state and prove now our first decomposition theorem for the representation $U(a,\Lambda)$ defined in Eq. (2.7).

Theorem 2.1: The representation $U(a,\Lambda)$ of $\mathscr{P}^{\uparrow}_+$ is the direct sum of two (reducible) subrepresentations

$$U = (U^{\mathrm{I}}, U^{\mathrm{II}}), \tag{2.29}$$

with the decomposition

$$\mathscr{H} = \mathscr{H}^{\mathrm{I}} \oplus \mathscr{H}^{\mathrm{II}}, \tag{2.30}$$

where $\mathscr{H}^{\mathrm{II}}$ consists of a direct integral of spaces corresponding to representations of $\mathscr{P}^{\uparrow}_+$ with vanishing and imaginary mass, while $\mathscr{H}^{\mathrm{I}}$ and U^{I} decompose into the direct integrals (up to isomorphism)

$$\mathscr{H}^{\mathrm{I}} = \int_{m>0}^{\oplus}\mathscr{H}(\Omega_m)d\sigma(m), \tag{2.31}$$

and

$$U^{\mathrm{I}}(a,\Lambda) = \int_{m>0}^{\oplus}U_m(a,\Lambda)d\sigma(m), \tag{2.32}$$

$\sigma(m)$ being the measure with

$$d\sigma(m) = \frac{c^2}{h^{1/2}}m^2dm. \tag{2.33}$$

Proof: Let $f \in \mathscr{H}$ and consider the Fourier transform

$$f(q,u) = \frac{1}{h^2}\int_{\tilde{\mathbb{R}}^4}e^{-(i/\hbar)kq}\tilde{f}(k,u)d^4k, \tag{2.34}$$

where $\tilde{\mathbb{R}}^4$ is the four-dimensional "momentum space." Denoting this Fourier isometry by $\mathscr{F}$ we have

$$\mathscr{F}f = \tilde{f}, \quad \mathscr{F}\mathscr{H} = \tilde{\mathscr{H}}, \tag{2.35}$$

and

$$\mathscr{F}U(a,\Lambda)\mathscr{F}^{-1} = \tilde{U}(a,\Lambda), \tag{2.36}$$

where clearly

$$(\tilde{U}(a,\Lambda)\tilde{f})(k,u) = \mathscr{D}^j(\Lambda_u^{-1}\Lambda\Lambda_{\Lambda^{-1}u})e^{(i/\hbar)ka}\tilde{f}(\Lambda^{-1}k,\Lambda^{-1}u). \tag{2.37}$$

Now $\tilde{\mathscr{H}}$ can be written as the direct sum

$$\tilde{\mathscr{H}} = \tilde{\mathscr{H}}^{\mathrm{I}} \oplus \tilde{\mathscr{H}}^{\mathrm{II}}, \tag{2.38}$$

corresponding to the (invariant) decomposition

$$\tilde{f} = \tilde{f}^{\mathrm{I}} \oplus \tilde{f}^{\mathrm{II}}, \tag{2.39}$$

of any vector $\tilde{f} \in \tilde{\mathscr{H}}$ into a part $\tilde{f}^{\mathrm{I}}(k,u)$ for which $k_0^2 > \mathbf{k}^2$ and into a part $\tilde{f}^{\mathrm{II}}(k,u)$ for which $k_0^2 \leqslant \mathbf{k}^2$. Thus, writing

$$f^{\mathrm{I}} = \mathscr{F}^{-1}\tilde{f}^{\mathrm{I}}, \quad f^{\mathrm{II}} = \mathscr{F}^{-1}\tilde{f}^{\mathrm{II}}, \tag{2.40}$$

we obtain Eq. (2.30). Further, U^{I} and U^{II} are just the restrictions of U to $\mathscr{H}^{\mathrm{I}}$ and $\mathscr{H}^{\mathrm{II}}$, respectively.

To analyze U^{I} further, we see that

$$(U^{\mathrm{I}}(a,\Lambda)f^{\mathrm{I}})(q,u) = \frac{1}{h^2}\int_{k_0^2 > \mathbf{k}^2} e^{-(i/\hbar)kq}(\tilde{U}^{\mathrm{I}}(a,\Lambda)\tilde{f}^{\mathrm{I}})(k,u)d^4k. \tag{2.41}$$

In Eq. (2.41), since $k_0^2 > \mathbf{k}^2$, let us write

$$k_0^2 = \mathbf{k}^2 + m^2c^2, \tag{2.42}$$

and change the integration over k_0 to one over m, where

$$0 < m < \infty. \tag{2.43}$$

A straightforward computation then yields the result

$$(U^{\mathrm{I}}(a,\Lambda)f^{\mathrm{I}})(q,u) = \frac{c^2}{h^2}\int_{m>0} m\,dm \int_{\mathscr{V}_m} e^{-(i/\hbar)kq}(\tilde{U}^{\mathrm{I}}_m(a,\Lambda)\tilde{f}^{\mathrm{I}}_m)(k,u)dv_m. \tag{2.44}$$

In Eq. (2.44), for each m, $\tilde{f}^{\mathrm{I}}_m$ is an element of the Hilbert space

$$\tilde{\mathscr{H}}_{m,0} \otimes L^2\left(\mathscr{V}, \frac{d^3\mathbf{u}}{u_0}\right) \otimes \mathscr{K}^j.$$

Also, for each m, we can replace this space by the space

$$\tilde{\mathscr{H}}(\Omega_m) = \tilde{\mathscr{H}}^+_{m,0} \otimes \tilde{\mathscr{H}}^+_{m,j} \oplus \tilde{\mathscr{H}}^-_{m,0} \otimes \tilde{\mathscr{H}}^-_{m,j}, \tag{2.45}$$

using the isometry $\tilde{f}^{\mathrm{I}}_m \mapsto \tilde{f}_m \in \mathscr{H}(\Omega_m)$ with

$$\tilde{f}_m(k,p) = \frac{1}{m}\tilde{f}^{\mathrm{I}}_m(k, p/m). \tag{2.46}$$

Then, as a consequence of Eq. (2.19) we see that Eq. (2.44) can be rewritten as

$$(U^{\mathrm{I}}(a,\Lambda)f^{\mathrm{I}})(q,u) = \frac{c^2}{h^{1/2}}\int_{m>0} m^2\,dm\,U_m(a,\Lambda)f_m(q,p), \tag{2.47}$$

from which Eqs. (2.31)–(2.33) follow.

The subrepresentation U^{II} can be analyzed after setting

$$k_0^2 = \mathbf{k}^2 - \rho^2c^2, \tag{2.48}$$

where now

$$0 \leqslant \rho^2 \leqslant \mathbf{k}^2/c^2, \tag{2.49}$$

and hence $i\rho$ plays the role of a vanishing or imaginary mass. Q.E.D.

The subrepresentation U^{II} will not be analyzed further here, since we shall mainly be interested in physical particles with positive, nonzero mass.

From the nature of the decomposition in Eqs. (2.31) and (2.32), it is seen that the U_m are only components in a direct integral decomposition, and not genuine subrepresentations. Further, they themselves are also reducible. However, it is these component representations that contain the irreducible subrepresentations U_e obtained in I. Also, the extension of U_e to a larger Hilbert space as defined in Eq. (3.31) of I is equivalent to the U_m just obtained. We proceed to explore this point in detail now.

Consider the sets

$$\Gamma^{\pm}_m = \mathbb{R}^3 \times \mathscr{V}^{\pm}_m. \tag{2.50}$$

In I it was pointed out that these sets are (topologically) isomorphic to the set

$$\mathscr{M} = \mathscr{P}^{\uparrow}_+/T \otimes \mathrm{SO}(3), \tag{2.51}$$

and carry the invariant measures

$$d\mu^{\pm}_m = d^3\mathbf{q}\,d^3\mathbf{p} \equiv d\mu_m. \tag{2.52}$$

It is important to notice in this context that elements $(\mathbf{q},\mathbf{p})$ of the spaces $\Gamma^{\pm}_m$ represent a parametrization of the coset space $\mathscr{M}$, and can, for example, be looked upon as the initial data for classical free particle trajectories. Thus, the space time point $q = (0,\mathbf{q})$ and the momentum $p = (p_0,\mathbf{p})$ together determine the coset representative $(\mathbf{q},\mathbf{p})$ in $\Gamma^{\pm}_m$; however, the translated space time and momentum point (through the group element $\{a,\Lambda\}^{-1}$)

$$(\tilde{q},\tilde{p}) = (\Lambda^{-1}(q-a),\Lambda^{-1}p), \tag{2.53}$$

corresponds to the $\Gamma^{\pm}_m$ space element $(\mathbf{q}',\mathbf{p}')$ with components

$$\begin{aligned} q'^k &= (\Lambda^{-1})^k_\nu(\mathbf{q}-a)^\nu - (\Lambda^{-1})^0_\nu \\ &\quad \times [(\Lambda^{-1})^k_\nu p^\nu/(\Lambda^{-1})^0_\nu p^\nu], \\ p'^k &= (\Lambda^{-1})^k_\nu p^\nu, \end{aligned} \tag{2.54}$$

[cf. Eqs. (2.10) and (2.16) in I], where $(\mathbf{q}-a)$ is the four vector $(-a_0, \mathbf{q}-\mathbf{a})$. Equations (2.54) may be written more conveniently as

$$\begin{aligned} (\mathbf{q}',\mathbf{p}') &= \{a,\Lambda\}^{-1}(\mathbf{q},\mathbf{p}) \\ &= (\tilde{\mathbf{q}} - \tilde{q}_0\tilde{\mathbf{p}}/\tilde{p}_0,\tilde{\mathbf{p}}). \end{aligned} \tag{2.55}$$

Let

$$\Gamma_m = \Gamma^+_m \cup \Gamma^-_m = \mathbb{R}^3 \times \mathscr{V}_m, \tag{2.56}$$

which also carries the invariant measure $d\mu_m$. The spaces $\Gamma^{\pm}_m$ were indentified in I with the physical (six-dimensional) phase spaces, which form special lightlike surfaces (corresponding to $t = 0$) in the seven-dimensional relativistic phase spaces $\Omega^{\pm}_m$. The representations of $\mathscr{P}^{\uparrow}_+$ over phase space were built on the Hilbert spaces

$$\mathscr{H}(\Gamma^{\pm}_m) = L^2(\Gamma^{\pm}_m, d\mu_m) \otimes \mathscr{K}^j. \tag{2.57}$$

Let

$$\mathscr{H}(\Gamma_m) = \mathscr{H}(\Gamma^+_m) \oplus \mathscr{H}(\Gamma^-_m) = L^2(\Gamma_m, d\mu_m) \otimes \mathscr{K}^j. \tag{2.58}$$

Elements in these three spaces will be denoted by $\psi^{\pm}_m$ and ψ_m, respectively. Changing the notation slightly in Eq. (3.31) of I, we see that the operators $W^{\pm}_m(a,\Lambda)$ implement a (reducible) representation of $\mathscr{P}^{\uparrow}_+$ in the manner

$$(W^{\pm}_m(a,\Lambda)\psi^{\pm}_m)(\mathbf{q},\mathbf{p})$$
$$= \left(\exp\left[\mp \frac{i}{\hbar c}\tilde{q}_0\left\{\frac{m^2c^3}{p'_0} + H^q_m - H^c_m\right\}\right]\right.$$
$$\left.\times \mathscr{D}^j(\Lambda_p^{-1}\Lambda\Lambda_{\Lambda^{-1}p})(\psi^{\pm}_m)\{a,\Lambda\}^{-1}(\mathbf{q},\mathbf{p})\right), \quad (2.59)$$

where

$$H^q_m = c[\hat{\mathbf{P}}^2 + m^2c^2]^{1/2}, \quad (2.60)$$

$$H^c_m = \frac{c}{p_0}[p\cdot\hat{\mathbf{P}} + m^2c^2], \quad (2.61)$$

and

$$\hat{\mathbf{P}} = -i\hbar\left(\frac{\partial}{\partial q^1}, \frac{\partial}{\partial q^2}, \frac{\partial}{\partial q^3}\right). \quad (2.62)$$

[If we also introduce on $\mathscr{H}(\Gamma^{\pm}_m)$ the operators

$$\hat{P}_0 = \pm \frac{1}{c}H^q_m, \quad (2.63)$$

then it is not hard to see that the operators $\hat{P}_0$ and $\hat{\mathbf{P}}$ together form a four-vector under Lorentz transformations Λ—a fact we shall use in the sequel.] Once again we define a representation W_m on $\mathscr{H}(\Gamma_m)$ as the direct sum

$$W_m = W^+_m \oplus W^-_m. \quad (2.64)$$

For each Λ in the Lorentz group $\mathscr{L}^{\uparrow}_+$, let us define an operator $\alpha(\Lambda)$ on $\mathscr{H}(\Gamma^{\pm}_m)$ as

$$(\alpha(\Lambda)\psi^{\pm}_m)(\mathbf{q},\mathbf{p}) = ([\hat{P}_0^{-1}(\Lambda\hat{P})_0 p_0(\Lambda p)_0^{-1}]^{1/2}\psi^{\pm}_m)(\mathbf{q},\mathbf{p}). \quad (2.65)$$

These operators display some very interesting properties which we will analyze somewhat later. Let us first use them to construct another set of representations of $\mathscr{P}^{\uparrow}_+$ on $\mathscr{H}(\Gamma^{\pm}_m)$ and $\mathscr{H}(\Gamma_m)$. Let

$$\hat{W}^{\pm}_m(a,\Lambda) = W^{\pm}_m(q,\Lambda)\alpha(\Lambda), \quad (2.66)$$

and

$$\hat{W}_m = \hat{W}^+_m \oplus \hat{W}^-_m. \quad (2.67)$$

We state next a theorem whose proof is somewhat lengthy and is given in Appendix A.

Theorem 2.2: The operators $\hat{W}^{\pm}_m$ and $\hat{W}_m$ implement [respectively, on $\mathscr{H}(\Gamma^{\pm}_m)$ and $\mathscr{H}(\Gamma_m)$] unitary (reducible) representations of $\mathscr{P}^{\uparrow}_+$ which are equivalent to the representations $U^{\pm}_m$ and U_m, respectively. These equivalences lead to the alternative (equivalent up to isomorphism) decompositions of $\mathscr{H}^1$ and U^1 in the manner

$$\mathscr{H}^1 = \hat{V}\left[\int^{\oplus}_{m>0}\mathscr{H}(\Gamma_m)d\sigma(m)\right], \quad (2.68)$$

and

$$U^1 = \hat{V}\left[\int^{\oplus}_{m>0}\hat{W}_m d\sigma(m)\right]\hat{V}^{-1}. \quad (2.69)$$

Here $\hat{V}$ is the unitary operator

$$\hat{V} = \int^{\oplus}_{m>0}\hat{V}(m)d\sigma(m), \quad (2.70)$$

with

$$\hat{V}(m) = (\hat{V}(m)^+, \hat{V}(m)^-), \quad (2.71)$$

and

$$(\hat{V}(m)^{\pm}\psi^{\pm}_m)(q,p) = \left([p_0^{-1}\hat{P}_0]^{-1/2}\exp\left[\mp\frac{i}{\hbar}H^q_m t\right]\psi^{\pm}_m\right)(\mathbf{q},\mathbf{p}), \quad (2.72)$$

defined initially on dense sets in $\mathscr{H}(\Gamma^{\pm}_m)$ and then extended by continuity.

As noted in the proof of this theorem, the equivalence of $\hat{W}^{\pm}_m$ with $U^{\pm}_m$ is achieved by applying the unitary maps $\hat{V}(m)^{\pm}$ defined in Eq. (2.72), so that

$$f^{\pm}_m(q,p) = (\hat{V}(m)^{\pm}\psi^{\pm}_m)(q,p). \quad (2.73)$$

Also, since the operators $\hat{P}_0$ and of multiplication by $[\pm p_0]^{-1/2}$ are well defined (in fact retaining also the same form) on $\mathscr{H}(\Omega^{\pm}_m)$, we have in addition the inverse relationships

$$\psi^{\pm}_m(\mathbf{q},\mathbf{p}) = ([\hat{V}(m)^{\pm}]^{-1}f^{\pm}_m)(\mathbf{q},\mathbf{p})$$
$$= \left([p_0^{-1}\hat{P}_0]^{1/2}\exp\left[\pm\frac{i}{\hbar}H^q_m t\right]f^{\pm}_m\right)(q,p). \quad (2.74)$$

Equations (2.74) show us that in going from the spaces $\Omega^{\pm}_m$ to $\Gamma^{\pm}_m$ we have to perform some kind of a mass (through p_0) and energy (through $\hat{P}_0$) renormalization.

It is with the representations $W^{\pm}_m$ that we are mainly concerned, since these are the ones which arise most naturally in the theory of relativistic quantum mechanics on stochastic phase spaces (cf. for example, Refs. 2 and 4). We shall prove below the unitary equivalence of $\hat{W}^{\pm}_m$ and $W^{\pm}_m$, and ultimately get a complete decomposition theory for $W^{\pm}_m$ and hence for the two other sets of equivalent representations $U^{\pm}_m$ and $\hat{W}^{\pm}_m$. However, it will be useful to have available the representations $\hat{W}^{\pm}_m$ in addition to the other two sets $U^{\pm}_m$ and $W^{\pm}_m$. Besides, as we shall see later, the representation $\hat{W}^{\pm}_m$ is a little "less relativistic" than $W^{\pm}_m$, in the sense that with respect to $\hat{W}^{\pm}_m$, the relativistic Newton–Wigner position and momentum operators[5] retain exactly the same form as in the nonrelativistic case.

We notice first that the operators $\alpha(\Lambda)$ are unitary, since $W^{\pm}_m$ and $\hat{W}^{\pm}_m$ are unitary in Eq. (2.66), and besides, for any pure rotation R,

$$\alpha(R) = I. \quad (2.75)$$

[The unitary of $\alpha(\Lambda)$ can also be proved directly; cf. Lemma A.2 in Appendix A.] Hence, it follows that for any element in $\mathscr{P}^{\uparrow}_+$ which is of the type $\{a,R\}$, we have

$$\hat{W}^{\pm}_m(a,R) = W^{\pm}_m(a,R), \quad (2.76)$$

and, writing $\Lambda = R\Delta_{p'}$,

$$\hat{W}^{\pm}_m(a,\Lambda) = W^{\pm}_m(a,\Lambda)\alpha(\Lambda_{p'}). \quad (2.77)$$

The next theorem, whose proof we again relegate to Appendix A, carries the relationships (2.76) and (2.77) further.

Theorem 2.3: The representations $\hat{W}^{\pm}_m$ and $W^{\pm}_m$ are unitarily equivalent, so that $W^{\pm}_m$ and W_m realize the alternative decomposition (up to isomorphism) of $\mathscr{H}^1$ and U^1 as

$$\mathscr{H}^1 = V\left[\int^{\oplus}_{m>0}\mathscr{H}(\Gamma_m)d\sigma(m)\right], \quad (2.78)$$

and

$$U^1 = V\left[\int_{m>0}^{\oplus} W_m d\sigma(m)\right]V^{-1}. \tag{2.79}$$

Here V is the unitary operator

$$V = \int_{m>0}^{\oplus} V(m)d\sigma(m), \tag{2.80}$$

with

$$V(m) = (V(m)^+, V(m)^-), \tag{2.81}$$

and

$$(V(m)^{\pm}\psi_m^{\pm})(q,p) = (mc)^{-1/2}\left([\Lambda_p^{-1}\hat{P}]_0^{1/2}\exp\left[\mp\frac{i}{\hbar}H^q{}_m t\right]\psi_m^{\pm}\right)(\mathbf{q},\mathbf{p}). \tag{2.82}$$

Once again we have the relations, analogous to Eqs. (2.73) and (2.74)

$$f_m^{\pm}(q,p) = (V(m)^{\pm}\psi_m^{\pm})(q,p), \tag{2.83}$$

and

$$\psi_m^{\pm}(\mathbf{q}\cdot\mathbf{p}) = ([V(m)^{\pm}]^{-1}f_m^{\pm})(\mathbf{q}\cdot\mathbf{p}) = (mc)^{1/2}\left([\Lambda_p^{-1}\hat{P}]_0^{-1/2}\exp\left[\pm\frac{i}{\hbar}H^q{}_m t\right]f_m^{\pm}\right)\times(q,p). \tag{2.84}$$

Let us note that if we work with the functions

$$\phi_m^{\pm}(q,p) = \left[\exp\left(\mp\frac{i}{\hbar}H^q{}_m t\right)\psi_m^{\pm}(\mathbf{q},\mathbf{p}),\right], \tag{2.85}$$

and identify the spaces of these functions once again (as can obviously be done) with $\mathscr{H}(\Gamma_m^{\pm})$, then in view of Eq. (2.84), the representations $W_m^{\pm}$ can be written very simply in the forms [analogous to Eq. (2.27)]

$$(W_m^{\pm}(a,\Lambda)\phi_m^{\pm})(q,p) = \mathscr{D}^j(\Lambda_p^{-1}\Lambda\Lambda_{\Lambda^{-1}p})\phi_m^{\pm}(\Lambda^{-1}(q-a),\Lambda^{-1}p). \tag{2.86}$$

As mentioned earlier, the representations $W_m^{\pm}$ were obtained in I by starting with the irreducible representations $\tilde{V}_{m,j}^{\pm}$ on $\tilde{\mathscr{H}}_{m,j}^{\pm}$ (cf. Eqs. (2.16) and (2.17)], embedding them isometrically in $\mathscr{H}(\Gamma_m^{\pm})$, and then extending the resulting expressions to the whole of $\mathscr{H}(\Gamma_m^{\pm})$. The procedure was clearly *ad hoc* and somewhat artificial. In this paper, we proceed now to follow the reverse route, namely, we decompose the representations $W_m^{\pm}$ into an infinite direct sum of irreducible representations which are unitarily equivalent to representations of the type $\tilde{V}_{m,j}^{\pm}$, thereby making the aforementioned embedding $\tilde{V}_{m,j}^{\pm}\to W_m^{\pm}$ much more natural.

Consider first the Hilbert spaces $\mathscr{H}(\Omega_m^{\pm})$ of functions $\tilde{f}_m^{\pm}$ and the representations $\tilde{U}_m^{\pm}$ defined in Eqs. (A16) and (A17). $\tilde{U}_m^{\pm}$ is unitarily equivalent to $U_m^{\pm}$ and hence to $\hat{W}_m^{\pm}$ and $W_m^{\pm}$. Let $\tilde{e}^{\pm}$ be a function in $\tilde{\mathscr{H}}_{m,0}^{\pm}$ satisfying

$$\tilde{e}^{\pm}(Rk) = \tilde{e}^{\pm}(k), \tag{2.87}$$

for all rotations R in SO(3) and almost all k, and

$$\int_{V_m^+}|\tilde{e}^{\pm}(k)|^2 d\nu_m^{\pm} = 1. \tag{2.88}$$

Consider now the set of functions in $\tilde{\mathscr{H}}(\Omega_m^{\pm})$ which are of the type

$$\tilde{f}_e^{\pm}(k,p) = \overline{\tilde{e}^{\pm}(\Lambda_p^{-1}k)}\,\mathscr{D}^j(\Lambda_k^{-1}\Lambda_p\Lambda_{\Lambda_p^{-1}k})^*\tilde{\phi}^{\pm}(k), \tag{2.89}$$

where $\tilde{\phi}^{\pm}\in\tilde{\mathscr{H}}_{m,j}^{\pm}$. Let us denote this set of functions by $\tilde{\mathscr{H}}(\Omega_e^{\pm})$.

Lemma 2.1: The set $\tilde{\mathscr{H}}(\Omega_e^{\pm})$ is a proper subspace $\tilde{\mathscr{H}}(\Omega_m^{\pm})$ which is isometric to $\tilde{\mathscr{H}}_{m,j}^{\pm}$ via Eq. (2.89). Furthermore, the irreducible representation $\tilde{V}_{m,j}^{\pm}$ of $\mathscr{P}_+^{\uparrow}$ on $\tilde{\mathscr{H}}_{m,j}^{\pm}$ induces, through this isometry, an irreducible representation $\tilde{U}_e^{\pm}$ in $\tilde{\mathscr{H}}(\Omega_e^{\pm})$ which is in fact a subrepresentation of $\tilde{U}_m^{\pm}$.

Proof: The mapping $\tilde{\phi}^{\pm}\mapsto\tilde{f}_e^{\pm}$ is clearly linear. To prove that it is an isometry, we have

$$\|\tilde{f}_e^{\pm}\|^2 = \int|\tilde{e}^{\pm}(\Lambda_p^{-1}k)|^2\|\tilde{\phi}^{\pm}(k)\|_j^2\frac{d^3\mathbf{k}}{|k_0|}\frac{d^3\mathbf{p}}{|p_0|}. \tag{2.90}$$

Now, if $\mathbb{1}_m$ is the four-vector $(mc,\mathbf{0})$, then

$$\begin{aligned}\Lambda_p^{-1}k &= \Lambda_p^{-1}\Lambda_k\mathbb{1}_m\\ &= \Lambda_p^{-1}\Lambda_k\Lambda_{\Lambda_k^{-1}p}^{-1}\Lambda_{\Lambda_k^{-1}p}\mathbb{1}_m\\ &= \Lambda_p^{-1}\Lambda_k\Lambda_{\Lambda_k^{-1}p}^{-1}\Lambda_k^{-1}p.\end{aligned} \tag{2.91}$$

However, $\Lambda_p^{-1}\Lambda_k\Lambda_{\Lambda_k^{-1}p}^{-1}$ is a rotation in SO(3), so that using Eqs. (2.91) and (2.87) we have

$$\tilde{e}^{\pm}(\Lambda_p^{-1}k) = \tilde{e}^{\pm}(\Lambda_k^{-1}p). \tag{2.92}$$

Putting this back into Eq. (2.90) and using the invariance of the measure $d^3\mathbf{p}/p_0$ together with Eq. (2.88), we see that

$$\|\tilde{f}_e^{\pm}\|^2 = \int_{V_m^+}\|\tilde{\phi}^{\pm}(k)\|_j^2 d\nu_m^{\pm} = \|\tilde{\phi}^{\pm}\|^2. \tag{2.93}$$

In view of this isometry it is clear that $\tilde{\mathscr{H}}(\Omega_e^{\pm})$ is also a proper subspace of $\tilde{\mathscr{H}}(\Omega_m^{\pm})$.

Let $\tilde{U}_e^{\pm}(a,\Lambda)$ be the image of $\tilde{V}_{m,j}^{\pm}(a,\Lambda)$ in $\tilde{\mathscr{H}}(\Omega_e^{\pm})$. Then,

$$\begin{aligned}(\tilde{U}_e^{\pm}(a,\Lambda)\tilde{f}_e^{\pm})(k,p) &= \overline{\tilde{e}^{\pm}(\Lambda_p^{-1}k)}\,\mathscr{D}^j(\Lambda_k^{-1}\Lambda_p\Lambda_{\Lambda_p^{-1}k})^*\\ &\quad\times(\tilde{V}_{m,j}^{\pm}(a,\Lambda)\tilde{\phi}^{\pm})(k),\\ &= \overline{\tilde{e}^{\pm}(\Lambda_p^{-1}k)}\,\mathscr{D}^j(\Lambda_k^{-1}\Lambda_p\Lambda_{\Lambda_p^{-1}k})^*\\ &\quad\times\mathscr{D}^j(\Lambda_k^{-1}\Lambda\Lambda_{\Lambda^{-1}k})e^{(i/\hbar)ka}\tilde{\phi}^{\pm}(\Lambda^{-1}k),\end{aligned} \tag{2.94}$$

by virtue of Eq. (2.17). Now,

$$\begin{aligned}\tilde{e}^{\pm}(\Lambda_p^{-1}k) &= \tilde{e}^{\pm}(\Lambda_p^{-1}\Lambda\Lambda_{\Lambda^{-1}p}\Lambda_{\Lambda^{-1}p}^{-1}\Lambda^{-1}k)\\ &= \tilde{e}^{\pm}(\Lambda_{\Lambda^{-1}p}^{-1}\Lambda^{-1}k),\end{aligned} \tag{2.95}$$

in view of Eq. (2.87) and the fact that $\Lambda_p^{-1}\Lambda\Lambda_{\Lambda^{-1}p}$ is a rotation. Also,

$$\begin{aligned}&\mathscr{D}^j(\Lambda_k^{-1}\Lambda_p\Lambda_{\Lambda_p^{-1}k})^*\mathscr{D}^j(\Lambda_k^{-1}\Lambda\Lambda_{\Lambda^{-1}k})\\ &\quad= \mathscr{D}^j(\Lambda_{\Lambda_p^{-1}k}^{-1}\Lambda_k\Lambda_p^{-1}\Lambda\Lambda_{\Lambda^{-1}k})\\ &\quad= \mathscr{D}^j(\Lambda_p^{-1}\Lambda\Lambda_{\Lambda^{-1}p})\mathscr{D}^j([\Lambda_{\Lambda^{-1}p}^{-1}\Lambda^{-1}\Lambda_p]\Lambda_{\Lambda_p^{-1}k}\\ &\qquad\times[\Lambda_p^{-1}\Lambda\Lambda_{\Lambda^{-1}p}]\Lambda_{\Lambda^{-1}p}^{-1}\Lambda_{\Lambda^{-1}k}).\end{aligned} \tag{2.96}$$

In Eq. (2.96) if we make use of the relation

$$R\Lambda_p R^{-1} = \Lambda_{Rp}, \tag{2.97}$$

for $R\in SO(3)$, we see that since $\Lambda_{\Lambda_k^{-1}\Lambda_p\Lambda^{-1}\Lambda_p}\in SO(3)$,

$$\mathcal{D}^j(\Lambda_k^{-1}\Lambda_p\Lambda_{\Lambda_p^{-1}k})^*\mathcal{D}^j(\Lambda_k^{-1}\Lambda\Lambda_{\Lambda^{-1}k}) = \mathcal{D}^j(\Lambda_p^{-1}\Lambda\Lambda_{\Lambda^{-1}p})\mathcal{D}^j(\Lambda_{\Lambda^{-1}k}^{-1}\Lambda_{\Lambda^{-1}p}\Lambda_{\Lambda_{\Lambda^{-1}p}^{-1}\Lambda^{-1}k})^*. \tag{2.98}$$

Hence, using Eqs. (2.95) and (2.98) in Eq. (2.94) and comparing with Eq. (A17), we see that

$$(\tilde{U}_e^{\pm}(a,\Lambda)\tilde{f}_e^{\pm})(k,p) = e^{(i/\hbar)ka}\mathcal{D}^j(\Lambda_p^{-1}\Lambda\Lambda_{\Lambda^{-1}p})\tilde{f}_e^{\pm}(\Lambda^{-1}k,\Lambda^{-1}p), \tag{2.99}$$

implying that $\tilde{U}_e^{\pm}$ is a subrepresentation of $\tilde{U}_m^{\pm}$. Q.E.D.

Lemma 2.2: The mapping

$$f_e^{\pm}(q,p) = \frac{1}{h^{3/2}}\int e^{-(i/\hbar)kq}\overline{\tilde{e}^{\pm}(\Lambda_p^{-1}k)} \times\mathcal{D}^j(\Lambda_k^{-1}\Lambda_p\Lambda_{\Lambda_p^{-1}k})^*\tilde{\phi}^{\pm}(k)\frac{d^3\mathbf{k}}{|k_0|}, \tag{2.100}$$

embeds $\tilde{\mathcal{H}}_{m,j}^{\pm}$ isometrically into a subspace $\mathcal{H}(\Omega_e^{\pm})$ of $\mathcal{H}(\Omega_m^{\pm})$ and the representation $\tilde{V}_{m,j}^{\pm}$ into an irreducible subrepresentation $U_e^{\pm}$ of $U_m^{\pm}$. The mapping

$$\hat{\psi}_e^{\pm}(\mathbf{q},\mathbf{p}) = \frac{1}{h^{3/2}}\int_{\gamma_m^{\pm}} e^{(i/\hbar)\mathbf{k}\cdot\mathbf{q}}\left(\frac{k_0}{p_0}\right)^{1/2}\overline{\tilde{e}^{\pm}(\Lambda_p^{-1}k)} \times\mathcal{D}^j(\Lambda_k^{-1}\Lambda_p\Lambda_{\Lambda_p^{-1}k})^*\tilde{\phi}^{\pm}(k)\frac{d^3\mathbf{k}}{|k_0|}, \tag{2.101}$$

and the mapping

$$\psi_e^{\pm}(\mathbf{q},\mathbf{p}) = \left(\frac{mc}{h^3}\right)^{1/2}\int_{\gamma_m^{\pm}} e^{(i/\hbar)\mathbf{k}\cdot\mathbf{q}}\frac{\overline{\tilde{e}^{\pm}(\Lambda_p^{-1}k)}}{|(\Lambda_p^{-1}k)_0|^{1/2}} \times\mathcal{D}^j(\Lambda_k^{-1}\Lambda_p\Lambda_{\Lambda_p^{-1}k})^*\tilde{\phi}^{\pm}(k)\frac{d^3\mathbf{k}}{|k_0|}, \tag{2.102}$$

embed $\tilde{\mathcal{H}}_{m,j}^{\pm}$ isometrically into a subspace $\hat{\mathcal{H}}(\Gamma_e^{\pm})$ and $\mathcal{H}(\Gamma_e^{\pm})$ of $\mathcal{H}(\Gamma_m^{\pm})$ and the representation $\tilde{V}_{m,j}^{\pm}$ into an irreducible subrepresentation $\hat{W}_e^{\pm}$ of $\hat{W}_m^{\pm}$ and $W_e^{\pm}$ of $W_m^{\pm}$.

The proof of this lemma follows immediately from Lemma 2.1 and Eqs. (A16), (A18), and (A23).

Equation (2.102) is a restatement of the isometry $\tilde{\psi}\mapsto\psi_e$ used in I (cf. Theorem 1 in that paper) and first established in Ref. 4. Actually, in both these papers the function

$$\hat{e}(k) = (mc)^{1/2}\frac{\tilde{e}^*(k)}{k_0^{1/2}}, \tag{2.103}$$

was used. It is then easy to verify that

$$\int_{\mathbb{R}^3}|\hat{e}(k)|^2 d^3\mathbf{k} = mc, \tag{2.104}$$

and

$$\int_{\mathbb{R}^3}|\hat{e}(\Lambda_p^{-1}k)|^2 d^3\mathbf{p} = k_0. \tag{2.105}$$

In Eq. (2.102) we have also used

$$(mc)^{1/2}\frac{\tilde{e}^{\pm}(\Lambda_p^{-1}k)}{|(\Lambda_p^{-1}k)_0|^{1/2}} = \hat{e}^{\pm}(\Lambda_p^{-1}k), \tag{2.106}$$

which being similar to $\hat{e}$ in Eq. (2.103) leaves the nature of the isometry unaltered. It is also clear from this discussion that the mapping $\tilde{f}_m^{\pm}\mapsto\psi_m^{\pm}$ in Eq. (A23) is a generalization of the isometry $\tilde{\psi}\mapsto\psi_e$ of I to the whole of $\mathcal{H}(\Gamma_m^{\pm})$.

To end this section we state our final theorem, which completes the decomposition of the representations $U_m^{\pm}$, $\hat{W}_m^{\pm}$, and $W_m^{\pm}$ and hence through Theorems 2.2 and 2.3 of U^{I}. The proof is given in Appendix A.

Theorem 2.4: The representations $W_m^{\pm}$ split into the infinite direct sums

$$W_m^{\pm}(a,\Lambda) = \oplus\sum_{n=1}^{\infty}\sum_{l=0}^{\infty}\sum_{\mu=|j-l|}^{j+l} W_{n,l,\mu}^{\pm}(a,\Lambda), \tag{2.107}$$

where, for each n,l and each integer, or half-odd integer μ, $W_{n,l,\mu}^{\pm}$ is a representatioan of $\mathcal{P}^{\uparrow}_{+}$ which is unitarily equivalent to the irreducible representation $\tilde{V}_{m,\mu}^{\pm}$. The index μ takes on the values $|j-l|$, $|j-l|+1$, $|j-l|+2$, etc.

It is now trivial to work out the analogous decomposition for $\hat{W}_m^{\pm}$ and $U_m^{\pm}$. Also, it is clear from Eq. (2.107) that Eq. (2.102) maps $\tilde{V}_{m,j}^{\pm}$ into a subrepresentation of the type $W_{n,0,j}^{\pm}$.

As a final remark, it can be seen from the proof of this theorem that the decomposition is somewhat similar to that of the direct product to two irreducible representations of $\mathcal{P}^{\uparrow}_{+}$ of the type $\tilde{V}_{m,j}^{\pm}$. Besides, the decomposition is effected by going to some sort of a centre of mass frame (cf. Ref. 6, Chap. 8), and comment at the end of Appendix A.

3. GENERATORS AND SYSTEMS OF IMPRIMITIVITY

In this section we first look at the infinitesimal generators of the representation $W_m^{\pm}$ and their decompositions into the generators for the sub-representations $W_{n,l,\mu}^{\pm}$. Secondly we look at some systems of imprimitivity associated with these representations as well as some resolutions of the identity on the subspaces $\mathcal{H}(\Gamma_e^{\pm})$. Finally, we discuss briefly the Newton–Wigner position and momentum operators in relation to these representations.

Consider the representation $\tilde{V}_{m,j}^{\pm}$ on $\tilde{\mathcal{H}}_{m,j}^{\pm}$. These have the generators (cf. Ref. 6)

$$\Pi_\mu = k_\mu, \tag{3.1}$$

$$\mathbf{M} = -i\hbar\mathbf{k}\times\nabla_{\mathbf{k}} + \mathbf{J}, \tag{3.2}$$

$$\mathbf{N} = +i\hbar k_0\nabla_{\mathbf{k}} - \frac{\mathbf{J}\times\mathbf{k}}{mc+k_0}. \tag{3.3}$$

In Eq. (3.1), Π_μ are the generators of the four space–time translations, while the three components of $\mathbf{M}$ in Eq. (3.2) are the generators of rotation (or angular momentum) and the three components of $\mathbf{N}$ in Eq. (3.3) are the generators of the pure Lorentz boosts. The vector $\mathbf{J}$ is the generator of spin, so that $\mathbf{J}^2$ has the eigenvalue $\hbar j(j+1)$.

From the nature of Eqs. (2.17) and (A16) it is then straightforward to verify that the representations $\tilde{U}_m^{\pm}$ on $\mathcal{H}(\Omega_m^{\pm})$ have the generators

$$\Pi_\mu = k_\mu, \tag{3.4}$$

$$\mathbf{M} = -i\hbar[\mathbf{k}\times\nabla_{\mathbf{k}} + \mathbf{p}\times\nabla_{\mathbf{p}}] + \mathbf{J}, \tag{3.5}$$

$$\mathbf{N} = +i\hbar[k_0\nabla_{\mathbf{k}} + p_0\nabla_{\mathbf{p}}] - \frac{\mathbf{J}\times\mathbf{p}}{mc+p_0}. \tag{3.6}$$

Using now the isometry in Eq. (A23), a straightforward computation yields the following generators for the representations $W_m^{\pm}$ on $\mathcal{H}(\Gamma_m^{\pm})$:

$$\Pi_\mu = \hat{P}_\mu = i\hbar\frac{\partial}{\partial q^\mu}, \tag{3.7}$$

$$\mathbf{M} = \mathbf{q}\times\hat{\mathbf{P}} - i\hbar\mathbf{p}\times\nabla_{\mathbf{p}} + \mathbf{J}, \tag{3.8}$$

$$\mathbf{N} = +[\mathbf{q} + i\hbar\mathbf{p}/(\Lambda_p^{-1}\hat{P})_0]\hat{P}_0 + i\hbar p_0[\nabla_{\mathbf{p}} + \hat{\mathbf{P}}/(\Lambda_p^{-1}\hat{P})_0] - \frac{\mathbf{J}\times\mathbf{p}}{mc + p_0}. \tag{3.9}$$

The decomposition of these generators into their irreducible subsectors can be achieved using Theorem 2.4. We only display here the form of these operators for the special case of $j = 0$ on the irreducible subspaces $\mathscr{H}(\Gamma_e^{\pm})$ defined in Lemma 2.2. Indeed, using Eqs. (3.1)–(3.3) in conjunction with the unitary map in Eq. (2.102), we obtain, after some computation, the following generators for $W_e^{\pm}$:

$$\Pi_\mu = \hat{P}_\mu = i\hbar\frac{\partial}{\partial q^\mu}, \tag{3.10}$$

$$\mathbf{M} = \mathbf{q}\times\hat{\mathbf{P}} + i\hbar p_0\hat{P}_0^{-1}\hat{\mathbf{P}}\times\nabla_{\mathbf{p}}, \tag{3.11}$$

$$\mathbf{N} = +\mathbf{q}\hat{P}_0 + i\hbar p_0\nabla_{\mathbf{p}}. \tag{3.12}$$

We proceed next to construct the projection operators $\mathbb{P}_e^{\pm}$ onto the subspaces $\mathscr{H}(\Gamma_e^{\pm})$ of $\mathscr{H}(\Gamma_m^{\pm})$. This will lead to the resolutions of the identity mentioned earlier. Consider the representation $\tilde{X}_m^{\pm}$ on $\tilde{\mathscr{H}}(\Omega_e^{\pm})$ defined in Eq. (A31) and let $\tilde{\mathbb{P}}_e^{\pm}$ be the projection operator $|\tilde{e}^{\pm}\rangle\langle\overline{\tilde{e}^{\pm}}|$ on $\tilde{\mathscr{H}}(\Omega_m^{\pm})$ corresponding to some vector $\tilde{e}_n^{\pm} = \tilde{e}^{\pm}$ in the decomposition in (A34), i.e.,

$$(\tilde{\mathbb{P}}_e^{\pm}\tilde{g}_m^{\pm})(k,p) = \int_{\gamma_m^{\pm}} \tilde{e}^{\pm}(p')\overline{\tilde{e}^{\pm}(p)}\tilde{g}^{\pm}(k,p')\frac{d^3p'}{|p_0'|}. \tag{3.13}$$

Let $\mathbb{P}_e^{\pm}$ be the unitary transform of $\tilde{\mathbb{P}}_e^{\pm}$ in $\mathscr{H}(\Gamma_m^{\pm})$. In view of Eqs. (A37) and (A42) it is then clear that $\mathbb{P}_e^{\pm}$ project onto the subspaces $\mathscr{H}(\Gamma_e^{\pm})$ and give rise to the subrepresentations $W_e^{\pm}$. We next write $\tilde{\mathbb{P}}_e^{\pm}$ in a somewhat different form. Let $\hat{e}^{\pm}$ be the vectors defined as in Eq. (2.103):

$$\hat{e}^{\pm}(k) = (mc)^{1/2}\frac{\tilde{e}^{\pm}(k)}{|k_0|^{1/2}}, \tag{3.14}$$

and set

$$\hat{\epsilon}^{\pm}(k,p) = h^{-3/2}\hat{e}^{\pm}(k)\overline{\tilde{e}^{\pm}(p)}, \tag{3.15}$$

and

$$\hat{\epsilon}_{\mathbf{q},\mathbf{p}}^{\pm}(k',p') = h^{-3/2}e^{-(i/\hbar)\mathbf{k}'\cdot\mathbf{q}}\hat{e}^{\pm}(\Lambda_p^{-1}k')\overline{\tilde{e}^{\pm}(p')}. \tag{3.16}$$

Actually, it is easy to verify, using Eq. (A31) and the rotational invariance of $\tilde{e}^{\pm}$, that $\hat{\epsilon}_{\mathbf{q},\mathbf{p}}^{\pm}$ is just the vector $\tilde{X}_{m,0}^{\pm}((0,\mathbf{q}),\Lambda_p)\hat{e}^{\pm}$, where $\tilde{X}_{m,0}^{\pm}$ is the representation (A31) for the case $j = 0$. The vectors $\hat{\epsilon}_{\mathbf{q},\mathbf{p}}^{\pm}$ then lead to the following resolutions of the identity:

$$\int_{\Gamma_m^{\pm}} |\hat{\epsilon}_{\mathbf{q},\mathbf{p}}^{\pm}\rangle\langle\hat{\epsilon}_{\mathbf{q},\mathbf{p}}^{\pm}|\,d^3\mathbf{q}\,d^3\mathbf{p} = \tilde{\mathbb{P}}_e^{\pm}, \tag{3.17}$$

where $|\hat{\epsilon}_{\mathbf{q},\mathbf{p}}^{\pm}\rangle\langle\hat{\epsilon}_{\mathbf{q},\mathbf{p}}^{\pm}|$ is the usual projection operator on $\tilde{\mathscr{H}}(\Omega_m^{\pm})$ corresponding to the function $\hat{\epsilon}_{\mathbf{q},\mathbf{p}}^{\pm}$.

The proof of this result is given in Appendix B. We may rewrite Eq. (3.17) on $\mathscr{H}(\Gamma_m^{\pm})$ to get the analogous result

$$\int_{\Gamma_m^{\pm}} |\hat{\eta}_{\mathbf{q},\mathbf{p}}^{\pm}\rangle\langle\hat{\eta}_{\mathbf{q},\mathbf{p}}^{\pm}|\,d^3\mathbf{q}\,d^3\mathbf{p} = \mathbb{P}_e^{\pm}. \tag{3.18}$$

Once again, a straightforward computation shows that $\hat{\eta}^{\pm}$ is the $(2j+1)\times(2j+1)$-matrix valued function

$$\hat{\eta}^{\pm}(\mathbf{q},\mathbf{p}) = \left(\frac{mc}{h^3}\right)^{1/2}\int_{\gamma_m^{\pm}} e^{(i/\hbar)\mathbf{k}\cdot\mathbf{q}}\mathscr{D}^j(\Lambda_k^{-1}\Lambda_p\Lambda_{\Lambda_p^{-1}k})^* \times[\hat{\epsilon}^{\pm}(k,\Lambda_k^{-1}p)/(\Lambda_p^{-1}k)_0^{1/2}]\frac{d^3k}{|k_0|}, \tag{3.19}$$

and, $\hat{\eta}_{\mathbf{q},\mathbf{p}}$ is the matrix valued function (using an obvious abuse of notation):

$$\hat{\eta}_{\mathbf{q},\mathbf{p}}^{\pm} = W_m^{\pm}((0,\mathbf{q}),\Lambda_p)\hat{\eta}^{\pm}. \tag{3.20}$$

With this notation, $|\hat{\eta}_{\mathbf{q},\mathbf{p}}^{\pm}\rangle\langle\hat{\eta}_{\mathbf{q},\mathbf{p}}^{\pm}|$ is the projection operator

$$|\hat{\eta}_{\mathbf{q},\mathbf{p}}^{\pm}\rangle\langle\hat{\eta}_{\mathbf{q},\mathbf{p}}^{\pm}|\psi)(\mathbf{q}',\mathbf{p}') = \int_{\Gamma_m^{\pm}} \hat{\eta}_{\mathbf{q},\mathbf{p}}^{\pm}(\mathbf{q}'',\mathbf{p}'')^*\hat{\eta}_{\mathbf{q},\mathbf{p}}^{\pm}(\mathbf{q}',\mathbf{p}')\psi(\mathbf{q}'',\mathbf{p}'')\,d\mu_m''.$$

The matrix functions $\hat{\eta}_{\mathbf{q},\mathbf{p}}^{\pm}$ will be called *resolution generators* for the irreducible representations $W_e^{\pm}$. They have the following useful properties (proved in Appendix B):

$$\hat{\eta}^{\pm}(R\mathbf{q},R\mathbf{p}) = \mathscr{D}^j(R)\hat{\eta}^{\pm}(\mathbf{q},\mathbf{p})\mathscr{D}^j(R)^*, \tag{3.21}$$

$$\hat{\eta}_{\mathbf{q},\mathbf{p}}^{\pm}(\mathbf{q}',\mathbf{p}') = \hat{\eta}_{\mathbf{q}',\mathbf{p}'}^{\pm}(\mathbf{q},\mathbf{p})^*, \tag{3.22}$$

$$\int_{\Gamma_m^{\pm}} \hat{\eta}_{\mathbf{q},\mathbf{p}}^{\pm}(\mathbf{q}',\mathbf{p}')^*\hat{\eta}_{\mathbf{q}'',\mathbf{p}''}^{\pm}(\mathbf{q}',\mathbf{p}')\,d^3\mathbf{q}'\,d^3\mathbf{p}' = \hat{\eta}_{\mathbf{q}'',\mathbf{p}''}^{\pm}(\mathbf{q},\mathbf{p}). \tag{3.23}$$

Equation (3.21) holds for all rotations $R\in$SO(3), while Eq. (3.23) is known as the *reproducibility property*.

Equation (3.18) suggests that there exists on $\mathscr{H}(\Gamma_e^{\pm})$ a positive operator valued system of imprimitivity (POVSI) with respect to the Euclidean subgroup of $\mathscr{P}_+^{\uparrow}$ (cf. Refs. 7 and 8 for relevant definitions of systems of imprimitivity, their relationship to stochastic phase spaces, and their use in defining extended localization of elementary systems). To explore this point further, let $\mathbb{E}^3$ denote the Euclidean subgroup of $\mathscr{P}_+^{\uparrow}$ (i.e., it is the subgroup $T^3\textcircled{S}$SO(3) of all spatial translations and rotations.] If $g\in\mathbb{E}^3$, then g has the form

$$g = ((0,\mathbf{q}'),R'), \tag{3.24}$$

where $\mathbf{q}'\in T^3$ and $R'\in$SO(3). Also, for the phase space element $(\mathbf{q},\mathbf{p})$ let

$$g^{-1}(\mathbf{q},\mathbf{p}) = (R'^{-1}(\mathbf{q}-\mathbf{q}'),R'^{-1}\mathbf{p}). \tag{3.25}$$

Then using Eq. (3.21) it is easy to see that

$$W_e^{\pm}(g)^*|\hat{\eta}_{\mathbf{q},\mathbf{p}}^{\pm}\rangle\langle\hat{\eta}_{\mathbf{q},\mathbf{p}}^{\pm}|W_e^{\pm}(g) = |\hat{\eta}_{g^{-1}(\mathbf{q},\mathbf{p})}^{\pm}\rangle\langle\hat{\eta}_{g^{-1}(\mathbf{q},\mathbf{p})}^{\pm}|. \tag{3.26}$$

Consider now the normalized POV measures $a^{\pm}(\Delta)$ defined on the Borel sets Δ of $\Gamma_m^{\pm}$:

$$a^{\pm}(\Delta) = \int_{\Delta} |\hat{\eta}_{\mathbf{q},\mathbf{p}}^{\pm}\rangle\langle\hat{\eta}_{\mathbf{q},\mathbf{p}}^{\pm}|\,d^3\mathbf{q}\,d^3\mathbf{p} \tag{3.27}$$

(cf. Ref. 7 for technicalities regarding the convergence, etc. of these integrals). Clearly $a^{\pm}(\Delta)$ satisfies

$$a^{\pm}(\Gamma_m^{\pm}) = \tilde{\mathbb{P}}_e^{\pm}, \tag{3.28}$$

$$a^{\pm}(\phi) = 0, \tag{3.29}$$

$$a^{\pm}\left(\cup\Delta_i\right) = \sum_i a^{\pm}(\Delta_i), \tag{3.30}$$

where ϕ denotes the null set and the Δ_i's are a countable set of mutually disjoint Borel sets of $\Gamma^{\pm}_m$. Also, in view of Eq. (3.26) we have

$$W^{\pm}_e(g)^* a^{\pm}(\Delta) W^{\pm}_e(g) = a^{\pm}(g^{-1}(\Delta)). \tag{3.31}$$

Equations (3.28)–(3.31) define the required POVSI. We state, without proof here, that this POVSI can be canonically embedded in a projection valued system of imprimitivity (PVSI) defined on the enlarged Hilbert space $\mathscr{H}(\Gamma^{\pm}_m)$. The result is analogous to that described in Sec. 3 of Ref. 8.

Finally, a word about the Newton–Wigner operators: It is well known that the Hilbert spaces $\mathscr{H}^{\pm}_{m,j}$ admit an irreducible representation of the canonical commutation relations via the "relativistic position and momentum" operators

$$\tilde{\mathbf{Q}} = i\hbar\left[\nabla_k - \frac{k}{2k_0^2}\right], \tag{3.32}$$

$$\mathbf{P} = \mathbf{k}. \tag{3.33}$$

Using Eq. (2.102), it is possible to map these operators into $\mathscr{H}(\Gamma^{\pm}_e)$. Consider, for simplicity the case of $j = 0$. In this case we get on $\mathscr{H}(\Gamma^{\pm}_e)$ the following equivalents of the Newton–Wigner operators:

$$\mathbf{Q} = \mathbf{q} + i\hbar[\, p_0 \hat{P}_0^{-1} \nabla_{\mathbf{p}} + \tfrac{1}{2}\hat{\mathbf{P}}\hat{P}_0^{-2}\,], \tag{3.34}$$

$$\mathbf{P} = -i\hbar\nabla_{\mathbf{q}}. \tag{3.35}$$

These operators can clearly be extended to the whole of $\mathscr{H}(\Gamma^{\pm}_m)$ giving there a reducible representation of the canonical commutation relations. The decomposition into irreducible components would then follow Theorem 2.4, the operators remaining form invariant on each subspace. The operators (3.34) and (3.35) do not have, however, the "proper" nonrelativistic limit. Roughly speaking, nonrelativistically, $p_0 \sim mc \sim \hat{P}_0$, and it is seen that the $\mathbf{Q}$ and $\mathbf{P}$ above do not go over into the nonrelativistic phase space position and momentum operators discussed in Ref. 8. Consider, however, the images of $\tilde{\mathbf{Q}}$ and $\tilde{\mathbf{P}}$ on $\hat{\mathscr{H}}(\Gamma^{\pm}_e)$ obtained via the isometry in Eq. (2.101). In this case (again for $j = 0$) we have

$$\hat{\mathbf{Q}} = \mathbf{q} - i\hbar p_0 \hat{P}_0^{-1} \nabla_{\mathbf{p}}, \tag{3.36}$$

$$\hat{\mathbf{P}} = -i\hbar\nabla_{\mathbf{q}}. \tag{3.37}$$

It is these latter operators that have the proper nonrelativistic limits. This does not necessarily imply, however, that for physical applications the representations $\hat{W}^{\pm}_m$ are to be preferred over $W^{\pm}_m$. In fact, quite the contrary is true, for the isometry (2.102) is directly amenable to physical interpretation in stochastic phase space terms.

4. CONCLUSION

The tremendous value of phase space representations of the Poincaré group has been amply underscored in a number of recent papers (cf. Ref. 9 and references cited therein). Physically, one is able to construct, using these representations, a consistent relativistic quantum mechanics of both bosons and fermions for which the wave functions involve only positive energy solutions. These wave functions have a proper probability interpretation and satisfy in addition the correct current conservation laws. Besides, they throw much light on the problem of sharp localizability of relativistic particles.

On the mathematical side, the phase space approach achieves a great unification of the description of physical particles, classical as well as quantum, at both relativistic and nonrelativistic velocities. The transition from the relativistic to nonrelativistic regimes should be effected by a contraction procedure on the representations of $\mathscr{P}^1_+$ described here, to go to the corresponding representations of the Galilei group. Such a procedure is well worth following, for it would also clarify the situation vis a vis the two sets of Newton–Wigner operators mentioned in Eqs. (3.34) and (3.35) and Eqs. (3.36) and (3.37).

Finally, on a more esoteric level, one can think in terms of a deformation theory in the sense of Bayen *et al.*,[10] to go from relativistic classical to quantum mechanics using phase space ideas on the representations of $\mathscr{P}^1_+$ developed in I and in the present paper.

APPENDIX A

To prove Theorem 2.2 we first need a technical lemma.

Lemma A.1: The operator $H^q_m \equiv c[\,-\hbar^2\nabla^2_{\mathbf{q}} + m^2c^2\,]^{1/2}$ is densely defined on each one of the spaces $\mathscr{H}(\Omega^{\pm}_m)$ and $\mathscr{H}(\Gamma^{\pm}_m)$. In each case, it has a completely positive spectrum extending from mc^2 to ∞ and hence it has a bounded inverse there. In addition, $[H^q_m]^{1/2}$ and $[H^q_m]^{-1/2}$ are well defined on both of these spaces.

Proof: Consider the spaces

$$\mathscr{H}(\Omega^{+}_m) = L^2(\mathbb{R}^4, \langle\ ,\ \rangle_+) \otimes L^2(\mathscr{V}^{+}_m, dv^{+}_m) \otimes \mathscr{K}^j, \tag{A1}$$

and

$$\mathscr{H}(\Gamma^{+}_m) = L^2(\mathbb{R}^6, d\mu_m) \otimes \mathscr{K}^j. \tag{A2}$$

In Eq. (A1), for functions ξ and ζ on $\mathbb{R}^4$, $\langle\ ,\ \rangle_+$ is a scalar product of the type

$$\langle \xi, \zeta \rangle_+ = \frac{i\hbar}{2c}\int_{\mathbb{R}^3}\left[\overline{\xi(x)}\,\frac{\partial}{\partial t}\zeta(x) - \overline{\left(\frac{\partial}{\partial t}\xi(x)\right)}\zeta(x)\right] d^3\mathbf{x}, \tag{A3}$$

and to be in $L^2(\mathbb{R}^4, \langle\ ,\ \rangle_+)$, a function ξ has to have the time dependence

$$\xi(x) = \xi(ct, \mathbf{x}) = \left[\exp\left(-\frac{i}{\hbar}H^q_m t\right)\xi\right](0, \mathbf{x}), \tag{A4}$$

with

$$H^q_m = c[\,-\hbar^2\nabla^2_{\mathbf{q}} + m^2c^2\,]^{1/2} = c\left[-\hbar^2\left(\frac{\partial^2}{\partial x_1^2}, \frac{\partial^2}{\partial x_2^2}, \frac{\partial^2}{\partial x_3^2}\right) + m^2c^2\right]^{1/2}, \tag{A5}$$

as follows from Eq. (2.19). Hence by Eq. (A1) H^q_m is densely defined on $\mathscr{H}(\Omega^{\pm}_m)$. Also, from Eq. (2.19) we see that

$$H^q_m = c\mathscr{F}^{-1}_m K^0_m \mathscr{F}_m, \tag{A6}$$

where K^0_m is the operator of multiplication by

$$k_0 = [\mathbf{k}^2 + m^2c^2]^{1/2} \tag{A7}$$

on $L^2(\mathscr{V}^{+}_m, dv^{+}_m)$. Since $0 \leqslant \mathbf{k}^2 < \infty$, it follows that H^q_m as an operator on $\mathscr{H}(\Omega^{+}_m)$ has a spectrum extending from mc^2 to ∞. Thus $[H^q_m]^{1/2}$ and $[H^q_m]^{-1/2}$ are operators having the same spectral projectors as H^q_m and with spectra ranging,

respectively, from $m^{1/2}c$ to ∞ and 0 to $m^{-1/2}c^{-1}$. In fact, $[H^q_m]^{-1}$ with spectrum ranging from 0 to $m^{-1}c^{-2}$ and $[H^q_m]^{-1/2}$ are both bounded operators on $\mathscr{H}(\Omega^+_m)$.

Consider now the set of Schwartz functions $\mathscr{S}(\mathbb{R}^3)$ of rapid decrease. For $g\in\mathscr{S}(\mathbb{R}^3)$ define

$$\xi(x)=\left[\exp\left(-\frac{i}{\hbar}H^q_m t\right)g\right](\mathbf{x}). \tag{A8}$$

Then the functions ξ form a dense set in $L^2(\mathbb{R}^4,\langle\ ,\ \rangle_+)$. Consider now functions g^+_m in $\mathscr{H}(\Omega^+_m)$ which are of the type

$$g^+_m=\xi\otimes\tilde{\phi}^+_m, \tag{A9}$$

with $\tilde{\phi}^+_m\in L^2(V^+_m,dv^+_m)\otimes\mathscr{K}^j$. These functions form a dense set in $\mathscr{H}(\Omega^+_m)$, as is clear from Eq. (A1). Finally, on this set of functions g^+_m consider the mapping

$$\psi^+_m(\mathbf{q},\mathbf{p})=(cp_0)^{-1/2}\left[[H^q_m]^{1/2}\exp\left(\frac{i}{\hbar}H^q_m t\right)g^+_m\right](q,p). \tag{A10}$$

A straightforward computation shows that

$$\|\psi^+_m\|^2_\Gamma=\|g^+_m\|^2_\Omega, \tag{A11}$$

$\|\cdots\|_\Gamma$ and $\|\cdots\|_\Omega$ being, respectively, the norms in $\mathscr{H}(\Gamma^+_m)$ and $\mathscr{H}(\Omega^+_m)$. Thus, the mapping in Eq. (A10) is an isometry from $\mathscr{H}(\Omega^+_m)$ into $\mathscr{H}(\Gamma^+_m)$ and hence can be extended to an isometry W defined over the whole of $\mathscr{H}(\Omega^+_m)$. Also it is clear from Eq. (A10) that the operator H^q_m on $\mathscr{H}(\Omega^+_m)$ is mapped into the same operator $[-\hbar^2\nabla^2_{\mathbf{q}}+m^2c^2]^{1/2}$ on the range of W in $\mathscr{H}(\Gamma^+_m)$. Thus, the inverse of W can be defined as

$$(W^{-1}\psi^+_m)(q,p)=(cp_0)^{1/2}\left[[H^q_m]^{-1/2}\exp\left(-\frac{i}{\hbar}H^q_m t\right)\psi^+_m\right](\mathbf{q},\mathbf{p}), \tag{A12}$$

and is a mapping which can be extended by continuity to the whole of $\mathscr{H}(\Gamma^+_m)$. Thus W is an isometric bijection. Let us write

$$\hat{V}(m)^*=W^{-1}. \tag{A13}$$

Then,

$$[\hat{V}(m)^*]^{-1}H^q_m\hat{V}(m)^*=H^q_m, \tag{A14}$$

and the lemma is proved for the space $\mathscr{H}(\Omega^+_m)$ and $\mathscr{H}(\Gamma^+_m)$. For the other two spaces $\mathscr{H}(\Omega^-_m)$ and $\mathscr{H}(\Gamma^-_m)$, the proof is similar.

Proof of Theorem 2.2: In view of the above lemma, to complete the proof of Theorem 2.2 it is only necessary to demonstrate that

$$[\hat{V}(m)^\pm]^{-1}U^\pm_m\hat{V}(m)^\pm=\hat{W}^\pm_m. \tag{A15}$$

Consider the Hilbert space $\tilde{\mathscr{H}}(\Omega^+_m)=\tilde{\mathscr{H}}^+_{m,0}\otimes\tilde{\mathscr{H}}^+_{m,j}$ [cf. Eq. (2.16)] obtained from $\mathscr{H}(\Omega^+_m)$ by the Fourier isometry [Eq. (2.19)]:

$$\tilde{f}^+_m(k,p)=\frac{k_0}{h^{3/2}}\int_{\mathbb{R}^3}e^{(i/\hbar)kq}f^+_m(q,p)d^3\mathbf{q}. \tag{A16}$$

On $\tilde{\mathscr{H}}(\Omega^+_m)$ the operators $U^+_m(a,\Lambda)$ transform to $\tilde{U}^+_m(a,\Lambda)$ with

$$(\tilde{U}^+_m(a,\Lambda)\tilde{f}^+_m)(k,p)=e^{(i/\hbar)ka}\mathscr{D}^j(\Lambda_p{}^{-1}\Lambda\Lambda_{\Lambda^{-1}p})\tilde{f}^+_m(\Lambda^{-1}k,\Lambda^{-1}p), \tag{A17}$$

and the operators H^q_m and $\hat{\mathbf{P}}$ become multiplication by ck_0 and $\mathbf{k}$ [cf. Eq. (A6)]. Next, it is straightforward to verify that the unitary map $[\hat{V}(m)^*]^{-1}$ in Eq. (A13) from $\mathscr{H}(\Omega^+_m)$ to $\mathscr{H}(\Gamma^+_m)$ transforms to the unitary map

$$\psi^+_m(\mathbf{q},\mathbf{p})=\frac{1}{h^{3/2}}\int_{\Gamma^+_m}e^{(i/\hbar)\mathbf{k}\cdot\mathbf{q}}\tilde{f}^+_m(k,p)\left(\frac{k_0}{p_0}\right)^{1/2}\frac{d^3\mathbf{k}}{k_0}, \tag{A18}$$

from $\tilde{\mathscr{H}}(\Omega^+_m)$ to $\mathscr{H}(\Gamma^+_m)$. Thus,

$$(\hat{W}^+_m(a,\Lambda)\psi^+_m)(\mathbf{q},\mathbf{p})=\frac{1}{h^{3/2}}\int_{\Gamma^+_m}e^{(i/\hbar)\mathbf{k}\cdot\mathbf{q}}(\tilde{U}^+_m(a,\Lambda)\tilde{f}^+_m)(k,p)\left(\frac{k_0}{p_0}\right)^{1/2}\frac{d^3\mathbf{k}}{k_0},$$

so that, using the invariance of the measure $d^3\mathbf{k}/k_0$

$$(\hat{W}^+_m(a,\Lambda)\psi^+_m)(\mathbf{q},\mathbf{p})=\frac{1}{h^{3/2}}\int_{\Gamma^+_m}e^{-(i/\hbar)k[\Lambda^{-1}(\mathbf{q}-a)]}\times\mathscr{D}^j(\Lambda_p{}^{-1}\Lambda\Lambda_{\Lambda^{-1}p})\tilde{f}^+_m(k,\Lambda^{-1}p)\left[\frac{(\Lambda k)_0}{p_0}\right]^{1/2}\frac{d^3\mathbf{k}}{k_0}.$$

Using Eqs. (2.53) and (2.54) this can be recast into the form

$$(\hat{W}^+_m(a,\Lambda)\psi^+_m(\mathbf{q},\mathbf{p})=\frac{1}{h^{3/2}}\int_{\Gamma^+_m}e^{-(i/\hbar)\tilde{q}_0[k_0-\mathbf{k}\cdot\mathbf{p}'/p'_0]}e^{(i/\hbar)\mathbf{k}\cdot\mathbf{q}'}\mathscr{D}^j(\Lambda_p{}^{-1}\Lambda\Lambda_{\Lambda^{-1}p})\times\left[\frac{(\Lambda k)_0(\Lambda^{-1}p)_0}{p_0k_0}\right]^{1/2}\tilde{f}^+_m(k,\Lambda^{-1}p)\times\left[\frac{k_0}{(\Lambda^{-1}p)_0}\right]^{1/2}\frac{d^3\mathbf{k}}{k_0},$$

which upon using Eqs. (2.60) and (2.61) becomes

$$(\hat{W}^+_m(a,\Lambda)\psi^+_m(\mathbf{q},\mathbf{p})=\exp\left[-\frac{i\tilde{q}_0}{\hbar c}\left\{\frac{m^2c^3}{p'_0}+H^q_m-H^c_m\right\}\right]\mathscr{D}^j(\Lambda_p{}^{-1}\Lambda\Lambda_{\Lambda^{-1}p})\times\frac{1}{h^{3/2}}\int_{\Gamma^+_m}e^{(i/\hbar)\mathbf{k}\cdot\mathbf{q}'}(\tilde{\alpha}(\Lambda)\tilde{f}^+_m)(k,\Lambda^{-1}p)\times\left[\frac{k_0}{(\Lambda^{-1}p)_0}\right]^{1/2}\frac{d^3\mathbf{k}}{k_0}. \tag{A19}$$

The operator $\tilde{\alpha}(\Lambda)$ in Eq. (A19) is defined as

$$(\tilde{\alpha}(\Lambda)\tilde{f}^+_m)(k,p)=\left[\frac{(\Lambda k)_0p_0}{(\Lambda p)_0k_0}\right]^{1/2}\tilde{f}^+_m(k,p). \tag{A20}$$

The isometry in Eq. (A15) is now an immediate consequence of Eqs. (A19), (2.59), (2.63), and (2.66). Similar considerations apply on the spaces $\mathscr{H}(\Omega^-_m)$ and $\mathscr{H}(\Gamma^-_m)$. Q.E.D.

Lemma A.2: The operators $\alpha(\Lambda)$ commute among themselves for all $\Lambda\in\mathscr{L}^\uparrow_+$ (the proper orthocronous Lorentz group) and are unitary.

Proof: Consider the representation $\tilde{\alpha}(\Lambda)$ of $\alpha(\Lambda)$ in the Fourier transformed spaces $\tilde{\mathscr{H}}(\Omega^\pm_m)$ as given by Eq. (A20). Since these are operators of multiplication, they commute for all $\Lambda\in\mathscr{L}^\uparrow_+$. To prove unitarity, we have (denoting the scalar product in $\mathscr{K}^j$ by $\langle|\rangle_j$)

$$\|\tilde{\alpha}(\Lambda)\tilde{f}_m^+\|^2$$
$$= \int \left[\frac{(\Lambda k)_0 p_0}{(\Lambda p)_0 k_0}\right]^{1/2} \langle \tilde{f}_m^+(k,p)|\tilde{f}_m^+(k,p)\rangle_j \frac{d^3\mathbf{k}}{k_0}\frac{d^3\mathbf{p}}{p_0}. \tag{A21}$$

Next, writing $\Lambda = R\Lambda_{p'}^{-1}$, we see that

$$\left[\frac{(\Lambda k)_0 p_0}{(\Lambda p)_0 k_0}\right]^{1/2} = \left[\frac{(p'k)p_0}{(p'p)k_0}\right]^{1/2} = \left[\frac{p_0' - \mathbf{p}'\cdot\mathbf{k}/k_0}{p_0' - \mathbf{p}'\cdot\mathbf{p}/p_0}\right]^{1/2}. \tag{A22}$$

Using Eq. (A22) and rewriting the integral in Eq. (A21) in the frame in which $\mathbf{p}' = 0$, and using the invariance of the measures $d^3\mathbf{k}/k_0$ and $d^3\mathbf{p}/p_0$, we get

$$\|\tilde{\alpha}(\Lambda)\tilde{f}_m^+\|^2$$
$$= \int \|\tilde{f}_m^+(\Lambda_{p'}^{-1}k, \Lambda_{p'}^{-1}p)\|_j^2 \cdot \frac{d^3\mathbf{k}}{k_0}\frac{d^3\mathbf{p}}{p_0}$$
$$= \|\tilde{f}_m^+\|^2. \qquad \text{Q.E.D.}$$

Proof of Theorem 2.3: Consider again the Hilbert space $\tilde{\mathscr{H}}(\Omega_m^{\pm})$ and the mapping $\tilde{f}_m^{\pm} \mapsto \psi_m^{\pm}$ from $\tilde{\mathscr{H}}(\Omega_m^{\pm})$ to $\mathscr{H}(\Gamma_m^{\pm})$ given by

$$\psi_m^{\pm}(\mathbf{q},\mathbf{p}) = \left(\frac{mc}{h^3}\right)^{1/2} \int_{\gamma_m^{\pm}} e^{(i/\hbar)\mathbf{k}\cdot\mathbf{q}} \frac{\tilde{f}_m^{\pm}(k,p)}{|(\Lambda_p^{-1}k)_0|^{1/2}} \frac{d^3\mathbf{k}}{|k_0|}. \tag{A23}$$

We prove first that this is an isometry. Indeed,

$$\|\psi_m^{\pm}\|^2 = \int_{\Gamma_m^{\pm}} \langle \psi_m^{\pm}(\mathbf{q},\mathbf{p})|\psi_m^{\pm}(\mathbf{q},\mathbf{p})\rangle_j \, d^3\mathbf{q}\, d^3\mathbf{p}$$
$$= \frac{mc}{h^3}\int e^{(i/\hbar)(\mathbf{k}-\mathbf{k}')\cdot\mathbf{q}} \times \langle \tilde{f}_m^{\pm}(k,p)|\tilde{f}_m^{\pm}(k',p)\rangle_j \times \left[|(\Lambda_p^{-1}k)_0(\Lambda_p^{-1}k')_0|\right]^{-1/2} \frac{d^3\mathbf{k}}{|k_0|}\frac{d^3\mathbf{k}'}{|k_0'|} d^3\mathbf{q}\, d^3\mathbf{p}.$$

For sufficiently smooth functions $\tilde{f}_m^{\pm}$, the $\mathbf{q}$ integration may be performed first and then the $\mathbf{k}'$ integration to yield

$$\|\psi_m^{\pm}\|^2 = mc\int \frac{\|\tilde{f}_m^{\pm}(k,p)\|_j^2}{(\Lambda_p^{-1}k)_0|k_0|}\frac{d^3\mathbf{k}}{|k_0|} d^3\mathbf{p}$$
$$= mc\int \frac{\|\tilde{f}_m^{\pm}(k,\Lambda_k p)\|_j^2}{p_0 k_0}(\Lambda_k p)_0 \frac{d^3\mathbf{k}}{|k_0|}\frac{d^3\mathbf{p}}{|p_0|}, \tag{A24}$$

where use has been made of the invariance of $d^3\mathbf{k}/k_0$. Since

$$mc\frac{(\Lambda_k p)_0}{p_0 k_0} = 1 + \frac{\mathbf{k}\cdot\mathbf{p}}{k_0 p_0}.$$

Eq. (A24) transforms into

$$\|\psi_m^{\pm}\|^2 = \|f_m^{\pm}\|^2 + \int \frac{\mathbf{k}\cdot\mathbf{p}}{k_0 p_0}\|\tilde{f}_m^{\pm}(k,\Lambda_k p)\|_j^2 \frac{d^3\mathbf{k}}{|k_0|}\frac{d^3\mathbf{p}}{|p_0|}. \tag{A25}$$

The integral on the right hand side of Eq. (A25) can now be shown to vanish. Indeed, one may write [cf. Eqs. (A29) and (A37) below] for a dense set of vectors $\tilde{f}_m^{\pm}$:

$$\tilde{f}_m^{\pm}(k,\Lambda_k p) = \sum_{n=1}^{\infty}\sum_{l=0}^{\infty} T_l \sum_{\mu=|j-l|}^{j+l} \beta_{n,l,\mu}\tilde{\phi}_{\mu}^{\pm}(k)\overline{\tilde{e}_n^{\pm}(p)},$$

where the $\beta_{n,l,\mu}$'s are constants, $\tilde{\phi}_{\mu}^{\pm} \in \tilde{\mathscr{H}}_{m,\mu}^{\pm}$, and $\tilde{e}_n^{\pm}$'s are a rotationally invariant orthonormal set [as defined in Eq. (A33) below] and the T_l's are isometries defined in Eq. (A34). Using the rotational invariance of the $\tilde{e}_n^{\pm}$'s, the vanishing of the integral in Eq. (A25) is immediate, establishing the required isometry.

In view of this isometry, and using the notation

$$p_0\hat{P}_0 - \mathbf{p}\cdot\hat{\mathbf{P}} = mc(\Lambda_p^{-1}\hat{P})_0, \tag{A26}$$

we see that Eq. (A16) implies the unitary map

$$f_m^{\pm}(q,p) = (V(m)^{\pm}\psi_m^{\pm})(q,p)$$
$$= (mc)^{-1/2}\left([\Lambda_p^{-1}\hat{P}]_0^{1/2} \times \exp\left[\mp\frac{i}{\hbar}H_m^q t\right]\psi_m^{\pm}\right)(\mathbf{q},\mathbf{p}), \tag{A27}$$

between $\mathscr{H}(\Omega_m^{\pm})$ and $\mathscr{H}(\Gamma_m^{\pm})$. Thus,

$$[V(m)^{\pm}]^{-1}U_m^{\pm}(a,\Lambda)V(m)^{\pm} = W_m^{\pm}(a,\Lambda), \tag{A28}$$

while the rest of the theorem follows from Theorem 2.1. Q.E.D.

Proof of Theorem 2.4: Consider the unitary map $T^{-1}: \tilde{f}_m^{\pm} \to \tilde{g}_m$, given by

$$\tilde{f}_m^{\pm}(k,p) = (T\tilde{g}_m^{\pm})(k,p)$$
$$= \mathscr{D}^j(\Lambda_k^{-1}\Lambda_p\Lambda_{\Lambda_p^{-1}k})^*\tilde{g}_m^{\pm}(k,\Lambda_k^{-1}p). \tag{A29}$$

It is straightforward to verify, using the methods of Lemma 2.1, that if

$$\tilde{U}_m^{\pm}(a,\Lambda) = T\tilde{X}_m^{\pm}(a,\Lambda)T^{-1}; \tag{A30}$$

then

$$(\tilde{X}_m^{\pm}(a,\Lambda)\tilde{g}_m^{\pm}(k,p)) = e^{(i/\hbar)ka}\mathscr{D}^j(\Lambda_k^{-1}\Lambda\Lambda_{\Lambda^{-1}k})\tilde{g}_m^{\pm}(\Lambda^{-1}k, \Lambda_{\Lambda^{-1}k}^{-1}\Lambda^{-1}\Lambda_k p). \tag{A31}$$

In Eq. (A31) we see that the variable p is acted upon by the rotation matrix $\Lambda_{\Lambda^{-1}k}^{-1}\Lambda^{-1}\Lambda_k$ under the effect of the group element (a,Λ). We therefore expand $\tilde{g}_m^{\pm}(k,p)$ as a function of p_0 and in the spherical harmonics of the unit vector $\mathbf{p}/|\mathbf{p}|$. Thus,

$$\tilde{g}_m^{\pm}(k,p) = \sum_{n=1}^{\infty}\sum_{l=0}^{\infty}\sum_{\mu=-l}^{l} \alpha_{n,l,\mu}\tilde{\phi}_j^{\pm}(k)\overline{\tilde{e}_n^{\pm}(p)}\, Y_{l,\mu}(\theta,\phi), \tag{A32}$$

where $\alpha_{n,l,\mu}$ are constants, $\tilde{\phi}_j^{\pm} \in \tilde{\mathscr{H}}_{m,j}^{\pm}$, $Y_{l,\mu}$ are spherical harmonics in the angles of $\mathbf{p}/|\mathbf{p}|$, and the $\tilde{e}_n^{\pm}$ are a set of orthonormal vectors in $\tilde{\mathscr{H}}_{m,0}^{\pm}$ which depend on p_0 alone and satisfy

$$\int_{\gamma_m^{\pm}} \overline{\tilde{e}_m^{\pm}(p)}\tilde{e}_n^{\pm}(p)\frac{d^3\mathbf{p}}{|p_0|} = \delta_{mn}. \tag{A33}$$

It is then clear that each $\tilde{e}_n^{\pm}$ is rotationally invariant and satisfies Eq. (2.87). Since for each l the $Y_{l,\mu}$'s form a basis in the $2l+1$ dimensional spinor space $\mathscr{K}^l$, we may unitarily map linear combinations of $Y_{l,\mu}$ into $\mathscr{K}^l$. Let T_l^{-1} denote this isometry. Then

$$\tilde{g}_m^{\pm}(k,p) = \sum_{n=1}^{\infty}\sum_{l=0}^{\infty} T_l\alpha_{n,l}\tilde{\phi}_j^{\pm}(k)\overline{\tilde{e}_n^{\pm}(p)}, \tag{A34}$$

where the $\alpha_{n,l}$'s are vectors in $\mathscr{K}^l$. Further, in view of Eq.

(A31) we see that under the action of the group (note that the $\tilde{e}_n^{\pm}$'s are unaffected)

$$(\tilde{X}_m^{\pm}(a,\Lambda)\tilde{g}_m^{\pm})(k,p)$$
$$= \sum_{n=1}^{\infty}\sum_{l=0}^{\infty} T_l \alpha_n e^{(i/\hbar)ka} \mathscr{D}^j(\Lambda_k^{-1}\Lambda\Lambda_{\Lambda^{-1}k})$$
$$\times \mathscr{D}^l(\Lambda_k^{-1}\Lambda\Lambda_{\Lambda^{-1}k})\tilde{\phi}_{j,l}^{\pm}(\Lambda^{-1}k)\overline{\tilde{e}_n^{\pm}(p)}. \quad \text{(A35)}$$

where $\tilde{\phi}_{j,l}^{\pm} \in \mathscr{K}^j \otimes \mathscr{K}^l \otimes \tilde{\mathscr{H}}_{m,0}^{\pm}$ and the α_n's are constants. The direct product representation $\mathscr{D}^j \otimes \mathscr{D}^l$ of SO(3) which appears in Eq. (A35) may now be decomposed, using standard techniques as

$$\mathscr{D}^j \otimes \mathscr{D}^l = \sum_{\mu=|j-l|}^{j+l} \mathscr{D}^{\mu}, \quad \text{(A36)}$$

where μ assumes the values $|j-l|$, $|j-l|+1$, $|j-l|+2,\ldots,j+l$. Hence, we may write

$$\tilde{g}_m^{\pm}(k,p) = \sum_{n=1}^{\infty}\sum_{l=0}^{\infty} T_l \sum_{\mu=|j-l|}^{j+l} \beta_{n,l,\mu}\tilde{\phi}_{\mu}^{\pm}(k)\overline{\tilde{e}_n^{\pm}(p)}, \quad \text{(A37)}$$

and

$$(\tilde{X}_m^{\pm}(a,\Lambda)\tilde{g}_m^{\pm})(k,p)$$
$$= \sum_{n=1}^{\infty}\sum_{l=0}^{\infty} T_l \sum_{\mu=|j-l|}^{j+l} \beta_{n,l,\mu}$$
$$\times [\mathscr{D}^{\mu}(\Lambda_k^{-1}\Lambda\Lambda_{\Lambda^{-1}k})e^{(i/\hbar)ka}\tilde{\phi}_{\mu}^{\pm}(\Lambda^{-1}k)]\overline{\tilde{e}_n^{\pm}(p)}, \quad \text{(A38)}$$

so that in fact we have the direct sum decomposition

$$\tilde{X}_m^{\pm}(a,\Lambda) = \oplus \sum_{n=1}^{\infty}\sum_{l=0}^{\infty} T_l\left[\sum_{\mu=|j-l|}^{j+l} \tilde{V}_{m,\mu}^{\pm}(a,\Lambda)\right]T_l^{-1}. \quad \text{(A39)}$$

Finally, combining the isometry T with the isometry (A23) connecting $\tilde{\mathscr{H}}(\Omega_m^{\pm})$ to $\mathscr{H}(\Gamma_m^{\pm})$, we get the stated result for $W_m^{\pm}(a,\Lambda)$. Explicitly, if we write

$$\tilde{\mathscr{H}}(\Omega_m^{\pm}) = \oplus \sum_{n=1}^{\infty}\sum_{l=0}^{\infty}\sum_{\mu=|j-l|}^{j+l} \tilde{\mathscr{H}}(\Omega_{n,l,\mu}^{\pm}), \quad \text{(A40)}$$

and

$$\mathscr{H}(\Gamma_m^{\pm}) = \oplus \sum_{n=1}^{\infty}\sum_{l=0}^{\infty}\sum_{\mu=|j-l|}^{j+l} \mathscr{H}(\Gamma_{n,l,\mu}^{\pm}), \quad \text{(A41)}$$

then $\tilde{\mathscr{H}}(\Omega_{n,l,\mu}^{\pm})$ consists of functions of the type

$$\tilde{f}_{n,l,\mu}^{\pm}(k,p) = T_l \mathscr{D}^{\mu}(\Lambda_k^{-1}\Lambda_p\Lambda_{\Lambda_p^{-1}k})^*\tilde{\phi}_{\mu}^{\pm}(k)\overline{\tilde{e}_n^{\pm}(\Lambda_k^{-1}p)}, \quad \text{(A42)}$$

[compare with Eq. (2.89)], and using Eq. (A23),

$$\psi_{n,l,\mu}^{\pm}(\mathbf{q},\mathbf{p})$$
$$= \left(\frac{mc}{h^3}\right)^{1/2}\int_{\gamma_m^{\pm}} e^{(i/\hbar)\mathbf{k}\cdot\mathbf{q}} T_l$$
$$\times \mathscr{D}^{\mu}(\Lambda_k^{-1}\Lambda_p\Lambda_{\Lambda_p^{-1}k})^*\tilde{\phi}_{\mu}^{\pm}(k)$$
$$\times \frac{\overline{\tilde{e}_n^{\pm}(\Lambda_p^{-1}k)}}{|(\Lambda_p^{-1}k)_0|^{1/2}}\frac{d^3\mathbf{k}}{|k_0|}, \quad \text{(A43)}$$

with

$$(W_{n,l,\mu}^{\pm}(a,\Lambda)\psi_{n,l,\mu}^{\pm})(\mathbf{q},\mathbf{p})$$
$$= \left(\exp\left[\mp\frac{i}{\hbar c}\mathbf{q}_0\left\{\frac{m^2c^3}{p_0'} + H_m^q - H_m^c\right\}\right]\right.$$
$$\left.\times \mathscr{D}^{\mu}(\Lambda_p^{-1}\Lambda\Lambda_{\Lambda^{-1}p})\psi_{n,l,\mu}^{\pm}\right)(\{a,\Lambda\}^{-1}(\mathbf{q},\mathbf{p})). \quad \text{(A44)}$$

Q.E.D.

Note that in view of the fact that

$$(k-p)^2 = 2m^2c^2 - 2kp = 2mc[mc - \Lambda_k^{-1}p)_0], \quad \text{(A45)}$$

the mapping T in Eq. (A29) is some sort of a center of mass transformation.

APPENDIX B

Proof of Eq. (3.17): From the definition in Eq. (3.16), we have, for sufficiently smooth vectors $\tilde{g}_m^{\pm}$ in $\tilde{\mathscr{H}}(\Omega_m^{\pm})$

$$\left(\int_{\Gamma_m^{\pm}} |\hat{\epsilon}_{\mathbf{q},\mathbf{p}}^{\pm}\rangle\langle\hat{\epsilon}_{\mathbf{q},\mathbf{p}}^{\pm}| d^3\mathbf{q}\, d^3\mathbf{p}\, \tilde{g}_m^{\pm}\rangle\right)(k'',p'')$$
$$= \frac{1}{h^3}\int e^{(i/\hbar)(\mathbf{k}''-\mathbf{k}')\cdot\mathbf{q}}\overline{\hat{e}^{\pm}(\Lambda_{k''}^{-1}p)}\,\tilde{e}^{\pm}(p')\hat{e}^{\pm}$$
$$\times(\Lambda_{k''}^{-1}p)\overline{\tilde{e}^{\pm}(p'')}\,\tilde{g}_m(k',p')\,d^3\mathbf{q}\,d^3\mathbf{p}\frac{d^3\mathbf{k}'}{|k_0'|}\frac{d^3\mathbf{p}'}{|p_0'|}$$
$$= \int_{\gamma_m^{\pm}} \tilde{e}^{\pm}(p')\overline{\tilde{e}^{\pm}(p'')}\,\tilde{g}_m^{\pm}(k'',p'')\frac{d^3\mathbf{p}'}{|p_0'|}. \quad \text{(B1)}$$

In the last step in Eq. (B1) use has been made of Eq. (2.105). If $\tilde{g}$ is an element in $\overline{\mathbb{P}}_e\,\tilde{\mathscr{H}}(\Omega_m^{\pm})$, i.e., if it has the form

$$\tilde{g}_m^{\pm}(k,p) = \tilde{\phi}^{\pm}(k)\overline{\tilde{e}^{\pm}(p)}, \quad \text{(B2)}$$

then using Eq. (2.105) once again, it is easy to see that the right hand side of Eq. (B1) reduces to $\tilde{g}_m^{\pm}(k'',p'')$. Further, applying the operator on the left hand side of Eq. (3.17) once more to Eq. (B1), its idempotence is straight-forward to prove. Self-adjointness is obvious from the definition, and hence the result. Q.E.D.

Proof of Eqs. (3.21)–(3.23): Equation (3.21) follows from (3.19) when combined with the rotation invariance of $\hat{e}^{\pm}$ and the invariance of the measure $d^3\mathbf{k}/k_0$. Next, combining Eqs. (3.19) and (3.20), we see that

$$\hat{\eta}_{\mathbf{q},\mathbf{p}}^{\pm}(\mathbf{q}',\mathbf{p}')$$
$$= \frac{1}{h^3}\int_{\gamma_m^{\pm}} e^{(i/\hbar)(\mathbf{k}\cdot(\mathbf{q}'-\mathbf{q}))}\mathscr{D}^j(\Lambda_k^{-1}\Lambda_{p'}\Lambda_{\Lambda_{p'}^{-1}k})^*$$
$$\times\mathscr{D}^j(\Lambda_k^{-1}\Lambda_p\Lambda_{\Lambda_p^{-1}k})\hat{e}(\Lambda_p^{-1}k)\overline{\hat{e}(\Lambda_{p'}^{-1}k)}\frac{d^3\mathbf{k}}{|k_0|}, \quad \text{(B3)}$$

from which Eq. (3.22) follows as an easy consequence.

Finally, to prove Eq. (3.23), we see using Eq. (B3) that

$$\int_{\Gamma_m^{\pm}} \hat{\eta}_{\mathbf{q},\mathbf{p}}^{\pm}(\mathbf{q}',\mathbf{p}')^*\hat{\eta}_{\mathbf{q}'',\mathbf{p}''}^{\pm}(\mathbf{q}',\mathbf{p}')\,d^3\mathbf{q}'\,d^3\mathbf{p}'$$
$$= \frac{1}{h^3}\int e^{-1/\hbar(\mathbf{k}\cdot(\mathbf{q}'-\mathbf{q}))}\mathscr{D}^j(\Lambda_k^{-1}\Lambda_p\Lambda_{\Lambda_{p''}^{-1}k})^*$$
$$\times\mathscr{D}^j(\Lambda_k^{-1}\Lambda_{p''}\Lambda_{\Lambda_{p''}^{-1}k})|\hat{e}^{\pm}(\Lambda_{p'}^{-1}k)|^2$$
$$\times\overline{\hat{e}^{\pm}(\Lambda_p^{-1}k)}\,\hat{e}^{\pm}(\Lambda_{p''}^{-1}k)\frac{d^3\mathbf{k}}{|k_0|}d^3\mathbf{p}'. \quad \text{(B4)}$$

Once again, in obtaining Eq. (B4) use has been made of the invariance of the measure $d\nu_m^{\pm}$. Using Eq. (2.105) in the right hand side of Eq. (B4) and applying Eq. (B3), the required result follows. Q.E.D.

As a corollary to Eqs. (3.18)–(3.23) it is easily deduced that

$$\begin{aligned}\psi_e^{\pm}(\mathbf{q},\mathbf{p}) &= (\mathbb{P}_e^{\pm}\psi_m^{\pm})(\mathbf{q},\mathbf{p}) \\ &= \int_{\Gamma_m'} \hat{\eta}_{\mathbf{q},\mathbf{p}}^{\pm}(\mathbf{q}',\mathbf{p}')^* \psi_m^{\pm}(\mathbf{q}',\mathbf{p}')\, d^3\mathbf{q}'\, d\mathbf{p}' \\ &= \int_{\Gamma_m'} \hat{\eta}^{\pm}(\mathbf{q}',\mathbf{p}')^* \\ &\times \left[W_m^{\pm}((0,\mathbf{q}),\Lambda_p)^* \psi_m^{\pm}\right](\mathbf{q}',\mathbf{p}')\, d\mu_m'. \end{aligned} \tag{B5}$$

[1] S.T. Ali, J. Math. Phys. **20**, 1385 (1979).
[2] S.T. Ali, "The Dirac equation on phase space" (to appear).
[3] E. Prugovečki, "A general framework for quantum field theory on relativistic stochastic phase spaces for massive particles," University of Toronto preprint (1978).
[4] E. Prugovečki, J. Math. Phys. **19**, 2260 (1978).
[5] T.D. Newton and E.P. Wigner, Rev. Mod. Phys. **21**, 400 (1949).
[6] N.N. Bogolubov, A.A. Logunov, and I.T. Todorov, *Introduction to Axiomatic Quantum Field Theory* (Benjamin, Reading, Mass., 1975).
[7] S.T. Ali and E. Prugovečki, J. Math. Phys. **18**, 219 (1977).
[8] S.T. Ali and E. Prugovečki, Physica (Utrecht) A **89**, 501 (1977).
[9] S.T. Ali, "Aspects of relativistic quantum mechanics on phase space," in Proceedings of the Informal Meeting on Differential Geometric Methods in Physics. Technische Universität Clausthal (July 1978) (to appear).
[10] F. Bayen, M. Flato, C. Fronsdal, A. Lichnerowicz, and D. Sternheimer, Ann. Phys. (N.Y.) **111**, 61 (1978).

The linear potential wavefunctions [a)]

Philip A. Maurone and Alain J. Phares
Department of Physics, Villanova University, Villanova, Pennsylvania 19085

(Received 15 May 1979; accepted for publication 21 December 1979)

We present an exact analytic solution to the Schrödinger equation for two particles interacting via a central linear potential of the from $V(r) = V_0 + kr$. The solution is given in terms of the generalized hypergeometric functions, and is especially useful when discussing problems where the radial parameter is small.

I. INTRODUCTION

In the past few years, the Schrödinger equation, with a central linear potential, has been of interest to many and has been studied extensively using various approximation techniques.[1] The purpose of this paper is to give the exact solution to this equation and then investigate this solution in the region where the radial parameter is small.

Basically, we arrive at the solution by solving, in a noniterative fashion, the three-term recursion relation that results from a power series solution to the Schrödinger equation.

In particular, we are interested in the nonrelativistic Schrödinger equation for two particles interacting through the potential

$$V = V_0 + kr\,. \tag{1}$$

The reduced equation for this potential is usually written in the form

$$\left(\frac{d^2}{d\rho^2} - \frac{l(l+1)}{\rho^2} - (\rho - t)\right)U_l(\rho, t) = 0\,, \tag{2}$$

where ρ and t are dimensionless distance and energy parameters.

The physically acceptable solutions to the above equation must satisfy the following boundary conditions[2,3]:

(i) $U_l(\rho,t)/\rho$ must be regular at the origin;

(ii) For large separations ρ, $U_l(\rho,t)$ should approach the well-behaved solution of the differential equation obtained by neglecting the centrifugal term in Eq. (2), namely

$$U'' - (\rho - t)U = 0\,. \tag{3}$$

The solution to this equation with the correct asymptotic behavior is the Airy function $\mathrm{Ai}(\rho - t)$, thus requiring

$$U_l(\rho, t) \underset{\rho\to\infty}{\sim} \mathrm{Ai}(\rho - t)\,. \tag{4}$$

The series solution to Eq. (2) satisfying condition (i) is

$$U_l(\rho, t) = \rho^{l+1} \sum_{m=0}^{\infty} b_m \rho^m\,. \tag{5}$$

The energy eigenvalues are generated by requiring $U_l(\rho,t)$ as expressed by Eq. (5) to be square integrable. This is equivalent to imposing the asymptotic condition of Eq. (4). The exact eigenvalue condition and the explicit eigenenergies for this potential have been presented in Ref. 2. These eigenvalues are used in the final forms of the wavefunction, derived in this paper to generate Figs. 1, 2, and 3, showing explicitly the required asymptotic fall off.

When Eq. (5) is used in Eq. (2), a three-term recursion relation on the expansion coefficients, b_m, is obtained

$$m(m + 2l + 1)\, b_m = -tb_{m-2} + b_{m-3} \quad \text{for } m > 0, \tag{6}$$

with the initial value conditions

$$b_m = \lambda_0 \delta_{m0} \quad \text{for } m \leqslant 0\,, \tag{7}$$

where λ_0 is fixed by requiring the wavefunction to be normalized.

We present two equivalent solutions to the recursion relation Eq. (6) that lead to two equivalent representations of the linear potential wavefunction in terms of generalized hypergeometric functions. This is done in Secs. II and III, respectively. Sec. IV gives the summary of our results and discusses the relevance of our wavefunction representation to the problem of quark confinement.[1]

II. SOLUTION

As usual, one can solve Eq. (6) for b_m iteratively, the

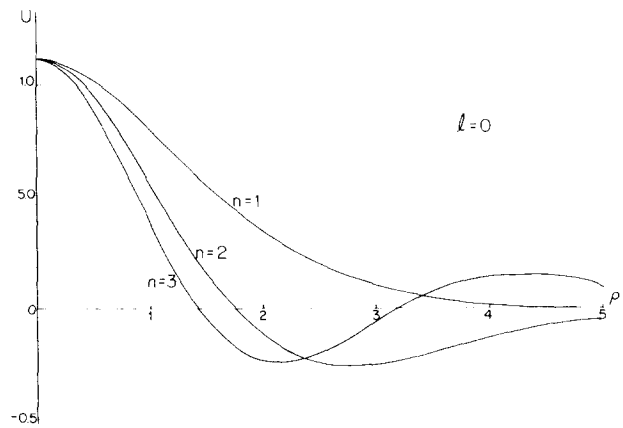


FIG. 1. Plot of the linear potential wavefunction for the orbital quantum number $l = 0$ and the principal quantum number $n = 1$, 2, and 3 and $\lambda_0 = 1$.

[a)]Work supported in part by a grant from the Office of Research and Patent Affairs (O.R.P.A.), Villanova University.

 0022-2488/80/040830-04$1.00

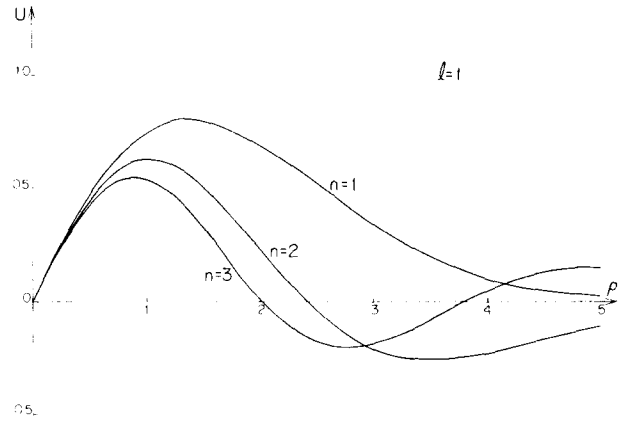


FIG. 2. Plot of the linear potential wavefunction for the orbital quantum number $l = 1$ and the principal quantum number $n = 1$, 2, and 3 and $\lambda_0 = 1$.

first few terms are

$$\begin{aligned} b_0 &= \lambda_0 , \\ b_1 &= 0 , \\ b_2 &= -t\lambda_0/2(3+2l) , \\ b_3 &= \lambda_0/3(4+2l) , \\ b_4 &= t^2\lambda_0/4(5+2l)2(3+2l) . \end{aligned} \tag{8}$$

However, the three-term recursion relation of Eq. (6) can be solved in an exact analytic fashion just as one would solve a simple two-term recursion relation.

A new combinatorics function technique[3] makes this possible. One possible representation of the solution is

$$\begin{aligned} b_m = \lambda_0 \tilde{\Gamma}(1) \sum_{n=n_{\min}}^{n_{\max}} & [(-t)^p/3^{2n}\tilde{\Gamma}(n+1+p/3)] \\ & \times \sum_{i_1=0}^{n-p} \cdots \sum_{i_p=0}^{i_{n-1}} \prod_{k=1}^{p} \tilde{\Gamma}\left(i_k + \frac{2p+2l-2k}{3}\right) \Big/ \\ & \tilde{\Gamma}\left(i_k + \frac{2p+3-2k}{3}\right), \end{aligned} \tag{9a}$$

where

$$\tilde{\Gamma}(a) = \Gamma(a)\Gamma\left(a + \frac{2l+1}{3}\right), \quad p = 3n - m , \tag{9b}$$

and Γ is the gamma function. The limits on the n summation in Eq. (9a) are given by

$$n_{\min} = \left[\frac{m+2}{3}\right] - \delta_{m1} , \quad n_{\max} = \left[\frac{m}{2}\right], \tag{9c}$$

where the square brackets refer to integer division and δ is Kronecker's symbol. Another representation of the solution to Eq. (6) will be given in Sec. III.

It is easy to show that the sequence in Eq. (8) is reproduced in a noniterative fashion. Using Eq. (9a) and some standard summation techniques,[4] the wavefunction $U_l(\rho,t)$ can be written in hypergeometric form as

$$\begin{aligned} U_l(\rho,t) = \lambda_0 \rho^{l+1} \sum_{n=0}^{\infty} (-t\rho^2/9)^n & \sum_{\{I_{n-1}\}} [\rho^3/9]^{I_{n-1}} \\ & \times {}_{2n}\tilde{F}_{2n+1}\left[\begin{matrix} c_1, \ldots, c_{2n} \\ d_1, \ldots, d_{2n+1} \end{matrix} ; \rho^3/9\right]. \end{aligned} \tag{10}$$

The c's and d's are given by

$$c_{2k-1} = I_{n-1} - I_{k-1} + 2(n-k+1) , \quad 1 \leqslant k \leqslant n , \tag{11a}$$

$$c_{2k} = c_{2k-1} + (2l+1)/3 , \quad 1 \leqslant k \leqslant n , \tag{11b}$$

$$d_{2k+1} = I_{n-1} - I_{k-1} + 2(n-k+l+2)/3 , \quad 0 \leqslant k \leqslant n , \tag{11c}$$

$$d_{2k} = d_{2k-1} - (2l+1)/3 , \quad 1 \leqslant k \leqslant n . \tag{11d}$$

In Eq. (10) we have made use of the shorthand notations

$$I_p \equiv i_0 + i_1 + \ldots + i_p , \quad p \geqslant 0 , \tag{12a}$$

$$I_{-1} \equiv 0 , \tag{12b}$$

$$\sum_{i_0=0}^{\infty} \sum_{i_1=0}^{\infty} \cdots \sum_{i_p=0}^{\infty} \equiv \sum_{\{I_p\}} . \tag{12c}$$

Equation (12c) continues to be meaningful for $p = -1$ and corresponds to no summation. Finally, ${}_p\tilde{F}_q$ as appearing in Eq. (10) is related to the standard hypergeometric function[4] ${}_pF_q$, according to

$${}_p\tilde{F}_q\left[\begin{matrix} c_1, \ldots, c_p \\ d_1, \ldots, d_q \end{matrix} ; z\right] = \frac{\Gamma(c_1)\ldots\Gamma(c_p)}{\Gamma(d_1)\ldots\Gamma(d_q)} {}_pF_q\left[\begin{matrix} c_1, \ldots, c_p \\ d_1, \ldots, d_q \end{matrix} ; z\right]. \tag{13}$$

The exact wavefunction is now given in terms of sums over generalized hypergeometric functions. The wavefunction is analytic in the variable $z = \rho^3/9$ since ${}_pF_q(z)$ is an entire function of z for $p \leqslant q$ which is our case here. It should be noted that for $\mathrm{Re}\, l = -1/2$ there is a cut in the complex l-plane. This is evident from the properties of ${}_p\tilde{F}_q(z)$ and Eqs. (11a)–(11d). This is not a surprise since such a property holds for all central potentials producing boundstate systems.

For the case $l = 0$ the $\{I_n\}$ summations can be done resulting in the hypergeometric form of the Airy function $\mathrm{Ai}(\rho - t)$ if one adds the energy eigenvalue condition, that is $\mathrm{Ai}(-t) = 0$.

On the other hand, the $\{I_{n-1}\}$ summations can also be done if one keeps just the first term in the hypergeometric series ${}_{2n}\tilde{F}_{2n+1}$. The result in this case is

$$U_l(\rho,t) = \lambda_0 \rho^{l+1} {}_0\tilde{F}_1[;l+3/2; -t\rho^2/4] , \tag{14a}$$

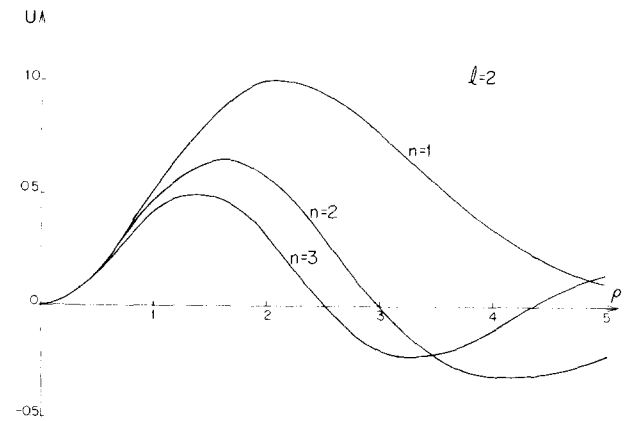


FIG. 3. Plot of the linear potential wavefunction for the orbital quantum number $l = 2$ and the principal quantum number $n = 1$, 2, and 3 and $\lambda_0 = 1$.

and can be written in terms of a Bessel function as

$$U_l(\rho,t)=\lambda_0\sqrt{z}J_\nu(z),\quad z=\rho\sqrt{t},\quad \nu=l+1/2. \tag{14b}$$

It is interesting to note that this is the regular solution of the truncated Schrödinger equation

$$\left(\frac{d^2}{d\rho^2}-\frac{l(l+1)}{\rho^2}+t\right)U_l(\rho,t)=0, \tag{15}$$

valid only for small values of the radial parameter ρ compared to the energy parameter t.

Equation (15) provides only the lowest order term contributing to the wavefunction for small ρ; however our solution to the nontruncated equation can provide corrections to any desired order. This is done by taking higher order terms in the hypergeometric series ${}_{2n}\tilde{F}_{2n+1}$. It is rather tedious, using Eq. (10), to collect these higher order terms in the correct sequence. In Sec. III we present another representation of the wavefunction that gives the correct sequence in a straightforward way.

III. ALTERNATE EXACT SOLUTION: A PERTURBATION APPROACH

As mentioned in Sec. II, there exists another representation to the solution of the three-term recursion relation of Eq. (6) and it is

$$b_m=\frac{\lambda_0\bar{\Gamma}(1)}{\bar{\Gamma}(1+m/2)}\sum_{n=n_{\min}}^{n_{\max}}\frac{(-t)^{3n-m}}{4^n}\times\sum_{i_1=0}^{3n-m}\sum_{i_2=0}^{i_1}\cdots\sum_{i_p=0}^{i_{p-1}}\prod_{k=1}^{p}\frac{\bar{\Gamma}[i_k+(3p-3k+3)/2\ ]}{\bar{\Gamma}[i_k+(3p-3k+2)/2\ ]}, \tag{16a}$$

where

$$\bar{\Gamma}(a)=\Gamma(a)\Gamma(a+l+1/2),\quad p=m-2n, \tag{16b}$$

and the limits on the n summation are the same as in Eq. (9c). Here again it is easy to show that Eq. (6a) generates the sequence in Eq. (8).[5]

The wavefunction expression resulting from this representation is

$$U_l(\rho,t)=\lambda_0\bar{\Gamma}(1)\sum_{m=0}^{\infty}(\rho^3/4)^m U_l^{(m)}(\rho,t), \tag{17a}$$

and

$$U_l^{(m)}(\rho,t)=\rho^{l+1}\sum_{\{I_{m-1}\}}(-t\rho^2/4)^{I_{m-1}}\times{}_{2m}\tilde{F}_{2m+1}\left[\begin{matrix}c'_1,\ldots,c'_{2m};\\ \\ d'_1,\ldots,d'_{2m+1};\end{matrix}\ -t\rho^2/4\right]. \tag{17b}$$

Here the arguments c' and d' of the hypergeometric function appearing in Eq. (17b) are slightly different from those of Sec. II., namely,

$$c'_{2k-1}=I_{m-1}-I_{k-1}+3(m-k+1)/2,\quad 1\leqslant k\leqslant m, \tag{18a}$$

$$c'_{2k}=c'_{2k-1}+l+1/2,\quad 1\leqslant k\leqslant m, \tag{18b}$$

$$d'_{2k+1}=I_{m-1}-I_{k-1}+l+3(m-k+1)/2,\quad 0\leqslant k\leqslant m, \tag{18c}$$

$$d'_{2k}=d'_{2k-1}-l-1/2,\quad 1\leqslant k\leqslant m. \tag{18d}$$

This new representation of the exact wavefunction is equivalent to the one presented in Sec. II. It has the advantage that, for small ρ, it can also be viewed as a perturbation expansion in the variable $(\rho^3/4)$, which gives all higher order terms directly and in the correct decreasing order of importance. In particular, it can be seen that the lowest order term obtained by setting $m=0$ in Eqs. (17a) and (17b) reproduces Eqs. (14a) and (14b). The second order term is simply obtained by setting $m=1$ again in Eqs. (17a), and (17b), whereas the second order term using the expression of Sec. II requires much more elaborate and tedious manipulations.[5]

IV. SUMMARY AND DISCUSSION

We have presented two exact and equivalent solutions to the Schrödinger equation with a central linear potential, for any orbital quantum number l. This was accomplished by solving, in a noniterative fashion, the three-term recursion relation that results among the expansion coefficients of the series representation of the wavefunction. Both solutions are written in terms of sums over generalized hypergeometric functions of the form ${}_pF_q(z)$ with $p<q$. Because of this we are able to make explicit the analytic properties of the wavefunction. In particular, the expected cut in the complex l plane at $\mathrm{Re}\,l=-1/2$ is made apparent.

Also, we have presented a new perturbation expansion, for small distances, where each term in the perturbation series is explicitly given for any l. This perturbation approach is most suitable, for example, in discussing quark confinement, when a linear potential is used as the confining mechanism.

We have plotted the wavefunction for various values of the principal and orbital quantum numbers, n and l, respectively. Figures 1, 2, and 3 have been generated using our perturbation expression. The agreement with the numerical solution of the Schrödinger equation is within less than 1% using only the first 5 order terms ($m=1$ to 5) in the expansion Eqs. (17a) and (17b).

We would like to thank Robert J. Meier, Jr. for his computational assistance.

[1]B.J. Harrington *et al.*, Phys. Rev. Lett. **34**, 168 (1975); 706 (1975); H.J.W. Müller-Kirsten, Phys. Rev. D **12**, 1103 (1975); P.L. Ferreira and J.A. Bastilho Alcarás, Lett. Nuovo Cimento **14**, 500 (1975); E. Eichten *et al.*, Phys. Rev. Lett. **34**, 370 (1975); **36**, 500 (1976); S.K. Bose *et al.*, Phys. Rev. D **12**, 1704 (1975); Phys. Rev. D **13**, 1704 (1976); S.K. Bose and H.J.W. Müller-Kirsten, Kaiserslautern Univ. prepr. (1978).

[2]A.F. Antippa and A.J. Phares, J. Math. Phys. **19**, 308 (1978); A.J. Phares, J. Math. Phys. **19**, 2239 (1978).

[3]A.F. Antippa and A.J. Phares, J. Math. Phys. **18**, 173 (1977); A.J. Phares, J. Math. Phys. **18**, 2214 (1977).

[4]See for example, *Handbook of Mathematical Functions*, edited by M. Abramowitz and I.A. Stegun (National Bureau of Standards, Washington, D.C., 1964); E.T. Copson, *An Introduction to the Theory of Functions of a Complex Variable* (Oxford U., Oxford, 1960); Y.L. Luke, *Mathematics in Science and Engineering* (Academic, London, 1969), Vol. 53-1 and Vol. II.

[5]The problem of writing down an explicit noniterative expression for the expansion coefficient b_m is intimately connected to the partitions of an interval of length m into parts of length 2 and 3. This connection is fully understood by means of the combinatorics function technique (CFT) (Ref. 3). Gamma-function expressions are then generated with arguments that are integer multiples of 1/2 or 1/3 depending on the summation choice. Equation (6a) is generated by singling out the parts of length 3, and Eq. (16a) is generated by singling out parts of length 2. More specific details of the calculations leading to the results stated throughout this paper will be made available in an unpublished Villanova preprint #TH 23-04 (1979).

Lower bounds for quantum energy levels through statistical mechanics methods

L. Nitti, M. Pellicoro, M. Villani
Istituto di Fisica, Università di Bari Istituto Nazionale di Fisica Nucleare, Sezione di Bari, I-70126 Bari, Italy

(Received 17 July 1979; accepted for publication 30 November 1979)

A perturbative variational method for the calculation of lower bounds to energy eigenvalues is given for a wide class of quantum mechanical systems. The method makes use of the quantum statistical partition function and exploits the inequalities satisfied by the generalized Padè approximants. The mathematical and physical content of our approach is illustrated through a standard model.

I. INTRODUCTION

In the calculations of quantum bound states, a very important role is played by variational methods. As it is well known, through the Rayleigh–Ritz's variational principle, approximate values for the energy levels, having the meaning of upper bounds, can be obtained. Still, the knowledge of lower bounds can be very useful, too. By using the latter, in fact, both error estimates and bounds for the energy level differences are easily obtainable.

A class of lower bounds can be calculated through the Weyl's inequalities, combined with intermediate Hamiltonians,[1] or by means of the Löwdin's bracket method.[2] These methods are particularly effective for simple quantum mechanical systems. Lower bounds can be also calculated in the context of the Brillouin–Wigner implicit perturbation expansion.[3] Moreover, in the particular case of unidimensional anharmonic oscillators, lower bounds for the energy eigenvalues can be computed by starting from the Rayleigh–Schrödinger (R–S) explicit perturbation expansion and by using the diagonal Padé approximants (PA).[4,5] Unfortunately, this approach requires peculiar analytic properties of the energy levels,[6] which have not yet been proved valid in other cases.

In this paper, for a wide class of quantum mechanical systems, a perturbative-variational method for the calculation of lower bounds is proposed. Instead of the R–S series, we make use of a perturbative expansion typical of the quantum statistical mechanics; in accordance with the functional integral approach,[7] we take into consideration the partition function. In a way our method is based on a Gaussian integration formula for functional integrals; furthermore, it exploits the usefulness of the partition function from a variational point of view.

We consider a quantum mechanical system with a finite number N of degrees of freedom, described by a Hamiltonian of the type

$$H = H_0 + \lambda V'(q_1, q_2, \ldots, q_N)\,, \tag{I.1}$$

where

$$H_0 = \sum_{j=1}^{N} \frac{p_j^2}{2m_j} + V(q_1, q_2, \ldots, q_N)\,, \quad p_j = -i\frac{\partial}{\partial q_j}\,.$$

The potentials V and V' are real function of $q \equiv (q_1, q_2, \ldots, q_N)$. Internal degrees of freedom and symmetry conditions are not taken into account. We consider the operator (I.1) in a region Ω, where Ω can be either a bounded region of R^N or the whole Euclidean space R^N. In this paper we assume the following:

(i) the coupling constant λ is positive;

(ii) $V(q) + \lambda V'(q) \geqslant K\,, |K| < +\infty\,, V'(q) \geqslant x_0\,, |x_0| < +\infty\,;$ (I.2)

(iii) $|V(q) + \lambda V'(q)|^p$ is locally integrable, where $p = 2N/(N+2)$, if $N \geqslant 3, p > 1$, if $N = 2$ and $p = 1$ if $N = 1$;

(iv) when Ω is bounded, the Dirichlet boundary condition is imposed.

Furthermore, we suppose that

$$Z(\lambda, t) = \mathrm{Tr}\exp\{-t(H_0 + \lambda V')\} < +\infty \tag{I.3}$$

for $t > 0$ and every $\lambda \geqslant 0$ and that the eigenvalues and the eigenvectors of H_0 are known.

It is well known that a first order expansion in λ of Eq. (I.3) provides an upper bound for the ground state energy. This result is achieved through the limit $t \to +\infty$ and by making use of the Gibbs–Bogolioubov inequality or the convexity inequality satisfied by the exponential function.[7]

We show that a lower bound for the ground state energy can be obtained from Eq. (I.3) if the second order term in λ is considered as well. Furthermore, this lower bound can be improved in a systematic way through the calculation of the higher order terms. Our results are drawn from the inequalities satisfied by the generalized Padé approximants (GPA), which are exploited in connection with a monotonicity property of $(-1/t)\ln Z(\lambda, t)$ as a function of t. Moreover, we show that the knowledge of upper bounds for the energy eigenvalues with the addition of their degeneracy allows us to obtain lower bounds to the excited levels, too.

The basic ideas will be explained in Sec. II. The meaning of our approach is illustrated in Sec. III through a standard example. The calculations show a remarkable feature of the method, which works well both in the weak coupling and in the strong coupling limits. Nearly uniform approximations for every value of the coupling constant involved in the model are obtained. Additional comments are made in the concluding Sec. IV.

 0022-2488/80/040834-06$1.00

II. PERTURBATIVE-VARIATIONAL LOWER BOUNDS

We assume for the time being that, besides the conditions (I.2) and (I.3), the Hamiltonian H is positive definite. The starting point is the following proposition:

Proposition A: "For every fixed $t>0$, we have the integral representation

$$Z(\lambda, t)=\int_{tx_0}^{+\infty} e^{-\lambda x} d\mu(x; t), \tag{II.1}$$

where

$$x_0=\inf_{q_1,\ldots,q_N} V'(q_1, q_2,\ldots,q_N), \quad |x_0|<+\infty,$$

and $d\mu(x; t)$ is a positive measure."

The proposition A has been proved in Ref. 8, on the basis of the Trotter's product formula. The latter holds under the hypotheses (I.2;ii,iii,iv).[9]

The representation (II.1) allows us to obtain upper bounds for $Z(\lambda, t)$,[8] by making use of its perturbative expansion in powers of λ. In this paper we will assume that the perturbative coefficients are all finite. They can be obtained by iterating the Bloch equation in integral form:

$$e^{-tH}=e^{-tH_0}-\lambda\int_0^t ds\, e^{-(t-s)H_0} V' e^{-sH}, \tag{II.2}$$

or by considering the functional integral expression for $Z(\lambda, t)$.

If the expansion (convergent or not) is given by

$$Z(\lambda, t)=c_0(t)+\lambda c_1(t)+\ldots+\lambda^{2N} c_{2N}(t)+\ldots, \tag{II.3}$$

then the GPA[10]

$$Z^{(N/N)}(\lambda, t)=w_0^{(N)}(t) e^{-\lambda t x_0}+\sum_{i=1}^{N} w_i^{(N)}(t) e^{-\lambda \xi_i^{(N)}(t)}, \tag{II.4}$$

$N=0, 1,\cdots$.

$\{w_0^{(N)}, w_1^{(N)},\ldots, w_N^{(N)}, \xi_1^{(N)},\ldots,\xi_N^{(N)}$ are determined univocally by the condition $Z(\lambda, t)-Z^{(N/N)}(\lambda, t)=O(\lambda^{2N+1})\}$ is an upper bound to $Z(\lambda, t)$. More precisely, due to Eq. (I.2;i), we have[10,11]

$$Z(\lambda, t)\leqslant\ldots\leqslant Z^{(N/N)}(\lambda, t)\leqslant\ldots\leqslant Z^{(1/1)}(\lambda, t)\leqslant Z^{(0/0)}(\lambda, t), \tag{II.5}$$

where

$$Z^{(0/0)}(\lambda, t)=c_0(t) e^{-\lambda t x_0}=\mathrm{Tr}(e^{-tH_0})\cdot e^{-\lambda t x_0}.$$

From Eq. (II.1) it follows that $\xi_1^{(N)},\ldots,\xi_N^{(N)}$ are distinct and are all greater than tx_0. Furthermore, $w_0^{(N)},\ldots,w_N^{(N)}$ are positive.

When the moment problem associated with the measure $d\mu(x; t)$ is determinate,[12] we have

$$\lim_{N\to\infty} Z^{(N/N)}(\lambda, t)=Z(\lambda, t). \tag{II.6}$$

Now let $E_0, E_1,\ldots,$ be the eigenvalues (distinct) of H, ordered in the following way:

$$0<E_0<E_1<\ldots, \tag{II.7}$$

and let

$$\epsilon_0<\epsilon_1<\cdots$$

be the eigenvalues (distinct) of H_0.

First of all we consider the calculation of a lower bound for E_0. From the spectral theorem

$$Z(\lambda, t)=\sum_{n=0}^{\infty}\mu_n e^{-tE_n}, \tag{II.8}$$

where the integers μ_n represent the degeneracy of E_n, one obtains the standard formula

$$E_0=\lim_{t\to+\infty}\left\{-\frac{1}{t}\ln Z(\lambda, t)\right\}. \tag{II.9}$$

On the other hand, from Eq. (II.5) it follows that

$$-\frac{1}{t}\ln Z(\lambda, t)\geqslant\cdots\geqslant-\frac{1}{t}\ln Z^{(N/N)}(\lambda, t)\geqslant\cdots$$
$$\geqslant-\frac{1}{t}\ln Z^{(1/1)}(\lambda, t)\geqslant-\frac{1}{t}\ln Z^{(0/0)}(\lambda, t) \tag{II.10}$$

for every $t>0$.

At first sight, formula (II.9) would lead us to carry out the limit $t\to+\infty$ in Eq. (II.10) and then to get lower bounds for E_0.[13] However, we will show at once that we achieve a proper calculation of lower bounds for E_0 in a different way. In fact, one must bear in mind the following proposition:

Proposition B: "For every fixed λ, $(-1/t)\ln Z(\lambda, t)$ is a monotonous increasing function of t, $(t>0)$".

Proof: Let us consider the function

$$\bar{Z}(\lambda, t)=\mu_0+\sum_{n=1}^{\infty}\mu_n e^{-t(E_n-E_0)}. \tag{II.11}$$

The derivative of Eq. (II.11) in the t variable can be done term by term. Then, it follows from Eq. (II.7) that $\bar{Z}'(\lambda, t)<0$. Furthermore, we have $\bar{Z}(\lambda, t)>1$. Thus,

$$\frac{d}{dt}\left\{-\frac{1}{t}\ln\bar{Z}(\lambda, t)\right\}=\frac{1}{t^2}\ln\bar{Z}(\lambda, t)-\frac{1}{t}\frac{\bar{Z}'(\lambda, t)}{\bar{Z}(\lambda, t)}>0.$$

Then the proposition follows from

$$-\frac{1}{t}\ln Z(\lambda, t)=E_0-\frac{1}{t}\ln\bar{Z}(\lambda, t). \tag{II.12}$$

The monotonicity property, stated by the previous proposition, allows us to carry out the limit $t\to+\infty$ for $(-1/t)\cdot\ln Z(\lambda, t)$ in Eq. (II.10), yielding

$$E_0\geqslant\ldots\geqslant-\frac{1}{t}\ln Z^{(N/N)}(\lambda, t)\geqslant\ldots\geqslant-\frac{1}{t}\ln Z^{(1/1)}(\lambda, t)$$
$$\geqslant-\frac{1}{t}\ln Z^{(0/0)}(\lambda, t). \tag{II.13}$$

It follows that the best lower bound for E_0, attainable through $Z^{(N/N)}(\lambda, t)$, is given by

$$E_0^{(N/N)}(\lambda)=\sup_{t\geqslant 0}\left(-\frac{1}{t}\ln Z^{(N/N)}(\lambda, t)\right). \tag{II.14}$$

It can be easily verified that $E_0^{(0/0)}(\lambda)=\epsilon_0+\lambda x_0$. Summarizing, we have

$$E_0(\lambda)\geqslant\ldots\geqslant E_0^{(N/N)}(\lambda)\geqslant\ldots\geqslant E_0^{(1/1)}(\lambda)\geqslant\epsilon_0+\lambda x_0 \tag{II.15}$$

for every fixed $\lambda>0$.

If Eq. (II.6) holds, then we have also

$$\lim_{N\to+\infty} E_0^{(N/N)}(\lambda)=E_0(\lambda). \tag{II.16}$$

The limit (II.16) replaces the double limit

$$\lim_{t\to+\infty}\lim_{N\to+\infty}\left(-\frac{1}{t}\ln Z^{(N/N)}(\lambda,t)\right)=E_0(\lambda)$$

by a single one. We have met a similar procedure for a class of power series expansion with divergent coefficients,[14,15] where a limit like Eq. (II.16) has been proved to be valid.

Now, let us suppose we know an upper bound U_0 for E_0 ($E_0\leqslant U_0$). We have then (we assume $\mu_0=1$)

$$\begin{aligned}Z(\lambda,t)-e^{-tE_0}&\leqslant Z^{(N/N)}(\lambda,t)-e^{-tE_0}\\&\leqslant Z^{(N/N)}(\lambda,t)-e^{-tU_0}.\end{aligned}\tag{II.17}$$

If $Z(\lambda,t)$ is replaced by $Z(\lambda,t)-e^{-tE_0}$, then the proposition B keeps its validity. The inequalities (II.7) and the limit

$$E_1=\lim_{t\to+\infty}\left(-\frac{1}{t}\ln\{Z(\lambda,t)-e^{-tE_0}\}\right)\tag{II.18}$$

lead to

$$\begin{aligned}E_1&\geqslant\ldots\geqslant-\frac{1}{t}\ln\{Z^{(N/N)}(\lambda,t)-e^{-tU_0}\}\geqslant\ldots\\&\geqslant-\frac{1}{t}\ln\{Z^{(1/1)}(\lambda,t)-e^{-tU_0}\}\\&\geqslant-\frac{1}{t}\ln\{Z^{(0/0)}(\lambda,t)-e^{-tU_0}\}\end{aligned}\tag{II.19}$$

for every $t>0$.

Therefore, the best lower bound for E_1, obtainable through $Z^{(N/N)}(\lambda,t)$ and U_0, is given by

$$\begin{aligned}E_1^{(N/N)}(\lambda)&=\sup_{t\geqslant0}\left(-\frac{1}{t}\ln\{Z^{(N/N)}(\lambda,t)-e^{-tU_0}\}\right)\\&\leqslant E_1(\lambda).\end{aligned}\tag{II.20}$$

If $\{U_0(N)\}$ is a sequence of upper bounds convergent to E_0 and if Eq. (II.6) holds, we have, replacing U_0 with $U_0(N)$ in Eq. (II.20),

$$\lim_{N\to+\infty}E_1^{(N/N)}(\lambda)=E_1(\lambda).\tag{II.21}$$

The results (II.20) can be easily generalized. Let us suppose we know the multiplicity $\mu_0,\mu_1,\ldots,\mu_{l-1}$ and some upper bounds $U_0,U_1,\ldots,U_{l-1}$ for $E_0,E_1,\ldots,E_{l-1}$. We have then

$$\begin{aligned}E_l^{(N/N)}&=\sup_{t>0}\left(-\frac{1}{t}\ln\{Z^{(N/N)}(\lambda,t)-\sum_{i=0}^{l-1}\mu_i e^{-tU_i}\}\right)\\&\leqslant E_l(\lambda).\end{aligned}\tag{II.22}$$

We can obtain $U_0,U_1,\ldots,U_{l-1}$, for example, through the usual Rayleigh–Ritz variational principle.

We remark that, when H is not definite positive, the previous results can be applied to the modified positive definite Hamiltonian: $H_0-\epsilon_0+\lambda(V'-x_0)$.

In the previously described procedure, the t variable, which has the physical meaning of the inverse of a temperature, plays the role of a variational parameter. However, in order to improve the method we can introduce additional variational parameters through suitable potentials, as it is usually done in statistical mechanics.

Let us start from the following Hamiltonian:

$$H=\sum_{j=1}^{N}\frac{p_j^2}{2m_j}+V(q_1,\ldots,q_N)+gU(q_1,\ldots,q_N)\qquad\left(p_j=-i\frac{\partial}{\partial q_j}\right)\tag{II.23}$$

and let us assume we know the eigenvalues and the eigenvectors of the unperturbed Hamiltonian

$$H_0=\sum_{j=1}^{N}\frac{p_j^2}{2m_j}+V(q_1,\ldots,q_N)+W(q_1,\ldots,q_N;\alpha),\tag{II.24}$$

where the potential $W(q_1,\ldots,q_N;\alpha)$ depends on a set of real parameters $\alpha\equiv(\alpha_1,\alpha_2,\ldots,\alpha_l)$, which belong to some region D. We consider, then, the Hamiltonian

$$\begin{aligned}H(\lambda;\alpha)&=H_0+\lambda(gU(q_1,\ldots,q_N)-W(q_1,\ldots,q_N;\alpha))\\&=H_0+\lambda V'(q),\end{aligned}\tag{II.25}$$

where $\lambda>0$. However, we are interested in the value $\lambda=1$, in order to recover our Hamiltonian (II.23): $H(1;\alpha)=H$.

Let us suppose that $H(\lambda;\alpha)$, for every α in D, satisfies our basic conditions (I.2) and (I.3), and that

$$Z(\lambda,t;\alpha)=\mathrm{Tr}\,e^{-tH(\lambda;\alpha)}\tag{II.26}$$

can be expanded in powers of λ with finite coefficients. Starting from this expansion, we can build up the upper bound

$$Z^{(N/N)}(\lambda,t;\alpha)\geqslant Z(\lambda,t;\alpha)\tag{II.27}$$

for every α in D and every $\lambda>0$. In particular, we have

$$Z^{(N/N)}(1,t;\alpha)\geqslant Z(1,t;\alpha)=\mathrm{Tr}\,e^{-tH}.\tag{II.28}$$

If H is positive definite, it follows from Eq. (II.28) that

$$\sup_{t\geqslant0,\alpha\in D}\left(-\frac{1}{t}\ln Z^{(N/N)}(1,t;\alpha)\right)\leqslant E_0,\tag{II.29}$$

where E_0 is the ground state energy.

The extension is straightforward to the excited levels and to the case of a not positive definite Hamiltonian.

III. AN EXAMPLE

The method given in the previous section will be illustrated through the standard unidimensional anharmonic oscillator

$$H=-\frac{1}{2}\frac{d^2}{dx^2}+\frac{1}{2}x^2+gx^4\quad(g>0).\tag{III.1}$$

The eigenvalues of the Hamiltonian (III.1) have been already calculated in the literature by several methods and with high accuracy. Our aim is not to reproduce results already known, but to see how our method works and to show its mathematical and physical meaning. We limit ourselves to consider the first nontrivial approximation of our approach, which needs the calculation of $Z^{(1/1)}(\lambda,t)$. We take as unperturbed Hamiltonian

$$H_0=-\frac{1}{2}\frac{d^2}{dx^2}+\frac{1}{2}(1+\omega^2)x^2\tag{III.2}$$

so that

$$H(\lambda)=H_0+\lambda\left(gx^4-\frac{1}{2}\omega^2x^2\right)=H_0+\lambda V'.\tag{III.3}$$

TABLE I. Comparison of our lower bounds for the ground state energy of the anharmonic oscillator (III.1) with the exact values.

g	$\omega_0(g)$	$t_0(g)$	$E_0^{(1/1)}(g)$	$E_0(g)$
0.05	0.526	8.30	0.5325	0.5326
0.50	1.381	3.80	0.6912	0.6962
7.00	3.560	1.50	1.3240	1.3517
$g\to+\infty$	$1.884\,g^{1/3}$	$2.91/g^{1/3}$	$0.6507\,g^{1/3}$	$0.6680g^{1/3}$

In this case the set α is reduced to the parameter $\omega\geqslant 0$.

In order to build up $Z^{(1/1)}(\lambda, t; \omega)$, we must calculate the first three coefficients of the expansion

$$\mathrm{Tr}\, e^{-tH(\lambda)} = c_0(t,\omega) + \lambda c_1(t,\omega) + \lambda^2 c_2(t,\omega) + \dots .$$

Using standard techniques, we obtain

$$c_0(t,\omega) = \frac{e^{-\alpha}}{1-e^{-2\alpha}},$$

$$c_1(t,\omega) = -tc_0(t,\omega)\frac{\coth\alpha}{4\sqrt{1+\omega^2}}\left(3g\frac{\coth\alpha}{\sqrt{1+\omega^2}} - \omega^2\right), \tag{III.4}$$

$$c_2(t,\omega) = \frac{t^2}{2}c_0(t,\omega)\left(g^2B_{22} - g\omega^2B_{21} + \frac{\omega^4}{4}B_{11}\right),$$

where

$$\alpha = \tfrac{1}{2}\sqrt{1+\omega^2}\,t, \quad \rho = \min(r,j),$$

$$B_{rj} = (2r)!(2j)!\left(\frac{\coth\alpha}{4\sqrt{1+\omega^2}}\right)^{r+j}\sum_{l=0}^{\rho}\frac{(\mathrm{sech}\alpha)^{2l}}{(r-l)!(j-l)!(2l)!} \times\left(\binom{2l}{l} + \frac{2}{\alpha}\sum_{i=0}^{l-1}\binom{2l}{i}\frac{\sinh 2(l-i)\alpha}{2(l-i)}\right).$$

The GPA $Z^{(1/1)}(\lambda, t; \omega)$ is given by

$$Z^{(1/1)}(\lambda,t;\omega) = c_0(t,\omega) \times (w_0(t,\omega)e^{-\lambda t x_0(\omega)} + w_1(t,\omega)e^{-\lambda\xi_1(t,\omega)}), \tag{III.5}$$

where

$$w_0 = \frac{2\tilde{c}_2 - \tilde{c}_1^2}{t^2x_0^2 + 2tx_0\tilde{c}_1 + 2\tilde{c}_2}, \quad w_1 = \frac{(tx_0+\tilde{c}_1)^2}{t^2x_0^2 + 2tx_0\tilde{c}_1 + 2\tilde{c}_2},$$

$$\xi_1 = -\frac{2\tilde{c}_2 + tx_0\tilde{c}_1}{\tilde{c}_1 + tx_0},$$

$$x_0 = \inf_x V'(x) = -\frac{\omega^4}{g}, \quad \tilde{c}_1 = c_1/c_0, \quad \tilde{c}_2 = c_2/c_0.$$

Previous formulas allow us to calculate the lower bound for the ground state energy $E_0(g)$:

$$E_0^{(1/1)}(g) = \sup_{t>0,\omega>0}\left\{-\frac{1}{t}\ln Z^{(1/1)}(1,t;\omega)\right\}. \tag{III.6}$$

In Table I we give $E_0^{(1/1)}(g)$ for some values of g, including the strong coupling limit $g\to+\infty$. The maximum on the right-hand side of Eq. (III.6) is attained for finite values of t and ω, which are functions of g and are denoted by $t_0(g)$ and $\omega_0(g)$, respectively. In the last column we report values of $E_0(g)$, calculated numerically with very high accuracy (see, for instance, Ref. 16).

We point out that our method, is able to provide lower bounds for $E_0(g)$, which also turn out to be very good approximations. The approximate values, of course, are better for small values of g, but the discrepancy keeps small also in the limit $g\to+\infty$.

It is worthwhile to show some remarkable properties of our results by the following comparisons:

Consider the usual R–S perturbative expansion for the ground state energy of the Hamiltonian (III.1):

$$E_0(g) = a_0 + a_1g + a_2g^2 + \cdots . \tag{III.7}$$

It has been proved[4,5] that the Padé approximants (PA) $[N/N](g)$ to the series (III.7) (a ratio of two polynomials in g of degree N) give lower bounds for $E_0(g)$ and that

$$\lim_{N\to+\infty}[N/N](g) = E_0(g)$$

for every fixed g. These results are a consequence of the analytical structure of $E_0(g)$, which is a Stieltjes function in the case of the anharmonic oscillator.

Now, let us compare our calculations with those reported in Ref. 5. If we take, for example, $g = 0.5$, we see that $E_0^{(1/1)}(0.5)$ is more accurate than the value obtained from the PA [3/3] to the series (III.7), which requires the knowledge of the expansion of $E_0(g)$ up to the order g^6: [3/3](0.5) $= 0.6869$. $E_0^{(1/1)}(0.5)$ is closer to the value given by the PA [4/4](0.5) $= 0.69188$, which is obtained through a perturbative expansion up to the order g^8. For greater values of g, $E_0^{(1/1)}(g)$ is exceptionally good when compared to the PA $[N/N](g)$. For example, $E_0^{(1/1)}(7.)$ is better than the PA [20/20] to Eq. (III.7), which requires the expansion of $E_0(g)$ up to the order g^{40}: [20/20](7.) $= 1.2785$.

If we consider the Padé–Borel approximants $[N/N]_B(g)$ to the series (III.7), we find that $[10/10]_B(7.) = 1.3417$.[17] We remark that $[10/10]_B(7.)$ is a better lower bound for E_0 than ours, but does not follow from any rigorous inequality.[17] On the other hand, the approximation $E_0^{(1/1)}(g)$ will certainly be more accurate than $[N/N]_B(g)$ for higher values of g. In fact, when $g\to+\infty$, the $[N/N]_B(g)$, as well as the $[N/N](g)$, become constants, while $E_0^{(1/1)}(g)/g^{1/3}$ has a finite limit, in accordance with the behavior of $E_0(g)$. It is well known[5] that the limit

$$E_0(g)/g^{1/3}\to\text{constant}$$

comes from a scaling property of the Hamiltonian (III.1). This property is recovered by $E_0^{(1/1)}(g)$ through the behavior of the maximum points $t_0(g)$ and $\omega_0(g)$ as a function of g. More precisely, it results that $\omega_0(g)$ increases with g without bounds. Then, for very large g, $t_0(g)$ and $\omega_0(g)$ can be deter-

TABLE II. Comparison of our lower bounds for the first excited level of the anharmonic oscillator (III.1) with the exact values.

g	$U_0(g)$	$\omega_1(g)$	$t_1(g)$	$E_1^{(1/1)}(g)$	$E_1(g)$
0.05	0.532643	0.564	3.55	1.6441	1.6534
0.50	0.696700	1.516	1.61	2.2412	2.3244
$g\to+\infty$	$0.680\,g^{1/3}$	$2.11\,g^{1/3}$	$1.13/g^{1/3}$	$2.1509\,g^{1/3}$	$2.3936\,g^{1/3}$

mined by replacing $(1+\omega^2)^{1/2}$ with ω ($\omega > 0$) in Eq. (III.4). After this change, if we introduce the parameters

$$\alpha = \frac{1}{2}\omega t\,, \quad y = \frac{g}{4\omega^3}\,, \tag{III.8}$$

it can be easily verified, through Eqs. (III.4) and (III.5), that

$$-\frac{1}{t}\ln Z^{(1/1)}(1, t; \omega) = -\frac{2g^{1/3}}{4^{4/3}\alpha y^{1/3}}\ln Z^{(1/1)}(\alpha, y)\,. \tag{III.9}$$

Therefore, for very large g, Eq. (III.6) leads to

$$E_0^{(1/1)}(g) = \frac{2g^{1/3}}{4^{4/3}} \sup_{y>0,\alpha>0}\left(-\frac{1}{y^{1/3}\alpha}\ln Z^{(1/1)}(\alpha, y)\right), \tag{III.10}$$

which shows clearly the above mentioned scaling property.

The expression at the right of Eq. (III.10) attains its maximum for $\alpha = 2.741$ and $y = 0.0374$; from these values and Eq. (III.8) we obtain $t_0(g)$ and $\omega_0(g)$ for very large g.

We consider, at last, the calculation of the lower bound $E_1^{(1/1)}(g)$ for the first excited level $E_1(g)$ of Eq. (III.1):

$$E_1^{(1/1)}(g) = \sup_{t>0,\omega\geq 0}\left(-\frac{1}{t}\ln\{Z^{(1/1)}(1, t; \omega) - e^{-tU_0}\}\right), \tag{III.11}$$

where $U_0(g)$ is an upper bound to $E_0(g)$. The results are given in Table II: $t_1(g)$ and $\omega_1(g)$ are the points where the right-hand side of Eq. (III.11) attains its maximum and the last column gives the exact values of $E_1(g)$.[16] As upper bounds for $U_0(g)$ we used very accurate values for small g.

The results show that our approximation for $E_1(g)$ behaves like the one obtained for $E_0(g)$, but it is less accurate, even if we use good values for $U_0(g)$. For instance, when $g = 0.5$, we have an error of 0.7% on $E_0^{(1/1)}$, against an error of 3.6% on $E_1^{(1/1)}$. Again for $g = 0.5$, the lower bound $E_2^{(1/1)}$ for the second excited level, calculated through good upper bounds for E_0 and E_1, gives a result with an error of 10%. We point out that this feature is common to other methods. In fact, a more accurate calculation of the excited levels would require either a higher order approximation or a clever choice of H_0.

We conclude this section with a comment on the physical meaning of our method. We limit ourselves to $E_0^{(1/1)}(g)$ and put Eq. (III.6) in the form

$$E_0^{(1/1)}(g) = \max_{t>0}\left(-\frac{1}{t}\ln Z^{(1/1)}(1, t; \omega_0(g))\right). \tag{III.12}$$

The maximum point is given by the condition

$$-\frac{1}{t}\ln Z^{(1/1)}(1, t; \omega_0(g)) + \frac{(d/dt)Z^{(1/1)}(1, t; \omega_0(g))}{Z^{(1/1)}(1, t; \omega_0(g))} = 0\,.$$

It follows then that

$$E_0^{(1/1)}(g) = -\left.\frac{(d/dt)Z^{(1/1)}(1, t; \omega_0(g))}{Z^{(1/1)}(1, t; \omega_0(g))}\right|_{t=t_0(g)}. \tag{III.13}$$

In other words, $E_0^{(1/1)}(g)$ has the meaning of a mean energy at the temperature $1/t_0(g)$ of a system described by the partition function $Z^{(1/1)}(1, t; \omega_0(g))$.

IV. CONCLUSIONS

The partition function has been the object of several investigations, due to its expression as a functional integral. Using classical methods, one can get out very useful results from it. Recently, the large orders of the R–S perturbative series have been calculated by applying the saddle point method.[18] Using the functional integral for the partition function with imaginary t ($t \to it$), the stationary phase method is able to recover the semiclassical part of the energy spectrum.[19]

In this paper it has been shown an approximation scheme very different from the classical ones and based on the partition function as Laplace transform of a positive measure. This property allows us to state rigorous inequalities and to draw out lower bounds for the energy eigenvalues. From the functional integral point of view, our method corresponds to splitting all the paths in families; in each one of them the potential term gives the same contribution.

Since in Euclidean field theory the partition function plays a very important role, it would be useful to extend our results to systems with infinite number of degrees of freedom. As a further extension, symmetry conditions have to be taken into account. Moreover, for quantum mechanical systems with a continuous part in the energy spectrum, one has to replace the partition function with a different statistical quantity. We think that the proper candidate would be the density matrix, which satisfies the proposition A in the coordinate representation.[8] However, we should probably replace or generalize proposition B with a new one.

[1] N.W. Bazley, D.W. Fox, Phys. Rev. **124**, 483 (1961).
[2] P.O. Löwdin, Phys. Rev. Sect. A **139**, 357 (1965); J. Chem. Phys. **43**, 175 (1965); C. Reid, J. Chem. Phys. **43**, 186 (1965).
[3] B.G. Giraud: "Lower bounds to bound state eigenvalues", Saclay preprint DPh-T/77/88 (1977); F. Brandas, D.A. Micha, J. Math. Phys. **13**, 155 (1972).
[4] J.J. Loeffel, A. Martin, B. Simon, and A.S. Wightmann, Phys. Lett. B **30**, 656 (1969).
[5] B. Simon, Ann. Phys. (Leipzig) **58**, 76 (1970).
[6] J.J. Loeffel and A. Martin, "Proprietés analytiques des niveaux de l'oscillateur anharmonique et convergence des approximants de Padé", *Cargèse Lectures in Physics*, edited by D. Bessis (Gordon and Breach, New York, 1972), Vol. 5.
[7] R.P. Feynman and A.R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).
[8] D. Bessis, P. Moussa, and M. Villani, J. Math. Phys. **16**, 2318 (1975).
[9] T. Kato, Ann. Scuola Norm. Super. Pisa, Cl. Sci. Seri. IV **5**, 105 (1978).
[10] G. Baker Jr., *Essentials of Padé Approximants* (Academic, New York, 1975).
[11] J.C. Wheeler and R.G. Gordon: "Bounds for Averages Using Moment Constraints", *Padé Approximants in Theoretical Physics*, edited by G. Baker and J. Gammel (Academic, New York, 1970).

[12] This requirement is fulfilled under broad conditions.[10]
[13] Concerning this limit, it follows easily from Eq. (II.10) that

$$\liminf_{t\to+\infty}\left(-\frac{1}{t}\ln Z^{(N/N)}(\lambda, t)\right) \geq \lim_{t\to+\infty}\left(-\frac{1}{t}\ln Z^{(0/0)}(\lambda, t)\right) = \epsilon_0 + \lambda x_0\,.$$

We note that $\epsilon_0 + \lambda x_0$ is a trivial lower bound for E_0 due to the obvious inequality $H_0 + \lambda V' \geq H_0 + \lambda x_0$.

[14]G. Fogli, M. Pellicoro, and M. Villani, Nuovo Cimento A **6**, 79 (1971).
[15]D. Bessis, L. Epele, and M. Villani, J. Math. Phys. **15**, 2071 (1974).
[16]F.T. Hioe and E.W. Montroll, J. Math. Phys. **16**, 1945 (1975).
[17]S. Graffi, V. Grecchi, and B. Simon, Phys. Lett. B **32**, 631 (1970).
[18]E. Brezin, J.G. Le Guillon, and J. Zinn-Justin, Phys. Rev. D **15**, 1544 (1977).
[19]R.F. Dashen, B. Hasslacher, and A. Neveu, Phys. Rev. D **10**, 4114 (1974).

Perturbation of self-adjoint operators by Dirac distributions

J. Zorbas
Department of Mathematics, The University of Alberta, Edmonton, Canada T6G 2G1

(Received 14 February 1979; accepted for publication 12 June 1979)

The existence of a family of self-adjoint Hamiltonians H_θ, $\theta \in [0, 2\pi)$, corresponding to the formal expression $H_0 + v\delta(x)$ is shown for a general class of self-adjoint operators H_0. Expressions for the Green's function and wavefunction corresponding to H_θ are obtained in terms of the Green's function and wavefunction corresponding to H_0. Similar results are shown for the perturbation of H_0 by a finite sum of Dirac distributions. A prescription is given for obtaining H_θ as the strong resolvent limit of a family of momentum cutoff Hamiltonians H^N. The relationship between the scattering theories corresponding to H^N and H_θ is examined.

I. INTRODUCTION

The assumption that the particles making up a quantum mechanical system interact by "zero-range potentials" (also referred to as "delta-function potentials") or "delta-shell potentials" provides a considerable simplification of the scattering and bound state problems. For example, the S-matrix has been obtained explicitly for various particular cases of N-particles interacting in R^1 by zero-range potentials.[1-4] In R^3 the scattering of two particles interacting by a zero-range potential can be solved exactly[5,6] which simplifies the corresponding three particle problem.[7] In the case of delta-shell potentials both the scattering and bound state problems can be solved exactly.[8,9]

In this paper we consider the problem of a particle interacting with N fixed particles, located at a_i, $1 \leqslant i \leqslant N$, by zero-range potentials plus a potential $V(x)$ with $V(x)$ a real function of $x \in R^n$. The formal Hamiltonian is given by

$$-\Delta + V(x) + \sum_{j=1}^{N} v_j \delta(x - a_j), \tag{1.1}$$

where

$$\Delta = \sum_{i=1}^{n} \frac{\partial^2}{\partial x_i^2}$$

and for each $1 \leqslant j \leqslant N$, $v_j \in R^1$ and $\delta(x - a_j)$ is the Dirac delta-function.

It is usually possible to associate with $-\Delta + V(x)$ a self-adjoint operator, denoted by H_0, with domain $D(H_0) \subset \mathscr{H} = \mathscr{L}^2(R^n; dx)$. If each $\psi \in D(H_0)$ is continuous at a_i, $i = 1,...,N$, then the operator H given by

$$(H\psi)\hat{}(p) = (H_0\psi)\hat{}(p) + (2\pi)^{-n/2} \sum_{j=1}^{N} v_j \exp(-ipa_j)\psi(a_j), \tag{1.2}$$

is well defined on $D(H_0)$. In general, however, $H\psi$ will not be in $\mathscr{H}$ for every $\psi \in D(H_0)$.

In this paper the following three methods for associating a self-adjoint operator with (1.1) will be considered:

Method 1: Let $\mathscr{D}$ be defined by

$$\mathscr{D} = \{\psi \in D(H_0) | \psi(a_i) = 0, \quad i = 1,...,N\}, \tag{1.3}$$

and assume that $\mathscr{D}$ is dense in $\mathscr{H}$. Define the Hamiltonian corresponding to H as a self-adjoint extension of $T = \overline{H_0 \upharpoonright \mathscr{D}}$.

Method 2: Assume H_0 is semibounded and let $\langle \phi, H_0\psi \rangle$ denote the quadratic form corresponding to H_0. Let q denote the quadratic form defined by

$$q(\phi, \psi) = \langle \phi, H_0\psi \rangle + \sum_{j=1}^{N} v_j \overline{\phi(a_j)}\psi(a_j). \tag{1.4}$$

If $\Sigma_{j=1}^{N} v_j \overline{\phi(a_j)}\psi(a_j)$ is relatively $\langle \phi, H_0\psi \rangle$-form bounded with relative form bound less than 1, then there exists a unique self-adjoint operator corresponding to q.[10]

Method 3: Consider the family of operators H^N defined by

$$(H^N\psi)\hat{}(p) = (H_0\psi)\hat{}(p) + v(N)(2\pi)^{-n} \sum_{j=1}^{N} \exp(-ipa_j)\chi_N(p) \times \int dq \exp(iqa_j)\chi_N(q)\hat{\psi}(q), \tag{1.5}$$

where $\hat{\psi}(q)$ is the Fourier transform of $\psi(x)$ and $\chi_N \in C^\infty(R^n)$ satisfies

$$\chi_N(p) = \begin{cases} 1 & \text{if } |p| \leqslant N \\ 0 & \text{if } |p| \geqslant N+1 \end{cases}.$$

For appropriate V the operator H^N is self-adjoint for each $N < \infty$. Define the Hamiltonian corresponding to H as the strong graph limit[11] of H^N (if this limit exists and is self-adjoint).

Method 1 has been used to give a rigorous definition of (1.1) for $H_0 = -\Delta$, $n = 3$.[6] The quadratic form approach has been used for $H_0 = -d^2/dx^2$, $n = 1$.[12-14] The definition of the Hamiltonian as the strong graph limit of momentum cutoff Hamiltonians is motivated by a standard prescription for dealing with delta-function potentials.[5-7,15] An alternative approach to the problem of delta-function potentials is given in Refs. 16 and 17.

In Sec. II of this paper a technical result is proven which allows us to associate a family of self-adjoint operators with the formal perturbation of a general self-adjoint operator by a finite sum of Dirac distributions. Various examples are given in Sec. III in the case of a single Dirac delta-function. The general case of a finite sum of Dirac distributions is discussed in Sec. V.

We now give a brief outline of the results contained in this paper concerning the self-adjointness of H for the case of a single Dirac delta-function with $H_0 = -\Delta$. Similar results

 0022-2488/80/040840-08$1.00

are also shown in the case $H_0 = -\Delta + V(x)$ for a general class of potentials $V(x)$.

For $n = 1$ each of the three methods described above leads to an infinite family H_θ, $\theta \in [0, 2\pi)$, of self-adjoint operators corresponding to H. In particular, for each fixed $\nu \in R^1$ there exists a $\theta \in [0, 2\pi)$ such that H_θ corresponds to the quadratic form (1.4) and is also the strong graph limit of H^N.

The form approach is shown to fail for $n \geqslant 3$. For $n \geqslant 4$ Methods 1 and 3 yield $-\Delta$ as the precise version of H. Thus it does not seem possible to define a zero-range potential for $n \geqslant 4$.

For $n = 3$ Method 1 yields an infinite family of self-adjoint operators H_θ, $\theta \in [0, 2\pi)$.[6] It is shown in Sec. VIII that if one "renormalizes" the coupling constant ν appearing in (1.5), that is, allows ν to have a particular N dependence, then H^N converges to H_θ in the strong graph limit. This result provides a rigorous relationship between the momentum cutoff dynamics and the operators H_θ.

In Sec. IX the wave operators $W_\pm(H^N, -\Delta)$ are shown to converge strongly in the limit $N \to \infty$ to the wave operators $W_\pm(H_\theta, -\Delta)$ for $n = 1$ and 3.

The results contained in Secs. VIII and IX concerning the large N behavior of H^N and $W_\pm(H^N, -\Delta)$ are motivated by the work of Berezin and Faddeev.[6,5,7] They showed, for $n = 3$ and $V(x) = 0$, the convergence of the wavefunction corresponding to H^N (with ν depending on N) to the wavefunction corresponding to H_θ in the limit $N \to \infty$. The analogous result for wave operators is verified in Sec. IX of this paper. In particular, the Hilbert space techniques used in this paper allow us to verify the convergence of $W_\pm(H^N, -\Delta)$ for a general class of potentials $V(x)$.

The paper concludes with a discussion of the validity of the stationary formalism when zero-range or delta-shell potentials are present. The applicability of the results of this paper to the problem of delta-shell potentials[8] and the modified Fermi potential[9] is discussed.

II. THE DEFICIENCY SUBSPACES

Let H_0 be a self-adjoint operator in $\mathscr{H}$ and $T = H_0 \upharpoonright \mathscr{D}$ where $\mathscr{D}$ is defined by (1.3). In the following a characterization of $\mathscr{K} = \mathrm{Ran}(T - z)^\perp$, $\mathrm{Im}z \neq 0$, is given which allows us to construct the deficiency subspaces of T explicitly.

In order to construct $\mathscr{K}$ we make the following assumptions concerning the Green's function and domain of H_0: $\mathscr{A}$.

There exists for each $z \in C$, $\mathrm{Im}z \neq 0$, and almost all $(x, y) \in R^n \times R^n$ a function $G(x, y; z)$ which satisfies:

(i) $((H_0 - z)^{-1}\psi)(x) = \int G(x, y; z)\psi(y)\, dy$, for each $\psi \in \mathscr{H}$.

(ii) $G(\cdot, y; z) \in \mathscr{H}$.

(iii) $G(x, y; z) = G(y, x, z)$.

(iv) $G(x, y; z) = \overline{G(x, y; \bar{z})}$.

$\mathscr{B}$. Each $\psi \in D(H_0)$ can be written as a sum of a continuous $\mathscr{L}^\infty$-function and a $\mathscr{L}^1$-function which is continuous at a_j; $j = 1,...,N$. Furthermore for each $a_j \in R^n$, $1 \leqslant j \leqslant N$, there exists $\eta_j \in D(H_0)$ with $\eta_j(a_j) = 1$ and $\mathrm{supp}\,\eta_i \cap \mathrm{supp}\eta_j = \emptyset$ for $i \neq j$.

Lemma 2.1: Let H_0 be a self-adjoint operator and assume $T = H_0 \upharpoonright \mathscr{D}$ is a closed symmetric operator in $\mathscr{H}$. If $\mathscr{A}$ and $\mathscr{B}$ are satisfied, then

$$\mathscr{K} = \left\{ \phi(x) \middle| \phi(x) = \sum_{j=1}^{N} c_j G(x, a_j; \bar{z}),\ c_j \in C,\ 1 \leqslant j \leqslant N \right\}. \tag{2.1}$$

Proof: Denote the right side of (2.1) by $\mathscr{K}'$. We first show that $\mathscr{K}' \subseteq \mathscr{K} = \mathrm{Ran}(T - z)^\perp$. It is sufficient to show

$$\langle G(\cdot, a_i; \bar{z}) | (H_0 - z)\phi \rangle = 0, \tag{2.2}$$

for each $\phi \in \mathscr{D}$, $1 \leqslant i \leqslant N$.

Let

$$g_\epsilon(x) = ((H_0 - z)^{-1}k_\epsilon)(x),$$

$$k_\epsilon(y) = \pi^{-n} \prod_{j=1}^{n} \frac{\epsilon}{(y_j - a_{ij})^2 + \epsilon^2},$$

where $y = (y_1,...,y_n)$, $a_i = (a_{i1},...,a_{in})$. Since each $\psi \in D(H_0)$ satisfies $\mathscr{B}$, we have

$$\lim_{\epsilon \to +0} \langle g_\epsilon | \phi \rangle = \lim_{\epsilon \to +0} \langle k_\epsilon | (H_0 - \bar{z})^{-1}\phi \rangle = \langle G(\cdot, a_i; z) | \phi \rangle,$$

for each $\phi \in \mathscr{H}$.

An application of the above result yields

$$\langle G(\cdot, a_i; \bar{z}) | (H_0 - z)\phi \rangle = \lim_{\epsilon \to +0} \langle k_\epsilon | \phi \rangle = \phi(a_i) = 0$$

for each $\phi \in \mathscr{D}$. Thus (2.2) is valid and $\mathscr{K}' \subseteq \mathscr{K}$.

We will now show that $\mathscr{K} \subseteq \mathscr{K}'$.

Let ψ be a fixed vector in $\mathscr{K}$. We first show that there exists $\lambda_i \in C$, $i = 1,...,N$, such that

$$\langle \psi | (H_0 - z)\phi \rangle = \sum_{i=1}^{N} \lambda_i \phi(a_i), \tag{2.3}$$

for each $\phi \in D(H_0)$. Let $\tilde{\phi}(x)$ be defined by

$$\tilde{\phi}(x) = \phi(x) - \sum_{i=1}^{N} \eta_i(x)\phi(a_i),$$

where the functions $\eta_i(x)$ are given in $\mathscr{B}$. Clearly $\tilde{\phi} \in \mathscr{D}$, and since $\psi \in \mathrm{Ran}(T - z)^\perp$, (2.3) follows with $\lambda_i = \langle \psi | (H_0 - z)\eta_i \rangle$.

We now note that there can be at most one function $\psi \in \mathscr{H}$, which satisfies (2.3). That is, if $\psi' \in \mathscr{H}$ satisfies

$$\langle \psi' | (H_0 - z)\phi \rangle = \sum_{i=1}^{N} \lambda_i \phi(a_i),$$

for all $\phi \in D(H_0)$, then

$$\langle \psi - \psi' | (H_0 - z)\phi \rangle = 0,$$

for all $\phi \in D(H_0)$. Since $\mathrm{Ran}(H_0 - z) = \mathscr{H}$, it follows that $\psi' = \psi$.

By the argument given in the first part of the proof the function

$$\psi(x) = \sum_{i=1}^{N} \lambda_i G(x, a_i; \bar{z}),$$

satisfies (2.3), and thus $\psi \in \mathscr{K}'$.

III. APPLICATIONS

In this section Lemma 2.1 is used to associate a family

of self-adjoint operators with

$$H_0 + v\delta(x-a), \tag{3.1}$$

for several choices of H_0. The perturbation of H_0 by a sum of delta-functions can be treated by an analogous argument or by repeated applications of Lemma 2.1 as discussed in Sec. V.

Denote the deficiency subspaces of $T = H_0 \upharpoonright \tilde{\mathscr{D}}$ by $\mathscr{K}_{\pm} = \mathrm{Ran}\,(T \pm i)^{\perp}$, where $\tilde{\mathscr{D}}$ is defined by (1.3) with $N = 1$ and $a_1 = a$. Under the hypothesis of Lemma 2.1 we have

$$\mathscr{K}_{\pm} = \{cG(x, a; \pm i), c \in C\},$$

where $G(x, y, z)$ denotes the Green's function corresponding to $(H_0 - z)^{-1}$. For each $\theta \in [0, 2\pi)$ define $\mathscr{D}_\theta$ by

$$\mathscr{D}_\theta = \{\psi(x) = \phi(x) + cG(x, a; i) + c\exp(i\theta)G(x, a; -i) | \phi(x) \in \tilde{\mathscr{D}}, c \in C\}.$$

The self-adjoint extensions of T denoted in the following by H_θ, $\theta \in [0, 2\pi)$, are given by (Theorem X.2, Ref. 12)

$$(H_\theta\psi)(x) = (H_0\phi)(x) + icG(x, a; i) - ic\exp(i\theta) \times G(x, a; -i), \tag{3.2}$$

where $\psi \in \mathscr{D}_\theta$. Thus under the hypothesis of Lemma 2.1 it is possible to associate an infinite family H_θ of self-adjoint operators with (3.1).

We now consider an example of the above construction. The formal Hamiltonian is given by

$$-\Delta + V(x) + v\delta(x), \tag{3.3}$$

where $x \in R^n$, $n = 1$ or 3. If $V(x)$ is $-\Delta$ -bounded with relative bound less than one then $H_0 = -\Delta + V(x)$ is self-adjoint with domain $D(-\Delta)$. In addition, each vector in $D(-\Delta)$ is a continuous $\mathscr{L}^\infty$-function[18,12] and $T = (-\Delta + V(x)) \upharpoonright \tilde{\mathscr{D}}$ is a closed symmetric operator. Furthermore, for a general class of potentials the Green's function corresponding to $-\Delta + V(x)$ exists and satisfies $\mathscr{A}$.[19-20] Thus for a general class of potentials a family H_θ, $\theta \in [0, 2\pi)$, of self-adjoint operators can be associated with (3.3).

Remarks: (i) When $n \geqslant 4$, $-\Delta$ is essentially self-adjoint on $C_0^\infty(R^n \setminus \{0\})$.[12] Thus it does not seem possible to define a delta-function potential for $n \geqslant 4$.

(ii) In the case $V(x) = 0$ the functions $G(x, 0; z)$ are given by

$$G(x, 0; z) = \begin{cases} -(2i\sqrt{z})^{-1}\exp[i\sqrt{z}|x|] & \text{if } n = 1, \\ (4\pi|x|)^{-1}\exp[i\sqrt{z}|x|] & \text{if } n = 3, \end{cases} \tag{3.4}$$

where $\mathrm{Im}\sqrt{z} > 0$. In this particular case we will use the symbol T_θ to denote the family of self-adjoint operators defined by (3.2) with $G(x, 0; z)$ given by (3.4)

(iii) A particularly interesting case is the formal Hamiltonian

$$-\Delta + \frac{\lambda}{|x|} + v\delta(x), \tag{3.5}$$

where $\lambda \in R^1$, $n = 3$. In this case the Green's function corresponding to $-\Delta + \lambda/|x|$ is known.[21] The functions $G(x, 0; z)$ are given by

$$G(x, 0; z) = (4\pi|x|)^{-1}\Gamma(1 + i\mu)W_{-i\mu;1/2}(-2i\sqrt{z}|x|), \tag{3.6}$$

where $W_{-i\mu;\gamma}(\omega)$ denotes the Whittaker function[22] and $\mu = \lambda/2\sqrt{z}$, $\mathrm{Im}\sqrt{z} > 0$. Thus a family of self-adjoint operators given by (3.2) with $G(x, 0; z)$ given by (3.6) can be associated with (3.5).

IV. THE GREEN'S FUNCTION

In this section we derive expressions for the Green's function and wavefunction corresponding to H_θ, $\theta \in [0, 2\pi)$, for a general class of self-adjoint operators H_0. The case $H_0 = -\Delta + \lambda/|x|$ is discussed in detail.

Theorem 4.1: Assume the hypothesis of Lemma 2.1 is satisfied and let H_θ, $0 \leqslant \theta < 2\pi$, be a self-adjoint extension of $T = H_0 \upharpoonright \tilde{\mathscr{D}}$. Then the Green's function $G_\theta(x, y; z)$ corresponding to $(H_\theta - z)^{-1}$ exists and is given by

$$G_\theta(x, y; z) = G(x, y; z) + \lambda(z, \theta)G(x, a; z)G(y, a; z), \tag{4.1}$$

where $\mathrm{Im}\sqrt{z} > 0$ and

$$\lambda(z, \theta) = (1 + \exp(i\theta))\Big[(i - z)\int dy\, G(y, a; z)G(y, a, i) - \exp(i\theta)(i + z)\int dy\, G(y, a; z)G(y, a; -i)\Big]^{-1}. \tag{4.2}$$

Proof: Let $\lambda(z, \theta)$ be a function of z and θ and define (see Sec. V.6 Ref. 7)

$$((H_\theta - z)^{-1}\psi)(x) = \int dy\, G(x, y; z)\psi(y) + \lambda(z, \theta)G(x, a; z) \times \int dy\, G(y, a; z)\psi(y), \tag{4.3}$$

for each $\psi \in \mathscr{H}$. In order to determine $\lambda(z, \theta)$ and, at the same time, verify that the operator (4.3) is the inverse of $H_\theta - z$ we must show for each $\psi \in D(H_0)$

$$((H_\theta - z)^{-1}(H_\theta - z)\psi)(x) = \psi(x). \tag{4.4}$$

It is straightforward to see that (4.4) is valid if $\lambda(z, \theta)$ is given by (4.2).

If $H_0 = -\Delta + V(x)$ with $V(x)$ a short-range potential then the Green's function (4.1) can be used to calculate the wavefunction, denoted $\phi_\theta(x, p)$, corresponding to H_θ. The usual prescription yields

$$\phi_\theta(x, p) = \phi(x, p) + \lambda^{(+)}(p^2; \theta)\phi(a, p)G^{(+)}(x, a; p^2), \tag{4.5}$$

where $\phi(x, p)$ is the wavefunction corresponding to H_0 and is given by

$$\phi(x, p) = \lim_{\epsilon \to +0}(-i\epsilon)(2\pi)^{-n/2} \times \int dy\, \exp(ipy)G(x, y; p^2 + i\epsilon). \tag{4.6}$$

In the case of long-range potentials the definition (4.6) is not valid. For example, if $H_0 = -\Delta + \lambda/|x|$, $x \in R^3$, the prescription (4.6) must be modified.[23] It is not hard to see however, that the modified prescription together with (4.1) also yields (4.5).

The Green's function given in part (a) of the following corollary has been applied to the study of spectral properties of singular perturbations.[24] The Green's function given in part (b) was first derived in Ref. 6. (see Ref. 7 for a detailed account.)

Corollary 4.2: (a) ($n = 1$) The resolvent $(T_\theta - z)^{-1}$ $0 \leqslant \theta < 2\pi$, is an integral operator for

$$z \notin \sigma(T_\theta) = \begin{cases} [0, \infty) \cup \{-\alpha^2/4\}, & \alpha < 0 \\ [0, \infty), & \alpha > 0 \end{cases},$$

$$\alpha = \frac{-2\cos(\theta/2)}{\cos((\theta/2) - (\pi/4))},$$

that is, for each $\psi \in \mathscr{L}^2(R^1)$,

$$((T_\theta - z)^{-1}\psi)(x) = \int dy\, G_\theta(x, y; z)\psi(y),$$

where

$$G_\theta(x, y; z) = -\frac{\exp[i\sqrt{z}|x - y|]}{2i\sqrt{z}} - \frac{i\alpha}{2\sqrt{z}(\alpha - 2i\sqrt{z})}\exp[i\sqrt{z}(|x| + |y|)], \tag{4.7}$$

with $\mathrm{Im}\sqrt{z} > 0$.

(b) ($n = 3$) The resolvent $(T_\theta - z)^{-1}$ is an integral operator for

$$z \notin \sigma(T_\theta) = \begin{cases} [0, \infty) \cup \{-\beta^2\}, & \beta > 0 \\ [0, \infty), & \beta < 0 \end{cases},$$

$$\beta = -\frac{\sin(\theta/2 - \pi/4)}{\cos(\theta/2)},$$

that is, for each $\psi \in \mathscr{L}^2(R^3)$

$$((T_\theta - z)^{-1}\psi)(x) = \int dy\, G_\theta(x, y; z)\psi(y),$$

where

$$G_\theta(x, y; z) = \frac{\exp[i\sqrt{z}|x - y|]}{4\pi|x - y|} - \frac{1}{4\pi(i\sqrt{z} + \beta)}\frac{\exp[i\sqrt{z}(|x| + |y|)]}{|x||y|}, \tag{4.8}$$

with $\mathrm{Im}\sqrt{z} > 0$.

Let H_θ^c denote the family of self-adjoint operators corresponding to (3.5). In this case (4.2), denoted by $\lambda_c(z, \theta)$, can be calculated and is given by

$$\lambda_c(z, \theta) = \left\{ -\frac{\lambda}{4\pi}\left[\psi\left(1 + \frac{i\lambda}{2\sqrt{z}}\right) + \ln\sqrt{z}\right] + \frac{\lambda i}{16} \times [2 + i\tan(\theta/2)] + \frac{\lambda}{4\pi}(1 + \exp(i\theta))^{-1} \times \left[\psi\left(1 + \frac{\sqrt{i}\lambda}{2}\right) + \exp(i\theta)\psi\left(1 + \frac{\lambda}{2\sqrt{i}}\right)\right] - (4\pi)^{-1}(i\sqrt{z} + \beta)\right\}^{-1}, \tag{4.9}$$

where $\psi(\omega)$ denotes Euler's psi function.[25]

Corollary 4.3: The resolvent $(H_\theta^c - z)^{-1}$, $\theta \in [0, 2\pi)$, is an integral operator for $\mathrm{Im}\sqrt{z} > 0$, that is, for each $\psi \in \mathscr{L}^2(R^3)$

$$((H_\theta^c - z)^{-1}\psi)(x) = \int dy\, G_\theta(x, y; z)\psi(y),$$

where $G_\theta(x, y; z)$ is defined by (4.1) with $\lambda(z, \theta)$ given by (4.9).

According to (4.5) the wavefunction coresponding to H_θ^c, denoted $\phi_\theta^c(x, p)$, is given by

$$\phi_\theta^c(x, p) = \phi_c(x, p) + \lambda_c^{(+)}(p^2, \theta)\phi_c(0, p)G_c^{(+)}(x, 0; p^2),$$

where $\phi_c(x, p)$ denotes the Coulomb wavefunction and $G_c(x, y; z)$ the Coulomb Green's function.

V. THE PROBLEM OF *N* DELTA-FUNCTIONS

In this section we outline a prescription for associating a family denoted $H(\theta_1,...,\theta_N)$, $\theta_i \in [0, 2\pi)$, $1 \leqslant i \leqslant N$, of self-adjoint operators with the formal expression

$$H_0 + \sum_{j=1}^{N} \nu_j \delta(x - a_j). \tag{5.1}$$

A procedure for obtaining the Green's function corresponding to $(H(\theta_1,...,\theta_N) - z)^{-1}$ is given.

In Sec. III a family of self-adjoint operators, denoted in this section by $H(\theta_1)$, $\theta_1 \in [0, 2\pi)$ was associated with

$$H_0 + \nu_1\delta(x - a_1),$$

for a general class of self-adjoint operators H_0. By Theorem 4.1 the Green's function corresponding to $(H(\theta_1) - z)^{-1}$ can be determined and will be denoted by $G_{\theta_1}(x, y; z)$.

For each fixed $\theta_1 \in [0, 2\pi)$ let

$$\mathscr{D}(\theta_1) = \{\psi(x) \in D(H(\theta_1)) | \psi(a_2) = 0\},$$

and assume the hypothesis of Lemma 2.1 is satisfied (with $H_0 = H(\theta_1)$ and $\mathscr{D} = \mathscr{D}(\theta_1)$). Then for each $\theta_2 \in [0, 2\pi)$ there is a self-adjoint operator $H(\theta_1, \theta_2)$ defined by

$$(H(\theta_1, \theta_2)\psi)(x) = (H(\theta_1)\phi)(x) + icG_{\theta_1}(x, a_2; i) - ic\exp(i\theta_2)G_{\theta_1}(x, a_2; -i),$$

for each $\psi \in D(H(\theta_1, \theta_2))$ where

$$D(H(\theta_1, \theta_2)) = \{\psi(x) = \phi(x) + cG_{\theta_1}(x, a_2; i) + c\exp(i\theta_2)G_{\theta_1}(x, a_2; -i) | \phi \in \mathscr{D}(\theta_1) \text{ and } c \in C\}.$$

By Theorem 4.1 the Green's function corresponding to $(H(\theta_1, \theta_2) - z)^{-1}$ is given by

$$G_{\theta_1,\theta_2}(x, y; z) = G_{\theta_1}(x, y; z) + \lambda(z, \theta_2)G_{\theta_1}(x, a_2; z)G_{\theta_1}(y, a_2; z).$$

It should be clear that by repeating this construction $N - 2$ times we can associate a family $H(\theta_1,...,\theta_N)$, $\theta_i \in [0, 2\pi)$. $1 \leqslant i \leqslant N$, of self-adjoint operators with (5.1) and at the same time construct the Green's function corresponding to $(H(\theta_1,...,\theta_N) - z)^{-1}$ for a general class of self-adjoint operators H_0.

VI. SELF-ADJOINTNESS OF $T_\theta + V(x)$

In Sec. III we associated a family of self-adjoint operators, denoted H_θ, with (3.3) by treating $\nu\delta(x)$ as a perturbation of $-\Delta + V(x)$. An alternative approach is to first associate a family T_θ of self-adjoint operators with $-\Delta + \nu\delta(x)$, as in remark (ii) of Sec. III, and then treat $V(x)$ as a perturbation of T_θ. In this section we show that the

second approach leads to an infinite family of self-adjoint operators $K_\theta = T_\theta + V(x)$, $\theta \in [0, 2\pi)$. Furthermore for a general class of potentials we show that both approaches are consistent, that is, they lead to the same family of self-adjoint operators.

Theorem 6.1: (a) ($n = 1$) Let $V = V_1(x) + V_2(x)$ with $V_1 \in \mathscr{L}^2(R^1)$ and $V_2 \in \mathscr{L}^\infty(R^1)$, then $T_\theta + V(x)$ is self-adjoint on $D(T_\theta)$ for each $\theta \in [0, 2\pi)$.

(b) ($n = 3$) Let $V = V_1(x) + V_2(x)$ with $V_1 \in \mathscr{L}^2(R^3; dx) \cap \mathscr{L}^2(R^3; |x|^{-2}dx)$ and $V_2 \in \mathscr{L}^\infty(R^3)$, then $T_\theta + V(x)$ is self-adjoint on $D(T_\theta)$ for each $\theta \in [0, 2\pi)$.

Proof: By Corollary 4.2 there exists a $\gamma_\theta < 0$ such that for all $-a^2 < \gamma_\theta$, $(T_\theta + a^2)^{-1}$ is an integral operator.

For $\psi \in D(T_\theta)$

$$\|V\psi\| \leqslant \|V(T_\theta + a^2)^{-1}\| [\|T_\theta \psi\| + a^2 \|\psi\|] .$$

It is easy to see that for large $|a|$, $\|V(T_\theta + a^2)^{-1}\| < 1$, and thus by the Kato–Rellich theorem $T_\theta + V(x)$ is self-adjoint on $D(T_\theta)$.

For potentials $\tilde{V}(x) = \gamma |x|^{-\alpha}$, $\gamma \in R^1$, $x \in R^3$, the above theorem requires $0 \leqslant \alpha < \frac{1}{2}$ in order that $T_\theta + \tilde{V}(x)$ be self-adjoint. In the following theorem we treat $V(x)$ as a form perturbation of T_θ. This allows us to define $T_\theta + V(x)$ as a self-adjoint operator for $V = V_1(x) + V_2(x)$, $V_1 \in R \cap \mathscr{L}^1(R^3; |x|^{-2}dx)$, $V_2 \in \mathscr{L}^\infty(R^3)$ where R denotes the class of Rollnick potentials.[19] In particular, $T_\theta + \tilde{V}(x)$ can be defined as a self-adjoint operator if $0 \leqslant \alpha < 1$.

Theorem 6.2: ($n = 3$) Let $V = V_1(x) + V_2(x)$ with $V_1 \in R \cap \mathscr{L}^1(R^3; |x|^{-2}dx)$ and $V_2 \in \mathscr{L}^\infty(R^3)$. Then there exists a unique self-adjoint operator K_θ for each $\theta \in [0, 2\pi)$ with $Q(K_\theta) = Q(T_\theta)$ and

$$\langle \varphi, K_\theta \psi \rangle = \langle \varphi, T_\theta \psi \rangle + \langle \varphi, V\psi \rangle ,$$

for each $\varphi, \psi \in Q(T_\theta)$.

Remark: The above theorem can be verified by showing that V is relatively form-bounded with respect to T_θ, with relative bound less than 1 (see Theorem 3.2, Ref. 10.) This can be verified by a similar argument as used to show relative form-boundedness of V with respect to $-\Delta$.[19]

The relationship between the operators H_θ and $T_\theta + V(x)$ is given in the following proposition.

Proposition 6.3: Assume $H_0 = -\Delta + V_1(x)$ where $V_1(x)$ satisfies the hypothesis of Theorem 6.1. Then

$$\{T_\theta + V_1(x) | \theta \in [\mathrm{L}0, 2\pi])\} = \{H_\theta | \theta \in [0, 2\pi)\} . \tag{6.1}$$

Proof: For each $\theta \in [0, 2\pi)$ $T_\theta + V_1(x)$ is a self-adjoint extension of $T = H_0 \upharpoonright \tilde{\mathscr{D}}$. Since every self-adjoint extension of T is a member of $\{H_\theta | \theta \in [0, 2\pi)\}$ we have $T_\theta + V_1(x) = H_{\theta'}$ for some $0 \leqslant \theta' < 2\pi$.

Let H_θ, $\theta \in [0, 2\pi)$, be a self-adjoint extension of T. Using Theorem (4.1) it is straightforward to show that V_1 is H_θ-bounded with relative bound less than 1. Thus H_θ-bounded with relative bound less than 1. Thus $H_\theta - V_1(x)$ is self-adjoint on $D(H_\theta)$. Furthermore $(H_\theta - V_1(x)) \upharpoonright \tilde{\mathscr{D}} = (-\Delta) \upharpoonright \tilde{\mathscr{D}}$ which implies $H_\theta - V_1(x) = T_{\theta'}$ for some $\theta' \in [0, 2\pi)$. Thus $H_\theta = T_{\theta'} + V_1(x)$ and (6.1) is valid.

VII. A QUADRATIC FORM APPROACH

In this section the quadratic form approach for associating a self-adjoint operator with $-\Delta + \nu\delta(x)$ is examined.

Theorem 7.1: (a) ($n = 1$) For each $\nu \in R^1$ there is a unique self-adjoint operator denoted H_ν corresponding to the quadratic form (1.4). If $\nu = \alpha$ (defined in Corollary 4.2) then $H_\nu = T_\theta$.

(b) ($n \geqslant 3$) There is no self-adjoint semibounded operator H_ν, $\nu > 0$, such that $q(\phi, \psi) = (\phi, H_\nu \psi)$ for all $\phi, \psi \in S(R^n)$.

Proof: (a) The existence of H_ν has been shown.[12-14]

An elementary calculation shows that each $\psi \in D(T_\theta)$ is contained in

$$D(H_\nu) = \left\{ \psi \in Q(p^2) \,|\, p^2\psi(p) + (2\pi)^{-1}\nu \int dq\, \psi(q) \in \mathscr{L}^2(R^1) \right\}$$

and $H_\nu \psi = T_\theta \psi$ for $\nu = \alpha$. Thus $H_\nu \supseteq T_\theta$, which implies $H_\nu = T_\theta$.

(b) If H_ν is a self-adjoint semibounded operator with an associated quadratic form $q(\phi, \psi)$ then $q(\phi, \psi)$ is closed.[11]

For $n \geqslant 3$ the form $q(\phi, \psi)$, given by (1.4) with $H_0 = -\Delta$, $N = 1$, and form domain $S(R^n)$, is not closable. In particular, the sequence $\phi_N(\mathbf{x}) = \exp[-(1/2)N|x|^2]$ is such that, $\phi_N \to 0$ in $\mathscr{L}^2(R^n)$ as $N \to \infty$, $q(\phi_N - \phi_M, \phi_N - \phi_M) \to 0$ as $N, M \to \infty$ but $q(\phi_N, \phi_N) \to \nu$ as $N \to \infty$. Thus, there is no semibounded self-adjoint operator H_ν such that $q(\phi, \psi) = (\phi, H_\nu \psi)$ for $\phi, \psi \in S(R^n)$, $n \geqslant 3$.

VIII. STRONG RESOLVENT CONVERGENCE OF H^N

In this section we examine the strong resolvent convergence of the family of operators H^N defined by

$$(H^N\hat{\psi})(p) = ((-\Delta + V(x))\psi)\hat{}\,(p) + (2\pi)^{-n}\nu\chi_N(p) \times \int dq\, \chi_N(q)\hat{\psi}(q) . \tag{8.1}$$

Theorem 8.1: (a) ($n = 1$) Let $V = V_1(x) + V_2(x)$ with $V_1 \in \mathscr{L}^2(R^1)$ and $V_2 \in \mathscr{L}^\infty(R^1)$. Then H^N converges to K_θ in the strong resolvent sense with $\nu = \alpha$.

(b) ($n = 3$) Let $V = V_1(x) + V_2(x)$ with $V_1 \in \mathscr{L}^2(R^3) \cap \mathscr{L}^2(R^3; |x|^{-2}dx)$ and $V_2 \in \mathscr{L}^\infty(R^3)$. For each fixed θ, $0 \leqslant \theta < 2\pi$, let ν appearing in (8.1) have the following N and θ dependence

$$\nu = -(2\pi)^3(4\pi N - 2\pi^2\beta + \gamma)^{-1} , \tag{8.2}$$

$$\gamma = (1 + \exp(i\theta))^{-1} \left\{ \int_{N \leqslant |q| \leqslant N+1} dq\, \chi_N^2(q)(q^2 - i)^{-1} + \exp(i\theta) \int_{N \leqslant |q| \leqslant N+1} dq\, \chi_N^2(q)(q^2 + i)^{-1} \right\} .$$

Then the momentum cutoff Hamiltonians H^N converge to K_θ in the strong resolvent sense.

(c) ($n \geqslant 4$) For each fixed $\nu \in R^1$, $H_0^N = H^N - V(x)$ converges to $-\Delta$ in the strong resolvent sense.

By Proposition 8.2 we must show that for each $\psi \in D(K_\theta)$ there exists $\{\psi_N\}$ with $\psi_N \in D(H^N)$ such that s-$\lim_{N\to\infty} \psi_N = \psi$ and s-$\lim_{N\to\infty} H^N\psi_N = K_\theta\psi$, where ν is related to N and θ by (8.2).

For each $\psi \in D(T_\theta)$ define ψ_N by

$$\hat{\psi}_N(p) = \hat{\phi}(p) + c_1\chi_N(p)(p^2 - i)^{-1} + c_1\exp(i\theta)$$

$$\times\chi_N(p)(p^2+i)^{-1},$$

where $\phi\in\tilde{\mathscr{D}}$. Clearly, $\psi_N\in D(H^N)$ and s-$\lim_{N\to\infty}\psi_N=\psi$.

In order to show $H^N\psi_N\to K_\theta\psi$, as $N\to\infty$ we consider

$$\begin{aligned}\|H^N\psi_N-K_\theta\psi\|&\leqslant(2\pi)^{-3}|\nu|\,\|\chi_N\int dq\,\chi_N(q)\hat{\phi}(q)\|\\&+|c_1|\,\|\chi_N\left\{(1+\exp(i\theta))+(2\pi)^{-3}\nu\int dq\,\chi_N^2(q)(q^2-i)^{-1}\right.\\&\left.+(2\pi)^{-3}\nu\exp(i\theta)\int dq\,\chi_N^2(q)(q^2+i)^{-1}\right\}\|\\&+\|V(\psi_N-\psi)\|+2|c_1|\left\{\int dp\,(1-\chi_N(p))^2(p^4+1)^{-1}\right\}^{1/2}.\end{aligned}\tag{8.3}$$

It is not difficult to show that the last two terms above converge to zero as $N\to\infty$.

The first term on the right side of the inequality (8.3) can be bounded as follows:

$$\begin{aligned}&(2\pi)^{-3}|\nu|\,\|\chi_N\int dq\,\chi_N(q)\hat{\phi}(q)\|\\&\leqslant|\nu|CN^{3/2}\left|\int(\chi_N(q)-1)\hat{\phi}(q)\,dq\right|\\&\leqslant|\nu|CN^{3/2}\int\frac{|\chi_N(q)-1|^{1/2}}{q^2+1}\,|\chi_N(q)-1|^{1/2}\\&\quad\times(q^2+1)|\hat{\phi}(q)|\,dq\\&\leqslant|\nu|C'N\left\{\int|\chi_N(q)-1|(q^2+1)^2|\,\hat{\phi}(q)|^2\,dq\right\}^{1/2},\end{aligned}$$

where C and C' are constants. Due to the N dependence of ν the above goes to zero in the limit $N\to\infty$.

The second term on the right side of the inequality (8.3) is bounded by

$$\begin{aligned}&|c_1|\,\|\chi_N\left\{(1+\exp(i\theta))+(2\pi)^{-3}\nu\left[4\pi\left(N+\frac{\pi i\sqrt{i}}{2}\right)\right.\right.\\&\quad\left.+\int_{N\leqslant|q|\leqslant N+1}dq\,\chi_N^2(q)(q^2-i)^{-1}\right]\\&\quad+(2\pi)^{-3}\nu\exp(i\theta)\left[4\pi\left(N-\frac{\pi\sqrt{i}}{2}\right)\right.\\&\quad\left.\left.+\int_{N\leqslant|q|\leqslant N+1}dq\,\chi_N^2(q)(q^2+i)^{-1}\right]\right\}\|\\&\quad+\pi^{-2}|c_1|\,|\nu|\,\|\chi_N\|\left|\int_N^\infty dp\,(p^2-i)^{-1}\right|.\end{aligned}$$

The first term above is zero due to the choice of ν and the second term goes to zero in the limit $N\to\infty$. Thus H^N converges to K_θ in the strong resolvent sense.

It is straightforward to show that $H_0^N\phi\to-\Delta\phi$, as $N\to\infty$, for each $\phi\in C_0^\infty(R^n\setminus\{0\})$, $n\geqslant4$. This verifies part (c) since $C_0^\infty(R^n\setminus\{0\})$ is a common core for H_0^N and $-\Delta$ for $n\geqslant4$.[12]

Proposition 8.2: Suppose that $\{A_n\}$ and A are self-adjoint operators. If for each $\psi\in D(A)$ there exist $\{\psi_n\}$, $\psi_n\in D(A_n)$, such that s-$\lim_{n\to\infty}\psi_n=\psi$ and s-$\lim_{n\to\infty}A_n\psi_n=A\psi$, then A_n converges to A in the strong resolvent sense.

Proof: Denote the graph of A by $\Gamma(A)$ and let Γ denote set of pairs $\langle\phi,\psi\rangle\in\mathscr{H}\times\mathscr{H}$ in the strong graph limit[11] of A_n

Since for each $\psi\in D(A)$ there exist $\{\psi_n\}$, $\psi_n\in D(A_n)$, $\psi_n\to\psi$ and $A_n\psi_n\to A\psi$ it follows $\Gamma(A)\subseteq\Gamma$. Furthermore (Theorem VIII.27, Ref. 11), Γ is the graph of a closed symmetric operator $\hat{A}$. Since $\hat{A}\supseteq A$ it follows that A is the strong graph limit of A_n. For self-adjoint operators strong graph convergence is equivalent to strong resolvent convergence which concludes the proof of the proposition.

IX. TIME-DEPENDENT SCATTERING FOR ZERO-RANGE POTENTIALS

In this section the usual time-dependent scattering formalism, which is based on the wave operator concept, is extended to Hamiltonians involving zero-range potentials. The relationship between the time-dependent scattering theories corresponding to H^N and K_θ is established.

The following theorem can be verified by a similar argument as used for the case of singular potentials.[26]

Theorem 9.1: ($n=1,3$) Let $V(x)$ be such that $K_\theta=T_\theta+V(x)$ is self adjoint, $0\leqslant\theta<2\pi$, with $K_\theta\restriction\tilde{\mathscr{D}}\cap S(R^n)=(-\Delta+V(x))\restriction\tilde{\mathscr{D}}\cap S(R^n)$. Furthermore assume that for some $\epsilon>0$ and $R>0$

$$\int_{|x|\geqslant R}|V(x)|^2(1+|x|)^{-(n-2)+\epsilon}\,dx<\infty.\tag{9.1}$$

Then the wave operators $W_\pm(K_\theta,-\Delta)$ defined by

$$W_\pm(K_\theta,-\Delta)=\operatorname*{s-lim}_{t\to\pm\infty}\exp(iK_\theta t)\exp(i\Delta t)$$

exist.

Remarks: (i) The above theorem is also valid with K_θ replaced by H_θ.

(ii) It is straightforward to extend the above results to include long-range potentials $V(x)$.

The results of Sec. VIII will now be used to verify the convergence of the wave operators $W_\pm(H^N,-\Delta)$ to the wave operators $W_\pm(K_\theta,-\Delta)$ in the limit $N\to\infty$.

Theorem 9.2: (a) ($n=1$) Let $V(x)\in\mathscr{L}^2(R^1)$ satisfy (9.1) and for each fixed θ, $0\leqslant\theta<2\pi$, let $\nu\in R^1$ satisfy $\nu=\alpha$. Then

$$\operatorname*{s-lim}_{N\to\infty}W_\pm(H^N,-\Delta)=W_\pm(K_\theta,-\Delta).$$

(b) ($n=3$) Let $V(x)=V_1(x)+V_2(x)$ satisfy (9.1), $V_1\in\mathscr{L}^2(R^3)\cap\mathscr{L}^2(R^3;|x|^{-2}dx)$, and $V_2\in\mathscr{L}^\infty(R^3)$. Furthermore, for each fixed θ, $0\leqslant\theta<2\pi$, let $\nu=-(2\pi)^3(4\pi N-2\pi^2\beta+\gamma)^{-1}$. Then

$$\operatorname*{s-lim}_{N\to\infty}W_\pm(H^N,-\Delta)=W_\pm(K_\theta,-\Delta).$$

Proof: Consider the inequality

$$\begin{aligned}&\|\{W_\pm(H^N,-\Delta)-W_\pm(K_\theta,-\Delta)\}\psi\|\\&\quad\leqslant\|[\exp(iH^Nt)-\exp(iK_\theta t)]\exp(i\Delta t)\psi\|\\&\quad+\|[W_\pm(H^N,-\Delta)-\exp(iH^Nt)\exp(i\Delta t)]\psi\|\\&\quad+\|[\exp(iK_\theta t)\exp(i\Delta t)-W_\pm(K_\theta,-\Delta)]\psi\|.\end{aligned}\tag{9.2}$$

If

$$W_{\pm}(H^N, -\Delta) = \underset{t\to\pm\infty}{\text{s-lim}} \exp(iH^N t)\exp(i\Delta t), \tag{9.3}$$

uniformly in N, $N>1$, then for sufficiently large $|t|$ the last two terms in (9.2) can be made arbitrarily small independent of N. By Theorem 8.1, H^N converges to K_θ in the strong resolvent sense and thus the first term on the right side of the inequality (9.2) goes to zero for each fixed $t \in R^1$ as $N\to\infty$. Thus, in order to verify the theorem we must show that (9.3) is valid independent of N, $N>1$.

Let $\chi_R(x)$ be a C^∞-function which satisfies

$$\chi_R(x) = \begin{cases} 0 \text{ if } |x| < R-\eta, \\ \qquad\qquad 0<\eta<R. \\ 1 \text{ if } |x| \geqslant R, \end{cases}$$

In order to verify (9.3) it is enough to show

$$W_{\pm}(H^N, -\Delta) = \underset{t\to\pm\infty}{\text{s-lim}} \exp(iH^N t)\chi_R \exp(i\Delta t),$$

uniformly in $N>1$. By a standard argument this will be true if for each $\psi \in \mathscr{K}$, with $\mathscr{K}$ dense in $\mathscr{L}^2(R^n)$ (The case $t<0$ is verified by an analogous argument.),

$$\lim_{t\to\infty} \int_t^\infty du\, \{I_1(u) + I_2(u) + I_3(u)\} = 0, \tag{9.4}$$

uniformly in $N>1$ where

$$I_1(u) = \|(-\Delta\chi_R + \chi_R\Delta)\exp(i\Delta u)\psi\|,$$
$$I_2(u) = \|V\chi_R \exp(i\Delta u)\psi\|,$$
$$I_3(u) = \|(H_0^N + \Delta)\chi_R \exp(i\Delta u)\psi\|.$$

Let $\mathscr{K}$ denote the set of all functions $\psi(x)$ such that their Fourier transforms are given by $\hat{\psi}(p) = (\Pi_{i=1}^n p_i) \times \exp[-p^2 - ip\cdot a]$, $a \in R^n$. This set of functions in dense in $\mathscr{L}^2(R^n)$ and satisfies[27]

$$(\exp(i\Delta u)\psi)(x) = \frac{C'}{(1+iu)^{3n/2}} \times \left\{\prod_{i=1}^n (x_i - a_i)\right\} \exp\left[-\frac{(x-a)^2}{4(1+iu)}\right], \tag{9.5}$$

where C' is a constant.

Using (9.5) it is straightforward to see that the u-integrals over $I_1(u)$ and $I_2(u)$ converge to zero uniformly in N for each $\psi \in \mathscr{K}$.

Using (9.5) for $n=1$ and performing several integration by parts yields

$$\begin{aligned} &\|(H_0^N + \Delta)\chi_R(\exp(i\Delta u)\psi)\| \\ &\leqslant \frac{\tilde{C}|\nu|N^{1/2}}{|1+iu|^{3/2}} \left| \int_{-\infty}^{+\infty} dy \left(\int_{-\infty}^{+\infty} dq \exp(iqy) \frac{\chi_N^{(4)}(q)}{q} \right) \right. \\ &\left. \times \frac{d}{dy}\left\{ y^{-4}\chi_R(y)(y-a)\exp\left[-\frac{(y-a)^2}{4(1+iu)}\right]\right\}\right| \\ &\leqslant \frac{CN^{1/2}}{|1+iu|^{3/2}} \log(1+N^{-1}), \end{aligned}$$

where C is a constant independent of $N>1$. Thus for each $\psi \in \mathscr{K}$ the u-integral of $I_3(u)$ converges to zero uniformly in $N>1$.

An argument along the lines of the above also verifies (9.4) for the case $n=3$.

Remark: Let $S^N = W_+^*(H^N, -\Delta)\, W_-(H^N, -\Delta)$ and $S^\theta = W_+^*(K_\theta, -\Delta)\, W_-(K_\theta, -\Delta)$. Under the hypothesis of Theorem 9.2, w-$\lim_{N\to\infty} S^N = S^\theta$.

X. CONCLUDING REMARKS

There does not seem to exist a rigorous argument showing the validity or lack of validity of the usual integral equations when Dirac delta-function potentials are present. For delta-shall potentials[8,9] and zero-range potentials in one dimension[28,4] these equations can be solved explicitly. However in the three-dimensional case the formal replacement of the potential appearing in these equations by a Dirac delta-function leads to a contradiction.

The results of Sec. VIII and IX of this paper provide a rigorous procedure for calculating the wavefunctions and S-matrix corresponding to $T_\theta + V(x)$, $\theta \in [0, 2\pi)$. One first calculates the wavefunctions and S-matrix for the momentum cutoff Hamiltonian with ν chosen so that H^N converges to $T_\theta + V(x)$ in the strong resolvent sense. For $N<\infty$ the usual integral equations are valid and thus in principle the momentum cutoff wavefunctions and S-matrix can be calculated. Since the wave operators corresponding to H^N converge to the wave operators corresponding to $T_\theta + V(x)$ as $N\to\infty$, it follows that the momentum cutoff wavefunctions and S-matrix converge to the corresponding wavefunctions and S-matrix of $T_\theta + V(x)$ in the sense of distributions.

In this paper we have concentrated on the problem of zero-range potentials in one and three dimensions. It is not hard to extend the results of this paper to include delta-shell potentials[8] and the modified Fermi potential.[9] In particular for the formal Hamiltonian

$$-\frac{d^2}{dx^2} + \frac{l(l+1)}{x^2} + \nu\delta(x-a), \tag{10.1}$$

where $a>0$, $\nu \in R^1$, and $l=0,1,\ldots$, it is possible to introduce momentum cutoff Hamiltonians which for each fixed value of ν (with ν independent of the cutoff) converge in the strong resolvent sense to a member of the family of self adjoint operators associated with (10.1) by Lemma 2.1. Similar results are valid for the formal Hamiltonian

$$-\frac{d^2}{dx^2} + \frac{l(l+1)}{x^2} + \frac{\lambda}{x} + \nu\delta(x-a),$$

which has been proposed as an alternative model for the Fermi contact Hamiltonian.[9]

ACKNOWLEDGMENT

I would like to thank J.A. Brooke for several useful conversations and in particular for suggesting the sequence used in the proof of Theorem 7.1.

[1] J.B. McGuire, J. Math. Phys. **5**, 622 (1964).
[2] E. Brezin and J. Zinn-Justine, C. R. Acad. Sci. Paris B **263**, 670 (1966).
[3] C.N. Yang, Phys. Rev. **168**, 1920 (1968).
[4] L.R. Dodd, J. Math. Phys. **11**, 207 (1970).
[5] Ya. B. Zel'dovich, Sov. Phys. (JETP) **11**, 594 (1960).
[6] F.A. Berezin and L.D. Faddeev, Sov. Math. (Doklady) **2**, 372 (1961).
[7] G. Flamand, "Mathematical Theory of Nonrelativistic Two- and Three-Particle Systems with Point Interactions" in *Applications of Mathematics to Problems in Theoretical Physics*, edited by F. Lurcat (Gordon and Breach, New York, 1967).
[8] A.K. Jain, S.K. Seb, and C.S. Shastry, Am. J. Phys. **46**, 147 (1978).

[9]S.M. Blinder, Phys. Rev. A **18**, 853 (1978).
[10]B. Simon, "An Introduction to the Self-Adjointness and Spectral Analysis of Schrödinger Operators" in The Schrödinger Equation, edited by W. Thirring and P. Urban (Springer–Verlag, Wien/New York, 1977).
[11]M. Reed and B. Simon, *Methods of Modern Mathematical Physics* I: *Functional Analysis* (Academic, New York, London, 1972).
[12]M. Reed and B. Simon, *Methods of Modern Mathematical Physics* II: *Fourier Analysis and Self-Adjointness* (Academic, New York, London, 1975).
[13]W.O. Amrein, J.M. Jauch, and K.B. Sinha, *Scattering Theory in Quantum Mechanics* (Benjamin, Reading, Mass., 1977).
[14]W.G. Faris, *Self-Adjoint Operators* (Springer–Verlag, New York, 1975).
[15]C.N. Friedman, J. Funct. Anal. **10**, 346 (1972).
[16]V.I. Karloukovski, Rep. Math. Phys. **10**, 87 (1976).
[17]S. Albeverio, R. Høegh-Krohn, and L. Streit, J. Math. Phys. **18**, 907 (1977).
[18]T. Kato, *Perturbation Theory for Linear Operators* (Springer–Verlag, New York, 1976).
[19]B. Simon, *Quantum Mechanics for Hamiltonians Defined as Quadratic Forms* (Princeton U.P., Princeton, New Jersey, 1971).
[20]E. Prugovečki, *Quantum Mechanics in Hilbert Space* (Academic, New York, 1971).
[21]L. Hostler, J. Math. Phys. **5**, 591 (1964).
[22]H. Buchholz, *The Confluent Hypergeometric Function* (Springer–Verlag, Berlin, 1969).
[23]J. Zorbas, J. Math. Phys. **18**, 1112 (1977).
[24]F. Zirilli, Boll. U.M.I., **15**-B, 386 (1978).
[25]I.S. Gradshteyn and I.M. Ryzhik, *Tables of Integrals, Series, and Products* (Academic, New York, 1965).
[26]J. Kupsch and W. Sandhas, Commun. Math. Phys. **2**, 147 (1966).
[27]S.T. Kuroda, Nuovo Cimento **12**, 431 (1959).
[28]I.R. Lapidus, Am. J. Phys. **37**, 1064 (1969).

Vertex function and crossing symmetry

W. D. Heiss
National Research Instutute for Mathematical Sciences of the CSIR, P.O. Box 395, Pretoria, South Africa

(Received 7 December 1978; accepted for publication 20 April 1979)

An integral equation is proposed for the vertex function occurring in the quantum-mechanical many-body problem, such that the s channel and the t channel are treated simultaneously on the same level. The equation thus provides an intricate blending of the two channels. Its solution is given in closed form, and although this is complicated, it allows the physically interesting features to be investigated numerically.

1. INTRODUCTION

In the Green's function formulation of the nonrelativistic quantum-mechanical many-body problem the vertex function plays an important role.[1] It is defined as consisting of all connected diagrams which begin and end with a two-body force[2] (we confine ourselves to one-body and two-body interactions). Physically, the vertex function is the scattering amplitude for two particles in an interacting medium.

We denote the vertex function by $\Gamma_{ijkl}(\omega,\omega',\omega)$, where the labels refer to a single particle basis. The energy parameters are three independent combinations of the four off-shell energies associated with the labels,

$$\widetilde{\omega} = \omega_i + \omega_j = \omega_k + \omega_l,$$
$$\omega' = \omega_k - \omega_i = \omega_j - \omega_l,$$
$$\omega = \omega_k .$$

This particular choice of variables is useful for formulating the s channel and t channel integral equations which must be satisfied by Γ. In the references[3,4] these two integral equations have been discussed at length and solutions have been found for specific approximations of the input, i.e., the respective irreducible vertex parts. While a great deal of insight is gained in the quoted papers, there is a certain drawback which originates from the fact that if, say, an s channel equation is solved for Γ, the t channel properties become very unsatisfactory (unless, of course, one could take the correct s channel irreducible part). This disadvantage is exhibited in the properties of the t channel two-particle propagator when it is calculated from a Γ which is the solution of an s channel equation with approximate input. Spurious poles are obtained; in particular the pole terms at the unperturbed energies reappear and the sum rule is not complied with. Likewise, a t channel integral equation furnishes an unsatisfactory s channel two-particle propagator.

It appears that it is too complicated a task to find a suitable input for the s channel (t channel) integral equation such that the t channel (s channel) two-particle propagator exhibits acceptable properties. The theoretical requirement of crossing symmetry which postulates that both the s channel *and* the t channel two-particle propagators should be calculated from a single vertex function seems to be beyond practical reach.

In this paper an attempt is made to implement crossing symmetry in an approximate way. We abandon the concept of a proper s channel or t channel equation. A single integral equation for Γ is suggested, which is neither an s channel nor a t channel equation but such that the two channels have the same footing, i.e., there is no bias towards a specific channel.

Denoting by $\omega_{ijkl}(\widetilde{\omega},\omega',\omega)$ and $\alpha_{ijkl}(\widetilde{\omega},\omega',\omega)$ the s channel and t channel irreducible parts of Γ, respectively, and by $\omega_{ijkl}(\widetilde{\omega},\omega',\omega)$ the contributions irreducible with respect to both channels, we may write down the equation

$$\Gamma = \omega + i\omega\cdot(GG)_s\cdot\Gamma - i\alpha^*(GG)_t{}^*\Gamma$$

in an abbreviated form, where dots and stars indicate s channel and t channel connections, respectively. Summation and integration over the appropriate labels and energy variables is understood. Replacing the single-particle Green functions G by their free counterparts, and the quantities ω, α and ω all by the same static term λ_{ijkl}, we arrive at the equation which is the subject of this paper, viz.,

$$\Gamma_{ijkl}(\widetilde{\omega},\omega',\omega) = \lambda_{ijkl} - \lambda_{ijmn}\int\frac{dz}{2\pi i}\frac{\Gamma_{mnkl}(\widetilde{\omega},z,\omega)}{(\omega - z - \epsilon_m)(\widetilde{\omega} - \omega + z - \epsilon_n)} + \lambda_{mjnl}\int\frac{dz}{2\pi i}\frac{\Gamma_{inkm}(z,\omega',\omega)}{(z - \omega - \epsilon_m)(z - \omega + \omega' - \epsilon_n)}. \tag{1.1}$$

Here and in the sequel the summation convention is used.

The approximations by which Eq. (1.1) is obtained seem to be very drastic. A major deficiency which emerges from this treatment is the lack of time-reversal symmetry. It leads to nonsymmetric two-particle propagators

$$K_{ij,kl}(\widetilde{\omega}) = K^{\text{free}}_{ij,kl} + [(GG)_s\cdot\Gamma\cdot(GG)_s]_{ij,kl}, \tag{1.2a}$$
$$F_{ik,lj}(\omega') = F^{\text{free}}_{ik,lj} - [(GG)_t{}^*\Gamma^*(GG)_t]_{ik,lj}, \tag{1.2b}$$

in the s channel and t channel, respectively. This defect could be overcome by properly symmetrizing Eq. (1.1). However, a very important attribute of Eq. (1.1) which would then be lost is its solvability in closed form. The advantage of having an analytic solution which is easily amenable to numerical investigation is here given higher priority than time-reversal symmetry, since the emphasis in this work lies on a study of the effects of implementing crossing properties within a simplified mathematical model. Note that if the last term in Eq. (1.1) we dropped, this would lead to a vertex function depending on $\widetilde{\omega}$ only, and from Eq. (1.2a) a traditional particle-particle RPA propagator would be obtained while $F(\omega')$

 0022-2488/80/040848-06$1.00

would exhibit spurious properties. Conversely, if the first integral in Eq. (1.1) were dropped, this would lead to a $\Gamma(\omega')$ and via Eq. (1.2b) to a traditional particle-hole RPA propagator, but $K(\tilde{\omega})$ would be spurious. The intricate blending of the two channels becomes obvous if one considers the Neumann series of Eq. (1.1).

The usefulness of Eq. (1.1) hinges to a large extent on the question of whether the propagators K and F have acceptable properties. The results of numerical investigations in which use was made within a schematic model of the expressions derived in this paper were in fact very satisfactory. The sum rules are satisfied to within about $\pm 1\%$ over a wide parameter range; inconsistencies[5] occur only for parameter values which closely approach those at which the stability limits of the usual RPA are reached. Further details are to be published elsewhere.[6]

The following two sections are devoted to the construction of a solution of Eq. (1.1). The final expressions appear in the form of rapidly converging matrix series. In order to facilitate the solution procedure we assume a degenerate particle and hole spectrum. The paper concludes with a discussion of the properties of the solution.

2. SOLUTION OF THE INTEGRAL EQUATION

We aim at a solution of Eq. (1.1) in a finite-dimensional space, i.e., each of the four labels runs over $N = N_h + N_p$ states corresponding to N_h hole states and N_p particle states. In accordance with the degeneracy assumption all particle states of the single-particle basis have the same energy denoted by ϵ, and δ denotes the degenerate hole spectrum. The singularities of the integral kernels of Eq. (1.1) are defined in the usual manner:

$$\frac{1}{z-\epsilon_k} = \frac{p_k}{z-\epsilon+i0} + \frac{h_k}{z-\delta-i0},$$

where here and in the following we denote the projection operators onto the particle space and hole space by p and h, respectively.

The structure of Eq. (1.1) implies that the solution $\Gamma(\tilde{\omega},\omega',\omega)$ cannot be increasing when $\tilde{\omega}$ or ω' approach plus or minus infinity along the real axis. While this is desirable from the physical point of view, it is of importance for our mathematical procedure. As a consequence, the first and the second integral of Eq. (1.1) vanish for $\tilde{\omega}\to\pm\infty$ and $\omega'\to\pm\infty$, respectively. Furthermore, we conclude from Eq. (1.1) that the solution Γ has the structure

$$\Gamma_{ijkl}(\tilde{\omega},\omega',\omega) = \lambda_{ijkl} + \Gamma^s_{ijkl}(\tilde{\omega},\omega) + \Gamma^t_{ijkl}(\omega',\omega). \quad (2.1)$$

Introducing this expression into Eq. (1.1) we obtain

$$\begin{aligned}
&\Gamma^s_{ijkl}(\tilde{\omega},\omega) + \Gamma^t_{ijkl}(\omega',\omega) \\
&= \lambda_{ijmn}\frac{p_m p_n - h_m h_n}{\tilde{\omega}-\epsilon_m-\epsilon_n}(\lambda_{mnkl} + \Gamma^s_{mnkl}(\tilde{\omega},\omega)) \\
&\quad - \lambda_{ijmn}\int\frac{dz}{2\pi i}\frac{\Gamma^t_{mnk}(z,\omega)}{(\omega-z-\epsilon_m)(\tilde{\omega}-\omega+z-\epsilon_n)} \\
&\quad + \lambda_{mjnl}\frac{p_n h_m - p_m h_n}{\omega'-\epsilon_n+\epsilon_m}(\lambda_{inkm} + \Gamma^t_{inkm}(\omega',\omega) \\
&\quad + \lambda_{mjnl}\int\frac{dz}{2\pi i}\frac{\Gamma^s_{inkm}(z,\omega)}{(z-\omega-\epsilon_m)(z-\omega+\omega'-\epsilon_n)}.
\end{aligned}$$

This equation is separable into two parts, one depending only on $\tilde{\omega}$ (the variable ω is a parameter) and the other depending on ω'. Each part must therefore be equal to a constant which must be zero owing to the fact that Γ^s and Γ^t must vanish for large $\tilde{\omega}$ and ω', respectively.

It is convenient to introduce the functions

$$\begin{aligned}
\bar{c}_{ijkl}(u) &= \frac{\Gamma^s_{ijkl}(u,\omega)}{u-\omega-\epsilon_l} \\
&\quad\times\left(\frac{p_j}{u-\epsilon_i-\epsilon_j+i0} + \frac{h_j}{u-\epsilon_i-\epsilon_j-i0}\right), \\
\bar{d}_{ijkl}(u) &= \frac{\Gamma^t_{ijkl}(u,\omega)}{u-\omega+\epsilon_i} \\
&\quad\times\left(\frac{p_j}{u-\epsilon_j+\epsilon_l+i0} + \frac{h_j}{u-\epsilon_j+\epsilon_l-i0}\right).
\end{aligned} \quad (2.2)$$

In terms of these functions we obtain the coupled system

$$\begin{aligned}
&(u-\omega-\epsilon_l)(u-\epsilon_i-\epsilon_j)\bar{c}_{ijkl}(u) \\
&\quad = \lambda_{ijmn}\frac{p_m p_n - h_m h_n}{u-\epsilon_m-\epsilon_n}\lambda_{mnkl} \\
&\qquad + \lambda_{ijmn}(p_m p_n - h_m h_n)(u-\omega-\epsilon_l)\bar{c}_{mnkl}(u) \\
&\qquad + \lambda_{ijmn}\int\frac{dz}{2\pi i}\frac{(z-\epsilon_n+\epsilon_l)\bar{d}_{mnkl}(z)}{z+u-\omega-\epsilon_n}, \\
&(u-\omega+\epsilon_i)(u-\epsilon_j+\epsilon_l)\bar{d}_{ijkl}(u) \\
&\quad = \lambda_{mjnl}\frac{p_n h_m - p_m h_n}{u-\epsilon_n+\epsilon_m}\lambda_{inkm} \\
&\qquad + \lambda_{mjnl}(p_n h_m - p_m h_n)(u-\omega+\epsilon_i)\bar{d}_{inkm}(u) \\
&\qquad + \lambda_{mjnl}\int\frac{dz}{2\pi i}\frac{(z-\epsilon_i-\epsilon_n)\bar{c}_{inkm}(z)}{z+u-\omega-\epsilon_n}.
\end{aligned} \quad (2.3)$$

The structure of Eqs. (2.3) suggests taking the Fourier transform, as the Fourier transform of a simple pole term is well known. From the asymptotic properties of Γ^s and Γ^t and from Eq. (2.2) it is clear that the Fourier transforms

$$\begin{aligned}
c_{ijkl}(\tau) &= \int\frac{du}{2\pi}e^{-iu\tau}\bar{c}_{ijkl}(u), \\
d_{ijkl}(\tau) &= \int\frac{du}{2\pi}e^{iu\tau}\bar{d}_{ijkl}(u),
\end{aligned}$$

exist, where the contragredient tranforms for c and d is taken for convenience. We obtain from Eqs. (2.3) the system of coupled differential equations

$$\begin{aligned}
&(id/d\tau-\omega-\epsilon_l)(id/d\tau-\epsilon_i-\epsilon_j)c_{ijkl}(\tau) \\
&\quad = -i\lambda_{ijmn}e^{-i(\epsilon_m+\epsilon_n)\tau}(\theta(\tau)p_m p_n \\
&\qquad + \theta(-\tau)h_m h_n)\lambda_{mnkl} \\
&\qquad + \lambda_{ijmn}(p_m p_n - h_m h_n)(id/d\tau-\omega-\epsilon_l)c_{mnkl}(\tau) \\
&\qquad + \lambda_{ijmn}(\theta(-\tau)h_n - \theta(\tau)p_n)
\end{aligned}$$

$$\times(-id/d\tau+\omega+\epsilon_l)e^{-i(\omega+\epsilon_n)\tau}d_{mnkl}(\tau)\,, \quad (2.4a)$$

$$(-id/d\tau-\omega+\epsilon_i)(-id/d\tau-\epsilon_j+\epsilon_l)d_{ijkl}(\tau)$$
$$=-i\lambda_{mjnl}\,e^{-(\epsilon_m-\epsilon_n)\tau}(\theta(\tau)p_m h_n+\theta(-\tau)p_n h_m)\lambda_{inkm}$$
$$+\lambda_{mjnl}(h_m p_n-h_n p_m)(-id/d\tau-\omega+\epsilon_i)d_{inkm}(\tau)$$
$$+\lambda_{mjnl}(\theta(\tau)h_n-\theta(-\tau)p_n)$$
$$\times(id/d\tau+\omega-\epsilon_i)e^{i(\omega+\epsilon_n)\tau}c_{inkm}(\tau). \quad (2.4b)$$

The explicit discontinuity of these coupled second-order equations, which appears on the right-hand side at $\tau=0$, rules out the possibility that the solution may be globally C^2. We therefore seek a strong solution within the two separate ranges $\tau>0$ (denoted by $\tau+$ in the following) and $\tau<0$ (denoted by $\tau-$), while we require the global solution to be C^1 throughout. We first treat the solution for $\tau>0$, while the range $\tau<0$ is discussed later.

A particular solution of the inhomogeneous system may be chosen as

$$\hat{c}_{ijkl}(\tau+)=-ip_i p_j\lambda_{ijkl}\frac{e^{-2i\epsilon\tau}}{\omega-2\epsilon+\epsilon_l}\,,$$

$$\hat{d}_{ijkl}(\tau+)=ih_j p_l\lambda_{ijkl}\frac{e^{-i(\epsilon-\delta)\tau}}{\omega+\epsilon-\delta-\epsilon_i}.$$

Four independent solutions of the homogeneous system are required in order to obtain the general solution of Eqs. (2.4a) and (2.4b). Two of these solutions are readily guessed, and we choose the form

$$c^{\mathrm{I}}_{ijkl}(\tau+)=i\alpha^{(+)}_{ijkl}e^{-i(\omega+\epsilon_l)\tau},$$
$$d^{\mathrm{I}}_{ijkl}(\tau+)=-ih_j p_l\alpha^{(+)}_{ijkl}e^{-i(\epsilon-\delta)\tau}, \quad (2.5)$$

and

$$c^{\mathrm{II}}_{cjkl}(\tau+)=-ip_i p_j\beta^{(+)}_{ijkl}e^{-2i\epsilon\tau},$$
$$d^{\mathrm{II}}_{ijkl}(\tau+)=i\beta^{(+)}_{ijkl}e^{-i(\epsilon_i-\omega)\tau}, \quad (2.6)$$

where the four-index quantities $\alpha^{(+)}$ and $\beta^{(+)}$ are integration constants. The two remaining solutions are harder to find. As verification is straightforward we only quote the results. In order to abbreviate the notation we introduce the quantities

$$1^s_{in,kl}=\delta_{ik}\delta_{jl},$$
$$1^t_{ik,lj}=\delta_{jk}\delta_{il},$$
$$S(\omega)=\omega 1^s-s,$$
$$s_{ij,kl}=(\epsilon_i+\epsilon_j)1^s+\lambda_{ijkl}(p_k p_l-h_k h_i),$$
$$T(\omega)=\omega 1^t-t,$$
$$t_{ik,lj}=(\epsilon_l-\epsilon_j)1^t+\lambda_{ijkl}(h_i p_k-p_i p_k).$$

The inverses of S and T are defined by the relations

$$[S(\omega)\cdot S^{-1}(\omega)]_{ijkl}=S_{ijmn}(\omega)S^{-1}_{mnkl}(\omega)=1^s,$$

$$[T(\omega)*T^{-1}(\omega)]_{ijkl}=T_{mjnl}(\omega)T^{-1}_{inkm}(\omega)=1^t.$$

A third solution of the homogeneous system is

$$c^{\mathrm{III}}_{ijkl}(\tau+)$$
$$=i\sum_s\sum_{n=0}^{\infty}c^{(n)}_{ijkl}(s+)\exp(-ip_s\tau)\exp[-in(\epsilon-\delta)\tau]\,, \quad (2.7)$$

$$d^{\mathrm{III}}_{ijkl}(\tau+)$$
$$=-i\sum_s\sum_{n=0}^{\infty}[T^{-1}(\omega+\delta-n(\epsilon-\delta)-\rho_s)*\lambda\,]_{mjrl}$$
$$\cdot h_r c^{(n)}_{irkm}(s+)\exp(-i(\rho_s+n(\epsilon-\delta)-\omega-\delta)\tau),$$

where

$$c^{(n)}_{ijkl}(s+)=-[S^{-1}(n(\epsilon-\delta)+\rho_s))\cdot\lambda\,]_{ijmr}p_r$$
$$\times[T^{-1}(\omega+\epsilon-n(\epsilon-\delta)-\rho_s)*\lambda\,]_{prql}$$
$$\cdot h_q c^{(n-1)}_{mqkp}(s+),$$
$$c^{(0)}_{ijkl}(s+)=\chi_{ij}(s)\bar{\chi}_{mn}(s)\gamma^{(+)}_{mnkl}.$$

The quantities $\chi(s)$ and $\bar{\chi}(s)$ are respectively left-hand and right-hand eigenvectors of $s_{ij,kl}$ associated with the eigenvalue ρ_s, i.e.,

$$S_{ijmn}(\rho_s)\chi_{mn}(s)=0,$$
$$\bar{\chi}_{mn}(s)S_{mnkl}(\rho_s)=0,$$

and γ^{+} is the integration constant. The fourth solution has an analogous structure,

$$c^{\mathrm{IV}}_{ijkl}(\tau+)=i\sum_t\sum_{n=0}^{\infty}[S^{-1}(\omega+\epsilon+n(\epsilon-\delta)-\sigma_t)\cdot\lambda\,]_{ijmr}$$
$$\cdot p_r d^{(n)}_{mrkl}(t+)\exp[-i(\omega+\epsilon$$
$$+n(\epsilon-\delta)-\sigma_t)\tau]\,,$$
$$d^{\mathrm{IV}}_{ijkl}(\tau+)=i\sum_t\sum_{n=0}^{\infty}d^{(n)}_{ijkl}(t+)e^{i\sigma_t\tau}e^{-in(\epsilon-\delta)\tau}, \quad (2.8)$$

where

$$d^{(n)}_{ijkl}(t+)=-[T^{-1}(\sigma_t-n(\epsilon-\delta))*\lambda\,]_{mjrl}h_r$$
$$\times[S^{-1}(\omega+\delta+n(\epsilon-\delta)-\sigma_t)\cdot\lambda\,]_{irpq}$$
$$\cdot p_q d^{(n-1)}_{pqkm}(t+)\,,$$
$$d^{(0)}_{ijkl}(t+)=\varphi_{jl}(t)\bar{\varphi}_{mn}(t)\xi^{(+)}_{inkm}.$$

This time we have defined φ, $\bar{\varphi}$ and σ_t by

$$T_{mjnl}(\sigma_t)\varphi_{nm}(t)=0,$$
$$\bar{\varphi}_{mn}(t)T_{inkm}(\sigma_t)=0,$$

and introduced the fourth integration constant $\xi^{(+)}$.

The same procedure is followed for $\tau<0$, with obvious modifications, giving rise to the integration constants $\alpha^{-},\ldots\xi^{-}$. The general solution of Eqs. (2.4) thus reads

$$c_{ijkl}(\tau\pm)=\hat{c}_{ijkl}(\tau\pm)+c^{\mathrm{I}}_{ijkl}(\tau\pm)+\cdots+c^{\mathrm{IV}}_{ijkl}(\tau\pm),$$
$$d_{ijkl}(\tau\pm)=\hat{d}_{ijkl}(\tau\pm)+d^{\mathrm{I}}_{ijkl}(\tau\pm)+\cdots+d^{\mathrm{IV}}_{ijkl}(\tau\pm).$$

3. THE INTEGRATION CONSTANTS

Having constructed a general solution of the coupled system of differential equations (2.4) we now proceeed to determine the integration constants in order to specify the solution in such a way that its Fourier transform fulfils Eqs. (2.3). The vertex function Γ which obeys Eq. (1.1) is then readily obtained by Eqs. (2.2) and (2.1)

Two different considerations yield the conditions for

determining the integration constants:

(i) The asymptotic boundary conditions requiring that $c(\tau\pm)$ and $d(\tau\pm)$ should vanish for $\tau\to\pm\infty$ to ensure that the Fourier transforms exist;

(ii) The matching conditions requiring that the solutions are C^1, i.e.,

$$c_{ijkl}(0+)=c_{ijkl}(0-),$$
$$\frac{d}{d\tau}c_{ijkl}(\tau)\Big|_{\tau=0+}=\frac{d}{d\tau}c_{ijkl}(\tau)\Big|_{\tau=0-}, \tag{3.1}$$

and correspondingly for d_{iikl}.

We turn to the boundary conditions first. As the infinitesimal imaginary parts of the particle and hole energies are negative and positive, respectively, we conclude from Eqs. (2.5) that

$$h_l\,\alpha^{(+)}_{ijkl}=0.$$

From the analogous expression for $c^{\mathrm{I}}_{ijkl}(\tau-)$ we obtain

$$p_l\,\alpha^{(-)}_{ijkl}=0.$$

Thus $\alpha^{(+)}$ and $\alpha(-)$ occur in mutually orthogonal subspaces. We may therefore deal with one integration constant α, remenbering that

$$\alpha^{(+)}_{ijkl}=p_l\,\alpha_{ijkl},$$
$$\alpha^{(-)}_{ijkl}=h_l\,\alpha_{ijkl}.$$

Similarly, from Eqs. (2.6) we introduce a single β with

$$\beta^{(+)}_{ijkl}=p_i\,\beta_{ijkl},$$
$$\beta^{(-)}_{ijkl}=h_i\,\beta_{ijkl}.$$

The asymptotic behavior of $c^{\mathrm{III}}_{ijkl}(\tau+)$ is determined by the infinitesimal imaginary part of the values ρ_s [Eqs. (2.7)] which are the eigenvalues[7] of s_{ijkl}. These eigenvalues may be classified in the following way:

$s\in S_1: N_p^2$ eigenvalues ρ_s with $\mathrm{Im}\,\rho_s<0$,

$s\in S_2: N_pN_h$ eigenvalues $\rho_s=\delta+\epsilon$ with $\mathrm{Im}\rho_s<0$, and $p_j\,\chi_{ij}(s)=\chi_{ij}(s)$,

$s\in S_3: N_h^2$ eigenvalues ρ_s with $\mathrm{Im}\,\rho_s>0$,

$s\in S_4: N_hN_p$ eigenvalues $\rho_s=\epsilon+\delta$ with $\mathrm{Im}\rho_s>0$, and $h_j\,\chi_{ij}(s)=\chi_{ij}(s)$.

Actually, the imaginary part of ρ_s for $s\in S_2\cup S_4$ is undefined as such. We define it in accordance with the first of Eqs. (2.2), where the convention is stipulated that the imaginary part of $(\epsilon_i+\epsilon_j)$ associated with $\bar c_{ijkl}$ is determined by the index j. In the case considered, the distinction between $\epsilon_i=\epsilon$, $\epsilon_j=\delta$, and $\epsilon_i=\delta$, $\epsilon_j=\epsilon$ is given by the second label of the eigenfunction $\chi_{ij}(s)$ associated with $\rho_s=\epsilon_i+\epsilon_j$. We conclude that only $s\in S_1\cup S_2$ may occur in the summation of Eqs. (2.7). The corresponding expressions for $\tau<0$ allow only $s\in S_3\cup S_4$. Since the projectors $\chi_{ij}(s)\bar\chi_{kl}(s)$ are mutually orthogonal for different s values,[8] we may again omit the superscripts $(+)$ and $(-)$ in the integration constants $\gamma^{(\pm)}_{ijkl}$ and use the single symbol γ_{ijkl}.

The eigenvalues σ_t of t_{ijkl}, which occur in Eqs. (2.8), are classified in a similar way:

$t\in T_1: N_pN_h$ eigenvalues σ_t with $\mathrm{Im}\sigma_t>0$,

$t\in T_2: N_h^2$ eigenvalues $\sigma_t=0$ *with* $Im\sigma_t>0$ and $h_j\,\varphi_{jl}=\varphi_{jl}$,

$t\in T_3: N_pN_h$ eigenvalues σ_t with $\mathrm{Im}\sigma_t<0$,

$t\in T_4: N_p^2$ eigenvalues $\sigma_t=0$ with $\mathrm{Im}\sigma_t<0$ and $p_j\,\varphi_{jl}=\varphi_{jl}$.

Concerning the infinitesimal imaginary part of the eigenvalues $\sigma_t=0$ we refer to the second of Eqs. (2.2), where the imaginary part of $(\epsilon_j-\epsilon_l)$ is defined by the index j of $\bar d_{ijkl}$ associated with it. Again, only $t\in T_1\cup T_2$ occurs in Eqs. (2.8), while only $t\in T_3\cup T_4$ occurs in the corresponding expressions for $\tau<0$. Owing to the orthogonality of the projectors $\varphi_{jl}(t)\bar\varphi_{ik}(t)$ we replace the integration constants $\zeta^{(\pm)}$ by ζ, as $\zeta^{(+)}$ and $\zeta^{(-)}$ appear in orthogonal subspaces.

The four four-index quantities $\alpha,\beta,\gamma,\zeta$ which play the role of integration constants in our procedure still depend on the parameter ω. They are uniquely determined by the linear equations expressing the matching conditions (3.1). To explicitly write down this linear system of four times N^4 equations requires some bookkeeping and is of little interest here. The coefficient matrix has full rank in general. The points ω_{pole} for which the determinant of the coefficient matrix vanishes are of physical interest. This aspect is discussed in Sec. 4. If $\alpha,\beta,\gamma,\zeta$ satisfy the linear equations (3.1), the solution of Eqs. (2.3) reads

$$\begin{aligned}\bar c_{ijkl}(\tilde\omega)=&\frac{h_l\,\alpha_{ijkl}}{\tilde\omega-\omega-\epsilon}-\frac{p_l\,\alpha_{ijkl}}{\tilde\omega-\omega-\epsilon}\\&+\frac{p_i\,p_j\,\beta_{ijkl}}{\tilde\omega-2\epsilon}-\frac{h_ih_j\,\beta_{ijkl}}{\tilde\omega-2\delta}\\&+\sum_{s\in S_1\cup S_4}\sum_{n=0}^{\infty}\frac{c^{(n)}_{ijkl}(s-)}{\tilde\omega+n(\epsilon-\delta)-\rho_s}\\&-\sum_{s\in S_1\cup S_2}\sum_{n=0}^{\infty}\frac{c^{(n)}_{ijkl}(s+)}{\tilde\omega-n(\epsilon-\delta)-\rho_s}\\&-\sum_{t\in T_1\cup T_2}\sum_{n=0}^{\infty}[S^{-1}(\omega+\epsilon+n(\epsilon-\delta)-\sigma_t)\cdot\lambda\,]_{ijm'n'}\\&\times\frac{p_{n'}d^{(n)}_{m'n'kl}(t+)}{\tilde\omega-\omega-\epsilon-n(\epsilon-\delta)+\sigma_t}\\&-\sum_{t\in T_3\cup T_4}\sum_{n=0}[S^{-1}(\omega+\delta-n(\epsilon-\delta)-\sigma_t)\cdot\lambda\,]_{ijm'n'}\\&\times\frac{h_{n'}d^{(n)}_{m'n'kl}(t-)}{\tilde\omega-\omega-\delta+n(\epsilon-\delta)+\sigma_t}\\&+\frac{p_ip_j+h_ih_j}{(\tilde\omega-\epsilon_i-\epsilon_j)(\omega-\epsilon_i-\epsilon_j+\epsilon_l)}\lambda_{ijkl},\end{aligned} \tag{3.2a}$$

$$\begin{aligned}\bar d_{ijkl}(\omega')=&\frac{p_jh_l\alpha_{ijkl}}{\omega'-\epsilon+\delta}-\frac{p_lh_l\,\alpha_{ijkl}}{\omega'-\delta+\epsilon}\\&+\frac{p_i\,\beta_{ijkl}}{\omega'-\omega+\epsilon}-\frac{h_i\,\beta_{ijkl}}{\omega'-\omega+\delta}\\&+\sum_{t\in T_1\cup T_2}\sum_{n=0}^{\infty}\frac{d^{(n)}_{ijkl}(t+)}{\omega'+n(\epsilon-\delta)-\sigma_t}\\&-\sum_{t\in T_3\cup T_4}\sum_{n=0}^{\infty}\frac{d^{(n)}_{ijkl}(t-)}{\omega'-n(\epsilon-\delta)-\sigma_t}\\&-\sum_{s\in S_1\cup S_2}\sum_{n=0}^{\infty}[T^{-1}(\omega+\delta-n(\epsilon-\delta)-\rho_s)^{*}\lambda\,]_{m'kn'l}\end{aligned}$$

$$\times \frac{h_{n'} c^{(n)}_{in'km'}(s+)}{\omega' - \omega - \delta + n(\epsilon - \delta) + \rho_s}$$

$$- \sum_{s \in S_1 \cup S_3} \sum_{n=0}^{\infty} [T^{-1}(\omega + \epsilon + n(\epsilon - \delta) - \rho_s)^* \lambda]_{m'kn'l}$$

$$\times \frac{p_{n'} c^{(n)}_{in'km'}(s-)}{\omega' - \omega - \epsilon - n(\epsilon - \delta) + \rho_s}$$

$$+ \frac{p_j h_l + p_l h_j}{(\omega' - \epsilon_j + \epsilon_l)(\omega - \epsilon_i - \epsilon_j + \epsilon_l)} \lambda_{ijkl}, \tag{3.2b}$$

from which the solution $\Gamma_{ijk}(\omega,\omega',\omega)$ of Eq. (1.1) can be obtained by using Eqs. (2.1) and (2.2).

4. DISCUSSION

Some of the properties of the solution of Eq. (1.1) may be read off immediately from Eqs. (3.2). While the poles at the unperturbed energies are cancelled by going from the pair $\bar{c},\bar{d}$ to Γ via Eq. (2.2), we find poles of Γ in the following combinations of the energy variables:

$$\tilde{\omega} = \rho_s + n(\epsilon - \delta), \quad s \in S_1,$$
$$\tilde{\omega} = \rho_s - n(\epsilon - \delta), \quad s \in S_3,$$
$$\omega' = \sigma_t + n(\epsilon - \delta), \quad t \in T_3,$$
$$\omega' = \sigma_t - n(\epsilon - \delta), \quad t \in T_1,$$

for $n = 0,1,2,\cdots$. We recall that $\tilde{\omega}$ and ω' are the s channel and t channel energies, respectively. It is obvious from the structure of the integral equation (1.1) that no explicit energy dependence occurs in the combination $\tilde{\omega} - \omega + \omega'$, which is the energy associated with the label j. Further poles occur in the combinations $\tilde{\omega} - \omega$ and $\omega - \omega'$, which are associated with the labels l and i, respectively

$$\tilde{\omega} - \omega = n(\epsilon - \delta) + \epsilon - \sigma_t, \quad t \in T_1,$$
$$\tilde{\omega} - \omega = n(\delta - \epsilon) + \delta - \sigma_t, \quad t \in T_3,$$
$$\omega - \omega' = n(\epsilon - \delta) - \delta + \rho_s, \quad s \in S_1,$$
$$\omega - \omega' = n(\delta - \epsilon) - \epsilon + \rho_s, \quad s \in S_3.$$

The positions of the poles listed so far are appropriate combinations of s channel (ρ_s) or t channel (σ_t) RPA energies with the unperturbed energies ϵ and δ. We still consider these energy combinations as essentially unperturbed in nature, since a proper blending of s channel and t channel features has not taken place.

A second class of poles occurs in the variable ω, however. These poles appear in the quantities $\alpha,\beta,\gamma,\zeta$ at those points at which the determinant of the linear system (3.1) vanishes (the case where no pole occurs in the solution although the determinant vanishes, would be purely incidental and is of no specific interest). The positions ω_p of these poles correspond to those values of ω for which the homogeneous part of Eq. (1.1) admits a nontrivial solution, i.e., they play the role of eigenvalues. From the physical point of view, these values correspond to an excitation spectrum which is generated by the specific model defined by Eq. (1.1). In terms of the single-particle basis the states occurring at the energy values ω_p consist of superpositions of n-particle, $(n-1)$-hole components or n-hole, $(n-1)$-particle components. Within the model considered it is in the variable ω that the blending mechanism of s channel and t channel becomes fully effective.

As stated in Sec. 3, the linear system (3.1) ensures that the expressions given in Eqs. (3.2) are in fact solutions of Eqs. (2.3). Exploiting (3.1) one finds as a consequence that

$$\lim_{\tilde{\omega} \to \pm \infty} \tilde{\omega}^2 \bar{c}(\tilde{\omega}) = 0,$$

$$\lim_{\omega' \to \pm \infty} \omega'^2 \bar{d}(\omega') = 0,$$

which is not obvious from the form in which $\bar{c}$ and $\bar{d}$ appear; it is consistent with the statement that the Fourier transform of $\bar{c}$ and $\bar{d}$ exists.

For zero interaction the determinant of the linear system (3.1) has zeros at ω equal to ϵ, δ, $2\epsilon - \delta$ and $2\delta - \epsilon$. To first order of the interaction parameter λ the solution is given by

$$\alpha = 0, \quad \beta = 0,$$

$$\gamma_{ijkl} = \frac{p_i p_j - h_i h_j}{\omega - \epsilon_i - \epsilon_j + \epsilon_l} \lambda_{ijkl}, \quad \zeta_{ijkl} = \frac{p_j h_l - h_j p_l}{\omega - \epsilon_i - \epsilon_j + \epsilon_l} \lambda_{ijkl}.$$

Using this result, it can be verified that $\bar{c}$ and $\bar{d}$ have no first-order term in λ, i.e., that the last term in Eqs. (3.2a) and (3.2b) in cancelled by the other terms. This property is consistent with Eqs. (2.3). As the interaction is turned on, the positions of the zeros of the determinant exhibit their parameter dependence. In addition to the zeros listed above, further zeros emerge from the unperturbed values at $n(\epsilon - \delta) + \epsilon$ and $n(\delta - \epsilon) + \delta$, $n = 2,3,\cdots$. The whole pattern, i.e., the dependence on the interaction and on the number of available particle and hole states, can be studied numerically. Since the rate of convergence of the series involved is governed by $1/(n!)^2$, such an investigation is easily implemented. While details of a calculation will be presented elsewhere,[6] we include here one further result which we consider to be important. The poles which occur in Γ in the variable $\tilde{\omega}$ are expected to appear in the four-point function $K(\tilde{\omega})$, since only integrations over ω' and ω are implied by going from Γ to K [see Eq. (1.2a)]. However, it turns out that the residues of K at the poles of $\tilde{\omega} = \rho_s \pm n(\epsilon - \delta)$ are very small for $n = 2,3,\cdots$. A new family of poles at $\tilde{\omega} = \omega_p^r + \epsilon$ and $\tilde{\omega} = \omega_p^l + \delta$ appears in turn with more significant residues (ω_p^r and ω_p^l are the poles of Γ in the variable ω, which were discussed above; the superscripts r and l indicate the negative and positive sign of the infinitesimal imaginary part, respectively). A corresponding statement holds for $F(\omega')$ [see Eq. (1.2b)].

We summarize as follows. Inorder to gain insight into the mechanism of implementation of crossing symmetry, an integral equation for the vertex funtion is considered, which brings about a considerable degree of genuine mixing between the s channel and the t channel. A major advantage in this mathematical study is solvability in closed form. In order to achieve this, simplifying assumptions are made which, we believe, are not crucial for the gross structure of the results such as the spectra displayed in $K(\tilde{\omega})$, $F(\omega')$ or the variable ω. Preliminary numerical calculations seem to confirm

the usefulness of this approach. While further studies should take into account time reversal as well as the u channel, only approximation schemes seem to be amenable to a solution. The methods used in this paper and the findings reached are expected to be of assistance in such further investigation.

ACKNOWLEDGMENT

A critical reading of the manuscript and some useful comments by Dr. R. Rösel are greatly appreciated.

[1]A.B. Migdal, *Theory of finite Fermi systems* (Interscience, New York, 1967). A.L. Fetter and J.D. Walecka, *Quantum theory of many-particle systems* (McGraw-Hill, New York, 1971).

[2]C.A. Engelbrecht, F.J.W. Hahne, and W.D. Heiss, Ann. Phys. (N.Y.) **104**, 221 (1977).

[3]F.J.W. Hahne, W.D. Heiss, and C.A. Engelbrecht, Ann. Phys. (N.Y.) **104**, 251 (1977).

[4]W.D. Heiss, C.A. Engelbrecht, and F.J.W. Hahne, Ann. Phys. (N.Y.) **104**, 274 (1977).

[5]F.J.W. Hahne, W.D. Heiss, and C.A. Engelbrecht, Phys.Lett. B **66**, 216 (1977).

[6]W.D. Heiss. Submitted for publication to S. Afr. J. Phys.

[7]We assume stability of the particle–particle and particle–hole RPA.

[8]In the degenerate cases $s \in S_2$ or $s \in S_4$ only the projectors $P_{S_a} = \Sigma_{s \in S_a} \chi_{ij}(s) \overline{\chi}_{kl}(s)$ are of interest ($a = 2$ or 4).

Independent subsystems in the problems of few-body interaction. Two-body scattering in force field

V. V. Komarov, A. M. Popova, and V. L. Shablov
Institute of Nuclear Physics, Moscow State University; Moscow 117234, USSR

(Received 9 February 1979; accepted for publication 23 April 1979)

A few-body quantum mechanical system is examined. It is assumed that the scattering operator of the system is a sum of two-body potential operators. Some odd ends in the sum commute. Four-dimensional perturbation potentials and second quantized formalism is used for a solution of the problem. The scattering of two distinct particles in a force field is considered as an example. For this case, the system of the integral equations for the scattering amplitudes are obtained. These amplitudes permit us to construct channel Möller operators. It has been shown that such system has unique solution equivalent to the solution of the corresponding Schrödinger equation.

1. INTRODUCTION

A method for solving the few-body scattering problems is proposed in the case where the interaction operator is given by a sum of operators, some of which commute with each other and with some of kinetic energy operators.

In our method, the second quantized formalism is used for an interaction operators definition. A scattering operator of the particles in the system is given by the series in the interaction representation, i.e., by the series of four-dimensional perturbation theory. The second quantitized formalism introduces one-particle Green's functions and dependence of the total scattering operator on the energy of all possible subsystems in the theory.[1,2]

As an example of how this method works, consider the problem of two distinct nonrelativistic particles scattering in a force field. For this case we will construct an integral equation set for scattering amplitudes which directly permit us to construct the channel Møller operators and to obtain a wave function with a real asymptotic behavior.

The equation set can be found by regrouping four-dimensional perturbation theory series and taking into account the commuting properties of the interacting operators. This set can be reduced to one equation containing in the kernel the total Green function of the system of two-particles in the force field, which are not interacting between themselves. The obtained set of equations is then investigated in this paper. It is shown that this set defines the system wave function which satisfies the corresponding Schroedinger equation. It is also shown that the homogeneous equation set corresponding to the one obtained in the present paper has no nontrivial solutions by values of the system energy which does not belong to the Hamiltonian energy spectrum.

In conclusion, the connections between different methods of deriving the many-particle scattering theory integral equations are considered.

2. THE REGROUPING OF THE FOUR DIMENSIONAL PERTURBATION THEORY SERIES

The Hamiltonian of a 2-particle in the force field is defined on the Hilbert space $H = L^2(R^6)$ in the form

$$H = H_0 + V, \qquad (1)$$

$$V = \sum_i V_i, \quad i = 1,2,12,$$

where $H_0 = H_{0_1} + H_{0_2}$ is the kinetic energy operator, V_i ($i = 1,2,12$) is a real function of i particle variables in $L^2(R^3)$ Hilbert space. It is known that in this case the operator H is self-adjoint on the domain of H_0. The operators V_1 and V_2 commute.

The presence of the commutating operators V_1 and V_2 in (1) distinguishes the solution of the problem with the Hamiltonian (1) from that of the three-body problem with finite masses.[3]

For solving the problem with Hamiltonian (1) the following scattering operator T in the interaction representation is used:

$$T = \sum_{n=1}^{\infty} \frac{(-i)^n}{n!} \int_{-\infty}^{\infty} dt_1 \cdots \int_{-\infty}^{\infty} dt_n P\{V(t_1)\cdots V(t_n)\}, \qquad (2)$$

because

$$T + 1 = \lim_{\substack{t \to \infty \\ t_0 \to -\infty}} U(t,t_0)$$

and $U(t,t_0)$ is evolution operator of the system. In Eq. (2) P is the Dyson operator and $V(t)$ is the particle interaction operator.

$$\begin{aligned} V(t) &= V_1(t) + V_2(t) + V_{12}(t), \\ V_i(t) &= e^{iH_{0i}t} V_i e^{-iH_{0i}t}, \quad i = 1,2, \\ V_{12}(t) &= e^{+iH_0 t} V_{12} e^{-iH_0 t}. \end{aligned} \qquad (3)$$

The scattering operator T defines all transitions in the given system from the initial state were the particles (1) and (2) are free (from the zero channel). Hence, the operator T is defined on the subspace $\mathscr{H}_0$ of the square-integrable functions and $\mathscr{H}_0$ does not contain the bound states state functions of the system.[4]

Using the full set of functions $|\mathbf{p}_1\, \mathbf{p}_2\rangle$, where $|\mathbf{p}_i\rangle$ is ith particle plane wave with the normalization of the form

$$\langle \mathbf{p}_i | \mathbf{q}_i \rangle = \delta(\mathbf{p}_i - \mathbf{q}_i),$$

 0022-2488/80/040854-08$1.00

one can represent

$$T|f_0\rangle = \int d\mathbf{p}_1 \int d\mathbf{p}_2 T(E_{p_i}+E_{p_2}+i0)|\mathbf{p}_1\mathbf{p}_2\rangle\langle\mathbf{p}_1\mathbf{p}_2|f_0\rangle,$$

and consider the operator T in the momentum space.

In this space the Hamiltonian (1) has the form:

$$\begin{aligned}\langle\mathbf{K}_1\mathbf{K}_2|H|\mathbf{q}_1\mathbf{q}_2\rangle &= E_{K_i}\delta(\mathbf{K}_1-\mathbf{q}_1)\delta(\mathbf{K}_2-\mathbf{q}_2)+E_{K_2}\delta(\mathbf{K}_1-\mathbf{q}_1)\delta(\mathbf{K}_2-\mathbf{q}_2)\\ &+\langle\mathbf{K}_1|V_1|\mathbf{q}_1\rangle\delta(\mathbf{K}_2-\mathbf{q}_2)+\langle\mathbf{K}_2|V_2|\mathbf{q}_2\rangle\delta(\mathbf{K}_1-\mathbf{q}_1)\\ &+\langle\mathbf{K}_{12}|V_{12}|\mathbf{q}_{12}\rangle\delta(\mathbf{K}_1+\mathbf{K}_2-\mathbf{q}_1-\mathbf{q}_2). \qquad (4)\end{aligned}$$

The operator $V(t)$ is defined in the second quantized formalism.

$$V_i(t)=\sum_{p,q_i} a_i^+(\mathbf{p}_i,t)\langle\mathbf{p}_i|V_i|\mathbf{q}_i\rangle a_i(\mathbf{q}_i,t), \quad i=1,2, \qquad (5)$$

$$\begin{aligned}V_{12}(t)=&\sum_{p_1p_2,q_1q_2} a_1^+(\mathbf{p}_1,t)a_2^+(\mathbf{p}_2,t)\langle\mathbf{p}_1\mathbf{p}_2|V_{12}|\mathbf{q}_1\mathbf{q}_2\rangle\\ &\times a_1(\mathbf{q}_1,t)a_2(\mathbf{q}_2,t). \qquad (6)\end{aligned}$$

It is known that the Green function of the particle (i) in second quantization[1,2] is determined as

$$\begin{aligned}g_{0i}(\mathbf{p}_i,t_1-t_2)&=\langle 0|P\{a_i(\mathbf{p}_i,t_1)a_i^+(\mathbf{p}_i,t_2)\}|0\rangle\\ &=\exp(-iE_{p_i}(t_1-t_2))\Theta(t_1-t_2)\\ &=\int_{-\infty}^{\infty}\frac{d\varepsilon_1}{-2\pi i}\frac{\exp(-i\varepsilon_i(t_1-t_2))}{\varepsilon_i-E_{p_i}+i0}, \qquad (7)\end{aligned}$$

where $|0\rangle$ denotes the vacuum state;

$$\begin{cases} a_i(\mathbf{p}_i,t) & \text{is the annihilation operator,}\\ a_i^+(\mathbf{p}_i,t) & \text{is the creation operator,}\\ \Theta(x) & \text{is the Heaviside function.}\end{cases} \qquad (8)$$

Further, we will use the time-independent Green function

$$g_{0i}(\varepsilon_i)=(\varepsilon_i-H_{0i}+i0)^{-1},$$

$$\langle\mathbf{p}_i|g_{0i}(\varepsilon_i)|\mathbf{q}_i\rangle=\frac{\delta(\mathbf{p}_i-\mathbf{q}_i)}{\varepsilon_i-E_{p_i}+i0}, \qquad (9)$$

which is an operator dependent on the variable ε_i.

Let us study the properties of the operator T given by the series (2) with the commuting operators $V_1(t)$ and $V_2(t)$.

It will be shown below that the presence of the commuting interaction operators leads to the special procedure of regrouping of the series (2). Such regrouping procedure may be treated as the procedure for selecting the asymptotics of the system wave function.[1,2]

This procedure may be demonstrated by transforming the third-order term of the series (3):

$$\begin{aligned}&\frac{1}{3!}\int_{-\infty}^{\infty}dt_1\int_{-\infty}^{\infty}dt_2\int_{-\infty}^{\infty}dt_3 P\Big\{\sum_\alpha V_\alpha(t_1)\\ &\times\sum_\alpha V_\alpha(t_2)\sum_\alpha V_\alpha(t_3)\Big\}\\ &=\frac{1}{3!}\sum_{\alpha_1\alpha_2\alpha_3}\int_{-\infty}^{\infty}dt_1\int_{-\infty}^{\infty}dt_2\int_{-\infty}^{\infty}dt_3 V_{\alpha_i}(t_1)\\ &\times V_{\alpha_2}(t_2)V_{\alpha_2}(t_3)\Theta(t_1-t_2)\Theta(t_2-t_1),\\ &\alpha_1,\alpha_2,\alpha_3=1,2,12. \qquad (10)\end{aligned}$$

The following cases are sufficient to consider in studying the properties of the expression (10).

First case: Let $V_1(t)=V_{\alpha_1}(t)=V_{\alpha_2}(t)$, $V_2(t)=V_{\alpha_3}(t)$ and the commutator $[V_1(t),V_2(t)]=0$ is equal to zero at any t_1 and t_2. In this case the term of the form

$$\begin{aligned}&\frac{1}{2!}\int_{-\infty}^{\infty}dt_1\int_{-\infty}^{\infty}dt_2\int_{-\infty}^{\infty}dt_3 P\{V_1(t_1)V_1(t_2)V_2(t_3)\}\\ &=\int_{-\infty}^{\infty}dt_1\int_{-\infty}^{\infty}dt_2\int_{-\infty}^{\infty}dt_3 V_1(t_1)V_1(t_2)V_2(t_3)\\ &\times\{\Theta(t_1-t_2)\Theta(t_2-t_3)+\Theta(t_1-t_3)\Theta(t_3-t_2)\\ &+\Theta(t_3-t_1)\Theta(t_1-t_2)\}, \qquad (11)\end{aligned}$$

is singled out in the sum (10). In (11), the position of the operator $V_2(t_3)$ relative to the operator $V_1(t)$ is of no importance because of the commutation property of these operators. Bearing in mind this fact and the following Θ-function property

$$\begin{aligned}&\Theta(t_1-t_2)\Theta(t_2-t_3)+\Theta(t_1-t_3)\Theta(t_3-t_2)\\ &+\Theta(t_3-t_1)\Theta(t_1-t_2)=\Theta(t_1-t_2), \qquad (12)\end{aligned}$$

the expression (11) will be reduced to the form

$$\begin{aligned}&\frac{1}{2!}\int_{-\infty}^{\infty}dt_1\int_{-\infty}^{\infty}dt_2\int_{-\infty}^{\infty}dt_3 P\{V_1(t_1)V_1(t_2)V_2(t_3)\}\\ &=\frac{1}{2!}\int_{-\infty}^{\infty}dt_1\int_{-\infty}^{\infty}dt_2 P\{V_1(t_1)V_1(t_2)\}\\ &\otimes\int_{-\infty}^{\infty}dt_3 V_2(t_3). \qquad (13)\end{aligned}$$

The symbol $\otimes$ designates the direct products of the operators acting on various arguments of $\langle p_1\, p_2\,|f_0\rangle$ so that the matrix element of such a product is equal to the product of the matrix elements in the momentum representation. In the case of higher-order terms, when only the commutating operators $V_1(t)$ and $V_2(t)$ are inserted after the sign P, the regrouping procedure is the same.

Second Case: Let

$$V_{\alpha_1}(t_1)=V_1(t_1), \quad V_{\alpha_2}(t_2)=V_2(t_2), \quad V_{\alpha_3}(t_3)=V_{12}(t_3).$$

The term

$$\begin{aligned}&\int_{-\infty}^{\infty}dt_1\int_{-\infty}^{\infty}dt_2\int_{-\infty}^{\infty}dt_3 V_1(t_1)V_2(t_2)V_{12}(t_3)\\ &\cdot\{\Theta(t_1-t_2)\Theta(t_2-t_3)+\Theta(t_2-t_1)\Theta(t_1-t_3)\}, \qquad (14)\end{aligned}$$

is singled out in (10). Since

$$\begin{aligned}&\Theta(t_1-t_2)\Theta(t_2-t_3)+\Theta(t_2-t_1)\Theta(t_1-t_3)\\ &=\Theta(t_1-t_3)\Theta(t_2-t_3), \qquad (15)\end{aligned}$$

the expression (14) is regrouped to the form

$$\begin{aligned}&\Big\{\int_{-\infty}^{\infty}dt_1 V_1(t_1)\otimes\int_{-\infty}^{\infty}dt_2^{V_2}(t_2)\Big\}\\ &\times\int_{-\infty}^{\infty}dt_3 V_{12}(t_3)\Theta(t_1-t_3)\Theta(t_2-t_3)\\ &=\Big\{\int_{-\infty}^{\infty}dt_1 V_1(t_1)\otimes\int_{-\infty}^{\infty}dt_2^{V_2}(t_2)\Big\}\\ &\times\int_{-\infty}^{\infty}dt_3 V_{12}(t_3)\Theta[\min(t_1,t_2)-t_3]. \qquad (16)\end{aligned}$$

Third case: Let

$$V_{\alpha_1}(t_1) = V_{12}(t_1), \quad V_{\alpha_2}(t_2) = V_1(t_2), \quad V_{\alpha_3}(t_3) = V_2(t_3).$$

Then the examined term from (10) is of the form

$$\int_{-\infty}^{\infty} dt_1 \int_{-\infty}^{\infty} dt_2 \int_{-\infty}^{\infty} dt_3 V_{12}(t_1)V_1(t_2)V_2(t_3) \times \{\Theta(t_1 - t_2)\Theta(t_2 - t_3) + \Theta(t_1 - t_3)\Theta(t_3 - t_2)\}. \quad (17)$$

Since

$$\Theta(t_1 - t_2)\Theta(t_2 - t_3) + \Theta(t_1 - t_3)\Theta(t_3 - t_2) = \Theta(t_1 - t_2)\Theta(t_1 - t_3), \quad (18)$$

regrouping of (17) gives the following result:

$$\int_{-\infty}^{\infty} dt_1 V_{12}(t_1) \times \left\{ \int_{-\infty}^{\infty} dt_2 V_1(t_2)\Theta(t_1 - t_2) \otimes \int_{-\infty}^{\infty} dt_3 V_2(t_3)\Theta(t_1 - t_3) \right\}. \quad (19)$$

3. THE INTEGRAL EQUATION SET FOR THE CHANNEL AMPLITUDES

The properties of the series (2), which are shown in Sec. 2 of the present paper, permit us to represent the scattering operator $T(E + i0)$ as the sum of the channel operators

$$T_1(E_1, E + i0), \quad T_2(E_2 = E - E_1, E + i0),$$
$$T_{1\otimes 2}(E_1, E - E_1, E + i0), \quad T_{12}(E_{12}, E + i0).$$

These operators are defined as

$$\left\langle \mathbf{K}_1\mathbf{K}_2 \left| \lim_{t\to\infty} W_1(t) \right| \mathbf{K}_1'\mathbf{K}_2' \right\rangle = -2\pi i\delta(E_{K_1} + E_{K_2} - E_{K_1'} - E_{K_2'}) \times \langle \mathbf{K}_1\mathbf{K}_2 | T_1(E_1, E + i0) | \mathbf{K}_1'\mathbf{K}_2' \rangle,$$

$$\left\langle \mathbf{K}_1\mathbf{K}_2 \left| \lim_{t\to\infty} W_2(t) \right| \mathbf{K}_1'\mathbf{K}_2' \right\rangle = -2\pi i\delta(E_{K_1} + E_{K_2} - E_{K_1'} - E_{K_2'}) \times \langle \mathbf{K}_1\mathbf{K}_2 | T_2(E_2 = E - E_1, E + i0) | \mathbf{K}_1'\mathbf{K}_2' \rangle,$$

$$\left\langle \mathbf{K}_1\mathbf{K}_2 \left| \lim_{t\to\infty} W_{1\otimes 2}(t) \right| \mathbf{K}_1'\mathbf{K}_2' \right\rangle = -2\pi i\delta(E_{K_1} + E_{K_2} - E_{K_1'} - E_{K_2'}) \times \langle \mathbf{K}_1\mathbf{K}_2 | T_{1\otimes 2}(E_1, E_2 = E - E_1, E + i0) | \mathbf{K}_1'\mathbf{K}_2' \rangle,$$

$$\left\langle \mathbf{K}_1\mathbf{K}_2 \left| \lim_{t\to\infty} W_{12}(t) \right| \mathbf{K}_1'\mathbf{K}_2' \right\rangle = -2\pi i\delta(E_{K_1} + E_{K_2} - E_{K_1'} - E_{K_2'}) \times \langle \mathbf{K}_1\mathbf{K}_2 | T_{12}(E_{12}', E + i0) | \mathbf{K}_1'\mathbf{K}_2' \rangle, \quad (20)$$

where the energies, E_α, $\alpha = (1,2,12)$, and E_{12} have the following meaning:

$$E_1 = E_{K_1} = K_1^2/2m_1, \quad E_2 = E_{K_2} = K_2^2/2m_2,$$

are the energies of the particles 1 and 2 in the initial state; $E_1' = E_{K_1'} = K_1'^2/2m_1$ and $E_2' = E_{K_2'} = K_2'^2/2m_2$ are the energies of the particles 1 and 2 in the final state, $E = E_1' + E_2' = E_{K_1'} + E_{K_2'}$ is the total energy and $E_{12}' = E - (K_1 + K_2)^2/2(m_1 + m_2)$ is the relative motion energy of the particles 1 and 2 in the final state. (The units adopted here are those in which $\hbar = 1$.) The operators $W_\alpha(t)$, $\alpha = 12$, $\alpha = (1),(2),(1\otimes 2)$, in (20), satisfy the equation set

$$W_i(t) = \sum_{n=1}^{\infty} (-i)^n \int_{-\infty}^{\infty} dt_1 \cdots \int_{-\infty}^{\infty} dt_n V_i(t_1)\cdots V_i(t_n) \cdot\Theta(t - t_1)\Theta(t_1 - t_2)\cdots\Theta(t_{n-1} - t_n) \times \{1 + W_{12}(t_n)\}, \quad i = 1,2,$$

$$W_{1\otimes 2}(t) = \sum_{n,m=1}^{\infty} (-i)^{n+m} \int_{-\infty}^{\infty} dt_1 \cdots \int_{-\infty}^{\infty} dt_n V_1(t_1)\cdots V_1(t_n) \times\Theta(t - t_1)\cdots\Theta(t_{n-1} - t_n) \otimes \int_{-\infty}^{\infty} d\tau_1 \cdots \times \int_{-\infty}^{\infty} d\tau_m V_2(\tau_1)\cdots V_2(\tau_m)\cdot\Theta(t - \tau_1)\cdots\Theta(\tau_{m-1} - \tau_m) \cdot\{1 + W_{12}[\min(t_n,\tau_m)]\},$$

$$W_{12}(t) = \sum_{n=1}^{\infty} (-i)^n \int_{-\infty}^{\infty} dt_1 \cdots \int_{-\infty}^{\infty} dt_n V_{12}(t_1)\cdots V_{12}(t_n) \times\Theta(t - t_n)\cdot \cdots \Theta(t_{n-1} - t_n)\cdot\{1 + W_1(t_n) + W_2(t_n) + W_{1\otimes 2}(t_n)\}. \quad (21)$$

It can be proved, based on the set (21), that the operator $U(t, t_0 = -\infty)$ given by the sum

$$U(t, t_0 = -\infty) = 1 + \sum_\alpha W_\alpha(t), \quad \alpha = (1),(2),(12),(1\otimes 2), \quad (22)$$

is the solution of the Schrödinger equation in the interaction representation

$$i\frac{\partial}{\partial t} U(t, t_0 = -\infty) = V(t)U(t, t_0 = -\infty), \quad U(t = -\infty, t_0 = -\infty) = 1. \quad (23)$$

This statement can be verified by differentiation.

This means that the result of the regrouping procedure, described in Sec. 2, is equivalent to the solution of the Schrödinger equation in the interaction representation. It follows from this in particular, that the Møller operator

$$\Omega_+ = U(t = 0, t_0 = -\infty),$$

is the isometric operator

$$\langle f_0' | \Omega_+^\dagger \Omega_+ | f_0 \rangle = \langle f_0' | f_0 \rangle.$$

We introduce the channel operators $T_\alpha(E + i0)$, where $\alpha = (1),(2),(1\otimes 2),(12)$, to define the transition of the system to all possible asymptotic states.

The operators $T_1(E + i0)$ and $T_2(E + i0)$ determine only the transitions when particle 1 or 2 in its final state is scattered by the force center while particle 2 or 1, respectively, is free. The operator $T_{1\otimes 2}(E + i0)$ corresponds to the transition when particles 1 and 2 in their states are simultaneously and independently scattered by the force center. The operator $T_{12}(E + i0)$ determines the transition when mutually interacting particles 1 and 2 are the final states.

Before deducing the set of integral equations for the scattering channel amplitudes it is necessary to give some definitions. The matrix element of the operator $t_i(E_i' + i0)$,

where $\alpha = 1,2,12$, is defined by the series

$$\left\langle \mathbf{K}_i \left| \sum_{n=1}^{\infty} (-i)^n \int_{-\infty}^{\infty} dt_1 \cdots \int_{-\infty}^{\infty} dt_n P\{V_i(t_1)\cdots V_i(t_n)\} \right| K_i' \right\rangle$$
$$= -2\pi i \delta(E_{K_i} - E_{K_i'}) \langle K_i | t_i(E_{K_i} + i0) | \mathbf{K}_i' \rangle;$$
$$E_i = K_i^2/2m_i. \tag{24}$$

In the case of $i = (12)$, E_i and $\mathbf{K}_i$ are the energy and momentum of the relative motion of particles 1 and 2.

It is easy to verify that the operators $t_1(E_1 + i0)$, $t_2(E_2 + i0)$, and $t_{12}(E_{12} + i0)$, satisfy the Lippman–Schwinger equations for the potentials V_1, V_2, V_{12}, respectively:

$$t_1(E_1 + i0) = V_1 + V_1 g_{01}(E_1 + i0) t_1(E_1 + i0),$$
$$t_2(E_2 + i0) = V_2 + V_2 g_{02}(E_2 + i0) t_2(E_2 + i0), \tag{25}$$
$$t_{12}(E_{12} + i0) = V_{12} + V_{12} \int_{-\infty}^{\infty} \frac{d\epsilon_1}{-2\pi i} g_{01}(\epsilon_1 + i0) \times g_{02}(E_{12} - \epsilon_1 + i0) t_{12}(E_{12} + i0).$$

One can see from (20), that the operators $T_1(E_1, E + i0)$, $T_2(E_2 = E - E_1, E + i0)$, $T_{1\otimes 2}(E_1, E_2 = E - E_1, E + i0)$, and $T_{12}(E_{12}, E + i0)$ depend on the energies of the real particles on the energy shell.

Consider further Eq. (21) and take into account Eqs. (24) and (25). This gives in the second quantized formalism for the operators $T_1(\epsilon_1, E + i0)$, $T_2(\epsilon_2 = E - \epsilon_1, E + i0)$, $T_{1\otimes 2}(\epsilon_1, \epsilon_2 = E - \epsilon_1, E + i0)$, and $T_{12}(\epsilon_{12}, E + i0)$, and the equation set:

$$T_1(\epsilon_1, E + i0) = t_1(\epsilon_1 + i0)\delta(\epsilon_1, E_1') + t_1(\epsilon_1 + i0) \times g_{01}(\epsilon_1 + i0) T_{12}(\epsilon_{12}, E + i0),$$
$$T_2(E - \epsilon_1, E + i0) = t_2(E - \epsilon_1 + i0)\delta(E - \epsilon_1, E_2') \times + t_2(E - \epsilon_1 + i0) g_{02}(E - \epsilon_1 + i0) T_{12}(\epsilon_{12}, E + i0),$$
$$T_{1\otimes 2}(\epsilon_1, E - \epsilon_1, E + i0) = -2\pi i \delta(\epsilon_1 - E_1') t_1(\epsilon_1 + i0) t_2(E - \epsilon_1 + i0) + t_1(\epsilon_1 + i0) t_2(E - \epsilon_1 + i0) g_{01}(\epsilon_1 + i0) \times g_{02}(E - \epsilon_1 + i0) T_{12}(\epsilon_{12}, E + i0),$$
$$T_{12}(\epsilon_{12}, E + i0) = t_{12}(\epsilon_{12} + i0)\delta(\epsilon_{12}, E_{12}') + t_{12}(\epsilon_{12} + i0) \cdot \int_{-\infty}^{\infty} \frac{d\varepsilon_1}{-2\pi i} g_{01}(\varepsilon_1 + i0) \cdot g_{02}(E - \varepsilon_1 + i0) \cdot [T_1(\varepsilon_1, E + i0) + T_2(E - \varepsilon_1, E + i0) + T_{1\otimes 2}(\varepsilon_1, E - \varepsilon_1, E + i0)]. \tag{26}$$

Here ε_1, ε_2, ϵ_1, ϵ_2 are the energies of the virtual particles off the energy shell.

The operators $T_1(\epsilon_1, E + i0)$, $T_2(\epsilon_2 = E - \epsilon_1, E + i0)$, and $T_{1\otimes 2}(\epsilon_1, E - \epsilon_1, E + i0)$ are the analytic continuation of the corresponding operators $T_1(E_1, E + i0)$, $T_2(E - E_1, E + i0)$, and $T_{1\otimes 2}(E_1, E - E_1, E + i0)$ over the variable E_1 and E_2 of the physical domain.

In Eq. (26) $\delta(\alpha_1\beta)$ is the Kroneker symbol, E_{12} is the relative motion energy of the real particles 1 and 2 in the final state, and ϵ_{12} is the relative motion energy of the virtual particles 1 and 2.

Let us note that in the matrix elements of the operators $T_{12}(\epsilon_{12}, E + i0)$ and $t_{12}(\epsilon_{12} + i0)$, which are

$$\langle \mathbf{q}_1 \mathbf{q}_2 | T_{12}(\epsilon_{12}, E + i0) | \mathbf{K}_1' \mathbf{K}_2' \rangle,$$

and

$$\langle \mathbf{q}_1 \mathbf{q}_2 | t_{12}(\epsilon_{12} + i0) | \mathbf{p}_1 \mathbf{p}_2 \rangle = \delta(\mathbf{q}_1 + \mathbf{q}_2 - \mathbf{p}_1 - \mathbf{p}_2) \langle \mathbf{q}_{12} | t_{12}(\epsilon_{12} + i0) | \mathbf{p}_{12} \rangle, \tag{27}$$

the energy ϵ_{12} and momentum q_{12} are

$$\epsilon_{12} = E - \frac{(\mathbf{q}_1 + \mathbf{q}_2)^2}{2(m_1 + m_2)}, \quad \mathbf{q}_{12} = \frac{\mathbf{q}_1 m_2 - \mathbf{q}_2 m_1}{m_1 + m_2}. \tag{28}$$

The one-particle Green functions (9) appears in the Eq. (26), obtained in second quantized formalism from Eq. (21), because of presence of the Θ functions in Eq. (21). Consider the operators $N_1(\epsilon_1, E + i0)$, $N_2(E - \epsilon_1, E + i0)$, $N_{1\otimes 2}(\epsilon_1, E - \epsilon_1, E + i0)$, and $N_{12}(\epsilon_2, E + i0)$, which are defined by the relations

$$T_1(\epsilon_1, E + i0) = t_1(\epsilon_1 + i0)\delta(\epsilon_1, E_1') + N_1(\epsilon_1, E + i0),$$
$$T_2(E - \epsilon_1, E + i0) = t_2(E - \epsilon_1 + i0)\delta(E - \epsilon_1, E_2') + N_2(E - \epsilon_1, E + i0),$$
$$T_{1\otimes 2}(\epsilon_1, E - \epsilon_1, E + i0) = -2\pi i \delta(\epsilon_1 - E_1') t_1(\epsilon_1 + i0) \cdot t_2(E - \epsilon_1 + i0) + N_{1\otimes 2}(\epsilon_1, E - \epsilon_1, E + i0),$$
$$T_{12}(\epsilon_{12}, E + i0) = t_{12}(\epsilon_{12} + i0)\delta(\epsilon_{12}, E_{12}') + N_{12}(\epsilon_{12}, E + i0). \tag{29}$$

The equations for these operations follow from Eq. (26):

$$N_1(\epsilon_1, E + i0) = t_1(\epsilon_1 + i0) g_{01}(\epsilon_1 + i0) t_{12}(E_{12}' + i0) + t_1(\epsilon_1 + i0) g_{01}(\epsilon_1 + i0) N_{12}(\epsilon_{12}, E + i0),$$
$$N_2(E - \epsilon_1, E + i0) = t_2(E - \epsilon_1 + i0) g_{02}(E - \epsilon_1 + i0) t_{12}(E_{12}' + i0) + t_2(E - \epsilon_1 + i0) g_{02}(E - \epsilon_1 + i0) N_{12}(\epsilon_{12}, E + i0),$$
$$N_{1\otimes 2}(\epsilon_1, E - \epsilon_1, E + i0) = t_1(\epsilon_1 + i0) t_2(E - \epsilon_1 + i0) \cdot g_{01}(\epsilon_1 + i0) \cdot g_{02}(E - \epsilon_1 + i0) t_{12}(E_{12}' + i0) + t_1(\epsilon_1 + i0) \times t_2(E - \epsilon_1 + i0) g_{01}(\epsilon_1 + i0) g_{02}(E - \epsilon_1 + i0) N_{12}(\epsilon_{12}, E + i0),$$
$$N_{12}(\epsilon_{12}, E + i0) = t_{12}(\epsilon_{12} + i0) g_{01}(E_1' + i0) t_1(E_1' + i0) + t_{12}(\epsilon_{12} + i0) \cdot g_{02}(E_2' + i0) t_2(E_2' + i0) + t_{12}(\epsilon_{12} + i0) g_{01}(E_1' + i0) g_{02}(E_2' + i0) \times t_1(E_1' + i0) t_2(E_2' + i0) + t_{12}(\epsilon_{12} + i0)$$

$$\int_{-\infty}^{\infty} \frac{d\varepsilon_1}{-2\pi i} g_{01}(\varepsilon_1 + i0)\, g_{02}(E-\varepsilon_1+i0)\cdot[N_1(\varepsilon_1,E+i0)+N_2(E-\epsilon_1,E+i0)$$
$$\times N_{1\otimes 2}(\varepsilon_1,E-\varepsilon_1,E+i0)]. \tag{30}$$

One can see that the matrix elements of the operators $N_\alpha(E+i0)$ define only connected amplitudes.

We note as a consequence that the S matrix of the studied system is the sum of all possible disconnected amplitudes and the completely bound amplitude with the singularity appearing due to the law of total energy conservation. In this case, the law of momentum conservation in the entire system is not satisfied. Thus, we get the following representation for the S matrix:

$$\begin{aligned}
&\langle \mathbf{K}_1\mathbf{K}_2|S|\mathbf{K}_1'\mathbf{K}_2'\rangle \\
&\quad = \delta(\mathbf{K}_1-\mathbf{K}_2')\delta(\mathbf{K}_2-\mathbf{K}_2') + (-2\pi i)\delta(E_{K_1}-E_{K_1'})\delta(\mathbf{K}_2-\mathbf{K}_2')\langle \mathbf{K}_1|t_1(E_{K_1'}+i0)|\mathbf{K}_1'\rangle + (-2\pi i) \\
&\qquad \times\delta(E_{K_2}-E_{K_2'})\delta(\mathbf{K}_1-\mathbf{K}_1')\langle \mathbf{K}_2|t_2(E_{K_2'}+i0)|\mathbf{K}_2'\rangle + (-2\pi i)^2\delta(E_{K_1}-E_{K_1'})\delta(E_{K_2}-E_{K_2'}) \\
&\qquad \cdot\langle \mathbf{K}_1|t_1(E_{K_1'}+i0)|\mathbf{K}_1'\rangle\cdot\langle \mathbf{K}_2|t_2(E_{K_2'}+i0)|\mathbf{K}_2'\rangle + (-2\pi i)\delta(E_{K_1}+E_{K_2}-E_{K_1'}-E_{K_2'}) \\
&\qquad \cdot\langle \mathbf{K}_1\mathbf{K}_2|N(E+i0)|\mathbf{K}_1'\mathbf{K}_2'\rangle,
\end{aligned} \tag{31}$$

where $E = E_{K_1'} + E_{K_2'}$ and

$$\begin{aligned}
\langle \mathbf{K}_1\mathbf{K}_2|N(E+i0)|\mathbf{K}_1'\mathbf{K}_2'\rangle &= \langle \mathbf{K}_1\mathbf{K}_2|N_1(E_1,E+i0)|\mathbf{K}_1\mathbf{K}_2\rangle + \langle \mathbf{K}_1\mathbf{K}_2|N_2(E-E_1,E+i0)|\mathbf{K}_1\mathbf{K}_2\rangle \\
&\quad + \langle \mathbf{K}_1\mathbf{K}_2|N_{1\otimes 2}(E_1,E-E_1,E+i0)|\mathbf{K}_1\mathbf{K}_2\rangle + \langle \mathbf{K}_1\mathbf{K}_2|N_{12}(E_{12},E+i0)|\mathbf{K}_1\mathbf{K}_2\rangle,
\end{aligned}$$

is the completely connected amplitude of particle scattering in the system, on the energy shell. The system (30) can be reduced to the one equation of the type

$$\begin{aligned}
N_{12}(\epsilon_1,E+i0) &= t_{12}(\epsilon_{12}+i0)\cdot[g_{01}(E_1'+i0)t_1(E_1'+i0) + g_{02}(E_2'+i0)t_2(E_2'+i0) + g_{01}(E_1'+i0)\,g_{02}(E_2'+i0) \\
&\quad \cdot t_1(E_1'+i0)\cdot t_2(E_2'+i0)] + t_2(\epsilon_2+i0)\cdot\int_{-\infty}^{\infty}\frac{d\varepsilon_1}{-2\pi i}[g_{01}(\varepsilon_1+i0)g_{02}(E-\epsilon_1+i0)] \\
&\quad \times t_1(\varepsilon_1+i0)g_{01}(\varepsilon_1+i0) + g_{01}(\varepsilon_1+i0)g_{02}(E-\varepsilon_1+i0)\,t_2(E-\varepsilon_1+i0)g_{02}(E-\varepsilon_1+i0) \\
&\quad + g_{01}(\varepsilon_1+i0)g_{02}(E-\varepsilon_1+i0)t_1(\varepsilon_1+i0)t_2(E-\varepsilon_1+i0)g_{01}(\varepsilon_1+i0)\cdot g_{02}(E-\varepsilon_1+i0)] \\
&\quad \cdot[N_{12}(\epsilon_{12},E+i0) + t_{12}(E_{12}'+i0)].
\end{aligned} \tag{32}$$

The set of integral equations for the amplitudes $N_\alpha(E+i0)$, $\alpha = (1),(2),(1\otimes 2),12$, is of the form

$$\begin{aligned}
\langle \mathbf{K}_1\mathbf{K}_2|N_1(\epsilon_1,E+i0)|\mathbf{K}_1'\mathbf{K}_2'\rangle &= \int d\mathbf{q}_1\,\frac{\langle \mathbf{K}_1|t_1(\epsilon_1+i0)|\mathbf{q}_1\rangle}{\epsilon_1-E_{q_1}+i0}\langle \mathbf{q}_1\mathbf{K}_2|t_{12}(E_{12}'+i0)|\mathbf{K}_1'\mathbf{K}_2'\rangle \\
&\quad + \int d\mathbf{q}_1\,\frac{\langle \mathbf{K}_1|t_1(\epsilon_1+i0)|\mathbf{q}_1\rangle}{\epsilon_1-E_{q_1}+i0}\left\langle \mathbf{q}_1\mathbf{K}_2\left|N_{12}\left(\epsilon_{12}=E-\frac{(\mathbf{q}_1+\mathbf{K}_1)^2}{2(m_1+m_2)},E+i0\right)\right|\mathbf{K}_1'\mathbf{K}_2'\right\rangle,
\end{aligned}$$

$$\begin{aligned}
\langle \mathbf{K}_1\mathbf{K}_2|N_2(E-\epsilon_1,E+i0)|\mathbf{K}_1'\mathbf{K}_2'\rangle &= \int d\mathbf{q}_2\,\frac{\langle \mathbf{K}_2|t_2(E-\epsilon_1+i0)|\mathbf{q}_2\rangle}{E-\epsilon_1-E_{q_2}+i0}\langle \mathbf{K}_1\mathbf{q}_2|t_{12}(E_{12}'+i0)|\mathbf{K}_1'\mathbf{K}_2'\rangle \\
&\quad + \int d\mathbf{q}_2\,\frac{\langle \mathbf{K}_2|t_2(E-\epsilon_1+i0)|\mathbf{q}_2\rangle}{E-\epsilon_1-E_{q_2}+i0}\left\langle \mathbf{K}_1\mathbf{q}_2\left|N_{12}\left(\epsilon_{12}=E-\frac{(\mathbf{q}_2+\mathbf{K}_1)^2}{2(m_1+m_2)},E+i0\right)\right|\mathbf{K}_1'\mathbf{K}_2'\right\rangle,
\end{aligned}$$

$$\begin{aligned}
\langle \mathbf{K}_1\mathbf{K}_2|N_{1\otimes 2}(\epsilon_1,E-\epsilon_1,E+i0)|\mathbf{K}_1'\mathbf{K}_2'\rangle &= \int d\mathbf{q}_1\int d\mathbf{q}_2\,\frac{\langle \mathbf{K}_1|t_1(\epsilon_1+i0)|\mathbf{q}_1\rangle}{\epsilon_1-E_{q_1}+i0}\,\frac{\langle \mathbf{K}_2|t_2(E-\epsilon_1+i0)|\mathbf{q}_2\rangle}{E-\epsilon_1-E_{q_2}+i0} \\
&\quad \cdot\langle \mathbf{q}_1\mathbf{q}_2|t_{12}(E_1'+i0)|\mathbf{K}_1'\mathbf{K}_2'\rangle + \int d\mathbf{q}_1\int d\mathbf{q}_2\,\frac{\langle \mathbf{K}_1|t_1(\epsilon_1+i0)|\mathbf{q}_1\rangle}{\epsilon_1-E_{q_1}+i0}\,\frac{\langle \mathbf{K}_2|t_2(E-\epsilon_1+i0)|\mathbf{q}_2\rangle}{E-\epsilon_1-E_{q_2}+i0} \\
&\quad \cdot\left\langle \mathbf{q}_1\mathbf{q}_2\left|N_{12}\left(\epsilon_{12}=E-\frac{(\mathbf{q}_1+\mathbf{q}_2)^2}{2(m_1+m_2)},E+i0\right)\right|\mathbf{K}_1'\mathbf{K}_2'\right\rangle,
\end{aligned}$$

$$\begin{aligned}
\left\langle \mathbf{K}_1\mathbf{K}_2\left|N_{12}\left(\epsilon_{12}=E-\frac{(\mathbf{K}_1+\mathbf{K}_2)^2}{2(m_1+m_2)},E+i0\right)\right|\mathbf{K}_1'\mathbf{K}_2'\right\rangle &= \int d\mathbf{q}_1\left\langle \mathbf{K}_1\mathbf{K}_2\left|t_{12}\left(E-\frac{(\mathbf{K}_1+\mathbf{K}_2)^2}{2(m_1+m_2)}+i0\right)\right|\mathbf{q}_1\mathbf{K}_2'\right\rangle \\
&\quad \times\frac{\langle \mathbf{q}_1|t_1(E_1'+i0)|\mathbf{K}_1'\rangle}{E_1'-E_{q_1}+i0} + \int d\mathbf{q}_2\left\langle \mathbf{K}_1\mathbf{K}_2\left|t_{12}\left(E-\frac{(\mathbf{K}_1+\mathbf{K}_2)^2}{2(m_1+m_2)}+i0\right)\right|\mathbf{q}_2\mathbf{K}_1'\right\rangle\frac{\langle \mathbf{q}_2|t_2(E_2'+i0)|\mathbf{K}_2'\rangle}{E_2'-E_{q_2}+i0} \\
&\quad + \int d\mathbf{q}_1\,d\mathbf{q}_2\left\langle \mathbf{K}_1\mathbf{K}_2\left|t_{12}\left(E-\frac{(\mathbf{K}_1+\mathbf{K}_2)^2}{2(m_1+m_2)}+i0\right)\right|\mathbf{q}_1\mathbf{q}_2\right\rangle\frac{\langle \mathbf{q}_1|t_1(E_1'+i0)|\mathbf{K}_1'\rangle}{E_1'-E_{q_1}+i0}\cdot\frac{\langle \mathbf{q}_2|t_2(E_2'+i0)|\mathbf{K}_2'\rangle}{E_2'-E_{q_2}+i0}
\end{aligned}$$

$$+\int d\mathbf{q}_1\int d\mathbf{q}_2\left\langle \mathbf{K}_1\mathbf{K}_2\left|t_{12}\left(E-\frac{(\mathbf{K}_1+\mathbf{K}_2)^2}{2(m_1+m_2)}+i0\right)\right|\mathbf{q}_1\mathbf{q}_2\right\rangle\cdot\int_{-\infty}^{\infty}\frac{d\varepsilon_1}{-2\pi i}\frac{1}{(\varepsilon_1-E_{q_1}+i0)}\cdot\frac{1}{(E-\varepsilon_1-E_{q_2}+i0)}$$
$$\cdot\langle \mathbf{q}_1\mathbf{q}_2\,|N_1(\varepsilon_1,E+i0)|\mathbf{K}_1'\mathbf{K}_2'\rangle+\langle q_1q_2\,|N_2(E_1-\varepsilon_1,E+i0)|\mathbf{K}_1'\mathbf{K}_2'\rangle]+\langle \mathbf{q}_1\mathbf{q}_2\,|N_{1\otimes 2}(\varepsilon_1,E-\varepsilon_1,E+i0)|\mathbf{K}_1'\mathbf{K}_2'\rangle]. \tag{33}$$

4. THE PROPERTIES OF THE INTEGRAL EQUATIONS FOR THE CHANNEL AMPLITUDES

In this section we prove that the analytical continuation of the operators $T_\alpha(E+i0)$ over the energy of particles, which was perfomed by obtaining Eq. (26), has been carried out correctly. It is necessary for this proof to construct the wave function of the system.

We insert the following notations:

$$g_i(\varepsilon_1+i\tau)=[\varepsilon_1-H_{0i}+i\tau]^{-1}$$

is a Green function of the ith particle ($i=1,2$):

$$\mathscr{G}_{0i}(E-H_{0j}+i\tau)=[(E-H_{0j})-H_{0i}+i\tau]^{-1}$$

is a Green function of the particle i ($i=1,2$) in the circumstance that particle $j\neq i$ ($j=2,1$) is on the energy shell, $\varepsilon_j=H_{0j}$, and $\mathscr{G}_0(E+i\tau)=(E-H_{01}-H_{02}+i\tau)^{-1}$ is a total Green function of the system of particles 1 and 2 with the following matrix element

$$\langle \mathbf{p}_1\mathbf{p}_2|\mathscr{G}_0(E+i\tau)|\mathbf{q}_1\mathbf{q}_2\rangle=\frac{\delta(\mathbf{p}_1+\mathbf{p}_2-\mathbf{q}_1-\mathbf{q}_2)\delta(\mathbf{p}_{12}-\mathbf{q}_{12})}{E-((\mathbf{p}_1+\mathbf{p}_2)^2/2(m_1+m_2))-(\mathbf{p}_{12}^2/2\mu_{12})+i\tau}. \tag{34}$$

The operators $\mathscr{G}_{0i}(E-H_{0j}+i0)$ and $\mathscr{G}_0(E+i0)$ are inverse to the operator $(E-H_{01}-H_{02})$.

The wave function of the system is determined as

$$|\psi\rangle=\Omega_+|\varphi_1\varphi_2\rangle=U(t=0,t_0=-\infty)|\varphi_1\varphi_2\rangle,$$

where $|\varphi_1\varphi_2\rangle$ is the initial state of the system. Further, according to the adiabatic hypothesis,[5] the wave function $|\psi\rangle=|\psi(E+i0)\rangle$ is to be constructed on the basis of the following limit process:

$$|\psi(E+i0)\rangle=\lim_{\tau\to 0}|\psi(E+i\tau)\rangle. \tag{35}$$

The convergence by $\tau\to 0$ in (35) is considered in the space of the generalized functions, because the vector (35) belongs to that space. The set (21) gives the representation of the wave function

$$|\psi(E+i\tau)\rangle=|\varphi_1\varphi_1\rangle+|\psi_1(E+i\tau)\rangle+|\psi_2(E+i\tau)\rangle+|\psi_{1\otimes 2}(E+i\tau)\rangle+|\psi_{12}(E+i\tau)\rangle. \tag{36}$$

The functions $|\psi_\alpha(E+i\tau)\rangle$, where $\alpha=(1),(2),(1\otimes 2)$, $\alpha=(12)$, are defined, by using (33) by the following expressions

$$|\psi_1(E+i\tau)\rangle=g_{01}(E_1+i\tau)t_1(E_1+i\tau)|\varphi_1\varphi_2\rangle+\int_{-\infty}^{\infty}\frac{d\varepsilon_1}{-2\pi i}g_{01}(\varepsilon_1+i\tau_1)g_{02}(E-\varepsilon_1-i\tau_2)\times N_1(\varepsilon_1,E+i\tau)|\varphi_1\varphi_2\rangle,$$

$$|\psi_2(E+i\tau)\rangle=g_{02}(E_2+i\tau)t_2(E_2+i\tau)|\varphi_1\varphi_2\rangle+\int_{-\infty}^{\infty}\frac{d\varepsilon_1}{-2\pi i}g_{01}(\varepsilon_1+i\tau_1)g_{02}(E-\varepsilon_1-i\tau_2)\times N_2(E-\varepsilon_1,E+i\tau)|\varphi_1\varphi_2\rangle,$$

$$|\psi_{1\otimes 2}(E+i\tau)\rangle=g_{01}(E_1+i\tau_1)t_1(E_1+i\tau_1)g_{02}(E_2+i\tau_2)t_2(E_2+i\tau_2)|\varphi_1\varphi_2\rangle\times\int_{-\infty}^{\infty}\frac{d\varepsilon_1}{-2\pi i}g_{01}(\varepsilon_1+i\tau_1)g_{02}(E-\varepsilon_1+i\tau_2)\cdot N_{1\otimes 2}(\varepsilon_1,E-\varepsilon_1,E+i\tau)|\varphi_1\varphi_2\rangle,$$

$$|\psi_{12}(E+i\tau)\rangle=\mathscr{G}_0(E+i\tau)t_{12}(E_{12}+i\tau)|\varphi_1\varphi_1\rangle+\mathscr{G}_0(E+i\tau)N_{12}(E_{12},E+i\tau)|\varphi_1\varphi_2\rangle,\quad \tau_1+\tau_2=\tau,\quad \tau_1,\tau_2>0. \tag{37}$$

The meanings of E_1, E_1 are

$$E_1+E_2=E,\quad H_{01}|\varphi_1\rangle=E_1|\varphi_1\rangle,\quad H_{02}|\varphi_2\rangle=E_2|\varphi_2\rangle.$$

It is emphasized that in compliance with the adiabatic hypothesis, we deal first with finite values of τ and only after that calculate the limit $\tau\to+0$ of the wave vectors, as was shown in (35).

We assume that the amplitudes t_1, t_2, t_{12} have poles only at real negative values of energy.[3] Therefore, the kernels of (37) are defined at all finite values of τ. Using the standard procedure

$$\frac{1}{x-a+i\tau}=\frac{1}{i}\int_0^\infty d\alpha\exp(i(x-a+i\tau)\alpha), \tag{38}$$

given in,[5] the set (37) is reduced to the form

$$|\psi_1(E+i\tau)\rangle=\mathscr{G}_{01}(E-H_{02}+i\tau)t_1(E-H_{02}+i\tau)|\varphi_1\varphi_2\rangle+\mathscr{G}_{01}(E-H_{02}+i\tau)t_1(E-H_{02}+i\tau)\times\mathscr{G}_{01}(E-H_{02}+i\tau)T_{12}(E+i\tau)|\varphi_1\varphi_2\rangle,$$

$$|\psi_2(E+i\tau)\rangle=\mathscr{G}_{02}(E-H_{01}+i\tau)t_2(E-H_{01}+i\tau)|\varphi_1\varphi_2\rangle+\mathscr{G}_{02}(E-H_{01}+i\tau)t_2(E-H_{01}+i\tau)\times\mathscr{G}_{02}(E-H_{01}+i\tau)T_{12}(E+i\tau)|\varphi_1\varphi_2\rangle,$$

$$|\psi_{1\otimes 2}(E+i\tau)\rangle=g_{01}(E_1+i\tau_1)t_1(E_1+i\tau_1)g_{02}(E_2+i\tau_2)\times t_2(E_2+i\tau_2)|\varphi_1\varphi_2\rangle+\int_{-\infty}^{\infty}\frac{d\varepsilon_1}{-2\pi i}g_{01}(\varepsilon_1+i\tau_1)g_{02}(E-\varepsilon_1+i\tau_2)\times N_{1\otimes 2}(\varepsilon_1,E-\varepsilon_1,E+i0)|\varphi_1\varphi_2\rangle,$$

$$|\psi_{12}(E+i\tau)\rangle=\mathscr{G}_0(E+i\tau)T_{12}(E+i\tau)|\varphi_1\varphi_2\rangle. \tag{39}$$

Now we prove two theorems.

Theorem I: The wave function $|\psi(E+i0)\rangle$ defined in (36) and (39) in the space of generalized function is a solution

of the Schrödinger equation

$$(E - H_{01} - H_{02} - V_1 - V_2 - V_{12})|\psi(E+i0)\rangle = 0,$$
$$\text{if } V_i(x_i) \in \mathscr{L}^2(R^2). \tag{40}$$

Proof: Let us substitute the wave function of the system, given in (36), in the Eq. (39). We have to calculate the results of the operator $(E - H_{01} - H_{02} - V_1 - V_2 - V_{12})$ acting on the function $|\psi_\alpha(E+i0)\rangle$, where $\alpha = (1,2,(1\otimes 2),12)$. For example, for the function $|\psi_1(E+i0)\rangle$ we have

$$\begin{aligned}
&(E - H_{01} - H_{02} - V_1 - V_2 - V_{12})|\psi_1(E+i0)\rangle \\
&\quad = V_1|\varphi_1\varphi_2\rangle - V_2 g_{01}(E_1+i0)t_1(E_1+i0)|\varphi_1\varphi_2\rangle \\
&\qquad - V_{12}g_{01}(E_1+i0)\cdot t_1(E_1+i0)|\varphi_1\varphi_2\rangle \\
&\qquad + V_1\mathscr{Y}_{01}(E - H_{02}+i0)T_{12}(E+i0)|\varphi_1\varphi_2\rangle \\
&\qquad - V_2\mathscr{Y}_{0i}(E - H_{02}+i0)t_1(E - H_{02}+i0) \\
&\qquad \times \mathscr{Y}_{01}(E - H_{02}+i0)T_{12}(E+i0)|\varphi_1\varphi_2\rangle \\
&\qquad - V_{12}\mathscr{Y}_{01}(E - H_{02}+i0)t_1(E - H_{02}+i0) \\
&\qquad \times \mathscr{Y}_{01}(E - H_{02}+i0)T_{12}(E+i0)|\varphi_1\varphi_2\rangle,
\end{aligned} \tag{41}$$

without taking into account the terms which are equal to zero when $\tau \to 0$. The expressions for the other wave functions are similarly calculated. Summation of the expressions of the type (41) gives the result.

Lemma: The homogeneous set of equations corresponding to the set (39) has no nontrivial solutions at the real energy E, except for the values in the point spectrum of the Hamiltonian H. The same class of potentials is considered as for the Theorem 1.

Proof: Consider the homogeneous set of equations corresponding to the set (39)

$$\begin{aligned}
|\tilde\psi_1(E+i0)\rangle &= \int_{-\infty}^{\infty} \frac{d\varepsilon_1}{-2\pi i} g_{01}(\varepsilon_1+i0) \\
&\quad \times g_{02}(E-\varepsilon_1+i0)\widetilde{T}_1(\varepsilon_1,E+i0)|f_0\rangle, \\
|\tilde\psi_2(E+i0)\rangle &= \int_{-\infty}^{\infty} \frac{d\varepsilon_1}{-2\pi i} g_{01}(\varepsilon_1+i0) \\
&\quad \times g_{02}(E-\varepsilon_1+i0)\widetilde{T}_2(E-\varepsilon_1,E+i0)|f_0\rangle, \\
|\tilde\psi_{1\otimes 2}(E+i0)\rangle &= \int_{-\infty}^{\infty} \frac{d\varepsilon_1}{-2\pi i} g_{01}(\varepsilon_1+i0)g_{02}(E-\varepsilon_1+i0) \\
&\quad \times \widetilde{T}_{1\otimes 2}(\varepsilon_1,E-\varepsilon_1,E+i0)|f_0\rangle, \\
|\tilde\psi_{12}(E+i0)\rangle &= \mathscr{Y}_0(E+i0)\widetilde{T}_{12}(E+i0)|f_0\rangle,
\end{aligned} \tag{42}$$

where the operators $\widetilde{T}_\alpha(E+i0)$ with fixed E act on an arbitrary vector to form $\mathscr{H}_0$. Here the operators $\widetilde{T}_\alpha(E+i0)$ satisfy the homogeneous equations obtained from the Eq. (26). Applying the same procedure as when proving the Theorem I, we get

$$(E - H_{01} - H_{02} - V_1 - V_2 - V_{12})|\tilde\psi(E+i0)\rangle = 0, \tag{43}$$

where

$$|\tilde\psi(E+i0)\rangle = \sum_\alpha |\tilde\psi_\alpha(E+i0)\rangle, \quad E \notin \text{spectrum of } H,$$
$$\alpha = (1),(2),(1\otimes 2),(12). \tag{44}$$

Theorem II. If the function $|\tilde\psi(E+i0)\rangle$ defined in (44) and (42) satisfies Eq. (43), all operators $\widetilde{T}_\alpha(E+i0)$ where $\alpha = 1,2,1\otimes 2,12$, and the homogeneous set of equations for the operators $\widetilde{T}_\alpha(E+i0)$ corresponding to the set (26) have only trivial solutions at the real total energy except the values in the spectrum of the Hamiltonian H.

Proof: Substituting $|\tilde\psi_{1\otimes 2}(E+i0)\rangle$ from (42) in the equation for the operator $T_{12}(E+i0)$ and bearing in mind (43) we get

$$\widetilde{T}_{12}(E+i0)|f_0\rangle = -t_{12}(E+i0)\mathscr{Y}_0(E+i0)T_{12}(E+i0)|f_0\rangle. \tag{45}$$

Multiplying (45) to the left by $[V_{12}\mathscr{Y}_0(E+i0)]$ we get

$$\widetilde{T}_{12}(E+i0)|f_0\rangle = 0 \quad \text{and} \quad \widetilde{T}_{12}(E+i0) = 0. \tag{46}$$

Hence, all the amplitudes $\widetilde{T}_{12}(E+i0)$ where $\alpha = 1,2,1\otimes 2,12$ are zero.

5. THE CORRESPONDENCE BETWEEN THE INTEGRAL EQUATIONS OBTAINED BY THREE AND FOUR DIMENSIONAL PERTURBATION THEORY

The rearrangement of the series of stationary (three-dimensional) perturbation theory also makes it possible to obtain a set of equations for the amplitudes of two-body scattering in central field which coincide with set (31) on the energy shell. Off the energy shell, however, the scattering amplitudes found in terms of four-dimensional perturbation theory do not prove to be as singular as the amplitude found in three-dimensional perturbation theory.

Indeed, consider the Lippman–Schwinger equation for the above mentioned problem:

$$\begin{aligned}
T(Z) &= (V_1 + V_2 + V_{12}) + (V_1 + V_2 + V_{12}) \\
&\quad \times G_0(Z)T(Z), \\
Z &= E + i0.
\end{aligned} \tag{47}$$

The operator $T(Z)$ will be presented as the sum

$$T(Z) = T_{1,2}(Z) + T_{12}(Z), \tag{48}$$

where the summed operators satisfy the equations of the form

$$\begin{aligned}
T_{12}(Z) &= t_{12}(Z) + t_{12}(Z)G_0(Z)T_{1,2}(Z), \\
T_{1,2}(Z) &= N_{1,2}(Z) + N_{1,2}(Z)G_0(Z)T_{12}(Z),
\end{aligned} \tag{49}$$

and where

$$t_{12}(Z) = V_{12} + V_{12}G_0(Z)t_{12}(Z), \tag{50}$$

$$N_{1,2}(Z) = (V_1 + V_2) + (V_1 + V_2)G_0(Z)N_{1,2}(Z), \tag{51}$$

and $G(Z)$ is the free Green function of the system.

As shown in Ref. 6, Eq. (51) fails to have a unique solution; then the operator $N_{1,2}(Z)$ should be determined using the total Green function $G_{1,2}(Z)$ corresponding to the Hamiltonian of the form

$$H_{1,2} = H_{01} + H_{02} + V_1 + V_2. \tag{52}$$

Such determination of $N_{1,2}(Z)$ is given by the relations

$$N_{1,2}(Z) = (V_1 + V_2) + (V_1 + V_2)G_{1,2}(Z)(V_1 + V_2), \tag{53}$$

$$N_{1,2}(Z) = (Z - H_0)G_{1,2}(Z)(Z - H_0) - (Z - H_0). \tag{54}$$

In this case, the Green function $G_{1,2}(Z)$ should be defined as

a convolution

$$G_{1,2}(Z) = \int_{-\infty}^{\infty} \frac{d\varepsilon_1}{-2\pi i} g_1(\varepsilon_1 + i0) g_2(E - \varepsilon_1 + i0),$$
$$Z = E + i0, \tag{55}$$

where $g_i(Z_i) = (Z_i - H_{0i} - V_i)^{-1}$, whence the operator $N_{1,2}(Z)$ may be determined as

$$\begin{aligned} N_{1,2}(Z = E + i0) = {} & t_1(E - H_{02} + i0) + t_2(E - H_{01} + i0) \\ & + \int_{-\infty}^{\infty} \frac{d\varepsilon_1}{-2\pi i} [g_{01}(\varepsilon_1 + i0) \\ & + g_{02}(E - \varepsilon_1 + i0)] t_1(\varepsilon_1 + i0) \\ & \times t_2(E - \varepsilon_1 + i0)[g_{01}(\varepsilon_1 + i0) \\ & + g_{02}(E - \varepsilon_1 + i0)]. \end{aligned} \tag{56}$$

el4

The operator $t_i(Z_i)$ in (56) can be found from the equation

$$t_i(Z_i) = V_i + V_i G_0(Z) t_i(Z_i),$$
$$Z_i = Z - H_{0j}, \quad i \neq j, \quad i,j = 1,2. \tag{57}$$

Bearing in mind the representation (56) for the operator $N_{1,2}(Z)$ Eqs. (49) may be rewritten as

$$\begin{aligned} T_1(Z) &= t_1(Z - H_{02}) + t_1(Z - H_{02}) G_0(Z) T_{12}(Z), \\ T_2(Z) &= t_2(Z - H_{01}) + t_2(Z - H_{01}) G_0(Z) T_{12}(Z), \\ T_{1\otimes 2}(Z) &= \mathcal{F}_{1\otimes 2}(Z) + \mathcal{F}_{1\otimes 2}(Z) G_0(Z) T_{12}(Z), \\ T_{12}(Z) &= t_{12}(Z) + t_{12}(Z) G_0(Z) \\ &\quad \times [T_1(Z) + T_2(Z) + T_{1\otimes 2}(Z)], \quad Z = E + i0, \end{aligned} \tag{58}$$

where

$$\begin{aligned} & \mathcal{F}_{1\otimes 2}(Z = E + i0) \\ & = \int_{-\infty}^{\infty} \frac{d\varepsilon_1}{-2\pi i} [g_{01}(\varepsilon_1 + i0) + g_{02}(E - \varepsilon_1 + i0)] \\ & \quad \times t_1(\varepsilon_1 + i0) t_2(E - \varepsilon_1 + i0)[g_{01}(\varepsilon_1 + i0) \\ & \quad + g_{02}(E - \varepsilon_1 + i0)], \end{aligned} \tag{59}$$

or

$$\begin{aligned} & \mathcal{F}_{1\otimes 2}(Z = E + i0) \\ & = \int_{-\infty}^{\infty} \frac{d\varepsilon_1}{-2\pi i} (E - H_0 + i0) g_{01}(\varepsilon_1 + i0) \\ & \quad \times g_{02}(E - \varepsilon_1 + i0) \cdot t_1(\varepsilon_1 + i0) t_2(E - \varepsilon_1 + i0) \\ & \quad \cdot g_{01}(\varepsilon_1 + i0) \cdot g_{02}(E - \varepsilon_1 + i0) \cdot (E - H_0 + i0). \end{aligned} \tag{60}$$

Obviously, the structures of sets (58) and (26) prove to be the same. Besides, the integral equations for scattering amplitudes written on the basis of (58) on the energy shell coincide with (31). Off the energy shell, however, the amplitude obtained from (31) does not prove to be as singular as those obtained from (58).

ACKNOWLEDGMENT

Authors (A.M.P. and V.V.K.) are pleased to thank Prof. T. Osborn for interesting discussions.

[1]V.V. Komarov and A.M. Popova, Phys. Elem. Part. Atom. Nucl. **5**, 1075 (1974).
[2]V.V. Komarov, A.M. Popova, and V.L. Shablov, *Secondary Quantization for the Few-Particle Scattering, Theory, Part* 1. (Moscow U., 1978).
[3]L.D. Faddeev, *Mathematical Aspects of the Three-Body Problem in Quantum Scattering Thoery*, (Davy, New York, 1965).
[4]C. Chandler, Nucl. Phys. A **301**, 1 (1978).
[5]N.I. Bogalyubov and D.V. Shirkov, *Introduction into Theory of Quantum Fields* (Nauka, Moscow, 1973).
[6]V.V. Komarov, A.M. Popova, and V.L. Shablov, Lett. Nuovo Cimento **20**, 649 (1977).

Null infinity and Killing fields

Abhay Ashtekar
Département de Physique, Université de Clermont, 63170 Aubière, France
B. G. Schmidt
Max-Planck Institut für Physik und Astrophysik, 8046 Garching bei München FRG

(Received 15 August 1979; accepted for publication 7 November 1979)

A recent investigation due to Ashtekar and Xanthopoulos of isometries compatible with asymptotic flatness at null infinity is extended by considering space–times which admit "only a piece of $\mathscr{I}$". This extension makes the interplay between asymptotic conditions and permissible isometries more transparent by bringing out how the various parts of the asymptotic structure individually lead to restrictions on isometries and by separating the consequences of field equations from those of purely geometric asymptotic conditions.

I. INTRODUCTION

There is available[1] in literature a detailed analysis of isometries admitted by space–times which satisfy certain asymptotic conditions at null infinity. These conditions require, in particular, that the null boundary $\mathscr{I}$ of the given space–time be topologically $S^2 \times R$ and that the generators of $\mathscr{I}$ be complete.[2] The purpose of this note is to extend this work by weakening the asymptotic requirements. More specifically, we shall proceed in the following steps: we will begin by examining space–times which admit only a "piece of $\mathscr{I}$," i.e., space–times for which $\mathscr{I}$ is neither topologically $S^2 \times R$ nor with complete generators, then focus on the case when $\mathscr{I}$ is $S^2 \times R$ but with generators which are not necessarily complete; and, finally, consider space–times with a "global" $\mathscr{I}$ as in Ref. 1. Such a step by step analysis serves two purposes. First, it makes the interplay between the structure of the permissible isometry groups and the imposed asymptotic conditions more transparent: one can now see, more clearly, how constraints on possible Killing fields arise from various individual pieces of the asymptotic structure. Second, the generality gained over Ref. 1 is useful from a rather practical viewpoint as well: Models of isolated systems with only a "piece of $\mathscr{I}$" will be substantially easier to construct than those admitting a global $\mathscr{I}$.[3] The present analysis may be useful in the investigation of global properties of known models as well as in the construction of new models of this type.

In Sec. 2, we consider space–times which admit only a "local $\mathscr{I}$," and , in Sec. 3, those with a $\mathscr{I}$ which is topologically $S^2 \times R$. It turns out that in spite of their apparent weakness, the asymptotic conditions involved do impose rather severe restrictions on the structure of isometry Lie algebras. Section 4 summarizes these results as well as those of Ref. 1, thereby providing a concise picture of the whole situation. For space–times with a global $\mathscr{I}$, the only essential gap in the analysis of Ref. 1 is the lack of an argument showing that if the asymptotic curvature does not fall off "too rapidly," a translational Killing field, if present, is necessarily timelike near $\mathscr{I}$. This gap is filled by the present investigation.

Throughout this paper, we use the same notation as in Ref. 1. However, unlike in Ref. 1, we do not include the validity of Einstein's vacuum equation near $\mathscr{I}$ as a condition in the definition of asymptotic flatness. Rather, at each stage in the analysis we explicityly impose additional conditions—whose satisfaction was ensured in Ref. 1 via Einstein's equation—which are directly required to obtain the desired results. Although this procedure makes the statements of some results more complicated, it has the advantage of separating the consequences of purely geometric asymptotic conditions from those of field equations. We hope that the separation of these two issues will be useful, e.g., in the construction of counterexamples to conjectures concerning isolated systems.

2. LOCAL $\mathscr{I}$

Consider a smooth (i.e., say c^∞) manifold $\hat{\mathbf{M}}$ without boundary equipped with a smooth metric $\hat{g}_{ab}$ of signature $(-+++)$.

Definition 1: $(\hat{\mathbf{M}}, \hat{g}_{ab})$ will be said to admit an *asymptote* provided there exists a manifold $\mathbf{M}$ with boundary $\mathscr{I}$, together with a smooth Lorentz metric g_{ab}, a smooth function Ω and a diffeomorphism from $\hat{\mathbf{M}}$ on to $\mathbf{M}$ (with which we identify $\hat{\mathbf{M}}$ with its image in $\mathbf{M}$) such that:

(i) on $\hat{\mathbf{M}}$, $g_{ab} = \Omega^2 \hat{g}_{ab}$,

(ii) on $\mathscr{I}$, $\Omega = 0$, $\nabla_a \Omega \neq 0$, $\nabla_a \Omega \, \nabla^a \Omega = 0$, and $\nabla_a \nabla_b \Omega = 0$, where ∇ is the derivative operator on $(\mathbf{M}, g_{ab})$.

$(\mathbf{M}, g_{ab})$ will be referred to as the asymptote of $(\hat{\mathbf{M}}, \hat{g}_{ab})$.

Remarks: (1) These asymptotic conditions are quite weak. In particular, there exist several physically interesting *nonstationary* space–times which admit an asymptote: examples are some of the Robinson–Trautman[4] space–times, the c metric[5] and Vaidya solutions.[6]

(2) The notation of an asymptote expressed by Definition 1 is somewhat stronger than that used by Geroch[7]: The requirement $\nabla_a \nabla_b \Omega = 0$ on $\mathscr{I}$ is omitted in Geroch's definition. In particular, although Friedmann space–times with flat 3-surfaces admits an asymptote in the sense of Geroch, it does not, apparently, admit one in the sense of Definition 1.[8]

(3) If Einsteins's vacuum equation holds in $\mathbf{N} \cap \hat{\mathbf{M}}$ for some neighborhood $\mathbf{N}$ of $\mathscr{I}$ in $\mathbf{M}$—or, more generally, if the stress–energy T_{ab} falls off in such a way that ΩT_{ab} admits a limit to $\mathscr{I}$—then the last two equations in condition (ii) are superfluous: If the other requirements of Definition 1 are satisfied, one can always find a conformal factor satisfying

 0022-2488/80/040862-06$1.00

$\nabla_a\Omega\ \nabla^a\Omega = 0$ and $\nabla_a\nabla_b\Omega = 0$ on $\mathscr{I}$

(4) For the class of space–times satisfying this definition, the BMS group need *not* be the group of asymptotic symmetries. Furthermore, the group of asymptotic symmetries can change from one space–time to another in this class.

Denote by $\mathscr{L}$ the Lie algebra of the isometry group of $(\hat{\mathbf{M}},\hat{g}_{ab})$. To investigate the structure of $\mathscr{L}$, we use, as in Ref. 1, the notion of conformal Killing data and the technique of conformal Killing transport. Let $\hat{\xi}^a$ denote a Killing field on $(\hat{\mathbf{M}},\hat{g}_{ab})$ and let the quadruplet $(\xi^a, F_{ab}, \phi, k_a)$ denote its conformal Killing data w.r.t. g_{ab}. (Recall: $F_{ab} := \nabla_{[a}\xi_{b]}$, $\phi := \frac{1}{4}\nabla_a\xi^a$, $k_a := \nabla_a\phi$. For details, see Refs. 1, and 9.) Then, using the same arguments as on page 2218 of Ref. 1 we have:

Lemma 1.1: $\hat{\xi}^a$ admits a smooth extension ξ^a to $\mathscr{I}$ which is a conformal Killing field on $(\mathbf{M},g_{ab})$.

Lemma 1.2: The conformal Killing data $(\xi^a, F_{ab}, \phi, k_a)$ at any point of $\mathscr{I}$ is tangential to $\mathscr{I}$. That is, $\xi^a n_a = 0, F_{ab}n^b = 0, K_a n^a = 0$ at any point of $\mathscr{I}$, where $n^a \equiv \nabla^a\Omega$.

Lemma 1.3: If ξ^a vanishes in an open neighborhood, *within* $\mathscr{I}$, of any point of $\mathscr{I}$, then ξ^a must vanish identically on $\mathbf{M}$.

These Lemmas assure us that we can investigate the structure of $\mathscr{L}$ by examining the behavior of various Killing fields at points of $\mathscr{I}$. We begin by introducing some notation. A Killing field $\hat{\xi}^a$ on $(\hat{\mathbf{M}},\hat{g}_{ab})$ will be said to be a *translational Killing field*, if the restriction of ξ^a to $\mathscr{I}$ is of the form $\xi^a|_{\mathscr{I}} = \alpha n^a \equiv \alpha\nabla^a\Omega$ for some function α. (Lemma 1.2 implies that α can have only isolated zeros.) What does the conformal Killing data of translations look like at a point of $\mathscr{I}$? It is easy to check that the data must satisfy the following conditions: $F_{ab}|_{\mathscr{I}} = v_{[a}n_{b]}$, $\phi|_{\mathscr{I}} = 0$, and $k_a|_{\mathscr{I}} = \beta n_a$, where v_a is a vector field on $\mathscr{I}$ satisfying $v_a n^a = 0$ and $\mathscr{L}_n v^a = 0$, and β is a scalar field on $\mathscr{I}$. Denote by τ the space of translational Killing fields. Then, from Lemma 1.3 and the form of the conformal Killing data of translations, it follows that dim $\tau \leqslant 4$, (where "dim" denotes dimension). Next, a simple calculation gives $\mathscr{L}_\xi n^a|_{\mathscr{I}} = -\phi n^a|_{\mathscr{I}}$ for *any* Killing field $\hat{\xi}$ on $(\hat{\mathbf{M}},\hat{g}_{ab})$. Hence $\mathscr{L}_\xi \alpha n^a|_{\mathscr{I}} = (\mathscr{L}_\xi\alpha - \alpha\phi)n^a|_{\mathscr{I}}$. Therefore, we have the following result:

Theorem 1: The space τ of translational Killing fields is an Abelian ideal of the isometry Lie algebra $\mathscr{L}$. Furthermore, dim $\tau \leqslant 4$ and dim $\mathscr{L}/\tau \leqslant 6$.

Remark: Denote by $\mathscr{S}$ the space of generators of $\mathscr{I}$ and by h_{ab} the natural metric thereon. [Since $\mathscr{L}_n g_{ab} \equiv 2\nabla_{(a}n_{b)}$ vanishes on $\mathscr{I}$, $\mathscr{S}$ inherits a natural metric. Since $\Omega = 0$ on $\mathscr{I}$ and since $n^a \equiv \nabla^a\Omega$ is nonzero and null there, h_{ab} has signature $(+\ +)$.] It is easy to show that each element of $\mathscr{L}/\tau$ induces a conformal isometry on $(\mathscr{S}, h_{ab})$. However, since $\mathscr{S}$ is two-dimensional and not necessarily compact, conformal isometries of $(\mathscr{S}, h_{ab})$ need not form a Lie group. Hence, for the class of space-times under consideration, the structure of the Lie algebra $\mathscr{L}/\tau$ can be quite arbitrary.

Next, we wish to restrict the number of permissible Killing fields by imposing suitable conditions on the fall-off of the Weyl curvature of $(\hat{\mathbf{M}},\hat{g}_{ab})$. For this purpose, we introduce a tensor field K_{abcd} which will be called the *asymptotic Weyl curvature*. Let C_{abcd} denote the Weyl tensor of g_{ab}. If C_{abcd} does not vanish identically on $\mathscr{I}$, set $K_{abcd} = C_{abcd}$. (Note: since $\mathscr{I}$ is not required to be topologically $S^2\times R$, and since the field equations have not been imposed, the usual argument[10] showing $C_{abcd}|_{\mathscr{I}} = 0$ does not go through.) If C_{abcd} does vanish identically on $\mathscr{I}$, $\Omega^{-1}C_{abcd}$ admits a smooth limit to $\mathscr{I}$.[11] If this limit itself does not vanish everywhere on $\mathscr{I}$, we set $K_{abcd} = \Omega^{-1}C_{abcd}$. If $\lim_{\to\mathscr{I}} \Omega^{-1}C_{abcd} = 0$, we consider $\Omega^{-2}C_{abcd}$ and so on. Thus, unless C_{abcd} and *all* its derivatives vanish identically on $\mathscr{I}$, there exists a tensor field K_{abcd}, representing the asymptotic curvature, which fails to vanish everywhere on $\mathscr{I}$. *Now onwards we assume that, for some (nonnegative) integer n,* $\lim_{\to\mathscr{I}} \Omega^{-n+1}C_{abcd}$ *vanishes on* $\mathscr{I}$ *but* $\lim_{\to\mathscr{I}} \Omega^{-n}C_{abcd}$ *fails to vanish identically there and set* $K_{abcd} = \Omega^{-n}C_{abcd}$. (By $K_{abcd}|_{\mathscr{I}}$ we shall always mean $\lim_{\to\mathscr{I}} \Omega^{-n}C_{abcd}$.) Note that this assumption is substantially weaker than that of Ref. 1: There, n was required to be 1. The intuitive idea behind this assumption is simply to exclude conformally flat space–times satisfying Definition 1 from the rest.

Fix a point p of $\mathscr{I}$ at which K_{abcd} fails to vanish. Since K_{abcd} is a smooth tensor field, there exists a neighborhood $\mathbf{N}_p$ of $\mathscr{I}$ such that K_{abcd} is nonzero everywhere in $\mathbf{N}_p$. Then, using the same arguments as in Lemma 2.1 of Ref. 1 with the trivial modifications in various numerical factors required by the fact that K_{abcd} equals $\Omega^{-n}C_{abcd}$, rather than $\Omega^{-1}C_{abcd}$ as in Ref. 1, we obtain:

Lemma 2.1: There exists a curvature scalar f on $\mathscr{I}$ and a point p' in $\mathbf{N}_p$ such that $f|_{p'} \neq 0$, and for *any* Killing field $\hat{\xi}^a$ on $(\hat{\mathbf{M}},\hat{g}_{ab})$, $\xi^m\nabla_m f = -(2n+4)\phi f$ in $\mathbf{N}_p$, where, as usual, $\phi \equiv \frac{1}{4}\nabla_m\xi^m$ is the third piece of the conformal Killing data of ξ^a w.r.t. g_{ab}.

Thus, the presence of the asymptotic Weyl curvature constrains the conformal Killing data of ξ^a at p': $\xi^a|_{p'}$ completely determines the value of $\phi|_{p'}$. Furthermore, since $\xi^m\nabla_m f = -(2n+4)\phi f$ *everywhere* in $\mathbf{N}_p$, the fourth piece, $k_a|_{p'}$, of the conformal Killing data is also constrained. We therefore have the following result: The conformal Killing data $(\xi^a, F_{ab}, \phi, k_a)$ at p' is one of the type

$$(\xi^a|_{p'}, F_{ab}|_{p'}, -(2n+4)^{-1}\xi^m(\nabla_m \log f)|_{p'},$$
$$-(2n+4)^{-1}[\xi^m\nabla_m\nabla_a \log f + F_a^m\nabla_m \log f$$
$$-(2n+4)^{-1}\xi^m(\nabla_m \log f)(\nabla_a \log f)]_{p'} + \gamma n^a|_{p'}),$$

for some real number γ. Thus, the presence of the asymptotic Weyl curvature implies that values of ξ^a and F_{ab} at p' themselves determine $\phi|_{p'}$ and $k^a|_{p'}$ up to a multiple of n^a. Using the fact that at any point of $\mathscr{I}$, every translational Killing field ξ^a has the data of the form $(\alpha n^a, n_{[a}v_{b]}, 0, \beta n^a)$, with $v^a n_a = 0$ and $\mathscr{L}_n v^a = 0$, we have:

Lemma 2.2: In the terminology of Theorem 1, dim $\mathscr{L}/\tau$ can not exceed three.

Thus, from Theorem 1 and Lemma 2.2 we know that the given space–time cannot admit more than seven Killing fields. A strong restriction is obtained by a further, apparently mild assumption. *Let us suppose that the asymptotic Weyl curvature* K_{abcd} *is such that* $K_{abcd}n^d$ *fails to vanish identically on* $\mathscr{I}$. Geometrically, this assumption requires that there is a point p of $\mathscr{I}$ at which n^a is not a fourfold degenerate princi-

pal null direction of K_{abcd} (For a physical motivation of this condition, See Ref. 1, page 2219). Now using the same arguments as in Lemma 2.3 of Ref. 1 we have:

Lemma 2.3: If $K_{abcd}n^d$ does not vanish identically on $\mathscr{I}$, dim $\tau \leqslant 1$.

Remark: There is an alternate, more geometric argument which makes the result of Lemma 2.3 more transparent. This will be given in Sec. 3 (after Lemma 3.6).

As a direct consequence of Lemmas 2.2 and 2.3 we have:

Theorem 2: If the asymptotic Weyl curvature K_{abcd} is such that $K_{abcd}n^d$ does not vanish identically on $\mathscr{I}$, the dimension of the isometry Lie algebra $\mathscr{L}$ of $(\mathbf{M},\hat{g}_{ab})$ cannot exceed four. Furthermore, dimension of the Lie ideal τ (of translations) of $\mathscr{L}$ cannot exceed one and that of $\mathscr{L}/\tau$ cannot exceed three.

Remark: Again, since $\mathscr{S}$, the space of generators of $\mathscr{I}$, is not necessarily a compact 2-manifold, we cannot make any statements regarding the structure of $\mathscr{L}/\tau$. In particular, $\mathscr{L}$ need *not* be a sub-Lie algebra of the Poincaré algebra! Note also that dim $\mathscr{L} = 4$ does not imply that $\mathscr{L}$ is isomorphic with the Lie algebra of Killing fields in the Schwarzschild space–time. Indeed, even if $\hat{g}_{ab}$ is required, in addition, to satisfy the vacuum equation near $\mathscr{I}$, $\mathscr{L}$ can differ from Schwarzschild Lie algebra: The Levi-Civita b metric[12] offers an example.

3. $\mathscr{I}$ WITH TOPOLOGY $S^2 \times R$

In this section we consider space–times satisfying stronger asymptotic requirements: We assume $(\hat{\mathbf{M}},\hat{g}_{ab})$ admits an asymptote $(\mathbf{M},g_{ab})$ such that the null boundary $\mathscr{I}$ is topologically $S^2 \times R$. Note, however, that we do *not* require the generators of $\mathscr{I}$ to be complete nor, in the first part of this section, $\hat{g}_{ab}$ to satisfy Einstein's vacuum equation to any asymptotic order. One expects there to exist a large class of space–times—not necessarily stationary—which admits an asymptote in this stronger sense. Indeed, analysis[5] in the linearized approximation strongly suggests that the c metric, which is a *vacuum space–time with gravitational radiation* belongs to this class.

Since $\mathscr{I}$ is topologically $S^2 \times R$, the Lie algebra of infinitesimal asymptotic symmetries is the BMS Lie algebra.[13] [We cannot yet consider the *group* of (finite) asymptotic symmetries since generators of $\mathscr{I}$ need not be complete.] Hence, it follows that $\mathscr{L}$ is a sub-Lie algebra of the BMS Lie algebra. (Lemma 1.1, of Ref. 1.) Furthermore, since the space $\mathscr{S}$ of generators of $\mathscr{I}$ is a 2-sphere and since the group of conformal isometries of a 2-sphere is the Lorentz group, it follows that $\mathscr{L}/\tau$ is a sub-Lie algebra of the Lorentz Lie algebra. This fact enables us to investigate the structure of permissible isometry Lie algebras.

From now onwards, we assume that the asymptotic curvature K_{abcd} is such that $K_{abcd}n^d$ does not vanish identically on $\mathscr{I}$. Then, from Theorem 2, we know that dim $\mathscr{L} \leqslant 4$, dim $\tau \leqslant 1$, dim $\mathscr{L}/\tau \leqslant 3$. We now consider the various possible cases.

Lemma 3.1: If dim $\mathscr{L} = 4$, τ generates a BMS time-translation on $\mathscr{I}$ and $\mathscr{L}$ is isomorphic with the Lie algebra of isometries on the Schwarzschild space–time. Furthermore $(\hat{\mathbf{M}},\hat{g}_{ab})$ is spherically symmetric in $\mathbf{N} \cap \hat{\mathbf{M}}$ for some neighborhood $\mathbf{N}$ of $\mathscr{I}$ in $\mathbf{M}$.

Proof: Since dim $\mathscr{L} = 4$, τ is necessarily one dimensional and $\mathscr{L}/\tau$ three dimensional. Let αn^a be the restriction to $\mathscr{I}$ of the translational Killing field. Then, the curvature scalar f of Lemma 2.1 satisfies $\alpha n^a \nabla_a f = 0$ on $\mathscr{I}$. Since, by Lemma 1.3, α cannot vanish in an open set within $\mathscr{I}$ and since f is smooth, f is a lift to $\mathscr{I}$ of a smooth function $\tilde{f}$ on the space $\mathscr{S}$ of generators of $\mathscr{I}$. Furthermore, since *every* Killing field $\hat{\xi}^a$ satisfies $\mathscr{L}_{\hat{\xi}} f = -(2n+4)\phi f$ on $\mathscr{I}$, with ϕ constant on each generator, each element of $\mathscr{L}/\tau$ gives rise to a conformal Killing field $\tilde{\xi}^a$ on $(\mathscr{S},h_{ab})$ satisfying $\mathscr{L}_{\tilde{\xi}} \tilde{f} = -(2n+4)\tilde{\phi}\tilde{f}$, where $\tilde{\phi}$ is the projection to $\mathscr{S}$ of ϕ on $\mathscr{I}$ and h_{ab} is the natural metric on $\mathscr{S}$ induced by the metric g_{ab} on $\mathscr{I}$. It is this last equation that constrains the Lie algebra structure of $\mathscr{L}/\tau$.

Since $\mathscr{L}/\tau$ is a sub-Lie algebra of the Lorentz Lie algebra, (i.e., of the quotient of the BMS Lie algebra by supertranslations), it follows that each element of $\mathscr{L}/\tau$ can be expressed as a 2-form on the Minkowskian vector space of BMS translations. Furthermore, one can always choose three basis vectors in $\mathscr{L}/\tau$ which are represented by *simple* 2-forms.[14] A simple 2-form represents either a spatial rotation or a boost or a null rotation. Let us now suppose that $\mathscr{L}/\tau$ is not the Lie algebra of SO(3). Then, one can always find an element of $\mathscr{L}/\tau$ which generates a rotation in a time-like 2-plane (i.e., a boost) or a null 2-plane.

Let us suppose that there is an element of $\mathscr{L}/\tau$ which generates a rotation in the "tz plane." Then, in the Bondi frame singled out by "the t axis," we have $\tilde{\xi}^a = -\sin\theta\partial/\partial\theta$ and $\tilde{\phi} = -\cos\theta$ whence the only regular solution to $\tilde{\xi}^a D_a \tilde{f} = -(2n+4)\tilde{\phi}\tilde{f}$ is $\tilde{f} = 0$, which contradicts the assumption. (Here, θ is the standard 2-sphere polar coordinate.) A similar argument shows that no element of $\mathscr{L}/\tau$ can generate a null rotation. Hence $\mathscr{L}/\tau$ must be the Lie algebra of SO(3).

Since τ is a Lie ideal of $\mathscr{L}$, it follows that the translation represented by τ is a *time-translation* (i.e., α is nowhere zero on $\mathscr{S}$). Consider the Bondi (conformal) frame in which $\mathscr{L}/\tau$ is the isometry Lie algebra of $(\mathscr{S}, h_{ab})$. [Such a frame must exist since $\mathscr{L}/\tau$ is isomorphic to the Lie algebra of SO(3).] In this frame, each Killing field $\tilde{\xi}^a$ on $(\mathscr{S}, h_{ab})$ must satisfy $\mathscr{L}_{\tilde{\xi}}\alpha = \kappa\alpha$, for some constant κ, since τ is a Lie ideal of $\mathscr{L}$. Choose a basis in $\mathscr{L}/\tau$ such that the corresponding $\tilde{\xi}$'s have closed orbits on $\mathscr{S}$. Then, since α is regular, the corresponding κ's must vanish, i.e., each of these $\tilde{\xi}$'s must satisfy $\mathscr{L}_{\tilde{\xi}}\alpha = 0$ on $\mathscr{S}$, and whence, $\mathscr{L}_{\tilde{\xi}}\alpha n^a = 0$ on $\mathscr{I}$. *Thus*, τ generates a BMS time translation which commutes with every Killing field in $\mathscr{L}$. Therefore, $\mathscr{L}$ is isomorphic with the Lie algebra of isometries in the Schwarzschild space–time. Since the orbits on $\mathscr{I}$ of $\mathscr{L}/\tau$ are 2-spheres, they must continue to be 2-spheres in a neighborhood. That is $(\hat{\mathbf{M}},\hat{g}_{ab})$ is static and spherically symmetric and in $\mathbf{N} \cap \hat{\mathbf{M}}$ for some neighborhood $\mathbf{N}$ of $\mathscr{I}$. □

Remark: Note that we have not imposed field equations at this stage. Hence $(\hat{\mathbf{M}},\hat{g}_{ab})$ need not be isometric to the Schwarzschild space–time in $\mathbf{N} \cap \hat{\mathbf{M}}$. Note also that, *a priori*,

we only knew that τ represents a BMS supertranslation. That it is a BMS *translation* is a consequence of the lemma.

Next, we consider the case when dim $\mathscr{L} = 3$. We have:

Lemma 3.2: If dim $\mathscr{L} = 3$, then dim $\tau = 0$ and hence $\mathscr{L}$ is a sub Lie algebra of the Lorentz Lie algebra.

Proof: Let us suppose dim $\tau = 1$, then dim $\mathscr{L}/\tau = 2$. Again, we can introduce a basis in $\mathscr{L}/\tau$ such that each basis vector represents a rotation in a spacelike, timelike, or null 2-plane in the Minkowskian vector space of BMS translations. Again, as in Lemma 3.2, the presence of a nonzero curvature scalar $\tilde{f}$ rules out timelike and null rotations. This is impossible since SO(3) does not admit a two-dimensional sub-Lie algebra. Hence, τ must be zero dimensional. □

Unfortunately, it is not easy to impose further restrictions on $\mathscr{L}$. The essential problem is that if dim $\tau = 0, f$ cannot, in general, be projected unambiguously on $\mathscr{S}$ and we must carry out the analysis on the orbits of the Killing fields on $\mathscr{I}$ itself. However, since the generators of $\mathscr{I}$ need not be complete, these orbits need not leave *any* 2-sphere cross section of $\mathscr{I}$ invariant. Hence, we need to impose additional conditions. We have:

Lemma 3.3: If there exists a cross section S of $\mathscr{I}$ which is left invariant by the action of $\mathscr{L}$, and if the curvature scalar f of Lemma 2.1 is not identically zero on S, then $\mathscr{L}$ is isomorphic to the Lie algebra of SO(3) and $(\hat{\mathbf{M}},\hat{g}_{ab})$ is spherically symmetric in $\mathbf{N}\cap\hat{\mathbf{M}}$ for some neighborhood $\mathbf{N}$ of $\mathscr{I}$ in $\mathbf{M}$.

Proof: Replace $\mathscr{S}$ by S in the proof of Lemma 3.1. □

Next, we consider the case dim $\mathscr{L} = 2$. We have:

Lemma 3.4: If dim $\mathscr{L} = 2$ and dim $\tau = 0$, there is no 2-sphere cross section S of $\mathscr{I}$ which is left invariant by $\mathscr{L}$ and on which the curvature scalar f does not vanish identically.

Proof: Replace $\mathscr{S}$ by S in the argument used in Lemma 3.2. □

Remark: The c metric does admit two (commuting) Killing fields neither of which is a translation. Thus, in this case, dim $\mathscr{L} = 2$, dim $\tau = 0$ and $\mathscr{L}$ belongs to a Lorentz sub-Lie algebra of the BMS group. If the generators of $\mathscr{I}$ are complete, every Lorentz subgroup of the BMS group leaves a cross section of $\mathscr{I}$ invariant. In the case of the linearized c metric, however, the generators are incomplete and the singularity structure is such that precisely that cross section of $\mathscr{I}$ "which would have been left invariant" by isometries (and the future or past of that cross section within $\mathscr{I}$) is missing. Lemma 3.4 implies that for the exact c metric, the situation must be similar. As one might expect, if the generators of $\mathscr{I}$ are complete and if the vacuum equation holds near $\mathscr{I}$, the cross section S of Lemmas 3.3 and 3.4 always exists. Hence, in this case, dim $\mathscr{L} = 3$ implies spherical symmetry and dim $\mathscr{L} = 2$ implies dim $\tau = 1$.

Next, let us suppose that dim $\tau = 1$. Then, we have the following lemma.

Lemma 3.5: If dim $\mathscr{L} = 2$ and dim $\tau = 1$, then $\mathscr{L}/\tau$ generates a SO(2) subgroup of the Lorentz group.

Proof: Since dim $\tau = 1$, the curvature scalar f has an unambiguous projection $\tilde{f}$ on the space $\mathscr{S}$ of generators of $\mathscr{I}$. Next, since $\mathscr{L}/\tau$ generates a (one-dimensional) subgroup of the Lorentz group, it determines, canonically, a 2-form in the Minkowskian vector space of BMS translations. If this two form is simple, by the argument in Lemma 3.1 $\mathscr{L}/\tau$ must generate a spatial rotation, i.e., a SO(2) subgroup of the Lorentz group. Let us suppose it is not simple. Then it must be[15] of the form $\lambda t_{[a}Z_{b]} + \mu X_{[a}Y_{b]}$ for some orthonormal tetrad (t_a timelike) in the space of BMS translations, with λ and μ some real constant. Let $\tilde{\xi}^a$ denote the conformal Killing field on $(\mathscr{S}, h_{ab})$ induced by $\mathscr{L}/\tau$. This $\tilde{\xi}^a$ satisfies $\mathscr{L}_{\tilde{\xi}}\tilde{f} = -(2n+4)\tilde{\phi}\tilde{f}$. In the Bondi frame singled out by the BMS time translation t^a, this equation becomes $-\lambda \sin\theta\, \partial f/\partial\theta + \mu\partial f/\partial\phi = (2n+4)\lambda\cos\theta f$, where θ and ϕ are the usual polar coordinates on $\mathscr{S}$. It is easy to show, e.g., by multiplying both sides by $f\cos\theta$, that the only everywhere regular solution to this equation is $f = 0$ on $\mathscr{S}$, which contradicts the initial assumption. Hence the 2-form representing $\mathscr{L}/\tau$ must be simple, whence $\mathscr{L}/\tau$ generates a SO(2) subgroup of the Lorentz group. □

Finally, we introduce the field equations: We restrict ourselves to space–times admitting asymptotes with $\mathscr{I} \approx S^2 \times R$ and which satisfy $\hat{R}_{ab} = 0$ in $\hat{\mathbf{M}}\cap\mathbf{N}$ for some neighborhood $\mathbf{N}$ of $\mathscr{I}$ in $\mathbf{M}$, where $\hat{R}_{ab}$ is the Ricci tensor of $\hat{g}_{ab}$. (Our discussion will also go through under the following weaker but somewhat more complicated condition: If $K_{abcd} = \Omega^{-n}C_{abcd}$ is the asymptotic Weyl curvature which admits smooth limits to $\mathscr{I}$, then $\Omega^{-n}R_{ab}$ must vanish in the limit to $\mathscr{I}$.) The vacuum equation gives rise to a number of simplifications. First, as already remarked, the condition that $\nabla^a\nabla_a\Omega\,|_{\mathscr{I}} = 0$ and $\nabla_a\nabla_b\Omega\,|_{\mathscr{I}} = 0$ of the definition are automatically satisfied. Second, the Weyl curvature C_{abcd} vanishes everywhere on $\mathscr{I}$.[10] Hence, the integer n appearing in the definition of K_{abcd} is strictly positive. Finally, from Lemma 1.4 of Ref. 1, one has:

Lemma 3.6: Let $K_{abcd} = \Omega^{-1}C_{abcd}$. Then, if $\hat{\xi}^a$ is a translational Killing field, i.e., if $\xi^a|_{\mathscr{I}} = \alpha n^a$, then $\xi^a|_{\mathscr{I}}$ is necessarily a BMS translation. Furthermore, in this case, the Bondi news (and hence also the radiation field) vanishes identically on $\mathscr{I}$.

Let us now assume that $K_{abcd}n^d$ does not vanish identically on $\mathscr{I}$ and explore the consequences of this assumption. A simple calculation shows that $\nabla_d\Omega^{-1}C_{abc}{}^d = 0$ in $\mathbf{N}$, whence, using regularity of K_{abcd} on $\mathscr{I}$, it follows that, if $n > 1$, $K_{abcd}n^d = 0$ on $\mathscr{I}$. Thus, for the class of space–times now under consideration, the condition that $K_{abcd}n^d$ should not vanish identically on $\mathscr{I}$ itself implies that $K_{abcd} = \Omega^{-1}C_{abcd}$ as in Ref. 1. For this class, we can now show that if there exists a translational Killing field, it is necessarily timelike in a neighborhood of $\mathscr{I}$.

Lemma 3.7: If $\hat{\xi}^a$ is a Killing field on $(\hat{\mathbf{M}},\hat{g}_{ab})$ with $\xi^a|_{\mathscr{I}} = \alpha n^a$, α is nowhere zero on $\mathscr{I}$.

Proof: Fix a point p of $\mathscr{I}$ at which $K_{abcd}n^d$ fails to vanish. Let us suppose that $\alpha|_p = 0$. Then the isometry generated by ξ^a leaves p fixed and induces a rotation in the tangent space T_p at p. Since $\mathscr{L}_\xi K_{abcd}|_{\mathscr{I}} = 0$ and since the rotation in T_p is a continuous one-parameter group of transformations, it must leave each principal null direction of $K_{abcd}|_p$ invariant. Now, since $K_{abcd}n^d|_p \neq 0$, $K_{abcd}|_p$ must admit at least one principal null direction which is *different* from n^a. However, since $\nabla_a\xi_b|_p = V_{[a}n_{b]}|_p$ for some vector V_a satisfying $V_an^a = 0$, the rotation generated by ξ^a in

TABLE I.

asymmetric conditions \ dim G	4	3	2	1
$\mathscr{I} \approx S^2 \times R$	Static spherically symmetric in $\mathbf{N} \cap \hat{\mathbf{M}}$	$G \subset L$ (radiation possible Example: Vaidya metric)	If dim $N = 1$, N need not even be a BMS translation	If dim $N = 1$, N need not even be a BMS translation
$\mathscr{I} \approx S^2 \times R$ $\hat{R}_{ab} = 0$ in $\mathbf{N} \cap \hat{\mathbf{M}}$	Isometric to Schwarzschild metric in $\mathbf{N} \cap \hat{\mathbf{M}}$	$G \subset L$ (radiation possible Example:?)	If dim $N = 1$, stationary axisymmetric (If dim $N = 0$, radiation possible. Example c metric)	If dim $N = 1$, stationary
$\mathscr{I} \approx S^2 \times R$ Generators complete $\hat{R}_{ab} = 0$ in $\mathbf{N} \cap \hat{\mathbf{M}}$	Isometric to Schwarzschild metric in $\mathbf{N} \cap \hat{\mathbf{M}}$	Isometric to Schwarzschild metric in $\mathbf{N} \cap \hat{\mathbf{M}}$	Stationary Axisymmetric[17]	If dim $N = 1$, stationary If dim $N = 0$, axisymmetric

T_p cannot leave invariant any null direction other than n^a unless $V_a \mid_p = 0$, i.e., unless $\nabla_a \xi_b \mid_p = 0$. Thus, if $\xi^a \mid_p = 0$, then $\nabla_a \xi_b \mid_p = 0$.

Let us suppose this is the case. Then the conformal Killing data of ξ^a at p are of the type $(0,0,0\beta n^a)$; ξ^a is a null translation. ξ^a satisfies, everywhere on $\mathbf{M}$, $\mathscr{L}_\xi K_{abcd} = \phi K_{abcd}$. By dividing the equation by Ω, taking the limit to p along any smooth curve in $\mathbf{M}$, and using the fact that $\xi^a \mid_p = 0$, and $\nabla_a \xi_b \mid_p = 0$, one obtains $K_{abcd} \mid_p = 0$, which contradicts the initial assumption. Thus, if $K_{abcd}\, n^d \mid_p \neq 0$, $\alpha \mid_p$ cannot vanish.

Finally, using Bianchi identities on K_{abcd}, one can show[16] that if there exists a point p of $\mathscr{I}$ with $K_{abcd}\, n^d{}_p \neq 0$, then $K_{abcd}\, n^d$ cannot vanish anywhere on $\mathscr{I}$. Hence the result. □

Remark: The above argument also brings out the geometric reason behind the result of Lemma 2.3. (The following discussion will hold even if $\mathscr{I}$ fails to be topologically $S^2 \times R$ and g_{ab} fails to satisfy Einstein's equation near $\mathscr{I}$; as in Lemma 2.3, we need only the existence of "local $\mathscr{I}$.") Supose ξ^a and ξ'^a are two translational Killing fields with $\xi^a \mid_{\mathscr{I}} = \alpha n^a$ and $\xi'^a \mid_{\mathscr{I}} = \alpha' n^a$. Then, given any point p of $\mathscr{I}$, we can always find a (constant-) linear combination of ξ^a and ξ'^a—i.e., a translational Killing field—which vanishes there. Choose p such that $K_{abcd}\, n^d \mid_p \neq 0$. Then, by the first part of the proof of Lemma 3.7, we arrive at a contradiction. Hence $(\hat{\mathbf{M}}, \hat{g}_{ab})$ cannot admit more than one translational Killing field.

As an immediate consequence of Lemma 3.7, we have:

Lemma 3.8: If dim $\mathscr{L} = 2$ and dim $\tau = 1$, then $\mathscr{L}$ is Abelian and the space–time is stationary and axisymmetric.[17]

Collecting the results of Lemmas 3.1–3.8 we obtain:

Theorem 3: Let $(\hat{\mathbf{M}}, \hat{g}_{ab})$ admit an asymptote with $\mathscr{I}$ topologically $S^2 \times R$. Let, furthermore, the asymptotic Weyl curvature K_{abcd} be such that $K_{abcd}\, n^d$ fails to vanish identically on $\mathscr{I}$. Then, the dimension of the isometry Lie algebra $\mathscr{L}$ cannot exceed four. If dim $\mathscr{L} = 4$, $(\hat{\mathbf{M}}, \hat{g}_{ab})$ is spherically symmetric and static in $N \cap \hat{\mathbf{M}}$ for some neighborhood $\mathbf{N}$ of $\mathscr{I}$ in $\mathbf{M}$. If dim $\mathscr{L} = 3$, the isometry group is a subgroup of the Lorentz group. If, in addition, $\hat{R}_{ab} = 0$ in $\mathbf{N} \cap \hat{\mathbf{M}}$, the translational Killing field, if present, is necessarily a BMS time translation and the Bondi news vanishes identically on $\mathscr{I}$. If dim $\mathscr{L} = 4$, $\hat{g}_{ab}$ is isometric to the Schwarzschild metric in $\mathbf{N} \cap \hat{\mathbf{M}}$; and, if dim $\mathscr{L} = 2$ and $\mathscr{L}$ contains a translational Killing field, $(\hat{\mathbf{M}}, \hat{g}_{ab})$ is stationary and axisymmetric.[17]

4. SUMMARY

The main results obtained here as well as those obtained in Ref. 1 may be summarized as follows. Let $(\hat{\mathbf{M}}, \hat{g}_{ab})$ be a space–time which admits an asymptote in the sense of Definition 1 of this paper.[18] Let, furthermore, the Weyl curvature be such that $\Omega^{-n} C_{abc}{}^d$ admits a smooth limit $K_{abc}{}^d$ to $\mathscr{I}$ ($n \geqslant 0$, integer). Finally, let $K_{abcd}\, n^d$ not vanish on $\mathscr{I}$ identically. Then the isometry group G of $(\hat{\mathbf{M}}, \hat{g}_{ab})$ is at most four dimensional. G can admit a translation subgroup N, with dim $N \leqslant 1$, this N is a normal subgroup of G with dim $G/N \leqslant 3$.

The structure of G can be restricted with further assumptions. The various possibilities are listed in Table I. The notation used is the following: $\mathbf{N}$ stands for a neighborhood of $\mathscr{I}$ in $\mathbf{M}$, L for Lorentz group, and , $G \subset L$ means "G is a subgroup of L." The curious possibility that has not been ruled out is the following. Let $(\hat{\mathbf{M}}, \hat{g}_{ab})$ be a vacuum, *radiating* space–time with $\mathscr{I}$ which is topologically $S^2 \times R$. In this case, $G \subset L$. If G were isomorphic with SO(3), or SO(2,1), one knows by (generalized) Birkhoff's theorem that there must also exist a translation in $(\mathbf{N}, \hat{g}_{ab})$, whence Bondi news must vanish. What if G is isomorphic to one of the remaining[14] three subgroups of L?

ACKNOWLEDGMENTS

One of us (A.A.) thanks the Munich relativity group for hospitality. We thank Lee Lindblom for pointing out several typographical errors in the manuscript.

[1]A. Ashtekar and B.C. Xanthopoulos, J. Math. Phys. **19**, 2216 (1978).
[2]Remark (2) in the note added in proof of Ref. 1 contains a minor error: as we shall see in Sec. 3 of this paper, although most results of Ref. 1 do continue to hold in "the infinitesimal form" if the generators of $\mathscr{I}$ fail to be complete, additional assumptions are needed to obtain the "infinitesimal analogs" of Lemmas 3.2 and 3.3 of Ref. 1.
[3]Indeed the present state of the art is that although several radiating solutions admitting a "piece of $\mathscr{I}$" are available, *no* such solution with a global $\mathscr{I}$ is known!
[4]I. Robinson and A. Trautman, Phys. Rev. Lett. **4**, 431 (1960); Proc. R. Soc. A London **265**, 463 (1962).
[5]W. Kinnersley and M. Walker, Phys. Rev. D **2**, 1359 (1970), and references therein.
[6]P.C. Vaidya, Nature **171**, 260 (1953).
[7]R. Geroch, the article in *Asymptotic structure of space–time*, edited by F.P. Esposito and L. Witten (Plenum, New York, 1977), p. 8.
[8]G.T. Horowitz, private communication.
[9]A. Ashtekar and A. Magnon-Ashtekar, J. Math. Phys. **19**, 1567 (1978).
[10]R. Penrose, Proc. R. Soc. A London **284**, 159 (1965).
[11]See, e.g., Ref. 7, p. 19.
[12]See, e.g., the article by J. Ehlers and W. Kundt in *Gravitation: An introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962).
[13]See, e.g., B.G. Schmidt, M. Walker, and P. Sommers, Gen. Rel. Grav. **6**, 489 (1975).
[14]See, e.g., J. Segerctrantz, Ann. Acad. Sci. Fennicae A **233** (1967); B.G. Schmidt, Dissertation submitted to the University of Hamburg (1968); R. Shaw, "The subgroup structure of the homogeneous Lorentz group," University of Hull, G.B. preprint (1969).
[15]See, e.g. J. Ehlers, W. Rindler, and I. Robinson, in *Perspectives in Geometry and Relativity*, edited by B. Hoffman (Indiana University, Bloomington, London, 1966), pp. 145–47; and J.E. Schremp, NRL Q. Nucl. Sci. Tech. **1** (1963).
[16]For the case when $K_{abcd} = \Omega^{-1} C_{abcd}$, see, E.T. Newman and R. Penrose, Proc. R. Soc. A London **305**, 175 (1968).
[17]By axisymmetry, we only mean that there is a SO(2) subgroup of the isometry group with spacelike orbits; the axis itself need not exist.
[18]The metric g_{ab} need not be C^∞ on $\mathscr{I}$. The entire analysis goes through if g_{ab} is sufficiently differentiable for the asymptotic curvature K_{abcd} to be only C^1 on $\mathscr{I}$.

Electromagnetic test fields in the Kerr–Newman metric

S.K. Bose
Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556

(Received 12 June 1979; accepted for publication 24 December 1979)

The form of the electromagnetic test fields in the Kerr–Newman background, with the background metric held fixed, is closedly evaluated.

There has been several attempts to study the behavior of perturbation equations in the Kerr–Newman metric.[1–7] Because of the complexity of the problem arising out of the coupled nature of the equations, it has been found possible to derive definite conclusions only in a few simple cases; as for instance, in the limiting case of stationary axially-symmetric perturbations to *black holes*[1,2] and that of perturbations to nonrotating black holes.[6] In this note we study another limiting, if admittedly oversimplified, case; namely, the behavior of electromagnetic *test fields* in the Kerr–Newman background. Somewhat unexpectedly, it turns out that the Maxwell's equations for the present problem can be integrated exactly and the solution, thus obtained, is unique. Physically, the solution corresponds to the addition of (in general, complex) charges to the source. Thus, only trivial fields are allowed in the test field approximation. It should be emphasized that no use is made of any boundary condition to derive this result.

In the test field approximation, the background metric and hence the Ricci tensor are held fixed and therefore, via the Einstein equations, so is the energy-momentum tensor. This last requirement severely restricts the admissible form of the test fields whenever the first order variation in the energy-momentum tensor does not vanish identically. In the present case, the foregoing restrictions [Eqs. (4) and (5) below] together with the Maxwell's equation completely determine [Eq. (18) below] the test field, which turns out to correspond to the addition of a magnetic monopole (more generally a complex charge) to the source.

In the Newman-Penrose formalism,[8] the electrovac Einstein equations are

$$\Phi_{ij} = 2\Phi_i \Phi_j^*, \quad (i,j = 0,1,2) \tag{1}$$

where Φ_{ij} are the usual tetrad projections of the Ricci and Φ_i those of the electromagnetic field tensor. Let us now write

$$\Phi_i = \Phi_i^{\mathcal{A}} + \Phi_i^{\mathcal{B}}, \tag{2}$$

where $\Phi_i^{\mathcal{A}}$ denote the unperturbed Kerr–Newman field and $\Phi_i^{\mathcal{B}}$ the test field. Equations (1) and (2) lead to

$$\Phi_i^{\mathcal{A}} \Phi_j^{*\mathcal{B}} + \Phi_j^{*\mathcal{A}} \Phi_i^{\mathcal{B}} = 0, \tag{3}$$

since Φ_{ij} is kept fixed at $\Phi_{ij}^{\mathcal{A}}$. Using Kinnersley's tetrad[9] and the resulting expressions[4] for $\Phi_i^{\mathcal{A}} (\Phi_0^{\mathcal{A}} = \Phi_2^{\mathcal{A}} = 0,\ \Phi_1^{\mathcal{A}} \neq 0)$ we obtain from (3)

$$\Phi_0^{\mathcal{B}} = \Phi_2^{\mathcal{B}} = 0, \tag{4}$$

and

$$\Phi_1^{\mathcal{A}} \Phi_1^{*\mathcal{B}} + \Phi_1^{*\mathcal{A}} \Phi_1^{\mathcal{B}} = 0. \tag{5}$$

Let us consider the sourceless Maxwell's equations. Utilizing Eqs. (2) and (4), these can be written, in a standard notation,[8] as

$$(D - 2\rho)\, \Phi_1^{\mathcal{B}} = 0, \tag{6}$$

$$(\Delta + 2\mu)\, \Phi_1^{\mathcal{B}} = 0, \tag{7}$$

$$(\delta - 2\tau)\, \Phi_1^{\mathcal{B}} = 0, \tag{8}$$

$$(\delta^* + 2\pi)\, \Phi_1^{\mathcal{B}} = 0, \tag{9}$$

thus, $\Phi_1^{\mathcal{B}}$ satisfies exactly the same equations as does $\Phi_1^{\mathcal{A}}$. Using the known[4] values of the spin coefficients, we now write out the above equations explicitly in the Boyer–Lindquist coordinates.[10] We thus obtain from (6) and (7):

$$(r^2 + a^2)(\partial/\partial t)\Phi_1^{\mathcal{B}} + a(\partial/\partial\phi)\Phi_1^{\mathcal{B}} = 0, \tag{10}$$

$$(\partial/\partial r)\Phi_1^{\mathcal{B}} = -\frac{2}{r_-}\Phi_1^{\mathcal{B}}. \tag{11}$$

Similarly, from (8) and (9) we obtain:

$$a \sin^2\theta\, (\partial/\partial t)\Phi_1^{\mathcal{B}} + (\partial/\partial\phi)\Phi_1^{\mathcal{B}} = 0, \tag{12}$$

$$(\partial/\partial\theta)\Phi_1^{\mathcal{B}} = -\frac{2ia\sin\theta}{r_-}\Phi_1^{\mathcal{B}}. \tag{13}$$

In the above, a denotes the angular momentum per unit mass of the source and $r_- = r - ia\cos\theta$. From Eqs. (10) and (12) it follows that

$$(\partial/\partial t)\Phi_1^{\mathcal{B}} = (\partial/\partial\phi)\Phi_1^{\mathcal{B}} = 0, \tag{14}$$

thus the test field must be stationary and axially-symmetric. Since $\Phi_1^{\mathcal{B}}$ can depend only on the variables r and θ, we are free to consider it to depend on r_- and r_+ (where $r_+ = r + ia\cos\theta$). We accordingly rewrite Eqs. (11) and (13) in the form:

$$(\partial/\partial r_-)\Phi_1^{\mathcal{B}} + (\partial/\partial r_+)\Phi_1^{\mathcal{B}} = (\partial/\partial r)\Phi_1^{\mathcal{B}} = -(2/r_-)\Phi_1^{\mathcal{B}}, \tag{15}$$

$$(\partial/\partial r_-)\Phi_1^{\mathcal{B}} - (\partial/\partial r_+)\Phi_1^{\mathcal{B}} = (1/ia\sin\theta)(\partial/\partial\theta)\Phi_1^{\mathcal{B}} = -(2/r_-)\Phi_1^{\mathcal{B}}. \tag{16}$$

From the above, it is evident that $\Phi_1^{\mathcal{B}}$ does not depend on r_+, i.e., $\partial\Phi_1^{\mathcal{B}}/\partial r_+ = 0$ and that:

$$(d/dr_-)\Phi_1^{\mathcal{B}} = -(2/r_-)\Phi_1^{\mathcal{B}}. \tag{17}$$

The solution of Eq. (17) is

$$\Phi_1^{\mathcal{B}} = \delta c / r_-^2, \tag{18}$$

where δc is a small constant parameter and the form of the solution (18) is obviously unique. Finally, equation (5) gives a constraint on δc:

$$e\delta c^* + e^*\delta c = 0, \tag{19}$$

since[4] $\Phi_1^{\mathcal{A}} = e/2r_-^2$. In the original Kerr–Newman solu-

 0022-2488/80/040868-02$1.00

tion,[11] the parameter e is real and stands for the electric charge of the source; consequently, c must be purely imaginary and the resulting test field corresponds physically to the addition of magnetic monopoles to the source. More generally, we are free to treat the parameter e in the unperturbed solution to be complex, corresponding to a source that carries both electric charge and magnetic monopoles. In the latter event c is complex satisfying the constraint (19). In any event, the solution (18) rules our nontrivial test fields. The foregoing result implies that nontrivial electromagnetic fields must show departure from test field behavior, that is, must be accompanied by perturbations in the metric.

[1]D.C. Robinson, Phys. Rev. D **10**, 458 (1974).
[2]S.K. Bose and M.Y. Wang, Phys. Rev. D **10**, 1675 (1974).
[3]D.M. Chitre, Phys. Rev. D **11**, 760 (1975); Phys. Rev. D **13**, 2713 (1976).
[4]S.K. Bose, J. Math. Phys. **16**, 772 (1975).
[5]C.H. Lee, J. Math. Phys. **17**, 1226 (1976).
[6]C.H. Lee and W.D. McGlinn, J. Math. Phys. **17**, 2159 (1976).
[7]D.N. Page, Phys. Rev. D **14**, 1509 (1976).
[8]E.T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
[9]W. Kinnersley, J. Math. Phys. **10**, 1195 (1969).
[10]R.H. Boyer and R.W. Lindquist, J. Math. Phys. **8**, 265 (1967).
[11]E.T. Newman, E. Couch, K. Chinapared, A. Exton, A. Prakash, and R. Torrence, J. Math. Phys. **6**, 918 (1965).

Curvatures of hypersurfaces of transitivity of Bianchi model universes

Alex Harvey
Physics Department, Queens College, City University of New York, Flushing, New York 11367

(Received 11 March 1979; accepted for publication 4 May 1979)

The Riemann curvatures of 2-surfaces are calculated for the spatial, 3-surfaces of transitivity of the various Bianchi models. It is found that some models are generally possessed simultaneously of both negative and positive curvature. The Bianchi V models are remarkable in that however large the degree of *metric* anisotropy, the curvatures are always isotropic. It is also found that Bianchi IX models, though invariably compact by virtue of the group structure, may have *negative* mean curvature.

INTRODUCTION

The evidence currently available,[1] i.e., isotropy of background microwave radiation, distribution of galaxies, and red-shift of galactic spectra, taken together with the Copernican priciple that these observations are not made from a special position in the universe, combines to suggest strongly that the universe, on sufficiently large scale, is and has been, for a substantial period of the time leading to the present epoch, of Friedmann–Robertson–Walker (FRW) character. Correctness of this view would rule out, *inter alia*, all the Bianchi types save the isotropic cases of Types I, V, and IX. (There is explicit recognition that use of such concepts as "sufficiently large scale" and "Bianchi models" permit gross simplifications.)

The nature of the universe at earlier times is much less well delimited. The observational data required to constrain the variety of possible models is not available. There are no more than hints in the current store of knowledge for any speculation beyond suggesting a primeval fireball and subsequent formation of elements and later galaxies. What model served as stage for this scenario can only be guessed. Indeed, Ellis[2] has emphasized that "... we may suspect that the universe was *not* very like a Robertson–Walker universe at early times··· ."

For reasons such as these as well as others there has been extensive interest in recent years in cosmological models not as highly symmetric as the FRW models. (The FRW models are maximally symmetric in that they are homogeneous and isotropic.) Among the many approaches avoiding this constraint is to set aside the requirement of isotropy. The metrics thereby admitted are the so-called Bianchi symmetry types,[3] and they may be used as either the basis for a model or for perturbation of a given FRW model. Whatever the usage, it is accomplished with surer sense of outcome if pursued with fuller understanding of these nonisotropic spaces. This paper presents the results of some studies undertaken to this end. The point of interest was the determination of the dependence of curvature on anisotropy.

Recall that the term "constant curvature" used in distinguishing FRW models refers to the (three-dimensional, spacelike) hypersurfaces of homogeneity and is a particular manifestation of isotropy. For the nonisotropy Bianchi types it may be expected that the curvature is also anisotropic. This turns out, with one remarakble exception, to be the case.

What is actually determined is the following: Three mutually orthogonal directions are assigned at a generic point of the hypersurface of homogeneity. By pairs these define 2-hypersurfaces, and for these the Riemann curvatures are calculated. The results, in general, (specific details are presented below) are that some of the types are distinguished by having simultaneously positive and negative curvatures. This is the case, for instance, for Brill's magnetic universe.[4] The remarkable exception noted earlier is the class of Bianchi V models which, independently of source or degree of *metric* anisotropy are of *constant*, i.e., isotropic (negative) curvature. The extent to which this should be considered when interpreting cosmic observations is not clear. Among other interesting results are those for Bianchi IX models; the completely anisotropic case may have negative scalar curvature, i.e., the sum of the Riemannian curvatures may be negative, thus furnishing an example of a compact space of negative curvature. A feature worth noting is that the technique is both general and simple in application.

THE RIEMANN CURVATURES

The Riemann curvature[5] is

$$K_{\xi\eta} = \frac{R_{ijkl}\,\xi^i\eta^j\xi^k\eta^l}{(g_{ij}\,g_{kl} - g_{lk}\,g_{jl})\xi^i\eta^k\xi^j\eta^l}, \tag{1}$$

where all quantities refer to the hypersurface of homogeneity. Indices range from 1 to 3 and R_{ijkl} is the Riemann–Christoffel curvature tensor. ξ^i and η^i are 3-vectors which define the 2-surface of which $K_{\xi\eta}$ is the Riemann curvature. $K_{\xi\eta}$ is an invariant and may be calculated in any coordinate system. This is most simply accomplished in a Cartan orthonormal frame.[3] In this frame

$$g_{ij} = \mathrm{diag}(1,1,1), \tag{2}$$

and with the choice of three unit vectors parallel to the basis 1-forms, viz.,

$$\xi^i = (1,0,0), \tag{3a}$$

$$\eta^i = (0,1,0), \tag{3b}$$

$$\mu^i = (0,0,1), \tag{3c}$$

the three curvatures defined by $\xi - \eta$, $\eta - \mu$, $\mu - \xi$ are

 0022-2488/80/040870-04$1.00

simply

$$K_{12} = R_{1212}, \tag{4a}$$

$$K_{23} = R_{2323}, \tag{4b}$$

$$K_{31} = K_{3131}. \tag{4c}$$

The calculation of R_{ijkl} for the 3-hypersurfaces is much simpler than for its four-dimensional counterpart because the calculations are for a fixed time. The general algorithm is outlined in Appendix A. Of the various models details are presented only for type only for type II and that only to illustrate the algorithm. For the rest only the exterior derivatives of the basis 1-forms and concomitant curvatures are included with the comments.

Bianchi I:

$$d\sigma^1 = d\sigma^2 = d\sigma^3 = 0. \tag{5a,b,c}$$

Consequently,

$$K_{12} = K_{23} = K_{31} = 0. \tag{6a,b,c}$$

This corresponds to the well-known situation that however anisotropic it may be, a Bianchi I model is always flat.

Bianchi II:

$$d\sigma^1 = d\sigma^2 = 0, \tag{7 a,b}$$

$$d\sigma^3 = \sigma^1 \wedge \sigma^2. \tag{7c}$$

For the orthonormal 1-forms the only nonvanishing exterior derivative is

$$d\omega^3 = cd\sigma^3 = c\sigma^1 \wedge \sigma^2 = \frac{c}{ab}\omega^1 \wedge \omega^2. \tag{8}$$

The corresponding structure coefficients are

$$\Omega_{12}{}^3 = -\frac{c}{ab} = -\Omega_{21}{}^3. \tag{9}$$

The three connection forms are

$$\omega_{12} = \frac{1}{2}(\Omega_{12k} + \Omega_{1k2} - \Omega_{2k1})\omega^k = -\frac{c}{2ab}\omega^3, \tag{10a}$$

and (similarly)

$$\omega_{23} = \frac{c}{2ab}\omega^1, \tag{10b}$$

$$\omega_{31} = \frac{c}{2ab}\omega^2. \tag{10c}$$

Finally, the curvature forms are

$$R_{12} = d\omega_{12} + \omega_{13} \wedge \omega_{12} = -\frac{3c^2}{4a^2b^2}\omega^1 \wedge \omega^2, \tag{11a}$$

and (similarly)

$$R_{23} = +\frac{1c^2}{4a^2b^2}\omega^2 \wedge \omega^3, \tag{11b}$$

$$R_{31} = +\frac{1c^2}{4a^2b^2}\omega^3 \wedge \omega^1. \tag{11c}$$

The corresponding relevant components of the Riemann–Christoffel tensor and the Riemann Curvatures are

$$K_{12} = R_{1212} = -\frac{3c^2}{4a^2b^2}, \tag{12a}$$

$$K_{23} = K_{31} = R_{2323} = R_{3131} = +\frac{c^2}{4a^2b^2}. \tag{12b,c}$$

These results are quite generally independent of the off-diagonal elements of the metic tensor if they exist.

Bianchi III:

$$d\sigma^1 = d\sigma^2 = 0, \tag{13a,b}$$

$$d\sigma^3 = \sigma^1 \wedge \sigma^3, \tag{13c}$$

$$K_{12} = -\frac{3g^2}{4a^2b^2}, \tag{14a}$$

$$K_{23} = \frac{g^2}{4a^2b^2}, \tag{14b}$$

$$K_{31} = -\frac{1}{a^2} + \frac{g^2}{4a^2b^2}. \tag{14c}$$

The curvatures depend directly on whether or not the off-diagonal term g is present. If it is absent then the 1-2 and 2-3 surfaces are flat. If it is present then the nature of the 1-3 surface depends on the inequality $1/a^2$ vs $g^2/4b^2$. The mean curvature is always negative.

Bianchi IV:

$$d\sigma^1 = 0, \tag{15a}$$

$$d\sigma^2 = \sigma^1 \wedge \sigma^2, \tag{15b}$$

$$d\sigma^3 = \sigma^1 \wedge \sigma^2 + \sigma^1 \wedge \sigma^3, \tag{15c}$$

$$K_{12} = -\left(\frac{1}{a^2} + \frac{3c^2}{4a^2b^2}\right), \tag{16a}$$

$$K_{23} = K_{31} = -\left(\frac{1}{a^2} - \frac{c^2}{4a^2b^2}\right). \tag{16b,c}$$

These results are quite interesting. It can be shown that a diagonal Bianchi IV metric does not exist. Nonetheless, the curvatures depend only on the diagonal elements.

The only known exact solutions[6] for this symmetry group have $b = c$ with the result

$$K_{12} = -\frac{5}{4a^2}, \tag{17a}$$

$$K_{31} = K_{23} = -\frac{3}{4a^2}. \tag{17b,c}$$

Bianchi V:

$$d\sigma^1 = 0, \tag{18a}$$

$$d\sigma^2 = \sigma^1 \wedge \sigma^2, \tag{18b}$$

$$d\sigma^3 = \sigma^1 \wedge \sigma^3, \tag{18c}$$

$$K_{12} = K_{23} = K_{31} = -1/a^2. \tag{19a,b,c}$$

This is most remarkable in that independent of the nature of the sources and consequent degree of metric anisotropy, the hypersurface of homogeneity is of uniform negative curvature.

Bianchi VI_h:

$$d\sigma^1 = 0, \tag{20a}$$

$$d\sigma^2 = h\sigma^1 \wedge \sigma^2, \tag{20b}$$

$$d\sigma^3 = \sigma^1 \wedge \sigma^3, \tag{20c}$$

where $h \neq 0,1$.

$$K_{12} = -\frac{h^2}{a^2} - \frac{3g^2}{4a^2b^2}(h-1)^2, \tag{21a}$$

$$K_{23} = -\frac{h^2}{a^2} + \frac{g^2}{4a^2b^2}(h-1)^2, \tag{21b}$$

$$K_{31} = -\frac{1}{a^2} + \frac{g^2}{4a^2b^2}(h-1)^2. \tag{21c}$$

If $h = 0$ or 1, the symmetry group becomes type III or V, respectively. In general, the behavior is much like a perturbed type V. It has the curious feature of being dependent on the off-diagonal element g and independent of the diagonal element c.

While the 1-2 surface is always of negative curvature, the other 2 may be either of positive or negative curvatures.

Bianchi VII_h:

$$d\sigma^1 = 0, \tag{22a}$$

$$d\sigma^2 = \sigma^1 \wedge \sigma^2 + (h-2)\sigma^1 \wedge \sigma^3, \tag{22b}$$

$$d\sigma^3 = \sigma^1 \wedge \sigma^3 + (h-1)\sigma^1 \wedge \sigma^3, \tag{22c}$$

where $h < 2$. For $h = 2$ the group becomes type IV.[7] (The choice of representation was such as to make obvious the relation between the two types.)

$$K_{12} = -\frac{1}{a^2}\left[1 + \frac{3}{4}\frac{c^2}{b^2} + (h-2)\left(2 - 2\frac{g}{c} - \frac{3}{2}\frac{g^2}{b^2}\right) + (h-2)^2\left(-\frac{1}{2}\frac{g}{c} - \frac{g^2}{c^2} + \frac{1}{2}\frac{b^2}{c^2} + \frac{3}{4}\frac{g^4}{b^2c^2}\right)\right], \tag{23a}$$

$$K_{23} = -\frac{1}{a^2}\left[+(h-1) - \frac{1}{4}\frac{c^2}{b^2} - \frac{1}{4}(h-2) \times\left(1 - \frac{g^2}{b^2} - \frac{cg}{b^2}\right) - \frac{1}{4}(h-2)^2 \times\left(\frac{g^4}{b^2c^2} + \frac{g^2}{b^2} + \frac{b^2}{c^2} + 2\frac{g^3}{b^2c} - \frac{g^2}{c^2} + 3\frac{g}{c}\right)\right], \tag{23b}$$

$$K_{31} = -\frac{1}{a^2}\left\{(h-1)^2 - \frac{1}{4}\frac{c^2}{b^2} + (h-2) \times\left[\frac{1}{2} + 2(h-1)\frac{g}{c} - \frac{1}{2}\frac{g}{c} + \frac{1}{2}\frac{cg}{b^2} + \frac{1}{2}\frac{g^2}{b^2}\right] + (h-2)^2\left(\frac{3}{4}\frac{b^2}{c^2} + \frac{1}{2}\frac{g^2}{c^2} - \frac{1}{4}\frac{g^4}{b^2c^2} - \frac{1}{4} - \frac{1}{2}\frac{g^3}{b^2c}\right)\right\}. \tag{23c}$$

There do not seem to be any general statements which are reasonable in this case.

Bianchi VIII

$$d\sigma^1 = -\sigma^2 \wedge \sigma^3, \tag{24a}$$

$$d\sigma^2 = \sigma^3 \wedge \sigma^1, \tag{24b}$$

$$d\sigma^3 = \sigma^1 \wedge \sigma^2; \tag{24c}$$

$$K_{12} = \frac{1}{a^2b^2c^2}[a^2e^2 - e^2f^2 - f^2a^2 + befg - \tfrac{3}{4}(f^2 - g^2 - c^2)^2 + \tfrac{1}{2}(-a^2 - e^2 + b^2) \times(-f^2 + g^2 + c^2) + \tfrac{1}{4}(-a^2 + e^2 - b^2)], \tag{25a}$$

$$K_{23} = \frac{1}{4a^2b^2c^2}[-3a^4 + 4e^2f^2 - 4efgb - 2a^2(g^2 + c^2 + f^2 + e^2 + b^2) + (e^2 - b^2 - f^2 + g^2 + c^2)^2], \tag{25b}$$

$$K_{31} = \frac{1}{4a^2b^2c^2}[-3(e^2 - b^2)^2 - 2e^2(a^2 - f^2 + g^2 + c^2) + 2b^2(-a^2 - f^2 + g^2 + c^2) - 4(f^2e^2 + g^2b^2 - a^2f^2 - 2efgb) + (a^2 - f^2 + g^2 + c^2)^2]. \tag{25c}$$

In distinction to earlier cases all elements of the metric influence the curvature. In the special case of a diagonal metric with two equal elements, i.e.,

$$g_{ij} = \mathrm{diag}(a,b,b),$$

it is easily determined that

$$K_{12} > 0, \tag{26a}$$

$$K_{23} < 0, \tag{26b}$$

$$K_{31} > 0. \tag{26c}$$

The mean curvature is always negative in this instance. Thus, it appears that Bianchi VIII models may generally be of mixed character.

Bianchi IX:

$$d\sigma^1 = \sigma^2 \wedge \sigma^3, \tag{27a}$$

$$d\sigma^2 = \sigma^3 \wedge \sigma^1, \tag{27b}$$

$$d\sigma^3 = \sigma^1 \wedge \sigma^2. \tag{27c}$$

The discussion of Bianchi IX models is considerably facilitated by exploiting the fact that the associated group of automorphisms[8] may be represented by orthogonal matrices. Hence, on a given hypersurface, g_{ij}, which is real and symmetric, may be diagonalized. (The diagonalization will not generally persist in time, but that is irrelevant here.) Thus, it suffices to consider diagonal metrics only.[9] Under this condition (and writing a_i for the three diagonal elements of g_{ij})

$$K_{ij} = \frac{1}{4a_i^2a_j^2a_k^2} \times(a_i^4 + a_j^4 - 3a_k^4 - 2a_i^2a_j^2 + 2a_j^2a_k^2 + 2a_k^2a_i^2). \tag{28a,b,c}$$

It is readily shown that the mean curvature may be positive or negative and thus, like the Bianchi VIII models, these too may be of mixed character.

For the special case of two equal metric elements, i.e., $g_{ij} = \mathrm{diag}(a,b,b)$,

$$K_{12} = K_{31} = \frac{a^2}{4b^4}, \tag{29a,b}$$

$$K_{23} = \frac{1}{b^2} - \frac{3a^2}{4b^4}. \tag{29c}$$

and the mean curvature is $K = (1 - a^2/4b^2)/b^2$. The character of K_{23} and K may change during the evolution of the model. For a reasonable range of parameters this is actually the case of Brill's electromagnetic universe.[4]

APPENDIX A

The calculation of the Riemann–Christoffel tensor is straightforward. The Bianchi models all admit a metric of the form[5]

$$ds^2 = -\sigma^0\sigma^0 + g_{ij}(t)\sigma^i\sigma^j, \tag{A1a}$$

where σ^0, σ^i are basis 1-forms. The metric in the orthonormal frame is

$$ds_2 = -\omega^0\omega^0 + \eta_{ij}\,\omega^i\omega^j, \tag{A1b}$$

and is obtainable from Eq. (2) by

$$\omega^0 = \sigma^0, \tag{A2a}$$

$$\omega^1 = a\sigma^1, \tag{A2b}$$

$$\omega^2 = e\sigma^1 + b\sigma^2, \tag{A2c}$$

$$\omega^3 = f\sigma^1 + g\sigma^2 + c\sigma^3. \tag{A2d}$$

This is much simpler to employ than the "square root" of g_{ij} and equally general. The coefficients a through g are functions of t only. The inverted set is

$$\sigma^1 = \frac{1}{a}\omega^1, \tag{A3a}$$

$$\sigma^2 = -\frac{c}{ab}\omega^1 + \frac{1}{b}\omega^2, \tag{A3b}$$

$$\sigma^3 = \frac{eg-bf}{abc}\omega^1 - \frac{g}{bc}\omega^2 + \frac{1}{c}\omega^3. \tag{A3c}$$

Because the calculation is confined to a hypersurface of constant t the exterior derivatives of the 1-forms are simply

$$d\omega^1 = ad\sigma^1, \tag{A4a}$$

$$d\omega^2 = ed\sigma^1 + bd\sigma^2, \tag{A4b}$$

$$d\omega^3 = fd\sigma^1 + gd\sigma^2 + cd\sigma^3. \tag{A4c}$$

This much is identical for all Bianchi models. At this point the $d\sigma^i$ must be replaced by the 1-forms characteristic of the Bianchi type in question, viz.,

$$d\sigma^i = \tfrac{1}{2}C_{jk}{}^i\sigma^j\wedge\sigma^k, \tag{A5}$$

where the $C_{jk}{}^i$ are the structure constants of the type group. The set of constants for a particular type are not unique, but depend on the choice of basis.[10]

With substitution of Eqs. (A5) into Eqs. (A4), replacement of σ^i by means of Eqs. (A3), and appropriate rearrangement of terms the exterior derivatives may be expressed as

$$d\omega^i = -\tfrac{1}{2}\Omega_{jk}{}^i\omega^j\wedge\omega^k, \tag{A6}$$

where the quantities $\Omega_{jk}{}^i$ are functions of the coefficients a–g, are denoted "structure coefficients," and, like the group structure constants, are antisymmetric in the lower two indices.

The structure coefficients are used to compute the connection forms. Firstly, note that the last index may be lowered by use of the 3-tensor η_{ij}, viz.,

$$\Omega_{jki} = \eta_{ih}\Omega_{jk}{}^h. \tag{A7}$$

Then the connection forms are given by

$$\omega_{ij} = \tfrac{1}{2}(\Omega_{ijk} + \Omega_{ikj} - \Omega_{jki})w^k. \tag{A8}$$

The curvature forms $\mathscr{R}_{ij}$ are given by the formula

$$\mathscr{R}_{ij} = d\omega_{ij} + \omega_{ik}\wedge\omega_{kj}\,. \tag{A9}$$

Inasmuch as the connection forms are linear combinations of 1-forms the $\mathscr{R}_{ij}$ are linear combinations of 2-forms,

$$\mathscr{R}_{ij} = \tfrac{1}{2}\mathscr{R}_{ijkl}\,\omega^k\wedge\omega^l,$$

the coefficients of which are just the components of the Riemann–Christoffel tensor for the hypersurface of transitivity.

[1]See, e.g., T.B. Criss *et al.* in *Proceedings of the Seventh International Conference on General Relativity and Gravitation*, edited by G. Shaviv and J. Rosen (Wiley, New York, 1975).

[2]G.F.R. Ellis in *Proceedings of the International School of Physics "Enrico Fermi,"* Course 47, General Relativity and Cosmology, edited by R.K. Sachs (Academic, New York, 1971), see p. 175.

[3]See, e.g., M.P. Ryan, Jr. and L.C. Shepley, *Homogeneous Relativistic Cosmologies* (Princeton U.P., Princeton, New Jersey, 1975).

[4]D.R. Brill, Phys. Rev. B **133**, 845 (1964).

[5]L.P. Eisenhart, *Riemannian Geometry* (Princeton U.P., Princeton, New Jersey, 1966), see p. 79.

[6]A. Harvey and D. Tsoubelis, Phys. Rev. D **15**, 2734 (1977).

[7]C. Ftaclas, Ph.D. thesis (unpublished).

[8]A. Harvey, J. Math. Phys. **20**, 251 (1979).

[9]N. Batakis and C. Ftaclas (private communication).

[10]M. Hamermesh, *Group Theory* (Addison–Wesley, Reading, Massachussetts, 1962).

A note on left-flat space-times

E. T. Newman[a)]
Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260
K. P. Tod
Mathematical Institute, Oxford, England

(Received 10 September 1979; accepted for publication 4 January 1980)

The theory of left-flat space-times is developed by introducing a null tetrad and imposing increasingly restrictive conditions on it. These conditions are justified from different points of view. In this way we present an elementary account of several of the different approaches to left-flat spaces.

I. INTRODUCTION

For many reasons (which we will not go into here) the subject of complex space-times satisfying the vacuum Einstein equations with a self-dual (or anti-self-dual) Weyl tensor has been of considerable interest in recent years.[1-11] We will refer to these complex space-times as left-flat (or right-flat). This paper is concerned only with the differential-geometric aspects of the subject and not with the physics. Basically, there have been three different geometric approaches to the subject. One indirect approach was based on the $\mathscr{H}$-space construction where the asymptotically shear-free complex light cones of an asymptotically flat space-time were identified with the points of the left-flat manifold.[1,2] A second and closely related approach due to Penrose involved the deformations of twistor space.[3] A more direct line due to Plebanski was to straight away try to solve the left-flat field equations.[4] (There is actually an elegant fourth approach due to Flaherty,[11] which is very close to that of Plebanski.)

In addition to the fact that these different approaches end with essentially the same results they share another common feature—namely it is difficult for the non-specialist to understand some of the remarkably beautiful differential-geometric results that are contained in this subject. It is a purpose of this note to try to partially remedy this situation. The results presented here are not essentially new, being basically implicit in earlier papers. We make them explicit and express them in more conventional notation. This results, we believe, in a deeper understanding.

We will consider a differential manifold with a complex structure so that local complex coordinates z^a can be introduced. As our results are purely local we will work in a fixed patch. We will choose on this patch a holomorphic metric tensor subject to certain conditions from which the theory of left-flat space-times follows in an obvious and largely trivial fashion. We will finally indicate how these conditions originate in the earlier approaches to the subject.

II. PRELIMINARIES

We begin by introducing four (in some sense, to be decieded later, privileged) holomorphic vector fields,

$$e_{\mu a}(z^b) = \{L_a, N_a, M_a, \tilde{M}_a\}, \tag{2.1}$$

with

$$e^{\mu}{}_a = \{N_a, L_a, -\tilde{M}_a, -M_a\}, \tag{2.2}$$

such that

$$e^{\mu}{}_a e_{\mu b} = g_{ab}(z^c) \tag{2.3}$$

and

$$e^{\mu}{}_a e^{\nu a} = \eta^{\mu\nu} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \end{pmatrix}. \tag{2.4}$$

We now define the *connection tensor* by

$$\Gamma_{cab} = e^{\mu}{}_a \nabla_c e_{\mu b}, \tag{2.5}$$

with ∇_c denoting the metric covariant derivative. Note that Γ_{cab} are *not* the Ricci rotation coefficients, which are scalars, but are the components of a legitimate tensor—they, however, have a nonhomogeneous transformation law under a change in the $e^{\mu}{}_a \to e'^{\mu}{}_a = L^{\mu}{}_{\nu} e^{\nu}{}_a$, $\eta^{\rho\sigma} L^{\mu}{}_{\rho} L^{\nu}{}_{\sigma} = \eta^{\mu\nu}$,

$$\Gamma'_{cab} = \Gamma_{cab} + e'^{\mu}{}_a e'^{\rho}{}_b \eta_{\mu\nu} L^{-1\sigma}{}_{\rho} \nabla_c L^{\nu}{}_{\sigma}.$$

From the fact that the connection (in 2.5) is metric one can show (by differentiating (2.3) and substituting in the Christoffel symbols) that

$$\Gamma_{cab} = e_{\mu a} e^{\mu}{}_{[b,c]} - e_{\mu b} e^{\mu}{}_{[a,c]} - e_{\mu c} e^{\mu}{}_{[a,b]}, \tag{2.6}$$

where the bracket denotes the skew part of the ordinary derivative.

Note two facts about Γ:

$$\Gamma_{cab} = -\Gamma_{cba} \tag{2.7}$$

and

$$\nabla_c e_{\mu b} = \Gamma_{cab} e_{\mu}{}^{a}. \tag{2.8}$$

From the Ricci Identity we have immediately

$$\tfrac{1}{2} R_{abcd} = \nabla_{[c} \Gamma_{d]ab} + \Gamma_{[c}{}^{e}{}_{a} \Gamma_{d]eb}. \tag{2.9}$$

Note again that every symbol here represents a tensor; no ordinary derivatives appear. (We point out in a brief digression an alternate meaning to the tensor Γ_{cab}. Since we have a privileged $e_{\mu a}$ we have a distant parallelism and can hence define a flat connection D, which preserves the metric but necessarily has torsion, by

$$D_a e^{\mu}{}_b = 0 = \frac{\partial e^{\mu}_b}{\partial x^a} - C^{c}_{ba} e^{\mu}{}_c. \tag{2.10}$$

Since

$$\nabla_a e^{\mu}{}_b = \Gamma_a{}^{c}{}_b e^{\mu}{}_c = \frac{\partial e^{\mu}_b}{\partial x^a} - \begin{Bmatrix} c \\ ba \end{Bmatrix} e^{\mu}{}_c, \tag{2.11}$$

a)This research has been partially supported by a NSF grant.

 0022-2488/80/040874-04$1.00

we see that Γ can be thought of as the difference between the flat connection and the Christoffel symbols.)

III. ASSUMPTIONS

We now make our first and most important restriction on the set $e_{\mu a}$ (and hence on the g_{ab}) namely, we assume

$$\nabla_a L_{[b} M_{c]} = 0\,, \tag{3.1a}$$

$$\nabla_a N_{[b} \tilde{M}_{c]} = 0\,. \tag{3.1b}$$

Note that

$$\begin{aligned} &L_{[a}M_{b]}\,,\\ &L_{[a}N_{b]} - M_{[a}\tilde{M}_{b]}\,,\\ &N_{[a}\tilde{M}_{b]}\,, \end{aligned} \tag{3.2a}$$

are a basis of anti-self-dual bivectors. That is, they satisfy

$$F^*_{ab} \equiv \epsilon_{abcd}F^{cd} = -iF_{ab}$$

with ϵ_{abcd} the alternating symbol with $\epsilon_{0123} = -\sqrt{g}$, while

$$\begin{aligned} &L_{[a}\tilde{M}_{b]}\,,\\ &L_{[a}N_{b]} + M_{[a}\tilde{M}_{b]}\,,\\ &N_{[a}M_{b]}\,, \end{aligned} \tag{3.2b}$$

are a basis of self-dual bivectors (i.e. $F^*_{ab} = +iF_{ab}$).

From (3.1), using (2.8), we have immediately

$$\Gamma_{ae[b}M_{c]}L^e - \Gamma_{ae[b}L_{c]}M^e = 0 \tag{3.3}$$

and

$$\Gamma_{ae[b}\tilde{M}_{c]}N^e - \Gamma_{ae[b}N_{c]}\tilde{M}^e = 0\,; \tag{3.4}$$

from which it follows that

$$\begin{aligned} &\Gamma_{abc}M^{[b}L^{c]} = 0\,,\\ &\Gamma_{abc}\tilde{M}^{[b}N^{c]} = 0\,,\\ &\Gamma_{abc}(L^{[b}N^{c]} + \tilde{M}^{[b}M^{c]}) = 0\,. \end{aligned} \tag{3.5}$$

Since (3.2) form a basis set for bivectors it easily is shown that (3.5) implies that Γ_{abc} has the form

$$\begin{aligned} \Gamma_{abc} = {} & G_a L_{[b}\tilde{M}_{c]} + H_a(L_{[b}N_{c]} + M_{[b}\tilde{M}_{c]})\\ & + I_a N_{[b}M_{c]} \end{aligned} \tag{3.6}$$

(i.e., it is self-dual on the last index pair). From (3.6) it follows that

$$\nabla_a(L_{[b}N_{c]} - M_{[b}M_{c]}) = 0 \tag{3.7}$$

[i.e. all the anti-self-dual bivector basis vectors (3.2a) are covariantly constant].

From (3.6) and (2.9) one can conclude (after a brief calculation involving the $\Gamma\Gamma$ term) that the Riemann tensor is self-dual, i.e.,

$$R_{abcd}{}^* = iR_{abcd}\,, \tag{3.8}$$

and hence

$$R_{ab} = 0,\ C_{abcd}{}^* = iC_{abcd}\,. \tag{3.9}$$

Since the anti-self-dual bivectors (3.2a) are covariantly constant, they are, in particular, closed. Every simple closed anti-self-dual bivector defines a family of integrable two-surfaces which fills the space–time. Further, these 2-surfaces are completely null, that is, any tangent vector to the two-surface is null and any two tangent vectors are orthogonal. This can be seen by consideration of the first bivector in the basis (3.2a) since the tangents to the surface defined by $L_{[a}M_{b]}$ at a point are linear combinations of L^a and M^a. The space of these two-surfaces for all possible simple anti-self-dual bivectors is the projective twistor space of Penrose.[3,7] It is a three-dimensional complex manifold and has further structures which can be deduced from the properties of the left-flat space. Penrose's construction of left-flat spaces[3,7] proceeds from a construction of twistor spaces as a deformation of the flat twistor space of Minkowski space. In the work of Plebanski, the completely null 2-surfaces are known as "null strings."

IV. RESULTS

We now assume [though it can be shown to follow from (3.1)] that one can choose L_a and M_a so that they are gradients, i.e.,

$$\begin{aligned} &L_a = L_{,a}\,,\\ &M_a = M_{,a}\,. \end{aligned} \tag{4.1}$$

From this we have (using (2.8), (3.6) and $\nabla_{[a}L_{b]} = \nabla_{[a}M_{b]} = 0$)

$$L^a\Gamma_{abc} = M^a\Gamma_{abc} = 0 \tag{4.2}$$

and

$$\Gamma_a{}^a{}_b = 0\,. \tag{4.3}$$

From (4.3), Eq. (3.6) takes the form

$$\begin{aligned} \Gamma_{abc} = {} & 2(\alpha L_a + \beta M_a)L_{[b}\tilde{M}_{c]} + 2(\beta L_a + \gamma M_a)(L_{[b}N_{c]}\\ & + M_{[b}\tilde{M}_{c]}) + 2(-\gamma L_a + \delta M_a)N_{[b}M_{c]}\,. \end{aligned} \tag{4.4}$$

In other words, the first index of the Γ (the 1-form index, $\Gamma_{bc} = \Gamma_{abc}dz^a$) holding the other two indices fixed, yields a vector which lies in the two surface defined by $L_{[a}M_{b]}$. From this it follows that

$$\Gamma_{[d}{}^e{}_{|a}\Gamma_{c]eb}$$

is anti-self-dual in the pair c,d and hence the equation for the curvature tensor can be split into self-dual and anti-self-dual parts,

$$\tfrac{1}{2}R_{abcd} = \nabla_{[c}\Gamma_{d],\,ab}\,, \tag{4.5a}$$

$$0 = \nabla_{[c}\Gamma_{d].\,ab} + \Gamma^e_{[c|a}\Gamma_{d]eb}\,, \tag{4.5b}$$

where the $+$ and $-$ mean respectively the self- and anti-self-dual parts. Notice that the curvature tensor is a linear function of Γ.

We now make our last assumption, namely that N_a and M_a have the form

$$N_a = X_{,a} + AL_a + BM_a\,, \tag{4.6a}$$

$$\tilde{M}_a = Y_{,a} - BL_a - CM_a\,. \tag{4.6b}$$

(Once more, these can be proven from our basic assumption (3.1), though we will not do so. This point will be discussed later.) If (4.1) and (4.6) are inserted into (2.6) and the results compared with (4.4), we obtain the following relations

$$L^a\nabla_a A = -\beta\,, \tag{4.7a}$$

$$M^a\nabla_a B = \beta\,, \tag{4.7b}$$

$$L^a\nabla_a B = -\gamma\,, \tag{4.7c}$$

$$M^a\nabla_a C = \gamma\,, \tag{4.7d}$$

$$M^a\nabla_a A = \alpha\,, \tag{4.8a}$$

$$L^a\nabla_a C = \delta\,, \tag{4.8b}$$

$$\tilde{M}^a\nabla_a A + N^a\nabla_a B = 0\,, \tag{4.9a}$$

$$N^a\nabla_a C + \tilde{M}^a\nabla_a B = 0\,. \tag{4.9b}$$

If we think of scalar functions $\phi(z^a)$ as functions of the four scalar fields L, M, X, and Y, i.e.,

$$\phi(L, M, X, Y)\,,$$

then

$$\phi_{,a} = \phi_{,L}L_a + \phi_{,M}M_a + \phi_{,X}X_{,a} + \phi_{,Y}Y_{,a}\,.$$

From the orthogonality properties of L_a, N_a, M_a, $\tilde{M}_a$ we have

$$L^a\phi_{,a} = \phi_{,X}\,, \tag{4.10}$$

$$M^a\phi_{,a} = -\phi_{,Y}\,,$$

and hence (4.7) and (4.8) become

$$A_{,X} = -\beta\,, \tag{4.11a}$$

$$B_{,Y} = -\beta\,, \tag{4.11b}$$

$$B_{,X} = -\gamma\,, \tag{4.11c}$$

$$C_{,Y} = -\gamma\,, \tag{4.11d}$$

$$A_{,Y} = -\alpha\,, \tag{4.11e}$$

$$C_{,X} = \delta\,. \tag{4.11f}$$

We now obtain in a straightforward manner the Plebanski θ function.[4] From (4.11a and b) we have

$$A = \Omega_{,Y}\,, \tag{4.12a}$$

$$B = \Omega_{,X}\,, \tag{4.12b}$$

while from (4.11c and d)

$$B = \Lambda_{,Y}\,, \tag{4.13a}$$

$$C = \Lambda_{,X}\,. \tag{4.13b}$$

Equations (4.12b) and (4.13a) imply

$$\Omega = \theta_{,Y}\,, \tag{4.14a}$$

$$\Lambda = \theta_{,X}\,, \tag{4.14b}$$

and thus we have the existence of $\theta(X, Y, L, M)$ such that

$$A = \theta_{,YY}\,,\quad B = \theta_{,XY}\,,\quad C = \theta_{,XX}\,, \tag{4.15}$$

$$\alpha = -\theta_{,YYY}\,,\quad \beta = -\theta_{,XYY}\,,\quad \gamma = -\theta_{,XXY}\,,$$

$$\delta = \theta_{,XXY}\,. \tag{4.16}$$

The remaining Eqs. (4.9) are conditions on θ and become

$$\Delta_{,X} = \Delta_{,Y} = 0 \rightarrow \Delta = \Delta(L, M)$$

with

$$\Delta = \theta_{,LX} - \theta_{,MY} - \theta_{,XX}\theta_{,YY} + \theta_{,XY}\theta_{,XY}\,.$$

With no loss in generality Δ may be taken to be zero by replacing, e.g.,

$$\theta \rightarrow \theta - X\int \Delta dL$$

and thus

$$\theta_{,LX} - \theta_{,MY} - \theta_{,XX}\theta_{,YY} + \theta^2_{,XY} = 0, \tag{4.17}$$

which is Plebanski's "second fundamental" equation.[4]

Returning to Eqs. (4.6) and (4.1) $g_{ab} = e^{\mu}_{a}e_{\mu b}$ becomes

$$g_{ab} = \eta_{ab} + 2H_{ab}\,, \tag{4.18}$$

$$\eta_{ab} = L_{,a}X_{,b} + L_{,b}X_{,a} - M_{,a}Y_{,b} - M_{,b}Y_{,a}\,, \tag{4.19}$$

$$H_{ab} = AL_aL_b + B(L_aM_b + M_aL_b) + CM_aM_b\,. \tag{4.20}$$

We conclude by showing the existence of a remarkable series of "Hertz-like" potentials and the relationships between them.[9]

Defining

$$\Theta_{abcd} = 4\Theta L_{[a}M_b\,L_{[c}M_{d]}\,, \tag{4.21}$$

$$\gamma_{abc} = 2L_{[b}M_{c]}(\Theta_Y L_a - \Theta_X M_a)\,, \tag{4.22}$$

one can prove, by directly applying the indicated operations and using relations already derived, the following

$$\gamma_{cab} = \nabla^d\Theta_{abcd}\,, \tag{4.23}$$

$$H_{ab} = \nabla^c\gamma_{abc}\,, \tag{4.24}$$

$$\nabla_{[c}H_{a]b} = \tfrac{1}{2}\Gamma_{bca} + H_{[a}{}^{e}\Gamma_{c]eb}\,, \tag{4.25a}$$

which decomposes into

$$\nabla_{[c}H_{a],\,b} = \tfrac{1}{2}\Gamma_{bca}\,, \tag{4.25b}$$

$$\nabla_{[c}H_{a]\,\,b} = H_{[a}{}^{e}\Gamma_{c]eb}\,, \tag{4.25c}$$

and finally

$$\tfrac{1}{2}R_{abcd} = \nabla_{[c}\Gamma_{d],\,ab}\,, \tag{4.5a}$$

$$\nabla_{[c}\Gamma_{d]\,\,ab} = \Gamma_{[d}{}^{e}|_a\Gamma_{c]eb}\,. \tag{4.5b}$$

V. DISCUSSION

For didactic reasons, we have given the development simply in terms of a series of assumptions, which are justified in different ways by the different approaches to the study of left-flat spaces. Penrose[3] and Plebanski[4] begin by assuming (3.9) and remark that this is the integrability condition for covariantly constant anti-self-dual bivectors. Penrose proceeds to construct a curved twistor space from the resulting completely null 2-surfaces. Plebanski justifies the choice of L_a and M_a as gradients, (4.1), and the choice (4.6) for N_a and M_a by verifying that the integrability conditions for this too are satisfied. In the $\mathscr{H}$-space approach[1,2,9,10] one is presented at once with a tetrad satisfying (4.1) and (4.6). These arise as follows: We recall that the $\mathscr{H}$-space of an asymptotically flat space-time M is the four-dimensional complex manifold of regular solutions to the good cut equation, a second-order nonlinear differential equation,

$$\eth^2 Z \equiv [(1 + \zeta\bar{\zeta})^2 Z_{,\zeta}]_{,\zeta} = \sigma^{\circ}(Z, \zeta, \bar{\zeta})\,. \tag{5.1}$$

The function $\sigma^{\circ}(u,\zeta,\bar{\zeta})$ is the data for the asymptotic radiation field of M.

Writing the general regular solution of (5.1) as $Z(z^a,\zeta,\bar{\zeta})$, [the (z^a) are coordinates on $\mathscr{H}$-space] a quadratic metric is defined on $\mathscr{H}$-space in terms of solutions of the linearization of (5.1).[1,2] Four covector fields are defined on $\mathscr{H}$-space by $Z_{,a}$, $\eth Z_{,a}$, $\bar{\eth}Z_{,a}$, and $\eth\bar{\eth}Z_{,a}$ at fixed but arbitrary ζ and $\bar{\zeta}$. These are all gradients and the first pair turns out to be null and orthogonal. Taking these for L_a and M_a satisfies (4.1). The tetrad can then be completed by taking M_a and N_a to be the second pair with appropriate linear combinations of

L_a and M_a. A different choice of ζ and $\tilde{\zeta}$ changes the tetrad while preserving conditions (4.1), (4.6), and the normalization. However, as ζ changes, the bivector $Z_{\cdot[a}\delta Z_{\cdot b]}$ sweeps out all possible anti-self-dual bivectors. Thus (3.1) and (3.7) will follow if $Z_{\cdot[a}\delta Z_{\cdot b]}$ is constant at all values of ζ. That this is constant is shown in Ref. 2.

In this paper we have attempted to give the relations between the various approaches to left-flat spaces in an elementary form. Penrose's construction gives all left-flat spaces at least locally. Similarly, all left-flat spaces are locally given by solutions of Plebanski's equations (4.17). It is an open question whether all left-flat spaces are locally $\mathscr{H}$-spaces [that is, whether they can be expressed in terms of the solution of a good-cut equation (5.1)].

[1]E.T. Newman in *General Relativity and Gravitation*, edited by G. Shaviv and J. Rosen, (Wiley and Sons, New York, 1975).
[2]M. Ko, E.T. Newman, and K.P. Tod, in *Asymptotic Structure of Space-Time*, edited by P. Esposito and L. Witten (Plenum Press, N.Y., 1976).
[3]R. Penrose, Gen. Relativ. Gravit. **7**, 31 (1976).
[4]J. Plebanski, J. Math. Phys. **16**, 2395 (1975).
[5]R.S. Ward, Proc. R. Soc. (Lond.), to appear.
[6]S.W. Hawking, Phys. Lett. A **60**, 81 (1977).
[7]K.P. Tod and R.S. Ward, Proc. R. Soc. (Lond.) to appear.
[8]J. Plebanski and S. Hacyan, J. Math. Phys. **16**, 2403 (1975).
[9]M. Ko, M. Ludvigsen, E.T. Newman, and K.P. Tod, "The Theory of $\mathscr{H}$ space," University of Pittsburgh preprint.
[10]R. Hansen, E.T. Newman, R. Penrose, and K.P. Tod, Proc. R. Soc. (Lond.) A **363**, 445 (1978).
[11]E. Flaherty, Gen. Relativ. Gravit. **9**, 961 (1978).

The conformal geometry of Minkowski space: Hyperbolic world lines

A. J. Briginshaw
Department of Mathematics, The City University, London EC1V 4PB, England

(Received 13 June 1979; accepted for publication 2 October 1979)

There is a discussion of (i) the transformation properties of hyperbolic hypersurfaces in Minkowski 4-space, (ii) the association of hyperbolic world lines in Minkowski 2-space with uniformly accelerated motion.

1. INTRODUCTION

Since the work of Cunningham and Bateman[1-3] it has been known that the full conformal group is the largest invariance group for a set of electromagnetic phenomena. From time to time there have been attempts to associate the same group with uniformly accelerated motion. An early attempt of Page[4] was subjected to criticism by Robertson,[5] who suggested that the correct framework for the investigation was that of general relativity. A paper of Page and Adams[6] concerning the transformation group for the electromagnetic field equations simply confirmed the results of Cunningham and Bateman, but Page's association of the special conformal group with the kinematics and dynamics of flat space–time was left open to question. In papers published in 1945 and 1947, Hill[7,8] reiterated the association of uniform acceleration with the special conformal group, but Kastrup[9,10] and Rohrlich[11] have given reasons for considering the association inappropriate. In the present work, we confirm that the special conformal group is an invariance group for the hyperbolic hypersufaces $x.x = x.b$. In order to obtain the widest possible invariance group which is physically relevant, we shall exclude the time reversal symmetry on the grounds that it is a global (and local) causality violator, and discuss the inclusion of the special conformal transformations which are global but not local causality violators. With a system that is specifically designed to cater for the world line of a particle, we are then able to confirm that a hyperbolic world line in two-dimensional Minkowski space is that appropriate to a uniformly accelerated material particle, and nominate the maximal invariance group for this behavior.

Notation: M_4 is the collection of quadruples of real numbers of the type (x_0, x_1, x_2, x_3) in which the inner product

$$x.y = x_0 y_0 - x_1 y_1 - x_2 y_2 - x_3 y_3 ,$$

is defined.

A hypersurface with equation $x.x = x.b$, where b has constant components is said to be hyperbolic; for reasons of physical interpretation we confine our attention to the case in which $b.b = -k^2$, $k \in R$. The inner product induces two relations[12] in M_4; we write $x < y$ iff $(x-y).(x-y) > 0$ and $x_0 < y_0$ ($<$ is a partial order); $x <. y$ iff $(x-y).(x-y) = 0$ and $x_0 < y_0$. ($<.$ is not a partial order.)

2. CONFORMAL MAPS OF M_4

A. The special conformal group and causality

If $f: M_4 \to \bar{M}_4$ (compactified M_4) and A is a region of M_4 which contains no singular points, f is said to be a causal bijection of A if f, f^{-1} preserve $<.$ (or $<$).[12] If A is the whole of M_4 it is known that f is an element of the affine group generated by the inhomogeneous orthochronous Lorentz group and dilatations.[12]

A map of M_4 defined by

$$x' = \frac{x + (x.x)a}{1 + 2(x.a) + (x.x)(a.a)} ,$$

is said to be a special conformal transformation.

The collection of such transformations is an Abelian subgroup of the conformal group.

If we write $\Delta_x \equiv 1 + 2(x.a) + (x.x)(a.a)$ it is easy to see that

$$(\xi' - \eta').(\xi' - \eta') = \frac{(\xi - \eta).(\xi - \eta)}{\Delta_\xi \Delta_\eta} .$$

Thus the map of a null line (punctured at a singular point) is another (punctured) null line; even so, if the points ξ, η are separated by the singular hypersurface $\Delta_x = 0$, the relation $\xi <. \eta$ is not preserved.

We may however speak of local preservation of $<$ within a neighborhood of $x = 0$. (Although there are difficulties with the preservation of $<.$ there, because for the space–time inversion that generates the special conformal transformation, the cone $x.x = 0$ is singular.)

3. BEHAVIOR OF THE HYPERBOLIC HYPERSURFACES UNDER LINEAR TRANSFORMATION

It is clear that the hypersurfaces $\{x : x.x = x.b, b.b = -k^2\}$ are mapped onto each other by elements of the group generated by the Lorentz transformations and dilatations. For reasons of causality preservation we shall exclude the nonorthochronous transformations and note that the sets $\{x : x.x = x.b, b.b = -k^2, x_0 \geqslant 0\}$ are similarly mapped onto each other using orthochronous Lorentz transformations and dilatations.

4. TRANSFORMATION PROPERTIES UNDER SPECIAL CONFORMAL TRANSFORMATION

Lemma: Under special conformal transformation, the set $\{x : x.x = x.b, \Delta_x \neq 0, b.b = -k^2\}$ is mapped onto $\{x' : x'.x' = x'.\beta$, where $\beta = (b/[1 + a.b])\}$, provided that $a.b \neq -1$.

Proof: We have

$$\Delta_x x'.x' = x.x ,$$

 0022-2488/80/040878-03$1.00

and

$$\Delta_x x'.b = x.b + (x.x)(a.b)\,;$$

thus, given that $a.b \neq -1$, the points of $x.x = x.b$ for which $\Delta_x \neq 0$ are mapped onto $x'.x' = x'.\beta$, where $\beta = (b/[1 + a.b])$, as required.
In addition,

$$\beta.\beta = \frac{b.b}{(1 + a.b)^2} < 0\,.$$

Remark: If we keep the restriction $a.b \neq -1$ and confine our attention to the region of the local causality preservation[13] we have the following corollary.

Corollary: The hypersurface fragments $\{x : x.x = x.b, b.b = -k^2\}$ which lie in the region of local causality preservation are mapped onto each other by elements of the special conformal group, provided that $a.b \neq -1$.

5. REMARK

If we choose a such that $a.b = -1$, then the set $\{x : x.x = x.b, \Delta_x \neq 0, b.b = -k^2\}$ is mapped onto $\{x' : x'.b = 0\}$.

It is clear, on the other hand, that every hyperplane $\{x : x.b = 0, b.b = -k^2\}$ is mapped onto $\{x' : x'.b' = 0, b'.b' = -k^2\}$ under a Lorentz transformation, with a similar preservation for dilatations. It is also easy to see that $\{x : x.b = 0, \Delta_x \neq 0, b.b = -k^2\}$ is mapped onto $\{x' : x'.x' = x'.b/a.b\}$ under special conformal transformation, provided that $a.b \neq 0$, and onto $\{x' : x'.b = 0, b.b = -k^2\}$ under a special conformal transformation for which $a.b = 0$.

We may refer to the set of all hyperbolic hypersurfaces and all homogeneous hyperplanes as H, and the collection of those parts of the ingredients of H for which $x_0 \geqslant 0$ as H^+.

6. THE MAXIMAL INVARIANCE GROUP, I

Let $L^\uparrow$ be the orthochronous Lorentz group for M_4, S_4 be the corresponding group of special conformal transformations, and G be the group generated by S_4 and $L^\uparrow$. Let I be the maximal invariance group for the collection $H^\uparrow$ which also effects local preservation for $x_0 \geqslant 0$; it is clear that $G \subset I$.

In addition every element of I preserves the asymptotic surface $x.x = 0$ and, from Bateman,[1] we have $I \subset G$. Thus we have the following theorem.

Theorem: The maximal invariance group for local preservation of the collection $H^\uparrow$ is that generated by S_4 and L

7. KINEMATICS OF ONE-DIMENSIONAL MOTION

A. Definitions

For a particle in one-dimensional motion we shall identify three systems of coordinates:

(i) a system in which the particle is instantaneously at rest is called a *rest-system*;[7]

(ii) a system in which the particle has zero acceleration instantaneously is called a *zero-acceleration system*;

(iii) a system in which, instantaneouly, the particle has zero acceleration and is also at rest is called a *proper-system*.[7]

We may use the definition given by Pauli[14] and Hill[7]: "A particle is said to be uniformly accelerated if its acceleration in every successive rest system is constant."

B. Interpretation of k

We may now confirm that the world line $x.x = x.b$ with $b.b = -k^2$ and $x_0 \geqslant 0$ is, as far as M_2 is concerned, the history of a uniformly accelerated material particle. Suppose that the particle passes through 0 with speed $u \neq 0$, then $b_0 \neq 0$ and $u = (b_0/b_1)$.

Now consider the Lorentz transformation

$$X_0 = \frac{x_0}{\sqrt{1-u^2}} - \frac{u}{\sqrt{1-u^2}} x_1\,,$$

$$X_1 = \frac{x_1}{\sqrt{1-u^2}} - \frac{u}{\sqrt{1-u^2}} x_0\,;$$

the map of $x.x = x.b$ is $X.X = X.B$, where $B_0 = 0$ and $B_1 = [\sqrt{(1-u^2)}]\, b_1 \neq 0$.

This is the form of the world line in Pauli,[14] and if K is the magnitude of the uniform acceleration, then

$$1/K^2 = \tfrac{1}{4}B_1^2 = \tfrac{1}{4}k^2\,,$$

so that $K = 2/k$.

However, if we consider a special conformal map of $x.x = x.b$, we obtain $x'.x' = x'.(b/[1 + a.b])$, in general. In this frame, the particle passes 0′ with speed (b_0/b_1), as before, but the magnitude of the uniform acceleration is now $K' = (2/k)(1 + a.b)$.

C. The zero-acceleration and proper systems

If $x.x = x.b$ is a given world line we may obtain the zero-acceleration system by using the special conformal transformation

$$X = \frac{(b.b)x - (x.x)b}{b.b - 2x.b + x.x}$$ (subject to the usual nonsingularity restrictions),

which gives a straight line history in the new system,

$$X.b = 0\,.$$

Thus, in the new system, the particle has constant speed (b_0/b_1), and zero acceleration. In addition, the proper system is obtained by applying the appropriate orthochronous Lorentz transformation to the zero-acceleration system.

D. Remarks

Unbounded uniform acceleration: It is clear from Sec. 3 that super light speeds are not attained for an unbounded uniform acceleration, since the limiting situation as $K \to \infty$ is the light cone as a degenerate hyperbola.

Zero uniform acceleration: We now examine the limiting case $K \to 0$; let $x.x = x.b$ be the usual hyperbolic world line, and let $u = (b_0/b_1)$ be the speed with which the particle passes 0. From $(x_0^2 - x_1^2)/b_1 = (b_0/b_1)\, x_0 - x_1$ we let $k \to \infty$ in such a way that $(b_0/b_1) = u$, and we obtain $x_1 = u x_0$ as the limiting equation for the world line, corresponding to the case of zero uniform acceleration.

Subgroups of the maximal invariance group: In Sec. 2 we have obtained the maximal invariance group for the local

preservation of the collection $H^{\dagger}$ of lines and hyperbolas. In the kinematics of one-dimensional motion, which is modelled by M_2, it is the nature of the whole group $R^{+} \circledS [L \circledS S_2]$ to preserve the local *uniformity* of acceleration; the whole group does not itself preserve *zeroness* and *nonzeroness*. To preserve zeroness and uniformity (and global causality) we are confined to the subgroup $R^{+} \circledS L$.

E. Interpretation of a

Let (x_0, ξ_1), with ξ_1 fixed, be the representative point of a given straight world line in the coordinate system associated with 0. It is not difficult to establish that the map of this line under

$$x' = \frac{x + (x.x)a}{1 + 2(x.a) + (x.x)(a.a)},$$

is the hyperbola (punctured)

$$(x'_0 - \beta_0)^2 - (x'_1 - \beta_1)^2 = -\mu^2,$$

where

$$\mu = \frac{1}{2[(a.a)\xi_1 + a_1]}, \quad \beta_0 = 2\mu a_0 \xi_1, \quad \beta_1 = \mu(2a_1\xi_1 - 1).$$

We may therefore identify this world line as being associated with a uniform acceleration $\pm 2[(a.a)\xi_1 + a_1]$ in the frame of 0′. Furthermore if we let $\xi_1 \to 0$ we obtain the uniform acceleration of 0 relative to 0′ viz $\pm 2a_1$. There is, however, no simple law of composition for uniform accelerations without certain conventions. For, if $x.x = x.b$ is the world line associated with a uniformly accelerated particle P in the frame 0, we have a transformation to the system in which the world line is $X.X = X.B$, with $B_0 = 0$ and $B_1 \neq 0$, so that the uniform acceleration of P relative to 0 is $(2/B_1)$. On the other hand, if we make the transformation

$$x' = \frac{x + (x.x)a}{1 + 2(x.a) + (x.x)(a.a)},$$

first, which is equivalent to giving 0 a uniform acceleration of magnitude $2a_1$ relative to 0′, then the uniform acceleration of P relative to 0′ is $(2/B_1)(1 + a.b)$. Only if 0, 0′ both agree to reduce a given particle P to zero initial speed before applying the special conformal transformation can we restrict our attention to world lines

$$x_0^2 - x_1^2 = -b_1 x_1,$$

and

$$(x'_0)^2 - (x'_1)^2 = -\beta_1 x'_1.$$

In this case, the acceleration of P relative to 0 is $K = (2/b_1)$, the acceleration of P relative to 0′ is $K' = (2/b_1)(1 - a_1 b_1)$, and we have the additive law $K' = K + 2a_1$. We must now adopt the convention that the acceleration of 0 relative to 0′ is $-2a_1$ (NOT $+2a_1$) to give a sensible interpretation of this result. With the same convention, we also have consistency regarding the proper system.

[1] H. Bateman, Proc. Lond. Math. Soc. **7**, 70 (1909).
[2] E. Cunningham and H. Bateman, Proc. Lond. Math. Soc. **8**, 77 (1910).
[3] H. Bateman, Proc. Lond. Math. Soc. **8**, 223, (1910).
[4] L. Page, Phys. Rev. **49**, 254 (1936).
[5] H. Robertson, Phys. Rev. **49**, 755 (1936).
[6] L. Page and N. Adams, Phys. Rev. **49**, 467 (1936).
[7] E.L. Hill, Phys. Rev. **67**, 358 (1945).
[8] E.L. Hill, Phys. Rev. **72**, 143 (1947).
[9] H.A. Kastrup, Phys. Rev. **143**, 1021 (1966).
[10] H.A. Kastrup, Phys. Rev. **150**, 1183 (1966).
[11] F. Rohrlich, Ann. Phys. **22**, 169 (1963).
[12] E.C. Zeeman, J. Math. Phys. **5**, 490 (1964).
[13] J. Wess, Nuovo Cimento **18**, 1090 (1960).
[14] W. Pauli, *Theory of Relativity* (Pergamon, New York, 1958), p. 75.

The Ising limit of the double-well model

Florin Constantinescu[a] and Berthold Ströter
Institut für Angewandte Mathematik der Johann Wolfgang Goethe-Universität, Frankfurt am Main 6, Frankfurt am Main, Robert-Mayer Str. 10, W. Germany

(Received 13 February 1979)

We prove that the pure phase correlation functions of the double-well model at small temperature in an appropriate limit case tend to the corresponding correlation functions of the classical Ising model in any dimension, extending results previously obtained by Rosen (see introduction). We use expansion methods proposed by Glimm, Jaffe, and Spencer.

1. INTRODUCTION

Let us consider a statistical lattice system in ν dimensions with continuous spin distribution given by $\lambda(x^2-1)^2$ with $\lambda>0$ (double-well potential function). Because of the relation

$$\lim_{\lambda\to\infty}\frac{\exp[-\lambda(x^2-1)^2]}{\int_{-\infty}^{\infty}\exp[-\lambda(x^2-1)^2\,dx]}=\tfrac{1}{2}[\delta(x+1)+\delta(x-1)]\,, \tag{1.1}$$

it is generally accepted that the double-well model goes into the Ising model in the large coupling constant limit $\lambda\to\infty$. Nevertheless, a rigorous proof of this result is not entirely trivial. Let us remark first that in a different context the statistical double-well model in one dimension was studied by Kac and Thomson.[1] A detailed discussion of the Ising and scaling limit of the double-well model in constructive quantum field theory was given by Glimm and Jaffe.[2] Based on a conjecture in Ref. 2 and using results from Ref. 1 Isaacson[3] proved that the scaling limit of the double-well model in one dimension is the continuous Ising model (Poisson processes). In the case of several dimensions Rosen[4] proved that in the presence of a nonzero external magnetic field or for large temperature the double-well model goes into the Ising model for $\lambda\to\infty$. In the case of zero magnetic field and small temperature, Rosen proved convergence of the two dimensional Dirichlet double-well model towards the Ising model.

In this paper we prove that the correlation functions of the pure $\pm$ phases of the double-well model converge to the corresponding correlation functions of the Ising model in any dimension. Our technique is the mean field (contour-cluster) expansion of Glimm, Jaffe, and Spencer.[5] Let us remark that from the technical point of view our expansion (on the lattice) is simpler than that in Ref. 5. In the process of expansion we translate the field exactly to one or another minimum of the potential. This fact simplifies the expansion. We expect further results concerning an expansion of the double-well model "near" the Isling model (expansion in powers of $1/\lambda$), Borel summability problems, etc.

2. THE DOUBLE-WELL MODEL

We consider the unit lattice $\mathbb{Z}^\nu$ in $\nu\geqslant 2$ dimensions. We leave out for the moment the temperature and consider the formal model

$$\exp[-(1/2)(q,(-\Delta)q)]\prod_i \exp[-P(q_i)\,dq_i]\,, \tag{2.1}$$

where

$$P(q)=\lambda(q^2-m^2/8\lambda)^2\,, \tag{2.2}$$

with λ and m^2 two positive constants. In (2.1) Δ is the lattice Laplacian and $(\cdot,\cdot)$ the scalar product in $l^2(\mathbf{Z}^\nu)$. Using standard notation (see Ref. 6) we have

$$\Delta(j_1,j_2)=\begin{cases}-2\nu, & j_1=j_2,\\ 1, & |j_1-j_2|=1,\\ 0, & |j_1-j_2|>1.\end{cases} \tag{2.3}$$

The minima of the spin distribution function $P(q)$ (double-well potential) are located at

$$\pm\xi=\pm m(1/8\lambda)^{1/2}.$$

The rigorous model with lattice cutoff Λ and $+$ boundary conditions is given by the measure on $l^2(\mathbf{Z}^\nu)$

$$dq_\Lambda=\exp[-P_\Lambda(q)+(m^2/2)(q-\xi)_\Lambda^2]\,d(q-\xi), \tag{2.4}$$

where

$$P_\Lambda(q)=\sum_{i\in\Lambda}P(q_i),\quad (q-\xi)_\Lambda^2=\sum_{i\in\Lambda}(q_i-\xi)^2\,,$$

and $d(q-\xi)$ is the Gaussian measure on $l^2(\mathbf{Z}^\nu)$ with mean ξ and covariance

$$A^{-1}=C=(-\Delta+m^2)^{-1}\,. \tag{2.5}$$

We take $\Lambda\subset\mathbf{R}^\nu$ such that $\partial\Lambda$ lies on the dual $(\mathbf{Z}^\nu)^*$. We remark that formally $d(q-\xi)$ contains a factor $\exp(-m^2/2)(q-\xi)_{\mathbf{Z}^\nu}^2$ which cancels $\exp(m^2/2)(q-\xi)_\Lambda^2$. We are left with the double-well potential factor inside Λ, whereas outside Λ there is a Gaussian tendency towards the $+$ minimum.

A similar definition can be given for the $-$ boundary conditions model.

Consider now an observable which is a product of monomials in q at some lattice points

$$A\equiv A(q).$$

We introduce the following standard notation

$$Z_\Lambda=Z(\Lambda)=\int dq_\Lambda\,,\quad Z(\Lambda)>0,$$

[a]Supported in part by the Deutsche Forschungsgemeinschaft.

 0022-2488/80/040881-10$1.00

$$\langle A \rangle_\Lambda = Z(\Lambda)^{-1} \int A \, dq_\Lambda \,, \tag{2.6}$$

where Λ contains the localization of A. In the first part of this paper we give a convergent expansion for $\langle A \rangle_\Lambda$ in the case of λ small and m^2 large.

3. PROJECTION ON THE CLASSICAL ISING VARIABLES AND THE CONTOUR EXPANSION

Let

$$\sigma_i = \operatorname{sgn} q_i, \quad i \in \Lambda. \tag{3.1}$$

Outside Λ we impose $+$ boundary conditions: $\sigma_i = +$, $i \notin \Lambda$. We denote by $\Sigma = \{\sigma_i\}$ a configuration on the classical Ising variables in Λ. The set of phase boundaries (contours) in the configuration Σ will be also denoted by Σ. Let

$$\chi_+(\zeta) \equiv \chi_{+1}(\zeta) = \chi_-(-\zeta) \equiv \chi_{-1}(\zeta) = \pi^{-1/2} \int_0^\infty \exp[-(\zeta - z)^2] \, dz. \tag{3.2}$$

We have

$$\chi_+(\zeta) + \chi_-(\zeta) = 1.$$

Let us define

$$\chi_\Sigma = \prod_{i \in \Lambda} \chi_{\sigma_i}(q_i). \tag{3.3}$$

It is easy to prove that

$$\sum_\Sigma \chi_\Sigma = 1. \tag{3.4}$$

For the convenience of the reader we remark that $\chi_\pm$ are regularized characteristic functions of the intervals $(0, \infty)$ and $(-\infty, 0)$. From (3.4) follows

$$dq_\Lambda = \sum_\Sigma dq_{\Lambda,\Sigma} \,, \tag{3.5}$$

where

$$dq_{\Lambda,\Sigma} = \chi_\Sigma \, dq_\Sigma. \tag{3.6}$$

This is the contour expansion (on the Ising variables). In Sec. 4 we prepare for the cluster expansion.

4. TRANSLATION OF COORDINATES

Let Σ be given. Our system consists of a set of islands of $+$ and $-$ spin coordinates separated by the contour Σ. On islands we translate the coordinates to the corresponding minimum of the potential. This produces sharp discontinuities on the phase boundaries. Let

$$h_i = \sigma_i \xi, \quad \psi_i = q_i - h_i. \tag{4.1}$$

The translated Guassian measure $d\psi = d(q - h)$ is equivalent to the Gaussian measure $d(q - \xi)$. For the Radon–Nykodym derivative we have the formula

$$\frac{d(\phi + f)}{d(\phi)} = \exp\{-[\tfrac{1}{2}(f, (-\Delta + m^2) f) + (\phi, (-\Delta + m^2) f)]\}.$$

From this formula we get

$$\frac{d(q - h)}{d(q - \xi)} = \exp\{-[\tfrac{1}{2}((\xi - h), A(\xi - h)) + ((q - \xi), A(\xi - h))]\}.$$

From (2.4) we get

$$dq_\Lambda = e^{-F} \exp[-P_\Lambda(q) + (m^2/2)(q - h)_\Lambda] \, d\psi = e^{-F} \exp(-\lambda \psi^4 - m\sqrt{2\lambda}\, \psi_\Lambda^3) \, d\psi, \tag{4.2}$$

where $d\psi$ is the Gaussian measure on $l^2(\mathbb{Z}^\nu)$, with mean zero and covariance $(-\Delta + m^2)^{-1}$ and

$$F = F(\mathbb{R}^\nu, \Sigma) = -(\psi, \Delta h) - \tfrac{1}{2}(h, \Delta h). \tag{4.3}$$

We introduce some more notation

$$Q = Q(\mathbb{R}^\nu, \Sigma) = V(\mathbb{R}^\nu, \Sigma) + F(\mathbb{R}^\nu, \Sigma),$$
$$V = V(\mathbb{R}^\nu, \Sigma) = P_\Lambda(q) - \tfrac{1}{2} m^2 (q - h)_\Lambda = \lambda \psi_\Lambda^4 + m(2\lambda)^{1/2} \psi_\Lambda^3 \,.$$

Then (4.2) gets the form

$$dq_\Lambda = e^{-Q(\mathbb{R}^\nu, \Sigma)} \, d\psi. \tag{4.4}$$

The Λ dependence of Q is not explicit.

5. THE CLUSTER EXPANSION

Let l, L be integers such that $L \gg l \gg 1$. For the purposes of this paper we choose (see Ref. 5)

$$l \approx |\log(\lambda / m^2)|^{1/4} \quad \text{and} \quad L \approx |\log(\lambda / m^2)|^2.$$

We perform the cluster expansion on islands of the lattice $l\mathbb{Z}^\nu$ far away (distance L) from phase boundaries. Let us give some details. With Σ given, the standard cluster expansion[7] is

$$dq_{\Lambda,\Sigma} = \sum_{\Gamma \subset \mathscr{B}(\Sigma)} dq_{\Lambda,\Sigma,\Gamma} \,, \tag{5.1}$$

$$dq_{\Lambda,\Sigma,\Gamma} = \int_0^1 ds_{\Gamma^c} \, \partial_s^{\Gamma^c \subset \mathscr{B}(\Sigma)} \chi_\Sigma \, e^{-Q(\mathbb{R}^\nu, \Sigma)} d\psi(s), \tag{5.2}$$

where

$$s_\Gamma = \{s_b \mid b \in \Gamma\},$$
$$ds_{\Gamma^c} = \prod_{b \in \Gamma^c} ds_b \Big|_{s_{\Gamma^c} = 0},$$
$$\partial_s^\Gamma = \prod_{b \in \Gamma} \frac{\partial}{\partial s_b},$$

and $d\psi(s)$ is the Gaussian measure on $l^2(\mathbb{Z}^\nu)$ with covariance

$$C(s) = \sum_{\Gamma \subset \mathscr{B}(\Sigma)} \left(\prod_{b \in \Gamma^c} s_b \right) \left(\prod_{b \in \Gamma} (1 - s_b) \right) C_\Gamma \,. \tag{5.3}$$

For given Σ, $\mathscr{B}(\Sigma)$ is the set of hyperbonds b of the dual lattice $(l\mathbb{Z}^\nu)^*$ on which we perform the cluster expansion. For $\Gamma \subset \mathscr{B}(\Sigma)$, C_Γ is defined as follows:

$$C_\Gamma = (-\Delta_\Gamma + m^2)^{-1}. \tag{5.4}$$

In (5.4) Δ_Γ is the lattice Laplacian in which we break connections across Γ (for some notational details see Ref. 8). Finally $s_b \in [0,1]$ are interpolating parameters, $s = \{s_b\}$ and $\Gamma^c = \mathscr{B}(\Sigma) - \Gamma$. Let us introduce some more notation. If W denotes a set in $\mathbb{R}^\nu$ whose boundary ∂W lies on the dual lattice $(\mathbb{Z}^\nu)^*$, then we denote by $W_\#$ the set of sites of $\mathbb{Z}^\nu$ inside W. Conversely, if $W_\#$ is a set of sites of $\mathbb{Z}^\nu$, we denote by W a set of smallest volume in $\mathbb{R}^\nu$ such that ∂W lies on $(\mathbb{Z}^\nu)^*$ and $W_\# \subset W$. Let $\partial_\#$ and $d_\#$ be the lattice boundary and the lattice distance, respectively. We define

$$X_{\#}^{\pm} = \{ i \in \mathbb{Z}^\nu | \sigma_i = \pm, d_{\#}(i, \partial_{\#} \Sigma) \geqslant L \}, \quad X = X^+ U X^-. \tag{5.5}$$

Now we can define the set $\mathscr{B}(\Sigma)$ on which the cluster expansion is performed. It consists of $(l\mathbb{Z}^\nu)^*$ lattice hyperbonds contained in the interior of X. In the Sec. 6 we give the full mean field (contour-cluster) expansion.

6. THE CONTOUR-CLUSTER EXPANSION

We combine the two expansions of Secs. 3 and 5 to get for the observable A:

$$\langle A \rangle_\Lambda = Z_\Lambda^{-1} \sum_{\Sigma} \sum_{\Gamma \subset \mathscr{B}(\Sigma)} R''(A, \mathbb{R}^\nu, \Sigma, \Gamma) \tag{6.1}$$

where

$$R''(A, \mathbb{R}^\nu, \Sigma, \Gamma) = \int_0^1 ds_{\Gamma^c} \, \partial_s^{\Gamma^c} \int \chi_\Sigma A e^{-Q(\mathbb{R}^\nu, \Sigma)} d\psi(s). \tag{6.2}$$

By the definition of ds_{Γ^c}, the measure $d\psi(s)$ has Dirichlet conditions on Γ.

The formula (6.1) is known as mean field (or contour-cluster) expansion and was introduced in the context of constructive quantum field theory by Glimm, Jaffe, and Spencer.[5] We note that in our case the measure $d\psi(s)$ has the lattice structure. We now perform two resummations in the spirit of Ref. 5. For the convenience of the reader we give brief details. Let $\{ Y_1, \ldots, Y_\mu \}$ denote connected components of $\mathbb{R}^\nu - \Gamma$. It is not difficult to show that there is factorization

$$R''(A, \mathbb{R}^\nu, \Sigma, \Gamma) = \prod_{k=1}^{\mu} R''(A_k, Y_k, \Sigma_k, \Gamma_k), \tag{6.3}$$

where $\Sigma_k = \Sigma \cap Y_k$, $\Gamma_k = \Gamma \cap Y_k$ and A_k has an obvious meaning. In (6.3) we have

$$R''(A_k, Y_k, \Sigma_k, \Gamma_k) = \int_0^1 ds_{\Gamma^c} \, \partial_s^{\Gamma^c \subset \mathscr{B}(\Sigma, Y_k)} \times \int \chi_{\Sigma_k} A_k e^{-Q(Y_k, \Sigma_k)} d\psi(s), \tag{6.4}$$

where $\mathscr{B}(\Sigma, Y_k) = \mathscr{B}(\Sigma) \cap Y_k$ and

$$Q(Y_k, \Sigma_k) = V(Y_k, \Sigma_k) + F(Y_k, \Sigma_k), \tag{6.5}$$

$$V(Y_k, \Sigma_k) = P_{\Lambda \cap Y_k}(q) - \tfrac{1}{2} m^2 (q - h)^2_{\Lambda \cap Y_k}, \tag{6.6}$$

$$F(Y_k, \Sigma_k) = -(\psi_{Y_k}, \Delta h) - \tfrac{1}{2}(h_{Y_k}, \Delta h). \tag{6.7}$$

We note that in (6.4) it is not necessary to restrict $d\psi(s)$ to Y_k because of the Dirichlet conditions on Γ.

Now let $Y = Y(\Gamma)$ be the closure of the union of Y_k which meets localization of A (in fact it is not necessary to take the closure as the reader can convince himself). We isolate A from the partition function in $Y^c = \mathbb{R}^\nu - Y$. Because of factorization we have

$$R''(A, \mathbb{R}^\nu, \Sigma, \Gamma) = R''(A, Y, \Sigma \cap Y, \Gamma \cap Y) = Z''(Y^c, \Sigma - Y, \Gamma - Y), \tag{6.8}$$

where

$$Z''(Y^c, \Sigma - Y, \Gamma - Y) = R''(A = I, Y^c, \Sigma - Y, \Gamma - Y). \tag{6.9}$$

We now perform two partial resummations of (6.1).

The first resummation: For Σ given we hold $Y = Y(\Gamma)$ fixed and sum up over all Γ compatible with $Y = Y(\Gamma)$:

$$\sum_{\Gamma \subset \mathscr{B}(\Sigma)} = \sum_{\partial Y} \sum_{\{ \Gamma \subset \mathscr{B}(\Sigma) | Y(\Gamma) = Y \}}$$

The inside sum (in Y) is

$$R'(A, Y, \Sigma \cap Y) = \sum_{\{ \Gamma \subset \mathscr{B}(\Sigma) | Y(\Gamma) = Y \}} R''(A, Y, \Sigma \cap Y, \Gamma \cap Y), \tag{6.10}$$

and the outside sum (in Y^c) is

$$Z'(Y^c, \Sigma - Y, \partial Y) = \sum_{\{ \Gamma \subset \mathscr{B}(\Sigma, Y^c) | Y(\Gamma) = Y \}} Z''(Y^c, \Sigma - Y, \Gamma - Y). \tag{6.11}$$

The outside sum can be evaluated in the manner of the cluster expansion and gives

$$Z'(Y^c, \Sigma - Y, \partial Y) = \int e^{-Q(Y^c, \Sigma - Y)} \chi_{\Sigma - Y^c} \, d\psi_{\partial Y}. \tag{6.12}$$

Here $d\psi_{\partial Y}$ is the Gaussian measure with Dirichlet conditions on ∂Y, and we used the fact that the constraint $Y(\Gamma) = \Gamma$ is equivalent to $\Gamma \supset \partial U$. After the first partial resummation we get

$$\langle A \rangle = \sum_{\Sigma} \sum_{\partial Y} R'(A, Y, \Sigma \cap Y) \frac{Z'(Y^c, \Sigma - Y, \partial Y)}{Z(\Lambda)}. \tag{6.13}$$

The second resummation: First we introduce some notation. Let

$$\partial Y^{\pm} = \partial Y \cap X^{\pm}.$$

We remark that each of $\partial Y, \partial Y^+$ and ∂Y^- consists of a finite number of nonoverlapping closed hypercontours. For a set $W_\#$ of sites in $\mathbb{Z}^\nu$, let $M(W_\#)$ be the set of sites in $\mathbb{Z}^\nu$ within distance L of W. Let

$$N(\partial Y^{\pm}) = M(\partial_\#(\partial Y^{\pm})).$$

The second resummation goes on with the constraint

$$\sum \restriction N(\partial Y^{\pm}) = \pm,$$

i.e.,

$$\sum_{\Sigma} \sum_{\partial Y} = \sum_{\partial Y'} \sum_{\{ \Sigma | \Sigma \restriction N(\partial Y') = \pm \}}$$

We define the inside sum

$$R(A, Y, \partial Y^{\pm}) = \sum_{\{ \Sigma \cap Y | \Sigma \restriction N(\partial Y') = \pm \}} R'(A, Y, \Sigma \cap Y), \tag{6.14}$$

and the outside sum

$$Z(Y^c, N(\partial Y^+), N(\partial Y^-), \partial Y) = \sum_{\{ \Sigma - Y | \Sigma \restriction N(\partial Y') = \pm \}} Z'(Y^c, \Sigma - Y, \partial Y) = \sum_{\{ \Sigma - Y | \Sigma \restriction N(\partial Y') = \pm \}} \int e^{-Q}(Y^c, \Sigma - Y) \, d\psi_{\partial Y}. \tag{6.15}$$

After the second partial resummation we get from (6.13)

$$\langle A \rangle = \sum_{\partial Y'} R(A, Y, \partial Y^{\pm}) \frac{Z(Y^c, N(\partial Y^+), N(\partial Y^-), \partial Y)}{Z(\Lambda)}. \tag{6.16}$$

Wet get estimates on R and on Z terms of this expansion in a series of Appendices. From the estimates of Appendices A–H we get the convergence of the mean field (contour-cluster) expansion.

Theorem 1: Let λ and m be positive constants such that $m \geqslant 1$ and $\lambda / m^2 < \epsilon$ where $\epsilon > 0$ is sufficiently small. Then the expansion (6.16) converges uniformly in Λ, λ, and m.

Proof: Follows as in Ref. 5, p. 635 from Proposition G2, Proposition H5, and a simple combinatoric estimate.

The convergence of the mean field expansion can now be used to obtain some other results on the lattice double-well model. The exponential clustering follows as usual from the convergence of the expansion by applying the Ginibre technique of duplicating variables (see Ref. 5, p. 668). From this result we get by standard methods (see Ref. 7, p. 162) the existence of the infinite volume correlation functions.

Theorem 2: Under the hypotheses of Theorem 1 the infinite volume correlation functions exist. This result is valid for the "plus" boundary condition as well as for the "minus" boundary condition double-well model on the lattice.

7. CONVERGENCE TOWARDS THE ISING MODEL

Instead of (2.1) let us consider the model

$$e^{-(1/2)\beta(q,(-\Delta)q)} \prod_i e^{-\nu^{-1}(\lambda q_i^2 - (1/4)m^2 q_i^2)}, \tag{7.1}$$

where $\beta = 1/kT$ is the Boltzmann factor and ν a positive constant. As in Sec. 2 we can start with (7.1) and define the rigorous model with finite cutoff and "plus" or "minus" boundary conditions. The results of Sec. 6 (combined with the estimates of A–H) show that for small temperature the correlation functions of the model (7.1) with "plus" or "minus" boundary conditions in Λ tend to the corresponding infinite volume correlations *uniformly in* ν if λ and m satisfy the conditions in Theorem 1. It is this uniformity in ν of the convergence which enables us to state the following result.

Theorem 3: Let λ and m satisfy the conditions in Theorem 1. Then the (infinite volume) correlation functions of the model $\nu^{-1}(\lambda q^4 - (m^2/4)q^2)$ at small temperature with "plus" or "minus" boundary conditions tend for $\nu \to 0$ to the corresponding correlation functions of the Ising model with spin $\sigma = \pm \xi$, $\xi = (m^2/8\lambda)^{1/2}$.

Proof: With finite cutoff Λ the convergence of the double-well model correlation functions to the correlation functions of the Ising model follows trivially from (1.1). The result for the infinite volume correlations follows now by an $\epsilon/3$ argument taking into account the uniformity in ν of the thermodynamic limit for correlation functions of the double-well model. We note that similar results for other boundary conditions can be obtained by applying the methods of Ref. 9 and of the present paper.

By using a scaling transformation we obtain convergence of the pure phases correlation functions of the double-well model with single spin distribution, given by $\nu^{-1}(q^2-1)^2$, $\nu \to 0$ towards the corresponding correlation functions of the Ising model at small temperature. Our finite result is then

Theorem 4: The infinite volume correlation functions of the double-well model at small temperature with single spin distribution given by $\nu^{-1}(q^2-1)^2$ and "plus" or "minus" boundary conditions converge for $\nu \to 0$ towards the corresponding correlation functions of the classical Ising model.

APPENDIX A: ESTIMATES ON R''

In this appendix we estimate

$$R''(A,Y,\Sigma,\Gamma) = \int_0^1 ds_{\Gamma^c}\, \partial_s^{\Gamma^c \subset \mathscr{B}(\Sigma,Y)} \times \int \chi_\Sigma A e^{-Q(Y,\Sigma)}\, d\psi(s),$$

where here A, Y, Σ, Γ stay for $A_k, Y_k, \Sigma_k, \Gamma_k$. The s integral causes no problems. For simplifying notation we change Γ in Γ^c and concentrate to estimate

$$\partial_s^{\Gamma^c \subset \mathscr{B}(\Sigma,Y)} \int e^{-Q(Y,\Sigma)} \chi_\Sigma A\, d\psi(s). \tag{A1}$$

The differentiation in s gives according to Ref. 7

$$\sum_{\pi \in \mathscr{P}(\Gamma)} \int \left[\prod_{\gamma \in \pi} \tfrac{1}{2}(\partial_s^\gamma C(s)) \cdot \Delta_\psi \right] e^{-Q(Y,\Sigma)} \chi_\Sigma A\, d\psi(s), \tag{A2}$$

where $\mathscr{P}(\Gamma)$ is the set of all possible partitions π of Γ. The elements γ of π are groups of hyperbonds of $(l\mathbb{Z}^\nu)^*$ in Γ. Further, we have

$$\tfrac{1}{2}\partial_s^\gamma C(s) \Delta_\psi = \sum_{j=(j_1,j_2)} \tfrac{1}{2}\partial_s^\gamma C(s)(j_1,j_2) \frac{d^2}{d\psi_{j_1} d\psi_{j_2}}. \tag{A3}$$

Using the notation

$$j_\gamma \equiv (j_{\gamma,1}, j_{\gamma,2}) \in \mathbb{Z}^\nu \times \mathbb{Z}^\nu \quad \text{for } \gamma \in \pi,$$

$$J \equiv J_\pi = \{ j_\gamma \}_{\gamma \in \pi},$$

we have to estimate

$$\sum_{\pi \in \mathscr{P}(\Gamma)} \sum_J \sum' \int A' W(\pi, J) e^{-Q(Y,\Sigma)}\, d\psi(s), \tag{A4}$$

where Σ' ranges over all possible derivatives of factors of A, χ_Σ, and $e^{-Q(Y,\Sigma)}$. Such factors are denoted by A' while

$$W(\pi, J) = \prod_{\gamma \in \pi} \tfrac{1}{2}\partial_s^\gamma C(s)(j_\gamma), \quad J = \{ j_\gamma \}. \tag{A5}$$

The Gaussian measure $d\psi(s)$ has Dirichlet conditions on Γ. We look first at the integral

$$\int A W(\pi, J) e^{-Q(Y,\Sigma)} d\psi(s). \tag{A6}$$

We apply Hölder's inequality with $1/p_1 + 1/p_1' = 1$ and get the following bound for (A6):

$$\| A' W(\pi, J) \|_{p_1} \| \chi_\Sigma' e^{-Q(Y,\Sigma)} \|_{p_1'}. \tag{A7}$$

The exact choice of p_1, p_1' will be given in Appendix B.

APPENDIX B: ESTIMATE ON $\| \chi_\Sigma' e^{-Q(Y,\Sigma)} \|_{p_1'}$

We have (see Appendix A)

$$\| \chi_\Sigma' e^{-Q(Y,\Sigma)} \|_{p_1'}^{p_1'}$$

$$= \int \exp(-p_1 [P_{\Lambda \cap Y} - \tfrac{1}{2} m^2 \psi^2_{\Lambda \cap Y} - \ln|\chi_\Sigma'| + F(Y,\Sigma)]) d\psi(s)$$

$$= \int \left\{ -p_1 \left[\left(P_{\Lambda \cap Y} - \frac{\eta}{2} m^2 \psi^2_{\Lambda \cap Y} - \ln|\chi_\Sigma'| \right) \right. \right.$$

$$+\left(-\frac{1-\eta}{2}m^2\psi^2_{\Lambda\cap Y}+F(Y,\Sigma)\right)\Big]\Big]d\psi(s) \tag{B1}$$

where η is a positive constant to be fixed later. The decomposition

$$\tfrac{1}{2}m^2\psi^2_{\Lambda\cap Y}=\frac{\eta}{2}m^2\psi^2_{\Lambda\cap Y}+\frac{1-\eta}{2}m^2\psi^2_{\Lambda\cap Y}$$

is useful because of the following two lemmas which have simple proofs.

Lemma B 1: Let $\eta<1/36$ and let λ/m^2 be sufficiently small. Then we have for $i\in\Lambda$:

$$P(q_i)-\frac{\eta}{2}m^2(q_i-h_i)^2-\ln\chi_{\sigma_i}(q_i)\geqslant 0. \tag{B2}$$

We remark that in the Lemma B1 χ is not differentiated. We have a better estimate for the case in which χ is differentiated.

Lemma B 2: Let $\eta<1/36$ and let λ/m^2 be sufficiently small. If n_i is the order of derivative of χ in $i\in\Lambda$, then there is a constant $a>0$ independent of λ and m such that

$$P(q_i)-\frac{\eta}{2}m^2(q_i-h_i)^2-\ln\chi'_{\sigma_i}(q_i)\geqslant a\frac{m^2}{\lambda}-\ln n_i!. \tag{B3}$$

The proofs of lemmas B1 and B2 are simple and are left to the reader (for proofs of similar results in constructive quantum field theory (see Ref. 5).

Now let us remember that

$$\chi_\Sigma=\prod_{i\in\Lambda\cap Y}\chi_{\sigma_i}(q_i).$$

We get for (B1) the following bound where $X'_\#$ is the subset of $\mathbb{Z}^\nu$ on which χ_Σ is differentiated (certainly $X'\subset\Lambda\cap Y$)

$$\left(\prod_{i\in X'_\#}n_i!\right)e^{-O(m^2/\lambda)(|X'|}e^{(p_1/2)(h,\Delta h)}$$
$$\times\exp\{p_1[(1-\eta/2)m^2\psi^2_{\Lambda\cap Y}+(\psi,\Delta h)]\}d\psi(s). \tag{B4}$$

We have to study the integral

$$\int\exp\{p_1[(\psi,\Delta h)+(1-\eta/2)m^2\psi^2_{\Lambda\cap Y}]\}d\psi(s), \tag{B5}$$

where the Gaussian measure $d\psi(s)$ has Dirichlet conditions on Γ. We are going to find an upper bound for the integral (B5). First we remark that we *increase* the integral in (B5) if we leave out the Dirichlet conditions on Γ and we increase the integral further by taking $s=1$. In fact, we have to estimate an integral of the form

$$\int\exp[(b,\psi)+\tfrac{1}{2}(\psi,B\psi)]d^{(A)}\psi, \tag{B6}$$

where $d^{(A)}\psi$ is a Gaussian measure on $l^2(\mathbb{Z}^\nu)$ with covariance $C=A^{-1}$. In our case the covariance C is not trace class such that the measure $d\psi^{(A)}$ is not σ additive on $l^2(\mathbb{Z}^\nu)$. It is easy to prove that

$$\int\exp[(b,\psi)+\tfrac{1}{2}(\psi,B\psi)]d^{(A)}\psi$$
$$=\int\exp[(A^{-1/2}b,\psi)+\tfrac{1}{2}(\psi,K\psi)]d^{(A=I)}\psi$$
$$=\exp[\tfrac{1}{2}(b,A^{-1/2}(I-K)^{-1}A^{-1/2}b)]$$
$$\times\exp[-\tfrac{1}{2}\mathrm{Tr}\ln(I-K)], \tag{B7}$$

where $K=A^{-1/2}BA^{-1/2}$ under the condition that K is a trace class operator and $\|K\|<1$. In our case these two conditions are satisfied. Indeed, K is trace class because $A^{-1/2}$ is bounded and B is a trace class operator. We remark that $\ln(I-K)$ is also trace class because K is trace class. The norm of K is

$$\|K\|\leqslant\|B\|\,\|A^{-1}\|\leqslant 1-\eta<1$$

because $\|A^{-1}\|\leqslant m^{-2}$ and $\|B\|=(1-\eta)m^2$. The two exponentials in (B7) can be easily estimated. Indeed, we have

$$-\mathrm{Tr}\ln(I-K)\leqslant\tfrac{3}{2}\mathrm{Tr}K$$

and

$$\mathrm{Tr}K\leqslant\|A^{-1}\|\mathrm{Tr}B\leqslant|Y\cap\Lambda|.$$

For the first factor we have

$$\frac{p_1^2}{2}(\Delta h,A^{-1/2}(I-K)^{-1}A^{-1/2}\Delta h)$$
$$\leqslant\frac{p_1^2}{2}\|A^{-1/2}(I-K)^{-1}A^{-1/2}(-\Delta)h\|(h,(-\Delta)h)$$
$$=\frac{p_1^2}{2}\|(A-B)^{-1}(-\Delta)\|(h,(-\Delta)h)\leqslant\frac{p_1^2}{2}$$
$$\times\frac{C}{C+\eta m^2}(h,(-\Delta)h).$$

For *all* $m^2>0$ we can choose $p_1=p_1(m^2)>1$ such that

$$C/(C+\eta m^2)<1/p_1$$

and this is actually a consequence of the fact that the lattice Laplacian is a *bounded* operator on $l^2(\mathbb{Z}^\nu)$. Collecting together all these estimates and taking into account (B4) we get the following

Theorem B 1: Let $\eta<1/36$ and let λ_0, m_0 be positive constants such that λ_0/m_0^2 is sufficiently small. Then for $\lambda\leqslant\lambda_0$ and $m\geqslant m_0$ we can find $p_1=p_1(m_0^2)>1$ such that

$$\|\chi'_\Sigma e^{-Q(Y,\Sigma)}\|_{p_1}\leqslant\left(\prod_{i\in X'_\#}n_i!\exp[-O(m^2/\lambda)(|\Sigma|+|X'|)]\right.$$
$$\left.+O(1)|Y\cap\Lambda|\right]. \tag{B8}$$

For completing the proof of this theorem we need only to remark that $(h,(-\Delta)h)$ is of the form $O(m^2/\lambda)|\Sigma|$. By denoting $Y_d=Y\cap\Lambda-X'$ and absorbing part of the divergent into the convergent factor we can write (B8) in the following form:

$$\|\chi'_\Sigma e^{-Q(Y,\Sigma)}\|_{p_1}\leqslant\left(\prod_{i\in X'_\#}n_i!\right)\exp[-O(m^2/\lambda)(|\Sigma|+|X'|)$$
$$+O(1)|Y_d|]. \tag{B9}$$

APPENDIX C: IMPROVED ESTIMATE ON $\|\chi'_\Sigma e^{-Q(Y,\Sigma)}\|_{p_1}$

In this appendix we show that the estimate on $\|\chi'_\Sigma e^{-Q(Y,\Sigma)}\|_{p_1}$ can be improved (see Ref. 5, p. 654). Let t_i be parameters between zero and one ($0\leqslant t_i\leqslant 1$) for all $i\in Y_d$ and let

$$t=\{t_i|i\in Y_t\},$$
$$Y_t=\{i\in Y_d|t_i\neq 0\}\subset Y_d.$$

Now let

$$V_i \equiv V(q_i) = P(q_i) - \tfrac{1}{2}m^2(q_i - h_i)^2,$$

$$V(t,Y) = \sum_{i\in Y_t} t_i V_i ,$$

$$Q(t,Y) \equiv Q(t,Y,\Sigma) = F(Y,\Sigma) + V(t,Y).$$

With the help of the identity

$$e^{-p_1 V_i} = 1 - p_1 V_i \int_0^1 e^{-p_1 t_i V_i} dt_i , \tag{C1}$$

we get

$$\int e^{-p_1 Q(Y,\Sigma)} |\chi'_\Sigma|^{p_1} d\psi(s)$$

$$\leqslant \sum_{Y_t \subset Y_d} \int_0^1 \int \left| \prod_{i\in Y_t} p_1 V(q_i) \right| \times e^{-p_1 Q(t,Y)} |\chi'_\Sigma|^p d\psi(s)\, dt.$$

Now we apply the Hölder inequality with indices $1/p_2 + 1/p'_2 = 1$. We have

$$\|\chi'_\Sigma e^{-Q(Y,\Sigma)}\|_{p_1}^{p_1}$$

$$\leqslant \sum_{Y_t \subset Y_d} \int_0^1 \left\{ \left[\int \left| \prod_{i\in Y_t} p_1 V(q_i) \right|^{p'_2} d\psi(s) \right]^{1/p'_2} \times \left[\int e^{-p_1 p_2 Q(t,Y)} |\chi'_\Sigma|^{p_1 p_2} d\psi(s) \right]^{1/p_2} \right\} dt$$

$$\leqslant \sum_{Y_t \subset Y_d} \left[\int \left(\prod_{i\in Y_t} p_1 V(q_i) \right)^{p'_2} d\psi(s) \right]^{1/p'_2} \times \sup_t \left[\int e^{-p_1 p_2 Q(t,Y)} |\chi'_\Sigma|^{p_1 p_2} d\psi(s) \right]^{1/p_2}. \tag{C2}$$

Let us look now at the integral

$$\int \left(\prod_{i\in Y_t} V(q_i) \right)^{p'_2} d\psi(s). \tag{C3}$$

We remove the Dirichlet conditions on Γ, increasing the integral. We further increase the integral by replacing the inverse covariance $A = C^{-1}$ in $d\psi(s)$ by $m^2 I$ (Neuman conditions). We end up with an upper bound for the integral (C3) which is a product of factors:

$$\frac{m}{(\pi)^{1/2}} \int_{-\infty}^{\infty} (V(q_i))^{p'_2} e^{-(m^2/2)\psi_i^2} d\psi_i , \tag{C4}$$

where

$V(q_i) = P(q_i) - (1/2)m^2(q_i - h_i)^2 = \lambda\psi_i^4 + m(2\lambda)^{1/2}\psi_i^3$.

Under the restriction $\lambda/m^4 < 1$ we get for (C4) the bound

$$K_1^{p'_2}(p'_2!)^2 \left(\frac{\lambda}{m^4}\right)^{p'_2/2},$$

where K_1 is a constant. Because we can find an upper bound on $p_1 = p_1(m^2)$ which is independent on m^2 and because p_1, p_2 can be chosen independent of m^2 if $m > m_0$ where m_0 is a given positive constant, we can find a constant K_2 such that

$$\left[\int \left(\prod_{i\in Y_t} V(q_i) \right)^{p'_2} d\psi(s) \right]^{1/p'_2} \leqslant \left(K_2 \left(\frac{\lambda}{m^4} \right)^{1/2} \right)^{|Y_t|}, \tag{C5}$$

there K_2 depends on m_0 but not λ, m^2 and Λ. We now study the second integral in (C2). We remark that in $t_i V_i$ the quotient $t_i \lambda/t_i m^2 = \lambda/m^2$ and therefore we can apply the same estimates (B2) and (B3) without change, i.e., we have

$$t_i P(q_i) - t_i (\eta/2)m^2(q_i - h_i)_2 - \ln\chi_{\sigma_i}(q_i) \geqslant 0,$$

$$t_i P(q_i) - t_i (\eta/2)m^2(q_i - h_i)_2 - \ln\chi'_{\sigma_i}(q_i) \geqslant am^2/\lambda - \ln n_i !.$$

Using also the inequality

$$\exp[p_1 t_i (1 - \eta/2)\psi^2_{Y_t}] \leqslant \exp[p_1 (1 - \eta/2)m^2 \psi^2_{\Lambda\cap Y}]$$

and applying Theorem B1 we get for (C2) the following estimate (valid for $m \geqslant m_0$)

$$\sum_{Y_t \subset Y_d} \left(K_2 \left(\frac{\lambda}{m^4} \right)^{1/2} \right)^{|Y_t|} \left(\prod_{i\in X'} n_i! \right) \times \exp[- O(m^2/\lambda)(|\Sigma| + |X'|) + O(1)|Y_t|] .$$

We absorb $e^{O(1)|Y_t|}$ into $K_2^{|Y_t|}$ and get for (C2) the bound

$$\sum_{Y_t \subset Y_d} \left(K_3 \left(\frac{\lambda}{m^2} \right)^{1/2} \right)^{|Y_t|} \left(\prod_{i\in X'} n_i! \right) \times \exp[- O(m^2/\lambda)(|\Sigma| + |X'|)]$$

$$= \left(\prod_{i\in X'} n_i! \right) \left(\prod_{i\in Y_d} \left(1 + K_3 \left(\frac{\lambda}{m^4} \right)^{1/2} \right) \right) \times \exp\left[- O(m^2/\lambda)(|\Sigma| + |X'|) \right]$$

$$\leqslant \left(\prod_{i\in X'} n_i! \right) \exp[- O(m^2/\lambda)(|\Sigma| + |X'|) + O((\lambda/m^4)^{1/2})|Y_d|].$$

We can therefore state the following results.

Theorem C 1: Let $\eta < 1/36$ and let λ_0, m_0 be positive constants such that λ_0/m_0^2 is sufficiently small and $\lambda_0/m_0^4 < 1$. Then, for $\lambda \leqslant \lambda_0$ and $m \geqslant m_0$, there exists $p_1(m_0^2) > 1$ such that for $p_1 < p(m_0^2)$ we have

$$\|\chi'_\Sigma e^{-Q(Y,\Sigma)}\|_{p_1} \leqslant \left(\prod_{i\in X'} n_i! \right) \times \exp[- O(m^2/\lambda)(|\Sigma| + |X'|) + O((\lambda/m^4)^{1/2})|Y\cap\Lambda|] . \tag{C6}$$

As in Appendix B we can write this estimate also in the following form:

$$\|\chi'_\Sigma e^{-Q(Y,\Sigma)}\|_{p_1} \leqslant \left(\prod_{i\in X'} n_i! \right) \times \exp[- O(m^2/\lambda)(|\Sigma| + |X'|) + O((\lambda/m^4)^{1/2})|Y_d|] . \tag{C7}$$

APPENDIX D: RANDOM WALK AND COVARIANCE OPERATORS

We use random walk methods in order to control the strong exponential decay of covariance operators and their derivatives. The results of this appendix are proved in Ref. 8. Let $A = C^{-1}$, $A_\Gamma = C_\Gamma^{-1}$. All these operators are bounded in $l_2(\mathbb{Z}^\nu)$. Let P be the transition operator in $l_2(\mathbb{Z}^\nu)$ of the

symmetric random walk on $\mathbb{Z}^\nu$. (For notations see Ref. 8). Let $p(n,x,y)$ be the transition probability to get from $x\in\mathbb{Z}^\nu$ to $y\in\mathbb{Z}^\nu$ in n steps. We have

$$p(n+m,x,y)=\sum_{z\in\mathbb{Z}^\nu}p(n,x,z)p(m,z,y),$$

$$P\varphi(x)=\sum_{y\in\mathbb{Z}^\nu}p(1,x,y)\varphi(y),\quad \varphi\in l^2(\mathbb{Z}^\nu),$$

and

$$P^n\varphi(x)=\sum_{y\in\mathbb{Z}^\nu}p(n,x,y)\varphi(y).$$

Let E be the identity operator in $l^2(\mathbb{Z}^\nu)$. We have $C=(1/2)(\lambda E-P)^{-1}$ and then

$$C=\frac{1}{2\nu\lambda}\sum_{n=0}^{\infty}\frac{1}{\lambda^n}P^n, \tag{D1}$$

where $\lambda=1+(m^2/2\nu)$. We want to get a similar formula for C_Γ. For given Γ we define

$$Q(x,y)\equiv q(1,x,y)=\begin{cases}p(1,x,y) & \text{if } \overline{xy}\cap\Gamma=\phi,\ |x-y|=1\\ 0 & \text{otherwise,}\end{cases}$$

$$Q\varphi(x)\equiv P_\Gamma\varphi(x)=\sum_{y\in\mathbb{Z}^\nu}q(1,x,y)\varphi(y),\quad \varphi\in l^2(\mathbb{Z}^\nu).$$

We define $q(n,x,y)$ recursively by

$$q(n+1,x,y)=\sum_{z\in\mathbb{Z}^\nu}q(n,x,z)q(1,z,y)$$

and

$$Q^n\varphi(x)=\sum_{y\in\mathbb{Z}^\nu}q(n,x,y)\varphi(y),\quad \varphi\in l^2(\mathbb{Z}^\nu).$$

We have $A_\Gamma=2\nu(\lambda E-P_\Gamma)$ and

$$C_\Gamma\equiv A_\Gamma^{-1}=\frac{1}{2\nu\lambda}\sum_{n=0}^{\infty}\frac{1}{\lambda^n}P_\Gamma^n. \tag{D2}$$

This is a useful formula which relates the Dirichlet covariance C_Γ to the symmetric random walk on the lattice.

APPENDIX E: ESTIMATES ON $\partial_s^\gamma C(s)(j_\gamma)$

The results in this Appendix are partially contained in Ref. 8, to which we refer the reader for details. Let Ω be the space of sample paths of the symmetric random walk on $\mathbb{Z}^\nu$. Let

$$\Omega^n_{x,y}=\{\omega\in\Omega\,|\omega(0)=x,\omega(n)=y\},$$

and let $dP(\cdot)$ be the measure on Ω induced by $p(n,x,y)$. Given a hyperbond b_0 on $(\mathbb{Z}^\nu)^*$, we define a measurable function

$$\chi_{b_0}(\omega)=\begin{cases}0 & \text{if } \exists m\in\mathbb{N}:\overline{\omega(m)\omega(m+1)}\cap b_0\neq\phi\\ 1 & \text{otherwise}\end{cases}.$$

It follows that

$$q(n,x,y)\equiv P_\Gamma^n(x,y)=\int_{\Omega^n_{x,y}}\prod_{b_0\in\Gamma}\chi_{b_0}(\omega)\,dP(\omega)$$

and we have from (D1)

$$C_\Gamma(x,y)=\frac{1}{2\nu\lambda}\sum_{n=0}^{\infty}\frac{1}{\lambda^n}\int_{\Omega^n_{x,y}}\prod_{b_0\in\Gamma}\chi_{b_0}(\omega)\,dP(\omega)$$

$$=\frac{1}{2\nu\lambda}\sum_{n=0}^{\infty}\frac{1}{\lambda^n}\int\prod_{b\in\Gamma}\Big(\prod_{b_0\in b}\chi_{b_0}(\omega)\Big)dP(\omega). \tag{E1}$$

Introducing (E1) into (5.3) we get

$$C(s)(x,y)=\frac{1}{2\nu\lambda}\sum_{n=0}^{\infty}\frac{1}{\lambda^n}\int_{\Omega^n_{x,y}}\prod_{b\in B}[s_b+(1-s_b)\chi_b(\omega)]dP(\omega), \tag{E2}$$

where

$$\chi_b(\omega)=\prod_{b_0\in b}\chi_{b_0}(\omega).$$

From (E2) we get

$$\partial^\gamma C(s)(x,y)\leqslant\frac{1}{2\nu\lambda}\sum_{n=0}^{\infty}\frac{1}{\lambda^n}\int_{\Omega^n_{x,y}}\prod_{b\in\gamma}(1-\chi_b(\omega))dP(\omega). \tag{E3}$$

We remind the reader that the cluster expansion is performed on a subset of $l\mathbb{Z}^\nu$ such that Γ and γ are sets of hyperbonds b of $(l\mathbb{Z}^\nu)^*$.

Obviously, the integral in (E3) is the probability that the sample path starting in x crosses all $b\in\gamma$ before reaching y in n steps.

Let $\mathcal{O}(\gamma)$ be the set of all linear orderings of hyperbonds b in γ. Let $W(\mathcal{O})$ be the set of paths which meet the elements b in the order given by $o\in\mathcal{O}(\gamma)$. Then we get from (E3)

$$\partial^\gamma C(s)(x,y)\leqslant\frac{1}{2\nu\lambda}\sum_{n=0}^{\infty}\frac{1}{\lambda^n}\sum_{o\in\mathcal{O}(\gamma)}\int dP(\omega). \tag{E4}$$

Let $b_1,b_2,\cdots$ be the elements of γ as ordered by o. We use separating parallel infinite $(\nu-1)$-dimensional hyperplanes on $\mathbb{Z}^\nu$ in order to project on the one-dimensional random walk (for details see Ref. 8). Suppose γ contains at least two elements. We note that the ordering o induces an ordering in the set of separating hyperplanes. Certainly some hyperplanes as ordered by o may coincide. We see that from the point of view of estimates there is an optimal choice of the hyperplanes. There are $m\geqslant 2$ separating hyperplanes.

Let

$$a_j=d_\#(H_{j+1},H_j),\quad 1\leqslant j\leqslant m-1.$$

The strong Markov property combined with an estimate on the generating function of the one-dimensional random walk gives[8]

$$\frac{1}{2\nu\lambda}\sum_{n=0}^{\infty}\frac{1}{\lambda^n}\int_{w(o)}dP(\omega)\leqslant\exp\Big(-\alpha\nu\sum_{j=1}^{m-1}a_j\Big), \tag{E5}$$

where

$$\alpha=\frac{1}{\nu}\log\frac{2\nu}{\alpha\nu+m^2-m(m^2+4\nu)^{1/2}}. \tag{E6}$$

By taking the geometric mean over all coordinate directions we get

$$\partial^\gamma C(s)(x,y)\leqslant\sum_{o\in\mathcal{O}(\gamma)}e^{-\alpha|o|}, \tag{E7}$$

where $|o|$ is the *hyperplane distance* of the o ordering of $b\in\gamma$. If $|\gamma=\{\phi\}$ or if γ contains only one bound b we take $|o|=0$. We give some simple properties of the coefficient α which are needed in the following. For $m>0$ small we have

$$\alpha = \frac{m}{\nu^{3/2}} + \frac{m^3}{24\nu^{5/2}} + O(m^4). \tag{E8}$$

For m large

$$\alpha = (2/\nu)\log m - (1/\nu)\log\nu + O(1/m). \tag{E9}$$

Finally, $\alpha = \alpha(m)$ is an increasing function of m on the interval $(0,\infty)$.

The same type of estimate gives[8]

$$\partial^\gamma C(s)(x,y) \leqslant K_4 e^{-\alpha d(j,\gamma)}, \tag{E10}$$

where $j = (j_1, j_2) \equiv (x,y)$ and $d(j,\gamma)$ is the maximal hyperplane distance from j to a line $b \in \gamma$. K_4 is a constant independent of s, j and γ. If $\gamma = \{\emptyset\}$ we take $d = d_{\#}(x,y)$. Obviously, there exist optimal (maximal) values for $|\sigma|$ and $d(j,\gamma)$ for given γ and σ

We now introduce the following notation. Let

$$|\gamma^*| = \min_{\sigma \in \mathcal{C}(\gamma)} |\sigma|. \tag{E11}$$

We prove the following

Lemma E 1: Let λ_0, m_0 be positive constants such that λ_0/m_0^2 is sufficiently small. There are constants $K_6(\gamma)$, K_7 and $\delta = \delta(m_0)$ such that for $\lambda \leqslant \lambda_0$ and $m \geqslant m_0$, we have

$$\partial^\gamma C(s)(j) \leqslant K_6(\gamma)\exp[-\alpha\delta(d(j,\gamma) + |\gamma^*|)], \tag{E12}$$

and

$$\sum_{\pi : \mathcal{P}(\Gamma)} \prod_{\gamma \in \pi} K_6(\gamma) \leqslant e^{K_7|\Gamma|}. \tag{E13}$$

Proof: The estimates (E7) and (E10) give

$$\partial^\gamma C(s)(j) \leqslant K_4^{1/2} e^{-(d(j,\gamma)/2)} \left(\sum_{\mathcal{C}} e^{-\alpha(|\mathcal{C}|/2)}\right)^{1/2}$$

$$\leqslant K_4^{1/2} e^{-\alpha[d(j,\gamma)/2]} \sum_{\mathcal{C}} e^{-(|\mathcal{C}|/2)}.$$

But

$$\sum_{\mathcal{C}} e^{-\alpha(|\mathcal{C}|/2)} = \sum_{\mathcal{C}} \exp\{-[\alpha(|\mathcal{C}|/3) + \alpha(|\mathcal{C}|/6)]\}$$

$$\leqslant e^{-\alpha(|\gamma^*|/6)} \sum_{\mathcal{C}} e^{-\alpha(|\mathcal{C}|/3)} = K_5(\gamma) e^{-(|\gamma^*|/6)},$$

where

$$K_5(\gamma) = \sum_{\mathcal{C}} e^{-\alpha(|\mathcal{C}|/3)}.$$

With $K_6 \equiv K_6(\gamma) = K_4^{1/2} K_5(\gamma)$ we get (E12) by taking into account the fact that $\alpha = \alpha(m)$ is an increasing function of m. We have to prove only (E13). Let us first call the reader's attention to the fact that $|\sigma|$ and $|\gamma^*|$ denote lengths while $|\gamma|$ and $|\Gamma|$ denote number of elements b in γ and Γ, respectively. With this remark (E13) follows by "scaling" the arguments in Ref. 7, pp. 230–31.

APPENDIX F: ESTIMATES ON $\|A'W(\pi,J)\|'_p$

In this appendix $J \equiv J_\pi = \{j_\gamma\}$, $\gamma \in \pi$ is a given set of indices. We have

$$W(\pi,J) = \prod_{\gamma \in \pi} \tfrac{1}{2}\partial_s^\gamma C(s)(j_\gamma). \tag{F1}$$

For a given j_γ we have to perform derivatives

$$\frac{d^2}{dq_{j_{\gamma,1}} dq_{j_{\gamma,2}}}$$

[see A(3)]. These derivatives act on factors of A, χ, and e^{-Q} and are indexed by elements $\gamma \in \pi$. We consider two distinct cases:

(i) $|\gamma| > 4$: Then there is a choice of the separating hyperplanes such that

$$|\gamma^*| \geqslant \text{const} l\,|\gamma|, \quad d(j_\gamma,\gamma) \geqslant l.$$

Thus, we may extract a convergent factor $e^{-\alpha\delta' l|\gamma|}$ from $\exp[-\alpha\delta(d(j_\gamma,\gamma) + |\gamma^*|)]$ of (E12). This factor controls the derivatives of A, χ, and e^{-V} (remember $Q = V + F$) but not the derivatives of e^{-F}. Indeed, $F = -(\psi, \Delta h) - (1/2)(h, \Delta h)$, and this derivative contribution of $\langle\psi, \Delta h\rangle$ gives a coefficient $O(m^2/\lambda)$ if $j_{\gamma,1}, j_{\gamma,2}$ are strictly concentrated near the phase boundary. But in this case $d(j_\gamma,\gamma) \geqslant L/2$. We then have

$$O\left(\frac{m^2}{\lambda}\right)\exp[-\alpha\delta(j_\gamma,\gamma)] \leqslant \exp[-\alpha\delta'' d(j_\gamma,\gamma)]. \tag{F2}$$

It follows that for each $\gamma \in \pi$ such that $|\gamma| > 4$ we can segregate in $\Pi_{\gamma\in\pi}$ a convergent factor

$$\exp[-\alpha\gamma(d(j_\gamma,\gamma) + l|\gamma|)], \tag{F3}$$

where $\delta > 0$ is a new constant.

(ii) $|\gamma| \leqslant 4$: The derivatives of e^{-V} are controlled by

$$O((\lambda/m^2)^{1/2}) \leqslant e^{-4\alpha\delta l} \leqslant e^{-\alpha\delta l|\gamma|}. \tag{F4}$$

If we have derivatives of $\chi_{j_{\gamma,1}}, \chi_{j_{\gamma,2}}$, then $j_{\gamma,1}, j_{\gamma,2}$ belong to X' and we use the inequality

$$1 \leqslant e^{-\alpha\delta l|\gamma|} e^{l}. \tag{F5}$$

We use the same inequalty for the differentiation of a linear factor of A. From the differentiation of e^{-F} we use the same inequality (F2) as in (i) together with the inequality (F5). We conclude that the factor (F3) also can be segregated from the product $\Pi_{\gamma\in\pi}$ in case (ii). From $\|A'W(\pi,J)\|_{p_i'}$ we can therefore extract a factor

$$\exp\left[-\alpha\delta l|r| - \alpha\delta\sum_{\gamma\in\pi} d(j_\gamma,\gamma)\right]. \tag{F6}$$

Now we come to factors in $\|A'W(\pi,J)\|_{p_i'}$ which do not depend on λ and m. They come out from the Gaussian integration of polynomials A'. Suppose the observable A has the form

$$A(\psi) = \prod_{i=1}^{n} \psi_i^{m_i}. \tag{F7}$$

Let

$$|||A||| = \prod_{i=1}^{n} m_i! \exp\left(Kl\sum_{i=1}^{n} m_i\right), \tag{F8}$$

where $K > 0$ is a constant. Using factorization of Gaussian integrals as in Appendix C and studying the factorials as in Ref. 5, p. 657 we get

Proposition F 1: Let λ_0 and $m_0 \geqslant 1$ be positive constants such that λ_0/m_0^2 is sufficiently small. Then for $\lambda \leqslant \lambda_0$ and $m \geqslant m_0$ there is a norm $|||A|||$ and a constant $\delta > 0$ indepen-

dent of λ, m and Λ such that

$$\|A'W(\pi,J)\|_{p_i'} \leqslant |||A|||e^{-\alpha\delta l|\Gamma|}e^{l|X'|} \times \prod_{\gamma\in\pi} K_6(\gamma)\exp[-\alpha\delta d(j_\gamma,\gamma)]. \quad \text{(F9)}$$

APPENDIX G: FINAL ESTIMATES ON R

The bound on (A4) is given by

$$\sum_{\pi\in\mathscr{P}(\Gamma)} \sum_J \sum' \|A'W(\pi,J)\|_{p_i'} \|\chi'_{\Sigma\cap Y} e^{-Q(Y,\Sigma)}\|_{p_1}. \quad \text{(G1)}$$

We introduce (F9) and (C4) in (G1). The divergent factor $e^{l|X'|}$ in (F9) is controlled by the convergent factor $e^{-O(m^2/\lambda)|X'|}$ if λ/m^2 is sufficiently small. The sums Σ' and Σ_J are controlled by scaling the Lemmas 10.1–10.2 of Ref. 1. The sum over $\pi\in\mathscr{P}(\Gamma)$ is dominated by (E13). We get

Proposition G 1: Under the hypotheses of Proposition F1 we have for an observable A of the form (F7):

$$\left|\partial_s^\Gamma \int e^{-Q(Y,\Sigma)}\chi_{\Sigma\cap Y} A\, d\psi(s)\right| \leqslant |||A|||\exp[-\alpha\delta l|\Gamma| - O(m^2/\lambda)|\Sigma\cap Y| + O(\lambda/m^4)^{1/2}|Y\cap\Lambda|]. \quad \text{(G2)}$$

Now suppose the observable A has the form

$$A \equiv A(q-\xi) = \prod_{i=1}^n (q_i-\xi)^{m_i}. \quad \text{(G3)}$$

We have

$$q_i - \xi = (q-h_i) + (h_i - \xi) = \psi + (h_i - \xi),$$

and

$$|h_i - \xi| \leqslant 2\xi < \left(\frac{m^2}{\lambda}\right)^{1/2}.$$

$A(q-\xi)$ can be written as a finite sum of terms C_j, each of the form (G3) with

$$|||C_j||| \leqslant \left(\frac{m^2}{\lambda}\right)^{1/2} |||A|||.$$

We apply Proposition G1 in each component Y_k of Y and get the following estimate on R''

$$|R''(A,Y,\Sigma,\Gamma)| \leqslant |||A|||\left(\left(\frac{m^2}{\lambda}\right)^{1/2}\right)^N e^{-\alpha\delta l|\Gamma^c|} \times \exp[-O(m^2/\lambda)|\Sigma\cap Y| + O(\lambda/m^4)^{1/2}|Y\cap\Lambda|]. \quad \text{(G4)}$$

In order to get an estimate on R we have to use (G4) together with (6.10) and (6.14), i.e., to sum up (G4) according to

$$\sum_{\{\Sigma\cap Y|\Sigma\restriction N(\partial Y^+)=\pm\}} \sum_{\{\Gamma\subset\mathscr{P}(\Sigma)|Y(\Gamma)=Y\}}$$

This can be done exactly as in Ref. 5, pp. 658–59 and we get

Proposition G 2: Let A be an observable of the form (G3). For λ and m as in Proposition F1 there exists a constant $\delta>0$ such that

$$|R(A,Y,\partial Y^\pm)| \leqslant \lambda^{-N/2}|||A|||e^{-(\delta/l)|Y\cap\Lambda|}. \quad \text{(G5)}$$

APPENDIX H: ESTIMATES ON Z

First we want to get a simple form of the outside sum (6.15). Changing for convenience Y^c into Y in (6.15) we get

$$Z(Y,N(\partial Y^+),N(\partial Y^-),\partial Y) = \sum_{\{\Sigma\cap Y|\Sigma\restriction N(\partial Y^+)=\pm\}} Z'(Y,\Sigma\cap Y,\partial Y) = \sum_{\{\Sigma\cap Y|\Sigma\restriction N(\partial Y^+)=\pm\}} \int e^{-Q(Y,\Sigma\cap Y)}\chi_{\Sigma\cap Y} d\psi_{\partial Y}. \quad \text{(H1)}$$

We simplify the notations as follow:

$$N(\partial Y^\pm) = N^\pm, \quad \partial Y = \Gamma, \quad \partial Y^\pm = \Gamma^\pm.$$

We say that Σ is compatible with $N^\pm$ if $\sigma\restriction N^\pm = \pm$. In analogy with (4.4) let us denote

$$dq_{Y\cap\Lambda,\Gamma} = e^{-Q(Y,\Sigma\cap Y)} d\psi_\Gamma. \quad \text{(H2)}$$

We have

Proposition H 1: The outside sum can be written as

$$Z(Y,N^+N^-,\Gamma) = \int \left[\prod_{i\in N^+}\chi_+(q_i)\right]\left[\prod_{i\in N^-}\chi_-(q_i)\right] dq_{Y\cap\Lambda,\Gamma}. \quad \text{(H3)}$$

The proof of this proposition follows from

Lemma H 2: Let A be an observable (function of q). For any two configurations Σ_1,Σ_2 compatible with $N^\pm$ we have

$$\int Ae^{-Q(Y,\Sigma_1)} d\psi_\Gamma = \int Ae^{-Q(Y,\Sigma_2)} d\psi_\Gamma. \quad \text{(H4)}$$

Proof: We have

$$d\psi_\Gamma = e^{-(1/2)(\psi,B\psi)}\, d\psi,$$

where B is concentrated on Γ, i.e.,

$$B(j_1,j_2) = \begin{cases} 1; & j_1 - j_2 = 1, \ \overline{j_1 j_2}\cap\Gamma \neq \phi, \\ 0; & \text{otherwise.} \end{cases}$$

Then

$$dq_{Y\cap\Gamma,\Gamma} = e^{-Q} d\psi_\Gamma = e^{-Q-(1/2)B} d\psi = e^{-(1/2)B} dq_{Y\cap\Lambda},$$

because [see (4.4)]

$$dq_{Y\cap\Lambda} = e^{-Q(Y,\Sigma\cap Y)}\, d\psi.$$

Furthermore, we have $\psi_i \equiv \psi_i(\Sigma_k) = q_i - h_i(\Sigma_k)$, for $k=1,2$. The compatibility conditions give $h_i(\Sigma_i) = h_i(\Sigma_2)$ and therefore $\psi_i(\Sigma_1) = \psi_i(\Sigma_2)$. The proof of (H3) is now trivial.

We introduce some more notation:

$$N_{\text{plus}} = \{N^+\cup N^-,\phi\},$$

$$Z(Y,N,\Gamma) = Z(Y,N^+,N^-,\Gamma).$$

The next result is valid for real coupling constant λ.

Lemma H 3: We have

$$\frac{Z(Y,N,\Gamma)}{Z(Y,N_{\text{plus}},\Gamma)} \leqslant 1. \quad \text{(H5)}$$

Proof: In order to specify the h dependence we introduce the notation

$$dq_{Y\cap\Lambda,\Gamma,h} \equiv dq_{Y\cap\Lambda,h}$$

and

$$Z(Y,h)=\int dq_{Y\cap\Lambda,\Gamma,h}\,, \tag{H6}$$

where $h=h_\Sigma$ and Σ are assumed to be compatible with N. We note that going from N to N_{plus} means going from h to ξ for all $i\in\mathbb{N}$. We claim first that

$$Z(Y,h)\leqslant Z(Y,\xi). \tag{H7}$$

Indeed, we have

$$Z(Y,h)=\int \exp[-(1/2)(q-h,B(q-h))]dq_{Y\cap\Lambda}\,.$$

The change of h affects only the linear term in $-\frac{1}{2}(q-h,B(q-h))$ which is of the form

$$\sum_{\substack{|i-j|=1\\ \overline{ij}\cap\Gamma\neq\emptyset}} q_i h_i\,.$$

We expand the exponentials, using the Griffiths' first inequality

$$\int q_i^k dq_{Y\cap\Lambda}\geqslant 0,\quad k=1,2,\cdots$$

and the inequality $h\leqslant\xi$ and resum in order to prove (H7). Now we introduce the notation

$$E_h(A(q))=Z^{-1}(Y,h)\int A dq_{Y\cap\Lambda,\Gamma}\,. \tag{H8}$$

From (H3) follows

$$E_h\left(\left[\prod_{i\in N_+}\chi_+(q_i)\right]\left[\prod_{i\in N_-}\chi_-(q_i)\right]\right)=Z'(Y,h)Z(Y,N,\Gamma)\,. \tag{H9}$$

The lemma follows if we can prove that

$$E_h\left(\left[\prod_{i\in N_+}\chi_+\right]\left[\prod_{i\in N_-}\chi_-\right]\right)\leqslant E_\xi\left(\prod_{i\in N_{\text{plus}}}\chi_+\right). \tag{H10}$$

In order to prove (H10) we apply the FKG inequalities four times! (See Ref. 5.)

Indeed, $\Pi\chi_+$ is monotone increasing and $\Pi\chi_-$ is monotone decreasing. Then FKG inequality gives

$$E_h\left(\left[\prod_{i\in N_+}\chi_+\right]\left[\prod_{i\in N_-}\chi_-\right]\right)\leqslant E_h\left(\prod_{N_+}\chi_+\right)E_h\left(\prod_{N_-}\chi_-\right).$$

A second application of the FKG inequality looking at $\Sigma g_i h_j$ as an external field shows that

$$E_h\left(\prod_{N_+}\chi_+\right)\leqslant E_\xi\left(\prod_{N_+}\chi_+\right).$$

In $E_h(\Pi_{N_-}\chi_-)$ we use the symmetry $q\to -q$, which gives

$$E_h\left(\prod\chi_-\right)=E_{-h}\left(\prod\chi_+\right).$$

From the FKG inequality as above

$$E_{-h}\left(\prod\chi_+\right)\leqslant E_\xi\left(\prod\chi_+\right),$$

and another application of the FKG inequality gives

$$E_h\left(\left[\prod_{N_+}\chi_+\right]\left[\prod_{N_-}\chi_-\right]\right)\leqslant E_\xi\left(\prod_{N_{\text{plus}}}\chi_+\right).$$

This completes the proof of the lemma H3. We need one more lemma.

Lemma H 4: Let Δ^l be an l-lattice square in Y. Then under the hypotheses of Proposition C1 with $m_0\geqslant 1$ we have

$$|Z(\Delta^l,N(\partial\Delta^l),\phi,\partial\Delta^l)|^{\pm}\leqslant\exp[O(\lambda/m^2)l^\nu]\,. \tag{H11}$$

This is a special case of Proposition C1 (see Ref. 5 and Ref. 9).

Now we get the final result of this appendix.

Proposition H 5: Under the hypotheses of Theorem 1 we have

$$\frac{Z(Y^c,N(\partial Y^+),N(\partial Y^-),\partial Y)}{Z(\Lambda)}\leqslant\exp[O(\lambda/m^2)|Y\cap\Lambda|]\,. \tag{H12}$$

Proof: The proof follows as in Ref. 5 or Ref. 9 by using the identity

$$\begin{aligned}\frac{Z(Y^c,N^+,N^-,\partial Y)}{Z(\Lambda)}&=\frac{Z(Y^c,N^+,N^-,\partial Y)}{Z(Y^c,N_{\text{plus}},\phi,\partial Y)}\\ &\times\frac{Z(\mathbb{R}^\nu,N_{\text{plus}}U(Y\cap\Lambda),\phi,(\Phi\cup B(\Lambda\cap Y)))}{Z(\Lambda)}\\ &\times\frac{1}{\Pi_{\Delta^l\subset Y}Z(\Delta^l,N(\partial\Delta^l),\phi,\partial\Delta^l)}\,,\end{aligned} \tag{H13}$$

by applying Lemmas H3 and H4, and noting that the second factor in (H13) is smaller than one.

[1] M. Kac and C. Thomson, "Phase transitions and eigenvalue degeneracy of a one-dimensional anharmonic oscillator," Stud. Appl. Math. **48**, 257–64 (1969).

[2] J. Glimm and A. Jaffe, "Critical Problems in Quantum Fields," in *Colloques Internationaux C.N.R.S.* (Marseille, France, 1975), p. 157–73.

[3] D. Isaacson, "The Critical Behavior of the Anharmonic Oscillator," preprint, Rutgers University, 1976.

[4] J. Rosen, "The Ising Model Limit of the ϕ^4 Lattice Fields," preprint, Rockefeller University, 1977.

[5] J. Glimm, A. Jaffe, and T. Spencer, "A Convergent Expansion about Mean Field Theory I, II," Ann. Phys. **101**, 610–30, 631–69 (1976).

[6] B. Simon, "The $P(\phi)_2$ Euclidean (Quantum) Field Theory" (Princeton U.P., Princeton, New Jersey, 1974).

[7] J. Glimm, A. Jaffe, and T. Spencer, "The particle structure of the weakly coupled $P(\phi)_2$ model and other applications of high temperature expansion. Part II. The cluster expansion, in *Constructive Quantum Field Theory*, Lectures notes in Physics, edited by G. Velo and A. Wightman (Springer-Verlag, Berlin, 1973), Vol. 25.

[8] F. Constantinescu and B. Ströter, "Exponential Decay of Derivatives of Covariance Operators on the Lattice," preprint, University of Frankfurt, 1978.

[9] B. Gidas, "The Glimm–Jaffe–Spencer Expansion for classical Boundary Conditions and Coexistence of Phases in the ϕ_2^4 Euclidean (Quantum) Field Theory," preprint, Rockefeller University, 1978.

Consequences of weakening the positivity property of Wightman quantum field theories

Glenn E. Johnson and Daniel I. Fivel
Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742

(Received 15 June 1979; accepted for publication 26 October 1979)

A weaker form of the notion of a Wightman theory is introduced. The weakened form embodies all the requirements of a Wightman theory. However, the consequences of local commutativity are obtained without an assertion that local commutation can be stated in terms of the field operators. Models are presented that illustrate the usefulness of the weakened requirements.

I. INTRODUCTION

The search for nontrivial Wightman field theories in three plus one dimensions has as yet been unsuccessul. In this paper we explore the consequences of modifying the notion of a Wightman theory in a mild way which does not alter its physical content. We shall illustrate the additional flexibility that results from this modificiation by showing that it facilitates a direct assault on the problem of constructing the Green's functions of the theory. Many simple Green's functions that must be discarded out-of-hand given the usual definitions are acceptable candidates in our modified version of the theory.

The possibility of modifying the definition of a Wightman theory without altering its physical content hinges upon the observation that the class of test functions used in the construction of the Hilbert space of physical states need not coincide with the space of test functions used to verify the local commutativity and spectral conditions. If one regards these latter two requirements on the support of the Green's functions as a condition that the Green's functions possess analyticity properties which guarantee the causal behavior of scattering amplitudes, then it is not necessary to demand that it is possible to test strictly local commutation rules for field operators. For example, if the Wightman distributions are defined on the Jaffe class,[1] but are only positive on a smaller class of functions rich enough to verify the asymptotic conditions, we will be able to ensure the probability interpretation of amplitudes and simultaneously guarantee the consequences of the local commutativity properties of fields. This will be accomplished without having to actually manufacture field operators smeared on compact space-time regions as an intermediate step. Indeed, if one eliminates the requirement that the Wightman distributions be positive on the entire Jaffe class, operators associated with compact space-time regions may not even exist because functions associated with bounded regions of space-time may not be included in the Hilbert space of physical states.

In a certain sense, one may regard the proposed weakened form of the Wightman theory as a hybrid of S-matrix theory and field theory in which causality is realized in a manner similar to the former while the other requirements can be stated in terms of the field operators.

We shall illustrate the flexibility gained by the suggested modification of the notion of a Wightman theory as follows: We show that it is possible to construct a class of distributions with a prescription ensuring the validity of the linear requirements (i.e., local commutativity, spectrum, Poincaré invariance and Hermiticity conditons) and the cluster requirement of a Wightman theory, but not the positivity property. We then observe that if the positivity requirement is weakened as described above, mechanisms exist whereby it may be possible to modify the model distributions to *make* them positive without disturbing the validity of the other properties. This possibility would be unavailable however, if one insisted on the usual version of the positivity property. We shall see that although these models are physically trivial, they have certain nontrivial aspects.

II. DEFINITION OF A MODIFIED WITHMAN THEORY

For the convenience of the reader we present some definitions: Let $F(R^{4n})$ by a nuclear,[2] countably normed space contained in $\mathcal{L}_2(R^{4n})$ whose elements are $f(x)_n$, where $(x)_n \equiv (x_1,x_2,...,x_n)$ in which $x_j \in R^4$. Let Ω_F be the set of all terminating sequences, $\underline{f} = (f_0 f(x)_1,...,f(x)_n,...)$, ($f_0 = \text{const.}$), in which addition and scalar multiplication are defined componentwise. Ω_F becomes a topological ring with unit element, $\underline{1} \equiv (1,0,0,...)$, under the sequential product $*$ defined by

$$\underline{f}*\underline{g} = \left\{ f_0 g_0,..., \sum_{k=0}^{n} f(x_1,...,x_k)\, g(x_{k+1},...,x_n),... \right\},$$

with an involution: $\underline{f} \to \underline{f}^* \equiv \{\bar{f}_0,...\bar{f}(x_n,...,x_1),...\}$. Ω_F is the union of nuclear spaces with the topology inherited from F.

A continuous linear functional, W, on Ω_F is said to be positive if $W(\underline{f}*\underline{f}^*) \geqslant 0$ for all $\underline{f} \in \Omega_F$ and such functionals define right ideals, I_W, in Ω_F. Namely: $I_W = \{\underline{f} | W(\underline{f}*\underline{f}^*) = 0\}$. (The fact that $W(\underline{f}*\underline{g}^*) = 0$ when $\underline{f} \in L_W$ for every $\underline{g} \in \Omega_F$ follows from Cauchy's inequality, $|W(\underline{f}*\underline{g}^*)|^2 = W(\underline{f}*\underline{g}^*)W(\underline{g}*\underline{f}^*)W(\underline{g}*\underline{g}^*)$, which holds only for positive functionals). Modding out I_W and noting that $(\underline{f},\underline{g}) \equiv W(\underline{f}*\underline{g}^*)$ is a sesquilinear, positive form, Ω_F/I_W becomes a pre-Hilbert space under the inner product, (,), and may be completed to a Hilbert space, $H(W,\Omega_F)$, under the corresponding norm.[3]

If G is a group of transformations of R^4 and $g \in G$, one defines $g\underline{f}$ by replacing each x_j by $g_j x$ in the arguments of the constituent functions. Ω_F is G-invariant if each $F(R^{4n})$ is G-invariant, and W is G invariant if $W(g\underline{f}) = W(\underline{f})$ for all $\underline{f} \in \Omega_F$ and $g \in G$.

Definition: A relativistic quantum "theory" is a rigged

 0022-2488/80/040891-05$1.00

Hilbert space $\Omega_F C H C \Omega'_F$ in which Ω_F and W and G-invariant, and G is the Poincaré group.

This definition is nonvacuous since the free scalar field satisfies its conditions. The Wightman reconstruction theorm (a variant of the GNS construction) allows one to define "field" operators whose vacuum (cyclic state) expectation values are identified with the Wightman function, $\langle 0,P(\phi,\underline{f})0\rangle = W(\underline{f})$.

Particular classes of relativistic quantum theories are obtained by imposing further restrictions on F and W (hence on $H(W,\Omega_F)$. The fact that events become uncorrelated as they are spacelike separated is reflected in the cluster condition which guarantees that the field operators are irreducible[4,5] and that the vacuum is unique.

Definition: The Wightman functional, W, is said to cluster if for every $\underline{f}, \underline{g}\in\Omega_F$, $(a)^2 = -1, \lim_{\lambda\to\infty} W(\underline{f}*T(\lambda a)\underline{g})_{n+m} = W(\underline{f})_n W(\underline{g})_m$ where $T(\lambda a)g(x)_m \equiv g(x_1-\lambda a,...,x_m-\lambda a)$.

This property is equivalent to requiring that the truncated n-point Wightman distributions decay in their spatial difference variables. The theory must be local and the spectrum of the energy-momentum operators must lie in the closed forward cone (e.g., be a particlelike spectrum). Thus we need the following:

Definition: W is said to be strictly localizable if it is simultaneously a continuous linear functional on Ω_J and Ω_F, where J is the Jaffe class. We remind the reader that the positivity requirement necessary to the construction of $H(W,\Omega_F)$ applies only to the elements of Ω_F.

Definition: A strictly localizable W is said to be local if, for every n, $W(f)_n = 0$ when $f(x)_n = g(...,x_i,x_{i+1},...) - g(...,x_{i+1},x_i,...)$ *and* $x_i,x_{i+1}\in$ support of g implies that $(x_i - x_{i+1})^2 < 0$.

Definition: A strictly localizable W obeys the spectral condition if, for every n, $W(f)_n = 0$ when the support of the Fourier transform of $f(x)_n$ does not contain a neighborhood of the point

$$p_1\in\overline{V}^+, \quad p_1+p_2\in\overline{V}^+,...,p_1+p_2+\cdots+p_n\in\overline{V}^+,$$

where $\overline{V}^+$ is the closed forward cone, $\overline{V}^+ = \{p\in R^4 | (p)^2\geqslant 0, o^0\geqslant 0\}$.

The fact that these support requirements hold in J plus the Lorentz-invariance of W imply that the Wightman distributions are analytic functions in the extended, permuted tube,[6] and these analyticity results are independent of any positivity property of W.

Scattering amplitudes of a theory are defined when an asymptotic condition is satisfied: The reconstructed field must approach the freefield associated with an irreducible representation of the Poincaré group in both the remote past and future. The asymptotic condition allows us to make a particle interpretation of the theory. Thus we need the following:

Definition: A one parameter family, $\underline{f}^{(t)}$, of elements of Ω_F, indexed by a real parameter, t, defines an asymptotic sequence if

$$\lim_{|t|\to\infty} W(\underline{f}^{(t)}) = W_0(\underline{f}), \tag{1}$$

where W_0 is the Wightman functional for a free scalar field, and $\underline{f}\in\Omega_F$.

We refer to the set of all asymptotic sequences as an asymptotic space, Ω_A.

This definition is nonvacuous; for example take

$$f^{(t)}(x)_n \in \{\text{span of } \prod_{k=1}^{n} f_k(x_k) | f_k(x) = \frac{1}{(2\pi)^{3/2}} \int d^4p h(p^2) e^{ipx} \times (f_+(p,t) + f_-(p,t))\},$$

where $f_\pm(p,t) = [(p_0 \pm w_p)/2w_p] \exp[i(p_0 \mp w_p)t] \times f_\pm(\mathbf{p})$, $h(m^2) = 1$, $h(\lambda)$ is C^∞ and of bounded support near $\lambda = m^2$, $f_\pm(\mathbf{p})\in J(R^3)$, the Jaffe class, and $w_p \equiv (m^2+\mathbf{p}^2)^{1/2}$. We refer to this as the Haag–Ruelle asymptotic class, and note that for any $t, f^{(t)}(x)_n \in J(R^{4n})$. It is known that the Haag–Ruelle class is an asymptotic space for any Wightman theory.[5] The fact that the Haag–Ruelle class is an asymptotic space does not depend on W being positive on Ω_J.

We defined the n-point Wightman distribution as $W(f)_n \equiv W(\underline{f})$ when $\underline{f} = \{0,...,0,f(x)_n,0,...\}$ and Eq. (1) is equivalent to

$$\lim_{|t|\to\infty} {}^TW(f^{(t)})_2 = \Delta(f), \quad \lim_{|t|\to\infty} {}^TW(f^{(t)})_n = 0, \quad n>2, \tag{2}$$

where Δ is the positive-frequency Pauli–Jordan distribution and ${}^TW_n$ are the n-point truncated[7] Wightman distributions. We note that, for the Haag–Ruelle asymptotic class,

$${}^TW(f^{(t)})_2 = \Delta(f), \tag{3}$$

independent of time, t.

Definition: If a strictly localizable, relativistic quantum theory is local, clusters, and satisfies the spectral conditon, and if $F = J$, then it is a usual Wightman theory. We remove the condition $F = J$, and refer to these as off-Wightman theories. If Ω_F contains an asymptotic space, Ω_A, the off-Wightman theory will be referred to as a modified Wightman theory.

In attempting to construct nontrivial Wightman theories, perhaps the most difficult requirement to implement is the positivity property. It is interesting to note that if $\Omega_F = \Omega_A$, an asymptotic positivity property follows from the other properties. The asymptotic property states that $W(\underline{f}^{(t)}) = W_0(f) + \epsilon(\underline{f}^{(t)})$, where $\lim_{|t|\to\infty} \epsilon(\underline{f}^{(t)}) = 0$ and $W_0(\underline{f}*\underline{f}^*)\geqslant 0$, since the freefield theory exists. Hence, for a sufficient large $|t|$, $W(\underline{f}^{(t)}*\underline{f}^{(t)*})\geqslant 0$. (For the Haag–Ruelle asymptotic class, $|\epsilon(\underline{f}^{(t)})\leqslant C(\underline{f})|t|^{-3/2}$ for large $|t|$.) However, asymptotic positivity is insufficient to construct Hilbert spaces, first of all because given any t, we are not guaranteed that there exists a (positivity-preserving) closure to the subset of Ω_F for which W is positive. We know that any two representations of the free, scalar field of mass m are unitarily equivalent, and therefore, formally, asymptotic positivity assures us that a unitary S-matrix exists. However, without the positivity property of W, that is, without an interpolating field, we cannot calculate the S-matrix elements.

The requirement that positivity hold only on some re-

stricted class of functions has immediate and far-ranging implications. For example, if we choose F as the Haag–Ruelle class, then we know that $W(f^{(\prime)})$ can be identified as the vacuum expectation values of a quasi-local field: $q(x) = (2\pi)^{-4}\int e^{ipx}h(p^2)\tilde{\phi}(p)\,d^4p$, smeared with smooth solutions of the Klein–Gordon equation.[8] Equation (3) states that the two-point function of this field is identical to the two-point function of a freefield, and normally we would conclude from this that the field was a free, trivial field. However, even though $\|(\Box + m^2)\phi(f^{(\prime)})|0\rangle\|^2 = 0$ we cannot use locality to conclude that $\phi(x)$ is a free field. The fact that $(\Box + m^2)\phi(f^{(\prime)})$ annihilates the vacuum is as much a property of the restricted functions, $f^{(\prime)}(x)$, as it is of the field, $\phi(x)$. If F is not sufficiently rich in functions, we cannot use functions of compact support to etablish that all vacuum expectation values of $(\Box + m^2)\phi(x)$ with fields $\phi(x_i)$ vanish on their Jost points and therefore vanish everywhere. J contains functions of compact support, however F need not. Therefore, in general, for a modified Wightman theory, $\phi(x)|0\rangle = 0$ does not imply $\phi(x) = 0$. This has wide-ranging implications (for example for the usual proof of the spin-statistics theorem) which we have not explored.

III. MODELS

The first model is entirely trivial but illustrates that there are realizations of the modified definitions that do not have the properties of a Wightman theory. We define the first model by setting all the truncated Wightman distributions to zero expect for

$$^{T}\widetilde{W}(p,q)_2 = \theta(p_0)\delta^4(p-q) \times\left[\delta(p^2-m^2) + \int_{4m^2}^{\infty} d\mu(a)\delta(p^2-a)\right],$$

where $d\mu(a)$ is any tempered measure. Because of (3), the theories defined by these Wightman distributions satisfy all our axioms if we take F as the Haag–Ruelle class. Of course, if $F = J$, and if $\mu(a)$ is a positive measure, these would be generalized free-field models.[9] All the scattering predictions of this model coincide with those of the free theory, yet it is not *the* free-field theory.

Next we consider a wide range of possible theories. Taking the familiar trivial theories as our guide, we construct our models entirely of products of two-point distributions, but in an attempt to escape the triviality of scattering amplitudes, we allow a wider range of distributions that that utilized in the familiar models.

We consider generalized Pauli–Jordan (GPJ) distributions, $D(x-y)$, that satisfy the conditions listed below:

(a) distribution, $D(x)\in J(R^4)$;
(b) invariance, $D(\Lambda x) = D(x)$ for every proper, orthochoronous Lorentz transformation, Λ;
(c) spectrum, $\widetilde{D}(p) = 0$ unless $p\in\overline{V}^{+}$;
(d) locality, $D(x) = D(-x)$ for $x^2 < 0$;
(e) cluster, $\lim_{\lambda\to\infty} D(x-\lambda a) = 0$ for $a^2 = -1$;
(f) hermiticity, $D(x) = \overline{D}(-x)$.

Note that the positive frequency Pauli–Jordan distribution satisfies these requirements in addition to the positivity property, $\Delta(f*\bar{f}) = \int\Delta(x-y)f(x)\bar{f}(y)\,dx\,dy \geqslant 0$. We consider the GPJ distributions which can be represented

$$\widetilde{D}(p) = (2\pi)^4\int_0^{\infty} d\mu(m^2)H(p^2)g(\Box p)\theta(p_0)\delta(p^2-m^2), \tag{4}$$

where $\mu(m^2)$ is a Jaffe class measure and both H and g are polynomials.[10]

Since D satisfies conditions (a)–(d), it has an analytic continuation to imaginary time. Its two-point Schwinger function is

$$D_e(x) = D(i|x^0|,\mathbf{x}) = \int d\mu(m^2) \times\int d^4p e^{ipx}H(|p|^2)\,g(\Delta\phi)(|p|^2+m^2)^{-1}\cdot(2\pi)^{-1},$$

where Δ is the Lapalacian in four dimensions, and $|x|^2$ is the Euclidean length squared of $x\in R^4$. We can make D_e Nelson–Symanzik positive, that is, $\int D_e(x-y)f(x)\bar{f}(y)\,dx\,dy = 0$. [For example, $H = (1-p^2)^2$, $g = 1 + \Box p$.] However, there is no physical positivity property of the GPJ two-point distributions for g not a constant. Indeed, no matter how one chooses H, $g(\Box_p)|\mathbf{f}(p)|^2$ is not positive as $\bar{f}$ runs through J, and therefore D is indefinite.[11]

n-point Wightman distributions satisfying the linear properties can easily be fashioned from the CPJ distributions. We choose an n-point distribution of the form

$$T_1(x)_n \equiv \prod_k D_k(x_{j_k} - x_{i_k}), \tag{5}$$

where the x_j are picked from $(x)_n$ according to a scheme that we refer to as the "organization" of the model. To ensure the distribution and spectral properties, each argument, $(x_j - x_i)$, must be ordered with $j > i$.

If each D_k is a GPJ distribution, the Poincaré invariance of T_1 is manifest. In order to prove the distribution property, we note that, because of the argument ordering condition, T_1 is expressed entirely in terms of sums of difference variables, $\epsilon_k \equiv x_{k+1} - x_k$. The convolution theorem for test functions then implies that $T_1(x)_n \in J(R^{4n})$ if $T_1(x)_n \equiv \tau(\epsilon)_{n-1} \in J(R^{4(n-1)})$.

$$\tau(f)_n = \int d^4\epsilon_n \prod_k \Delta(\epsilon_{j_k,} + \epsilon_{j_k} + \cdots \epsilon_{i_k}) \times g_k(-(\epsilon_{j_{k-1}} + \cdots + \epsilon_{i_k})^2) \times H_k(-\Box_{(\epsilon_{j_k-1}+\cdots+\epsilon_{i_k})})f(\epsilon),n,$$

where f is C^{∞}, and polynomials are multipliers in J, so $\Pi_k g_k H_k f \equiv F \in J$. Therefore,

$$|\tau(f)_n| = (2\pi)^{-4n}|\tilde{\tau}(\tilde{f})| \leqslant \int\prod_k d^4p_k\,\theta(p_k^0)\delta(p_k^2-m^2) \times\exp\left[-QW\left(\sum_{j=1}^{n}\left|\sum_{l_j} p_{l_j}\right|^2\right)\right] \cdot\sup_{(q)_n}\left|\exp\left[Qw\left(\sum_{j=1}^{n}|q_j|^2\right)\tilde{F}(q)_n\right]\right|,$$

where Q is a sufficiently large integer, and w is a Jaffe indicator function.[1] Because each ϵ_j appeared with a positive relative sign, this bound is finite,[12] and the definition of Jaffe norms provides that

$$|\tau(f)_n| \leqslant C\|f\|_J,$$

which is the distribution property. The spectral property now follows immediately from

$$\tilde{\tau}(q)_n = \int \prod_k d^4p_k \, \tilde{D}_k(p_k) \sum_{j=1}^{n} \delta^4\Big(q_j - \sum_{l_j} p_{i_j}\Big),$$

and the fact that $\bar{V}^+$ is a cone. Using the hermiticity of the GPJ distributions, we establish the hermiticity of T_1:

$$T_1(x)_n = \prod_k D_k(x_{j_k} - x_{i_k}) = \prod_k \bar{D}_k(-x_{j_k} + x_{i_k})$$
$$= \bar{T}_1(x_n,...,x_1) = \bar{T}_1(-x_1,..., -x_n).$$

This last equality reflects the validity of the PCT theorm.

The locality property of the GPJ distributions now allows us to implement locality simply by appropriately summing different distributions of the form 5).

$$T(x)_n \equiv \sum_{\pi} T_1(x_{\pi(1)},...,x_{\pi(n)}), \tag{6}$$

where the sum is over all permutations of the arguments of T_1, and the argument ordering condition is applied to each term. This prescription preserves the validity of the other properties and institutes the proper symmetry for T.

If we choose each truncated n-point Wightman distribution to be a linear combination (with real coefficients) of the distributions, $T(x)_n$, then the Wightman distributions themselves satisfy all the linear properties. (The relation between the truncated and Wightman distributions[6] displays $W(x)_n$ itself in the form of a linear combination of $T(x)_n$ distributions.) In each term of ${}^TW(x)_n$ is connected, i.e., if it is not composed of two or more factors that depend on distinct space-time variables, then the cluster property is satisfied. We note that each connected $T_1(x_1,...,x_i,x_{i+1} - \lambda a,...,x_n - \lambda a)$ has at least one factor of $D(x_{j_k} - x_{i_k} - \lambda a)$ that vanishes as $\lambda \to \infty$, and thus, ${}^TW_n \to 0$ as its arguments are space-like separated.

It should be evident from the above construction that we can form new distributions satisfying the linear and cluster properties by adding the truncated distributions of any two theories that satisfy the linear and cluster properties. In particular, we can add the distributions of a positive theory to any set of truncated distributions satisfying the linear and cluster properties, and thereby hope to make an indefinite theory less negative.[12]

We have now constructed a class of distributions satisfying all but the positivity properties of a Wightman theory. The class includes the familiar Wick monomial theories, in addition to a large number of models which are manifestly nonpositive because of the distributional derivatives involved in the definition of the GPJ distributions.

The only positive candidate distributions available are the physically trivial theories constructed from the Pauli–Jordan distribution and its derivatives. If we wish to render a more general set of distributions positive by adding in the trivial but positive distributions, the first requirement is that the magnitudes of the positive and indefinite terms be comparable. The indefinite terms involve x^{2n} times the test functions $f(x)(\Box_p^n \tilde{f}(p))$ while the positive terms involve only $f(x)(\tilde{f}(p))$. Therefore, by choosing $f(x)$ such that $\tilde{f}(p)$ is small with respect to $\Box_p \tilde{f}(p)$, the indefinite terms can always be made to dominate. This is the case if we choose $F = J$. *However, if we restrict our choice of functions to the Haag–Ruelle class, we get much stronger bounds on the magnitudes of general terms.* This is essentially because the derivatives are applied to a fixed mollifier function, $h(p^2)$, and therefore the positive and general terms can be comparable in magnitude. We now derive such a bound[13]; take:

$${}^TW(f)_n \equiv \int \prod_{k=1}^{n} f_k(x_k) \overset{\leftrightarrow}{\partial}_{0k} {}^TM(x)_n \, d\mathbf{x}_k \Big|_{x_k^0 = 5},$$

where ${}^TM(x)_n$ is the distributional Fourier transform of $\Pi_{k=1}^n h(p_k^2)\tilde{W}(p)_n$, and each f_k is a smooth solution of the Klein–Gordon equation. (f is an element of the Haag–Ruelle class.) We now take advantage of the cluster property to write

$$|{}^TW(f)_n| \leqslant \int \Big| \overset{0}{f}_1(\mathbf{y},t)\, d\mathbf{y} \prod_{k=2}^{n} \sup_{\mathbf{x}_k} \Big| \overset{0}{f}_k(\mathbf{x}_k,t) \Big| \, \beta^{n-1} C \Big|,$$

where 0 may denote time differentiation, $\beta = \int d\epsilon \, (1+\epsilon^2)^{-2}$, and $\Big| \Pi_{k=2}^n (1+\epsilon_k^2)^2 \overset{0}{{}^TM}(x)_n \Big| \leqslant C$. C is independent of t because of translation invariance. Hence, we see that $W(f)_n$ will be bounded by expressions involving only f and its integrals and derivatives, and not derivatives of its Fourier transform! This allows up to hope that trivial terms may be sufficiently positive that they can make the more general theories positive. Therefore, one may attempt to institute positivity for this wide range of models simply by adding in the truncated Wightman distributions of a trivial theory and by choosing F in a sufficiently clever manner.

These GPJ distribution models are interesting because they display nontrivial scattering in the following sense:

$$S \equiv \int \tilde{T}\tilde{W}(p_1,...,p_j, -p_{J+1},..., -p_n)$$
$$\times \prod_{k=1}^{n} (p_k^2 - m^2) f_k(p_k)\, d^4p_k,$$

in the limit each $f_k(p_k)$ becomes supported on $p_k^0 = w_k$, is nonvanishing. $TW(x)_n$ is the time-ordered Wightman distribution. Of course, given a Wightman theory, this is the expression for an S-matrix element.[13] However, the models, if they can be made positive, are actually trivial if we use the Haag–Ruelle asymptotic condition. In order for the theory to be nontrivial, the operator,

$$A \equiv \int [a_{\text{in}}(\mathbf{p}) + a_{\text{out}}(\mathbf{p})] f(\mathbf{p})\, d\mathbf{p},$$

must not vanish. Given the Haag–Ruelle asymptotic condition, matrix elements of A are

$$\langle (p)_i, A(p_{i+2},...,p_n) \rangle$$
$$= \lim_{t\to\infty} N \int \frac{\sin(p_i^0 - w_{p_i})t}{(p_i^0 - w_{p_i})\, 2wp_i}$$
$$\times (p_i^2 - m^2) h(p_i^2) \tilde{W}(p)_n f(\mathbf{p}_i)\, d^4p_i.$$

N is a normalizing factor. After we apply the distributional derivatives of $\tilde{W}_n$, nontriviality is equivalent to

$$\lim_{t\to\infty} (t)^m \int \exp[i(p_i^0 - w_{p_i})t\,](p_i^0 + w_{p_i}) \frac{1}{2w_{p_i}}$$

$$\times h(p_i^2)\widetilde{W}(p)_n f(\mathbf{p}_i)\, d^4 p_i \neq 0,$$

where $\widetilde{W}_0(p)_n$ has no principal value singularities at $p_i^0 = w_{p_i}$ since it is composed entirely of ordinary Pauli–Jordan distributions. However, we know that these limits vanish.[14] We wish to emphasize that an asymptotic condition involves both the properties of W and the sequence of functions $f_-^{(t)}$, and therefore it is conceivable that an asymptotic condition could be formualted in such a manner that these models would be nontrivial in the conventional sense. However, that raises questions beyond the scope of this paper.

Apart from the possibility of opening new directions in the search for nontrivial quantum field theories, there is an intrinsic interest to the elimination of $F = J$ as a requirement.[15] The Jaffe class contains functions, such as functions of bounded support in space–time, which cannot be realized as particle states. The positivity property is needed only to ensure the probability interpretation of the theory which is most relevant to particlelike states. Hence, we should not require the positivity property except for such states as the physics absolutely demands.

[1]A. Jaffe, Phys. Rev. **158**, 1454 (1967).

[2]A nuclear space is chosen in order to ensure the existence of (generalized) eigenstates and the validity of the spectral theorem when applied to the physical operators that generate the Poincaré group.

[3]This is a rigged ("equipped") Hilbert space: $\Omega_F \prec H(W,\Omega_F) \prec \Omega'_F$. The eigenstates of physical operators are, in general, generalized functions in Ω'_F.

[4]H.J. Borchers, Nuovo Cimento **24**, 214 (1962).

[5]D. Ruelle, Helv. Phys. Acta. **35**, 147 (1962).

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

[7]N.N. Bogolubov, A.A. Logunov, and I.T. Todorov, *Introduction to Axiomatic Quantum Field Theory*, trans. by Stephen Fulling and Ludmilla Popova (Benjamin, Reading, Mass., 1975), pp. 272–274.

[8]Ref. 7, Chap. 13.

[9]O.W. Greenberg, Ann. Phys. **16**, 158 (1959).

[10]g must be a polynomial in order that the distribution D be of finite order, and there is no gain in generality by allowing H to be a more general function.

[11]$D(x)$ can be displayed as the difference of two positive, thought not invariant distributions by using the identity:

$$\Box_p |\tilde{f}(p)|^2 = |(1+\Box_p)\tilde{f}|^2 - |\Box_p \tilde{f}|^2 - |\tilde{f}|^2 + \left|\frac{d}{dp^0}\tilde{f}\right|^2 - \sum_{i=1}^{3}\left|\frac{d}{dp_i}\tilde{f}\right|^2.$$

[12]G.E. Johnson, Doctoral thesis (University of Maryland, (1978).

[13]Ref. 7, Chap. 13, theorem 10.4, and lemma 3.2.

[14]K. Hepp, Commun. Math. Phys. **1**, 95 (1965).

[15]See note at end of Sec. 15.5, Ref. 6.

Scale limit and low momentum behavior of Euclidean fields. II

Z. Haba[a)]
Department of Theoretical Physics, University of Bielefeld, Bielefeld, Germany

(Received 19 December 1978; accepted for publication 4 April 1979)

The problem of the limit behavior of a rescaled random field $\epsilon^{\alpha}\varphi(x/\epsilon)$ as $\epsilon\to 0$ is considered as an analog for random fields of the classical limit theorems for sums of random variables. If the random field φ does not depend on the scale ϵ, the limit is Gaussian in Euclidean field theory with a mass gap. If φ itself is a function of ϵ (giving a double sequence scheme), any infinitely divisible distribution can appear as the limit. In this formulation the limit of $\epsilon^{-d}\varphi(x/\epsilon)$ as $\epsilon\to 0$ is discussed. It is proved that the limit field (if it exists) is a random field with values independent at every point. A correspondence between this scale limit and low momentum behavior is revealed. It is shown also that the scale limit may be considered as a limit of field theory when $m\to\infty$ and the coupling constant $g\to\infty$ in superrenormalizable theories and $m\to\infty$, $g\to 0$ for nonrenormalizable interactions. The scale limit of some ultraviolet regularized Euclidean fields is calculated. As an application some functional integrals are evaluated in the Ginzburg–Landau model in the low momentum limit.

I. INTRODUCTION AND FORMULATION OF THE PROBLEM

The behavior of sums of a large number of arbitrarily distributed independent[1] or weakly dependent[2] random variables is determined by the classical limit theorems. These sums should have infinitely divisible distributions[1] and under some additional restrictions only the normal distribution is allowed. The limit theorems are of paramount importance as they provide some knowledge about the behavior of the sum disregarding the peculiar distributions of the summands. These classical limit theorems should also govern the behavior of random fields because a random field $\varphi(x)$ may be considered as a collection of (infinite number) random variables fixed at various points x. It would be of some importance for Euclidean field theory to separate this model independent large number behavior from the true dynamical problems.

It has been noticed recently[3,4] that scale transformations of random fields are analogs of sums of random variables. In fact, let us define

$$T_{\alpha}(\lambda)\varphi(x)=\lambda^{\alpha}\varphi(x),$$
$$S(\lambda)\varphi(x)=\varphi(x/\lambda), \tag{I.1}$$

and consider the action of $T_{\eta-d}(\epsilon)S(\epsilon)$ (d is the space dimension) on $\varphi(x)$ smeared out over the region Δ:

$$\varphi_{\Delta}=\int_{\Delta}\varphi(x)\,dx. \tag{I.2}$$

Then we have for $1/\epsilon^{d}=N$,

$$T_{\eta-d}(\epsilon)S(\epsilon)\varphi_{\Delta}=\epsilon^{\eta}\int_{\Delta}\epsilon^{-d}\varphi(x/\epsilon)dx=\epsilon^{\eta}\int_{\Delta/\epsilon}\varphi(x)dx=\sum_{i=1}^{N}\epsilon^{\eta}\varphi_{\Delta_i}. \tag{I.3}$$

Hence, as $\epsilon\to 0$ we deal with an infinite sum of random variables. The normalization factor ϵ^{η} is needed to make $\epsilon^{\eta}\varphi_{\Delta}$ small and the infinite sum convergent. The central limit theorem now requires $\eta=d/2(\epsilon^{d/2}=1/N^{1/2})$ according to the square root law of fluctuations. This central limit theorem is expected to hold (see Refs. 2 and 5) in field theory with a mass gap, i.e., with short range correlations.

Another scale limit has been discussed by Glimm and Jaffe.[6,7] Their scale transformations also transform the bare mass and coupling constant. It can be formulated in another way. We have a Euclidean field φ depending on some parameters (e.g., the bare mass and coupling constant in superrenormalizable field theories), then we perform a scale transformation with the scale parameter ϵ being a function of the parameters κ [we may assume $\epsilon(\kappa)\to 0$ as $\kappa\to\infty$]. In Refs. 6 and 7 $\epsilon\to 0$ when $m_0\to\infty$, $g_0\to\infty$ with the dimensionless coupling constant fixed. This scale limit is an anlog of the double sequence scheme[1] related to the Poisson's limit theorem. In such a case we consider sequences of sums

$$s_n=\xi_{1n}+\cdots+\xi_{r_n n} \tag{I.4}$$

of independent (in each sequence) random variables with $r_n\to\infty$. It is known that any *infinitely divisible* random variable may be the limit of this sequence.[1] Let us note that now the normalization of the sum (I.3) is different. In particular, no normalization factor ϵ^{η} is needed if $\varphi_{\kappa}\to 0$ in the sense that the characteristic function $S_{\kappa}[f]$ of $\int\varphi_{\kappa}(x)f(x)\,dx$ tends to 1 and the characteristic function $S_{\epsilon(\kappa),\kappa}[f]$ of $\int dx f(x)\epsilon^{-d}(\kappa)\varphi_{\kappa}(x/\epsilon(\kappa))$ has a nontrivial limit $S_0[f]$ when $\kappa\to\infty$. This assumption means that when $\kappa\to\infty$ $S_{\kappa}[f]$ tends to 1 as $(S_{\infty}[f])^{\epsilon^{d}(\kappa)}$ (an example of such a behavior of $S_{\kappa}[f]$ as a function of the coupling constant κ is provided by a one-dimensional model considered in Ref. 8). $S_{\kappa}[f]$ should behave in this way if the limit of $S_{\kappa,\epsilon(\chi)}[f]$ is to exist because the characteristic function $S_{\epsilon}[f]$ of $\epsilon^{-d}\varphi(x/\epsilon)$ tends to zero as $(S_0[f])^{\epsilon^{d}}$ [there are ϵ^{-d} summands in the sum (I.3)]. Hence, the limit of $(S_{\epsilon}[f])^{\epsilon^{d}}$ is expected to exist also and to be related to the limit of $S_{\kappa,\epsilon(\kappa)}[f]$.

[a)]On leave of absence from Institute of Theoretical Physics, University of Wroclaw, Wroclaw/Poland.

 0022-2488/80/040896-06$1.00

In this paper we will continue investigation of the scale limit $\epsilon \to 0$ of $(S_\epsilon[f])^{\epsilon^d}$ extending results of the previous paper[9] (hereafter quoted as I). It will be shown in Sec. II that if so formulated a limit of $\epsilon^{-d}\varphi(x/\epsilon)$ exists as a random field it should be a random field with values independent at every point.[10]

The fields with values independent at every point have no continuation to Minkowski space. Nevertheless, they can be of some importance in Euclidean field theory for some approximations in the limit of small (Euclidean) momentum or large mass and coupling constant (Secs. II and IV). In statistical mechanics a system of noninteracting Bosons at zero temperature consists of (infinite number) uncorrelated parts. The density distribution is then a random field with values independent at every point.[11]

II. THE SCALE LIMIT OF EUCLIDEAN FIELDS

Let $S_\epsilon[f]$ be the characteristic function of $\int dx f(x)\epsilon^{-d}\varphi(x/\epsilon)$. Owing to the discussion in Sec. I we are interested in the limit $\epsilon \to 0$ of $(S_\epsilon[f])^{\epsilon^d}$ or in terms of the truncated (connected) Green functions in the limit

$$G_n^{(0)}(x_1,\ldots,x_n) = \lim_{\epsilon\to 0} \epsilon^{d i^{-n}} \frac{\delta}{\delta f(x_1)} \cdots \frac{\delta}{\delta f(x_n)} \ln S_\epsilon[f]$$
$$= \lim_{\epsilon\to 0} \epsilon^d \epsilon^{-nd} G_n\left(\frac{x_1}{\epsilon},\ldots,\frac{x_n}{\epsilon}\right). \tag{II.1}$$

Let us separate the δ function (resulting from the translational invariance) from the Fourier transform of the truncated Green functions denoting

$$\tilde{G}_n(p_1,\ldots,p_n) = (2\pi)^{d/2}\delta(p_1+\cdots+p_n)F_n(p_2,\ldots,p_n). \tag{II.2}$$

It is an "axiomatic" result[12] (cf. also Ref. 13) that $F_n(p_2,\ldots,p_n)$ has no singularities in Euclidean domain in Wightman field theory with a *mass gap*. Therefore, taking the Fourier transform of Eq. (II.1) we get

$$\tilde{G}_n^{(0)}(p_1,\ldots,p_n) = \lim_{\epsilon\to 0} \epsilon^d (2\pi)^{d/2}\delta(\epsilon p_1+\cdots+\epsilon p_n) \times F_n(\epsilon p_2,\ldots,\epsilon p_n)$$
$$= (2\pi)^{d/2}\delta(p_1+\cdots+p_n)F_n(0). \tag{II.3}$$

So the limit (II.1) of the truncated Green functions exists and is in general non-Gaussian ($F_n \neq 0$ for $n>2$). If we considered instead of (II.1) the limit of $\epsilon^{-d/2}\varphi(x/\epsilon)$ with φ independent of ϵ [cf. the discussion below Eq. (I.3)] we would find that only the two-point function $(2\pi)^{d/2}\delta(p_1+p_2)F_2(0)$ remained after taking the limit. This limit behavior is a manifestation of the central limit theorem for random fields. [This result has been obtained before by Dr. V. Enss (unpublished). The author wishes to thank Dr. V. Enss for showing him that the limit of $\epsilon^{-d/2}\varphi(x/\epsilon)$ is Gaussian.] The definition (II.1) of the scale limit allows us to pick higher order Green functions. From the truncated Green functions $G_n^{(0)}$ one can get the Schwinger functions $S_n^{(0)} = \Sigma G_{k_1}^{(0)}\cdots G_{k_n}^{(0)}$ as a sum over partitions of $[1,\ldots,n]$. However, such $S_n(0)$ will not in general be moments of an Euclidean measure. In order to determine when this is the case let us compare $G_n^{(0)}$ with the truncated functions of random fields values independent at every point. The characteristic function of these fields is given by (Ref. 10)

$$S[f] = \exp\left\{\int_{|\lambda|>0}[e^{i\lambda f(x)} - 1 - i\lambda f(x)]d\sigma(\lambda)dx + \mathrm{im}\int f(x)dx - \frac{a}{2}\int f^2(x)dx\right\}, \tag{II.4}$$

where $a \geqslant 0$ and $d\sigma(\lambda)$ is a positive measure such that

$$\int_{|\lambda|>1} d\sigma(\lambda) + \int_{0<|\lambda|<1} \lambda^2 d\sigma(\lambda) < \infty. \tag{II.5}$$

We can then show:

Theorem: The functions $G_n^{(0)}$ (II.3) are truncated Green functions of an Euclidean field if and only if a positive measure $d\sigma(\lambda)$ exists such that

$$F_n = \int \lambda^n d\sigma(\lambda), \quad \text{for } n>2. \tag{II.6}$$

If these conditions are fulfilled, then the scale limit (II.3) determines a random field with values independent at every point with the characteristic function (II.4).

Proof: Taking the Fourier transform of Eq. (II.3) we find that $G_n^{(0)}(x_1,\ldots,x_n)$ is proportional to $\delta(x_2-x_1)\cdots\delta(x_n-x_1)$. From this it follows that $\int G_n^{(0)}(x_1,\ldots,x_n)f(x_1)\cdots f(x_n)\,dx_1\cdots dx_n \sim \int f^n(x)dx$ and next that $S_0[f] = \exp\{\int R(f(x))dx\}$. Then it is proved in Ref. 10 that such a functional $S_0[f]$ is a Fourier transform of a positive measure if and only if it has the form (II.4). Comparing Eq. (II.4) with Eq. (II.3) we get the conditions (II.6) [Eq. (II.4) gives no restriction on $F_1(0)$ and $F_2(0)>0$]

Corollary: The scale limit of an infinitely divisible field always exists and is a random field with values independent at every point.

Proof: For an infinitely divisible field the function $(S_\epsilon[f])^{\epsilon^d}$ is again a characteristic function of a random field. Then the infinitely divisible fields are closed under weak limits.[14] Hence, the limit $\epsilon \to 0$ is again a random field and from the theorem it follows that this limit field has values independent at every point.

It has been shown[15] that the Wick square: φ^2 : of the free Euclidean field is infinitely divisible. So the existence of the scale limit follows from the corollary. We have proved this in another way in Ref. 10. The moment problem (II.6) has been solved with the result (here $d<4$)

$$S_0[f] = \lim_{\epsilon\to 0}\left\{\int d\mu_{m_0}(\varphi) \times \exp\left[\frac{i}{2}\int dx f(x)\epsilon^{-d}{:}\varphi^2{:}(x/\epsilon)\right]\right\}^{\epsilon^d}$$
$$= \int d\mu_0(\Phi)\exp\left[\frac{i}{2}\Phi(f)\right]$$
$$= \exp\left\{\frac{1}{2(2\pi)^{d/2}}\int_0^\infty ds e^{-m_0^2 s} s^{-1-d/2} \times \int d^d x[e^{isf(x)} - 1 - isf(x)]\right\}, \tag{II.7}$$

where μ_{m_0} is the Gaussian measure with covariance

$$\int \varphi(x)\varphi(y)\,d\mu_{m_0}(\varphi) = \frac{1}{(2\pi)}d\int \frac{e^{ip(x-y)}}{p^2+m_0^2}dp. \tag{II.8}$$

We will next show that the scale limit in superrenormalizable filed theories constructed by means of the perturbation $\exp[-g_0^2 V(\varphi)]$ of the free measure (II.8) may be considered as a large mass and large coupling constant limit. Whereas for *nonrenormalizable* interactions this is a *small coupling* (but large mass) limit. Let us consider first the free Euclidean field. From Eq. (II.8) it can be easily seen that the transformation

$$U(\lambda): \varphi(x) \to \lambda^{1-d/2}\varphi(x/\lambda), \tag{II.9}$$

is a unitary transformation from $L^2_{\mu_{m_0}}$ onto $L^2_{\mu_{m_0/\lambda}}$. Hence, the scale transformation,

$$\varphi(x) \to \epsilon^{-d}\varphi(x/\epsilon) = \epsilon^{-1-d/2} U(\epsilon)\varphi(x), \tag{II.10}$$

renormalizes the field and transforms the mass m_0 into $m_0/\epsilon \to \infty$ as $\epsilon \to 0$. So in Eq. (II.7) instead of taking the limit of $\epsilon^{-d} : \varphi^2 : (x/\epsilon)$ we could show that the limit of $m^2/m_0^2 : \varphi^2_{m_0} : (x)$ exists when $m_0 \to \infty$ and equals the r.h.s. of Eq. (II.7) (with $m_0 \to m$). Consider now the interaction $g_0^2 \int : \varphi^k : (x) d^d x$. From the unitary equivalence (II.9) we get

$$\int d\mu_{m_0}(\varphi) \exp\left[-g_0^2 \int : \varphi^k : (x) dx\right] F(\epsilon^{1-d/2}\varphi(x/\epsilon))$$
$$= \int d\mu_{m_0/\epsilon}(\varphi) \exp[-g_0^2 \epsilon^{k(d/2-1)-d}$$
$$\times \int : \varphi^k : (x) dx] F(\varphi(x)). \tag{II.11}$$

We can see that the scale limit $\epsilon \to 0$ on the l.h.s. of Eq. (II.11) is equivalent to the limit of large mass and coupling constant going to

$$g^2 = g_0^2 \epsilon^{k(d/2-1)-d} \to \begin{cases} \infty & \text{if } k(d/2-1)-d < 0 \text{ (superrenormalizable theory),} \\ 0 & \text{if } k(d/2-1)-d > 0 \text{ (nonrenormalizable theory).} \end{cases} \tag{II.12}$$

The result that an independent value field theory is a small coupling limit of a nonrenormalizable theory suggests that this field theory can be a starting point for perturbation expansion in g_0 for nonrenormalizable interactions. [It was suggested first by Klauder[16] that independent value field can be a gradient-free model of a nonrenormalizable field theory. However, it should be emphasized that our independent value field and Klauder's are different (cf. the discussion of this point in I).] An example of such an expansion was considered in I. That a large mass and large coupling limit of superrenormalizable theory can give a model of unrenormalizable field theory has been advocated by Parisi.[17] There are known examples of such a behavior on the Lagrangian level. The limit of infinite mass of the intermediate boson leads to the Fermi current–current interaction and the limit of the interaction $g_0^2(\varphi^2 - a^2)^2$ as $g_0 \to \infty$ gives the σ model constraint $|\varphi| = a$.

The facts that the scale limit is at the same time a large mass and a small momentum limit [Eq. (II.3)] are clearly not independent. From dimensional analysis we find that G_n should have the form

$$G_n(p_1,\ldots,p_n) = m_0^d u(g_0^2 m_0^{1/\delta}) g_n(p_1/m_0, p_1 g_0^{2\delta},\cdots), \tag{II.13}$$

where

$$1/\delta = k(d/2-1) - d,$$

hence the limit $m_0 \to \infty$, $g_0^2 m_0^{k(d/2-1)-d} = \text{const}$ of $m_0^{-d} G_n(p_1,\ldots,p_n)$ is finite and coincides with the limit (II.3). Let us remark that the expansion in powers of momenta of the functions $F_n(p_1,\ldots,p_n)$ [Eq. (II.2)] is then an expansion in powers of $1/m_0$ and g_0^{δ}.

We have shown (Corollary) that the scale limit of an infinitely divisible Euclidean field is again an Euclidean field. In Sec. III we will give some examples of limits of fields which are not infinitely divisible. Nevertheless, the conditions (II.6) are rather restrictive and it is not difficult to give an example of an Euclidean field whose truncated Green functions have the scale limit (II.1) but the conditions (II.6) are not fulfilled. It is known that the truncated four point Green function of the field φ in Euclidean φ^4 field theory is nonpositive.[18,19] So $F_n(0)$ is nonpositive and $F_n(0)$, being the physical coupling constant in φ^4 (Ref. 19), is expected to be strictly negative. Hence, the condition (II.6) is violated.

If the numbers $F_n(0)$ are known then the moment problem (II.6) has a solution[20] if and only if for all N and all complex $\alpha_1,\ldots,\alpha_N$

$$\sum_{m,n=1}^{N} \bar{\alpha}_m \alpha_n F_{n+m}(0) \geqslant 0.$$

It is possible to express $F_n(0)$ in terms of the effective potential.[21] Define first the generating functional for vertex functions $\Gamma[\varphi_c]$

$$\Gamma[\varphi_c] = \ln S[-iJ] - \int dx J(x)\varphi_c(x), \tag{II.14}$$

where

$$\varphi_c(x) = \frac{1}{S[J]} \frac{\delta S[-iJ]}{\delta J(x)}, \tag{II.15}$$

then from (II.14) we have

$$J(x) = -\frac{\delta \Gamma[\varphi_c]}{\delta \varphi_c(x)}. \tag{II.16}$$

It is assumed that $\Gamma(\varphi_c)$ can be expanded in terms of φ_c and its derivatives (this is equivalent to an expansion of the vertex functions $\Gamma(p_1,\ldots,p_n)$ in powers of p_μ around $p = 0$):

$$-\Gamma[\varphi_c] = \int dx[V_e(\varphi_c(x)) + \tfrac{1}{2} Z(\varphi_c(x))(\nabla\varphi(x))^2 \cdots]. \tag{II.17}$$

The first term in the expansion is called the effective potential. Now we take the scale limit (II.1) of Eq. (II.14). In this limit the derivatives in Eq. (II.17) vanish. In such a case

$$J(x) = V_e'(\varphi_c(x)). \tag{II.18}$$

Equation (II.6) is then equivalent to the following equation:

$$iuV_c'(u) - iV_c(u) = \int (e^{i\lambda V_c'(u)} - 1 - i\lambda V_c'(u)) d\sigma(\lambda)$$

$$-a^2/2(V_\epsilon'(u))^2 + imV_\epsilon''(u). \quad \text{(II.19)}$$

In the one-loop (tree) approximation the effective potential coincides with the classical one. Then Eq. (II.19) has no solution for polynomial interactions. However, already in the two-loop approximation nonpolynomial terms appear (cf. Ref. 21). Nevertheless, the loop expansion seems useless for solution of Eq. (II.19). It would be possible to answer the question about the existence of solutions of Eq. (II.19) if either σ of V_ϵ were known exactly.

We could also compute $F_n(0)$ in the conventional perturbation theory. Then $F_{2k}(0)$ is negative in the lowest order in g_0^2 for the polynomial interaction $g_0^2 : \varphi^{2k} :$. So Eq. (II.6) fails. The situation looks more promising if we consider $: \varphi^2 :$ in the field theory with interaction. Then Eq. (II.13) is fulfilled in each order of perturbation expansion in g_0 for g_0 sufficiently small and therefore in each order the moment problem (II.6) has a solution. For φ^2 we will derive some exact results in Sec. III.

III. A SCALE LIMIT OF REGULARIZED $(: \varphi^2 :)^r$

We define first the regularized φ^{2^k}

$$\vdots \varphi_\rho^{2^k} \vdots (x) = \Big(\int \cdots \Big(\int \Big(\int : \varphi^2 : (y_1) \rho_{y_2}(y_1) dy_1 \Big)^2 \times \rho_{y_3}(y_2) dy_2 \Big)^2 \cdots \rho_x(y_{k-1})\, dy_{k-1} \Big)^2, \quad \text{(III.1)}$$

where $\rho_k(y')$ has its support concentrated at $y - y' \sim 0$. Then

$$\exp\Big[- g_0^2 \int \vdots \varphi_\rho^{2^k} \vdots (x) h(x) dx \Big] = \int d\mu_1(\psi) \exp\Big[2i g_0 \int \vdots \varphi_\rho^{2^{k'}} \vdots (y) \times \rho_k(y) \sqrt{h(x)}\, \psi(x)\, dx\, dy \Big]; \quad \text{(III.2)}$$

here $d\mu_1$ is the Gaussian measure with covariance I. We shall denote by $\vdots \varphi_{\rho,\epsilon}^{2^k} \vdots (x)$ the field obtained by rescaling $: \varphi^2 : (x) \to \epsilon^{-d} : \varphi^2 : (x/\epsilon)$. We are going to compute the limit

$$Z_{0,\rho}[f,h] = \lim_{\epsilon \to 0} \Big\{ \int d\mu_{m_\epsilon}(\varphi) \exp\Big[- g_0^2 \int \vdots \varphi_{\rho,\epsilon}^{2r} \vdots h(x)\, dx \Big] \times \exp\Big[- \frac{1}{2} \int \epsilon^{-d} : \varphi^2 : (x/\epsilon) f(x)\, dx \Big] \Big\}^{\epsilon^d}, \quad \text{(III.3)}$$

where $r = 2^{k-1}$. Let us note that the interaction φ^{2^k} in Eq. (III.3) depends on ϵ. We will get rid of this ϵ in the definition of the interaction later (Sec. IV). Now the computation of the limit (III.3) can be reduced, by a successive use of the formula (III.2), to the limit of φ^2, which has been computed previously [Eq. (II.7)]. This can be easily seen if we investigate both limits (II.7) and (III.3) in terms of the truncated Green functions. After performing the ϵ limit we evaluate again the ψ integrals [Eq. (III.2)] and we get the result

$$Z_{0,\rho}[f,h] = \int d\mu_0(\Phi) \exp\Big[- g_0^2 \int \Phi_\rho^r\, h(x)\, dx \Big] \times \exp[- \tfrac{1}{2} \Phi(f)], \quad \text{(III.4)}$$

where $\Phi_\rho(x) = \int \rho_x(y) \Phi(y)\, dy$, and μ_0 is the measure corresponding to the independent value field defined by the acteristic function (II.7). Next the regularization ρ can be removed. Namely, Eq. (III.4) has a finite limit (Ref. 22) if we choose

$$\rho_x^r(y) \to (b/2) 2^{-r} \delta(x - y). \quad \text{(III.5)}$$

Here b is a dimensional constant. In the limit (III.5) we get

$$Z_0[f,h] = \exp\Big[\frac{1}{2(2\pi)^{d/2}} \int_0^\infty ds e^{-m_0^2 s} s^{-1-d/2} \times \int dx \exp(- s f(x) - \tfrac{1}{2} b g_0^2 s^r h(x)) - 1 + s f(x) \Big]. \quad \text{(III.6)}$$

For a more detailed derivation of the formula (III.6) in the case of φ^4 we refer to I. If the power series resulting from the expansion of the regularized $\exp[- g_0^2 \int (: \varphi^2 :)^r]$ is Borel summable (it is for $r = 2$,[23,24]) then the formula (III.6) can be proved for all integer r. In such a case we do not need the method of ψ integrals in order to prove that the ϵ limit results in the replacement $: \varphi_{\rho,\epsilon}^2 : \to \Phi_\rho$, where Φ is the independent value field (II.7). As a special case of Eq. (III.6) we get the characteristic function of φ^2 in φ^{2r}

$$S_0[f] = \exp\Big[\frac{1}{2(2\pi)^{d/2}} \int_0^\infty ds \exp(- m_0^2 s - \tfrac{1}{2} b g_0^2 s^r) \times s^{-1-d/2} \int dx\, (e^{isf(x)} - 1 - isf(x)) \Big] \times \exp\Big[- \frac{i}{2(2\pi)} d/2 \int_0^\infty ds e^{-m_0^2 s} s^{-d/2} \times (1 - e^{-1/2b\, g_0^2 s^r}) \int dx f(x) \Big], \quad \text{(III.7)}$$

and the characteristic function of φ^{2r}

$$S_0^{(r)}[h] = \exp\Big[\frac{1}{2(2\pi)^{d/2}} \int ds \exp(- m_0^2 s) s^{-1-d/2} \times \int dx\, (\exp(i/2b\, g_0^2 s^r h(x)) - 1) \Big]. \quad \text{(III.8)}$$

Equation (III.7) makes sense only in $d < 4$ dimensions, whereas (III.8) is finite if $d < 2r$. The characteristic function (III.8) has a limit when $m_0 \to 0$ which is the scale invariant stable distribution with $|S_0^r[h]| = \exp[- a \int |h(x)|^{d/2r} dx]$ (see Ref. 25 for the relation of stable distributions to long range order).

The meaning of the limit (III.3) will be further clarified in Sec. IV. In particular, we shall show that $\vdots \varphi_{\rho,\epsilon}^{2r} \vdots (x)$ in Eq. (III.3) may be replaced by $: \varphi^2 : (x) (\int : \varphi^2 : (y)\, dy)^{r-2}$, leading to the same final results [(III.6)–(III.9)].

IV. AN APPLICATION: LOW MOMENTUM LIMIT OF GINZBURG–LANDAU MODEL

In this section we will investigate the meaning of the limit (III.3) in the momentum space and relate it to the low momentum limit discussed in Sec. II. The Euclidean field theory has physical meaning as a model of statistical mechanics. Therefore, we shall discuss the low momentum limit in this physical context. A large class of phenomena in statistical mechanics can be described by the Ginzburg–Landau partition function[26,27]

$$Z = \int D\varphi(x) e^{-\beta \mathcal{F}[\varphi]}$$
$$= \int D\varphi(x) \exp\left(-\tfrac{1}{2} A \int [(\nabla \varphi(x))^2 + m_0^2 \varphi^2(x)] dx \right)$$
$$\times \exp\left(K \left[\frac{\sigma}{2} \int \varphi^2(x)\, dx - g_0^2 \int \varphi^4(x)\, dx - R(\varphi) \right] \right), \tag{IV.1}$$

where $\varphi(x)$ may be considered as an amplitude of fluctuations of the order parameter and integration is over all configurations. The square $\varphi^2(x)$ can be related to the density of these fluctuations. At small temperature there is no exchange of energy and the densities $\varphi^2(x)$ at different points are uncorrelated (cf. Refs. 11, 28). In momentum space this means that the correlation functions of $\tilde{\varphi}^2(q)$ do not depend on q (except of δ function ensuring the conservation of momentum). We can then approximate $\tilde{\varphi}^2(q)$ in Eq. (IV.1) by $\tilde{\varphi}^2(0) = \int dx \varphi^2(x)$ and further $\int \varphi^4(x)\, dx$ $= \int \tilde{\varphi}^2(-k) \tilde{\varphi}^2(k)\, dk$ by $c\varphi^2(0)^2 = c(\int \varphi^2(x) dx)^2$. In such a case we are able to calculate the partition function (IV.1) or rather the pressure (Z behaves as $e^{p|\Omega|}$ as $\Omega \to R^d$):

$$p = \lim_{\Omega \to R^d} \frac{1}{|\Omega|} \ln \left\{ \int d\mu_{m_0}(\varphi) \exp\left[\frac{\sigma}{2} \int_\Omega :\varphi^2:(x)\, dx \right.\right.$$
$$\left.\left. \times -c g_0^2 \left(\int_\Omega :\varphi^2:(x)\, dx \right)^2 + c g_0^2 \left\langle \left(\int :\varphi^2: \right)^2 \right\rangle \right] \right\}. \tag{IV.2}$$

Here we set $c = 8b(2\pi)^{d/2}$. We shall show that p coincides with

$$\lim_{\Omega \to R^d} \lim_{\epsilon \to 0} \frac{\epsilon^d}{|\Omega|} \ln \left\{ \int d\mu_{m_0}(\varphi) \right.$$
$$\times \exp\left[\frac{\sigma}{2} \epsilon^{-d} \int_\Omega :\varphi^2: \left(\frac{x}{\epsilon} \right) dx - g_0^2 \right.$$
$$\left.\left. \times \int_\Omega \vdots \varphi^4_{\rho,\epsilon} \vdots\, dx + g_0^2 \left\langle \int \varphi^4_{\rho,\epsilon} \right\rangle \right] \right\}. \tag{IV.3}$$

For this purpose we need an expansion of the exponentials in both formulas into power series. It is known that after integration over φ the series are divergent. However, it has been shown[23,24] that such series are Borel summable if $d \leqslant 2$. Therefore, it is sufficient to take the Borel transform of the series in order to get a convergent perturbation expansion. Let us consider a typical term in the expansion of (IV.3):

$$\frac{1}{m!n!} \left(-\frac{\sigma}{2} \right)^m (-g_0^2)^n \int_\Omega dx_1 \cdots dx_m\, dy_1\, dy_1' \cdots dy_n\, dy_n'\, dz_1$$
$$\cdots dz_n\, \rho_{z_1}(y_1) \rho_{z_1}(y_1') \cdots \rho_{z_n}(y_n) \rho_{z_n}(y_n')$$
$$\times \int d\mu_{m_0}(\varphi) \epsilon^{-d} :\varphi^2: \left(\frac{x_1}{\epsilon} \right) \cdots \epsilon^{-d} :\varphi^2: \left(\frac{x_m}{\epsilon} \right)$$
$$\times \epsilon^{-d} :\varphi^2: \left(\frac{y_1}{\epsilon} \right) \epsilon^{-d} :\varphi^2: \left(\frac{y_1'}{\epsilon} \right) \cdots \epsilon^{-d} :\varphi^2: \left(\frac{y_n}{\epsilon} \right)$$
$$\times \epsilon^{-d} :\varphi^2: \left(\frac{y_n'}{\epsilon} \right). \tag{IV.4}$$

Let $\rho_y(y') \to \rho(y - y')$ and denote the Fourier transform of $\rho(y)$ as $\tilde{\rho}(p)$. Then in momentum space (IV.4) equals

$$\int d\mu_{m_0}(\varphi) \int_{\Omega/\epsilon} dx_1 \ldots dx_m :\varphi^2:(x_1) \cdots :\varphi^2:(x_m)$$
$$\times \int d p_1\, d p_1' \cdots d p_n\, d p_n'\, dz_1 \cdots dz_n e^{-i(p_1 + p_1')z_1}$$
$$\cdots e^{-i(p_n + p_n')z_n} \tilde{\rho}(-p_1) \tilde{\rho}(-p_1') \cdots \tilde{\rho}(-p_n) \tilde{\rho}(-p_n')$$
$$\times :\tilde{\varphi}^2:(\epsilon p_1) :\tilde{\varphi}^2:(\epsilon p_1') \cdots :\tilde{\varphi}^2:(\epsilon p_n) :\tilde{\varphi}^2:(\epsilon p_n'). \tag{IV.5}$$

In the limit $\epsilon \to 0$: $\tilde{\varphi}^2:(\epsilon p) \to :\tilde{\varphi}^2:(0)$, and by assumption (III.5) $\int \tilde{\rho}(-p_1) \tilde{\rho}(p_1)\, dp_1 p_1 = \int \rho^2(x)\, dx = b/8$. Hence, (IV.5) is equal to

$$\left(\frac{b}{8} \right)^n (2\pi)^{n d/2} \int_{\Omega/\epsilon} dx_1 \cdots dx_m :\varphi^2:(x_1) \cdots :\varphi^2:(x_m)$$
$$\left(\int_{\Omega/\epsilon} dz_1 :\varphi^2:(z_1) \right)^2 \cdots \left(\int_{\Omega/\epsilon} dz_n :\varphi^2:(z_n) \right)^2,$$

which coincides with the expansion of Eq. (IV.2).

The expression (IV.3) is known from the calculations in Sec. III [Eqs. (III.3)–(III.6)]. We set $r = 2, f = -\sigma, h = 1$ in Eq. (III.6) and subtract the mean value $\langle :\varphi^2: \rangle$ $\sim \langle \Phi^2 \rangle$, finally getting for p [Eq. (IV.2)] the result

$$p = \frac{1}{2(2\pi)^{d/2}} \int_0^\infty ds e^{-m_0^2 s} s^{-1-d/2}$$
$$\times (e^{s\sigma - 1/2 b g_0^2 s^2} - 1 - s\sigma + \tfrac{1}{2} b g_0^2 s^2). \tag{IV.6}$$

It can be easily be seen that the argument above can be repeated for the Green functions G_r of $\int \epsilon^{-d} :\varphi^2:(x/\epsilon) f(x)$ $\times dx = \int :\tilde{\varphi}^2:(\epsilon q) \tilde{f}(-q)\, dq$. These terms inserted in Eq. (IV.5) give in the limit $\epsilon \to 0$ as a result $\delta(q_1 + \cdots + q_r)$ times Green functions at zero momentum, i.e., the scale limit (II.3). So we get for the generating functional

$$\lim_{\epsilon \to 0} Z^{-1} \left(\int d\mu_{m_0}(\varphi) \right.$$
$$\times \exp\left[\frac{\sigma}{2} \int :\varphi^2:(x)\, dx - c g_0^2 \left(\int :\varphi^2:(x)\, dx \right)^2 \right] \Big)$$
$$\left. \times \exp\left[\frac{i}{2} \int \epsilon^{-d} :\varphi^2:(x/\epsilon) f(x)\, dx \right] \right)^{\epsilon^d}$$
$$= \exp\left[\frac{1}{2(2\pi)^{d/2}} \int_0^\infty ds \exp(-m_0^2 s + \sigma s \right.$$
$$\left. - 1/2 b g_0^2 s^2) s^{-1-d/2} \int dx (e^{i s f(x)} - 1 - i s f(x)) \right]$$
$$\times \exp\left[-\frac{i}{2(2\pi)^{d/2}} \int_0^\infty ds e^{-m_0^2 s} s^{-d/2} \right.$$
$$\left. \times (1 - e^{\sigma s - 1/2 b g_0^2 s^2}) \int f(x)\, dx \right] = S_0[f], \tag{IV.7}$$

where $c = 8b(2\pi)^{d/2}$ because the l.h.s. of Eq. (IV.7) coincides (apart from the normalization) with the limit (III.3) (for $k = 2$) which is equal to (III.7) (with $r = 2$). Let us recall (Sec. II) that the limit on the l.h.s. of Eq. (IV.7) has the sense of a low momentum limit of the field theory with $(\int :\varphi^2:)^2$ interaction.

In Eq. (IV.7) the term $\exp(-m_0^2 s + \sigma s - 1/2 b g_0^2 s^2)$ has maximum at $s_0 = 2(\sigma - m_0^2)/b g_0^2$. If we approximate the integral (IV.7) by this saddle point we get for the generating functional of the correlation functions of the density $:\varphi^2:(x)$ the expression

$$S_0[f] = \exp\Big[A \int (e^{is_0 f(x)} - 1 - is_0 f(x))dx + iB \int f(x)\,dx\Big], \qquad \text{(IV.8)}$$

which coincides with the formula for density distribution of Bose gas at zero temperature derived in Ref. 11. This approximation is justified for higher order ($n > 2$) truncated functions when the contribution to the integral over s in Eq. (IV.7) from the point $s = 0$ is negligible. In general Eq. (IV.7) can be interpreted as a density distribution of Bosons with an internal degree of freedom (cf. Klauder's Fock space construction[29,11]).

V. CONCLUDING REMARKS

We have discussed a scale limit of Euclidean field theory which we consider as a special case of a more general formulation of the limit scale behavior of random fields. In this limit only random fields with values independent at every point can appear as a result. It seems that in more general formulation any infinitely divisible field could be a limit. If the limit of infinite physical mass (zero correlation length) exists at all as a random field, it should be a random field with values independent at every point. We have shown that our scale limit is at the same time a large mass and large coupling limit (for superrenormalizable interactions). So, if the limit of superrenormalizable theory for large (bare) coupling and fixed physical mass m exists (cf. Refs. 17, 6) it should coincide at $p = 0$ (i.e., when $m \to \infty$) with the limit considered in this paper. This suggests that the independent value field can be a good starting point for a construction of such singular models of field theory (cf. the expansion in part I). So far we were able to calculate the scale limit only in some models for the φ^2 field. This was due to infinite divisibility of φ^2 in free theory. The calculation of this limit has, then, some technical relevance for computations of certain expressions which are translation invariant and therefore depend only on Green functions at $p = 0$ e.g., the pressure in the infinite volume limit.

ACKNOWLEDGMENT

A financial support of Gesellschaft für Kernforschung m.b.H., Karlsruhe is gratefully acknowledged.

[1]B.V. Gnedenko and A.N. Kolmogorov, *Limit Distirbutions for Sums of Independent Random Variables* (Addison–Wesley, Reading, Massachussetts, 1968); E. Lukacs, *Characteristic Functions*, 2nd ed. (Hafner, New York, 1970); J.R. Klauder, in Boulder Lectures (Gordon and Breach, New York, 1973).
[2]I.A. Ibragimov and Yu.V. Linnik, *Independent and Stationary Sequences of Random Variables* (Wolters–Noordhoff, Gronigen, The Netherlands, 1971).
[3]R.L. Dobrushin, Avtomodelnost i Renorm-grupa Obobshchenikh Sluchaynikh Poley (to be published in Ann. Probability).
[4]J. Sinai, Theor. Prob. Appl. **21**, 64 (1976).
[5]G. Gallavotti and G. Jona-Lasinio, Commun. Math. Phys. **41**, 301 (1975).
[6]J. Glimm and A. Jaffe, Phys. Rev. D **11**, 2816 (1975).
[7]J. Glimm and A. Jaffe, in *Quantum Dynamics; Models and Mathematics*, edited by L. Streit (Springer, Berlin and New York, 1976).
[8]Z. Haba, Remarks on the stochastic process corresponding to interaction, preprint BI-TP 78/20, to appear in Rep. Math. Phys.
[9]Z. Haba; J. Math. Phys. **20**, 1896 (1979).
[10]I.M. Gel'fand and N.Y. Vilenkin, *Generalized Functions* (New York, 1964) Volume 4.
[11]G.A. Goldin, J. Grodnik, R.P. Powers, and D.H. Sharp, J. Math. Phys. **15**, 88 (1974).
[12]D. Ruelle; Nuovo Cimento **19**, 356 (1961).
[13]J.P. Eckmann, H. Epstein, and J. Fröhlich, Ann. Inst. H. Poincaré **25**, 1 (1976).
[14]G.C. Hegerfeldt, Commun. Math. Phys. **45**, 137 (1975).
[15]K. Baumann, G.C. Hegerfeldt, J. Math. Phys. **18**, 171 (1977).
[16]J.R. Klauder, "Soluble models and the meaning of nonrenormalizability," in *Proc. Int. Symp. on Math. Problems in Theor. Phys.*, edited by H. Araki (Springer, Berlin and New York, 1975); see also B. De Facio and C.L. Hammer, J. Math. Phys. **18**, 1216 (1977).
[17]G. Parisi, "On nonrenormalizable interactions," in *New Developments in QFT and Stat. Mechanics*, edited by M. Levy and P. Mitter (Plenum, New York, 1977).
[18]Ch.M. Newman, Commun. Math. Phys. **41**, 1 (1975).
[19]J. Glimm and A. Jaffe, Ann. Inst. H. Poincaré **22**, 1 (1975).
[20]J.A. Shohat and J.D. Tamarkin, "The problem of moments" (Am. Math. Soc., Providence, Rhode Island, 1943).
[21]S. Coleman and E. Weinberg, Phys. Rev. D **7**, 1888 (1973).
[22]G.C. Hegerfeldt and J.R. Klauder, Nuovo Cimento A **19**, 153 (1974).
[23]L. Rosen and B. Simon, Trans. Ann. Math. Soc. **165**, 365 (1972).
[24]J.P. Eckmann, J. Magnen, and R. Sénéor, Commun. Math. Phys. **39**, 251 (1975).
[25]G. Jona-Lasinio, Nuovo Cimento B **26**, 99 (1975).
[26]L.D. Landau and E.M. Lifschitz, *Statistical Physics* (Pergamon, New York, 1958).
[27]P. Pfeuty and G. Toulouse, *Introduction to the Renormalization Group and to Critical Phenomena* (Wiley, New York, 1977).
[28]H. Araki and E.J. Woods, J. Math. Phys. **4**, 637 (1963).
[29]J.R. Klauder, Acta Phys. Austr. **41**, 237 (1975).

jmp4 00423A001
03-MAR-80 20:01:34

Charge conservation as a concomitant of conformal motions coupled to generalized gauge transformations

Gerald H. Katzin
Department of Physics, North Carolina State University, Raleigh, North Carolina 27650
Jack Levine
Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27650

(Received 24 October 1979; accepted for publication 21 December 1979).

By using a Lagrangian for an electromagnetic field with sources (in which the four-current is assumed to be specifiable without consideration of any dependence upon a fundamental charge carrying field) we have established that in any conformally flat space–time (which includes Minkowski space–time) charge conservation may be obtained as a concomitant of infinitesimal Noether symmetry mappings that are not gauge transformations of the second kind. These Noether mappings, which couple geometrical and internal symmetries, are a combination of point mappings which define conformal motions in space–time with associated generalized gauge transformations of the four-vector potential.

I. INTRODUCTION

With respect to a given Riemannian[1] space–time with metric tensor $g_{ij}(x)$ referred to general coordinates x^i, charge conservation in the form[2]

$$J^i_{:i} = 0, \tag{1.1}$$

follows immediately from Maxwell's equations

$$F^{ij}_{:j} = (1/\epsilon_0)J^i, \tag{1.2}$$

where J^i is the current four-vector, ϵ_0 is the permittivity of vaccum, and F_{ij} is the electromagnetic field tensor which is expressible in terms of the four-vector potential ϕ_i by the relation

$$F_{ij} \equiv \phi_{j\cdot i} - \phi_{i\cdot j} \quad (F^{ij} = g^{ia}g^{jb}F_{ab}). \tag{1.3}$$

As a consequence any meaningful method which establishes charge conservation as the result of the existence of a symmetry mapping must be based upon a procedure which avoids use of Maxwell's equations. Distinguished by the manner in which charge is introduced into the theory two such procedures are well known:

(i) If $J^i = J^i(x)$ is assumed to be specifiable without consideration of its structural dependence on any fundamental charge carrying field, then to obtain charge conservation in the form (1.1) as a concomitant of a symmetry transformation it is sufficient to demand the invariance of the action integral, based upon the Lagrangian ($g \equiv |\det g_{ij}|$)

$$\mathscr{L} = -(\epsilon_0/4)g^{1/2}g^{ia}g^{jb}F_{ab}F_{ij} + g^{1/2}J^i\phi_i, \tag{1.4}$$

with respect to an infinitesimal gauge transformation of the second kind.[3] This type of internal symmetry is defined by a deformation in the four-vector potential in which

$$\phi_i(x) \to \bar{\phi}_i(x) = \phi_i(x) + \delta\phi_i(x), \quad \delta\phi_i(x) \equiv \psi(x)_{,i}\delta a, \tag{1.5}$$

where $\psi(x)$ is an essentially arbitrary scalar.

(ii) If J^i is assumed to be defined in terms of a charge carrying field, one may obtain charge conservation as a concomitant of a symmetry mapping by demanding the invariance of an action integral based upon a more complicated Lagrangian which couples the electromagnetic field to the charge carrying field.[4] For this more complicated case charge conservation may also be associated with an internal symmetry in which the four-vector potential is again altered by a gauge transformation of the second kind (1.5), but this deformation is now accompanied by an associated deformation in the charge carrying field (gauge transformation of the first kind)[4].

For each of the cases (i) and (ii) we propose to examine whether charge conservation may be considered a concomitant of Noether symmetry mappings in which the deformation of the four-vector potential is not a gauge transformation of the second kind. To increase the generality of the study we shall assume the space–time to be Riemannian,[1] and consider infinitesimal symmetry transformations which couple point mappings in space–time with " generalized gauge changes."

Although case (ii) is of more fundamental interest, in this paper we shall restrict our analysis to case (i). We examine this case first because the essentially arbitrary structure of $J^i(x)$ in this case does not impose any *a priori* restrictions on the allowed symmetry mappings. Hence we may regard the symmetries obtained in case (i) as fiducial with respect to the more limited symmetries which we anticipate obtaining in case (ii) where the current J^i has a complex structure dependent upon the field variables.

Hence in this paper we will consider the most general Noether symmetries [based upon Lagrangian (1.4)] in which the infinitesimal transformations are of the form

$$\bar{x}^i = x^i + \delta x^i, \quad \delta x^i \equiv \xi^i[\phi_j(x), x^j]\delta a, \tag{1.6}$$

$$\bar{\phi}_i = \phi_i + \delta\phi_i, \quad \delta\phi_i \equiv \eta_i[\phi_j(x), x^j]\delta a. \tag{1.7}$$

We shall show that in any conformally flat space–time (such space–times include Minkowski space–time) charge conservation [for case (i)] is a consequence of the existence of certain groups of conformal motions which occur as part of these general Noether symmetry mappings.[5]

 0022-2488/80/040902-07$1.00

In Sec. II we formulate the Noether symmetry conditions (2.5) based upon mappings of the type (1.6) and (1.7). For the electromagnetic Lagrangian (1.4) we find the Noether symmetry conditions require that the point mapping vector $\xi^i = \xi^i(x)$ define a conformal motion (2.26) in the Riemannian space–time, and that the symmetry deformation (2.28) of the four-vector potential ϕ_i denoted by the vector η_i consists of two parts: a "pure" gauge change denoted by $N_i(x)$ plus an additional part dependent upon the point mapping and the four-vector potential. The pure gauge change N_i is in algebraic relation to the current J^i (2.39). In addition the dynamics of N_i is determined by a Maxwell-like equation (2.38) which has as its "four-current" a function which is essentially the Lie deformation of the current four-vector J^i with respect to the conformal motion vector ξ^i.

In Sec. III we obtain from the Maxwell-like equation (2.38) mentioned above a relation (3.3) between the conformal motion vector ξ^i and $J^i_{;i}$ the divergence of the current four-vector. To proceed with the analysis we then assume the space–time to be conformally flat so that the conformal motion vectors ξ^i required by the Noether symmetry condition will exist. *It then follows that charge conservation is a consequence of the existence of either of two isomorphically related 5-parameter conformal motion groups which occur as part of the Noether symmetry requirement.* For the case in which the space–time is Minkowski one of these groups consists of four inversions and a scale change and the other of four translations and a scale change.

Based upon this same procedure it is also shown that for any Riemannian space–time charge conservation cannot be deduced as a concomitant *solely* of motion vectors which may occur as part of the Noether symmetry condition.

In Sec. IV it is shown for any conformally flat space–time that corresponding to each conformal motion vector $\xi^i_{(\alpha)}$ ($\alpha = 1,...,15$) a solution to the Noether symmetry condition exists in which the generalized gauge change vector $\eta_{(\alpha)i}$ is dependent upon the $\xi^i_{(\alpha)}$.

Section V is a brief conclusion which summarizes several of the interesting features of this new method for associating charge conservation with the existence of symmetry mappings.

II. NOETHER SYMMETRIES BASED ON THE LAGRANGIAN FOR THE ELECTROMAGNETIC FIELD WITH SOURCES

Based on mappings of the type (1.6) and (1.7) we define

$$\delta\phi_{b\cdot i} \equiv [d\bar{\phi}_b(\bar{x})/d\bar{x}^i] - [d\phi_b(x)/dx^i] = (\eta_{b\cdot i} - \phi_{b\cdot j}\xi^j_{\cdot i})\delta a. \tag{2.1}$$

For any Lagrangian $\mathscr{L}(\phi_{b\cdot i}, \phi_b, x^i)$ it can be shown that the following Noether identity holds[6]

$$\delta\mathscr{L} + \mathscr{L}\xi^i_{\cdot i}\delta a \equiv \frac{d}{dx^j}\left[\frac{\partial\mathscr{L}}{\partial\phi_{b\cdot j}}\eta_b - \left(\frac{\partial\mathscr{L}}{\partial\phi_{b\cdot j}}\phi_{b\cdot i} - \delta^j_i\mathscr{L}\right)\xi^i\right]\delta a - [(\eta_b - \phi_{b\cdot i}\xi^i)\delta a]\Lambda^b(\mathscr{L}), \tag{2.2}$$

where

$$\delta\mathscr{L} \equiv \frac{\partial\mathscr{L}}{\partial\phi_{b\cdot i}}\delta\phi_{b\cdot i} + \frac{\partial\mathscr{L}}{\partial\phi_b}\delta\phi_b + \frac{\partial\mathscr{L}}{\partial x^i}\delta x^i, \tag{2.3}$$

and where the Lagrange operator

$$\Lambda^a(\mathscr{L}) \equiv \frac{d}{dx^j}\frac{\partial\mathscr{L}}{\partial\phi_{a\cdot j}} - \frac{\partial\mathscr{L}}{\partial\phi_a}. \tag{2.4}$$

If for some functions $\tau^j(\phi_a, x^i)$ there exists ξ^i and η_j such that

$$\delta\mathscr{L} + \mathscr{L}\xi^i_{\cdot i}\delta a + \tau^i_{\cdot i}\delta a = 0, \tag{2.5}$$

then (1.6) and (1.7) define a *Noether symmetry mapping*. Note that the Noether symmetry condition (2.5) *does not assume that Lagrange's equation are satisfied*, that is $\Lambda^a(\mathscr{L})$ is not zero, and hence the $\phi_a(x)$ may be considered as arbitrary in obtaining solutions of (2.5).

If we assume (i) that the Noether symmetry condition (2.5) is satisfied and (ii) that the $\phi_i(x)$ satisfy Lagrange's equations $\Lambda^i(\mathscr{L}) = 0$, then it follows from the identity (2.2) that for each Noether symmetry mapping there will exist an associated conservation law[7]

$$\frac{d}{dx^j}\left[\frac{\partial\mathscr{L}}{\partial\phi_{a\cdot j}}\eta_a - \left(\frac{\partial\mathscr{L}}{\partial\phi_{a\cdot j}}\phi_{a\cdot i} - \delta^j_i\mathscr{L}\right)\xi^i + \tau^j\right] \doteq 0. \tag{2.6}$$

The two assumptions (i) and (ii) when used in the identity (2.2) will be referred to as the *standard Noether procedure* for obtaining a conservation law as a concomitant of a symmetry mapping. Since charge conservation in the form (1.1) follows directly from Lagrange's equation based on the Lagrangian (1.4) it would be inappropriate to use this standard Noether procedure [because of assumption (ii)] to attempt to show charge conservation to be a concomitant of a Noether symmetry. We shall show however that the existence of certain Noether symmetries by themselves (which are not gauge transformations of the second kind) imply charge conservation. We therefore proceed with the determination of the Noether symmetries.

If we expand (2.5) by use of (1.4), (1.6), (1.7), (2.1), and (2.3) the result may be expressed (after a lengthy calculation) as a cubic polynomial in the $\phi_{a\cdot i}$

$$\phi_{a\cdot b}\phi_{i\cdot j}\phi_{m\cdot c}B^{abijmc} + \phi_{a\cdot b}\phi_{i\cdot j}B^{abij} + \phi_{i\cdot j}B^{ij} + B = 0, \tag{2.7}$$

where

$$B^{abijmc} \equiv \epsilon_0 g^{1/2}[(g^{bj}g^{mi} - g^{bi}g^{mj})\xi^c_{,\phi_a} - \tfrac{1}{2}(g^{jb}g^{ia} - g^{ib}g^{ja})\xi^c_{,\phi_m}], \tag{2.8}$$

$$B^{abij} \equiv \epsilon_0 g^{1/2}[(g^{bi}g^{cj} - g^{bj}g^{ci})\eta_{c,\phi_a} + (g^{ej}g^{ai} - g^{ei}g^{aj})\xi^b_{,e} + \tfrac{1}{2}(g^{jb}g^{ja} - g^{jb}g^{ia})\xi^c_{,c}] + (\epsilon_0/2)[(g^{1/2}g^{ib}g^{ja})_{,c}\xi^c - (g^{1/2}g^{jb}g^{ia})_{,c}\xi^c], \tag{2.9}$$

$$B^{ij} \equiv \epsilon_0 g^{1/2}(g^{ci}g^{cj} - g^{cj}g^{ci})\eta_{c,e} + g^{1/2}\xi^j_{,\phi_i}J^k\phi_k + \tau^j_{,\phi_i}, \tag{2.10}$$

$$B \equiv g^{1/2}J^c\eta_c + (g^{1/2}J^k)_{,c}\xi^c\phi_k + g^{1/2}\xi^c_{,c}J^k\phi_k + \tau^j_{,j}. \tag{2.11}$$

By symmetrization (2.7) may be expressed in the form

$$\tfrac{1}{6}\phi_{a\cdot b}\phi_{i\cdot j}\phi_{m\cdot c}(B^{abijmc} + B^{abmcij} + B^{ijabmc} + B^{mcabij} + B^{ijmcab} + B^{mcijab}) + \tfrac{1}{2}\phi_{a\cdot b}\phi_{i\cdot j}(B^{abij} + B^{ijab}) + \phi_{i\cdot j}B^{ij} + B = 0. \tag{2.12}$$

We wish to determine conditions on the $\xi^i(\phi_j, x^k)$ and $\eta_i(\phi_j, x^k)$ (independent of $\phi_{i,j}$) inorder to satisfy (2.12), wherein the $\phi_i(x)$ are regarded as arbitrary. Hence to obtain these conditions on ξ^i and η_i we set the coefficients of the $\phi_{i,j}$ in (2.12) equal to zero, and regard the resulting differential equations in the unknowns ξ^i, η^i (and τ^i) as functions of the independent variables ϕ_i and x^i.

Carrying out this procedure we obtain from (2.12) the following four equations

$$\begin{aligned}&(g^{hj}g^{mi} - g^{hi}g^{mj})\xi^{c}_{,\phi_{a}} - (g^{jb}g^{ia} - g^{ib}g^{ja})\xi^{c}_{,\phi_{m}} \\ &\quad + (g^{bc}g^{im} - g^{bm}g^{ic})\xi^{j}_{,\phi_{a}} - (g^{cb}g^{ma} - g^{mb}g^{ca})\xi^{j}_{,\phi_{i}} \\ &\quad + (g^{jb}g^{ma} - g^{ja}g^{mb})\xi^{c}_{,\phi_{i}} + (g^{cb}g^{ia} - g^{ca}g^{ib})\xi^{j}_{,\phi_{m}} \\ &\quad + (g^{jc}g^{am} - g^{jm}g^{ac})\xi^{b}_{,\phi_{i}} - (g^{cj}g^{mi} - g^{mj}g^{ci})\xi^{b}_{,\phi_{a}} \\ &\quad + (g^{cj}g^{ai} - g^{ci}g^{aj})\xi^{b}_{,\phi_{m}} = 0, \end{aligned} \tag{2.13}$$

$$\begin{aligned}&-\eta_{c,\phi_{a}}g^{hj}g^{ci} - \eta_{c,\phi_{i}}g^{jb}g^{ca} + \eta_{c,\phi_{a}}g^{bi}g^{cj} + \eta_{c,\phi_{i}}g^{ja}g^{cb} \\ &\quad + \xi^{b}_{,c}g^{ej}g^{ai} + \xi^{j}_{,c}g^{eb}g^{ia} - \xi^{b}_{,e}g^{ei}g^{aj} - \xi^{j}_{,e}g^{ea}g^{ib} \\ &\quad - g^{jb}g^{ia}\xi^{c}_{,c} + g^{ib}g^{ja}\xi^{c}_{,c} + g^{-1/2}[(g^{1/2}g^{ib}g^{ja})_{,c}\xi^{c} \\ &\quad - (g^{1/2}g^{ia}g^{jb})_{,c}\xi^{c}] = 0, \end{aligned} \tag{2.14}$$

$$\epsilon_0 g^{1/2}(g^{ci}g^{cj} - g^{cj}g^{ci})\eta_{c,c} + g^{1/2}\xi^{j}_{,\phi_i}J^{k}\phi_k + \tau^{j}_{,\phi_i} = 0, \tag{2.15}$$

$$\begin{aligned}&g^{1/2}J^{c}\eta_{c} + (g^{1/2}J^{k})_{,c}\xi^{c}\phi_k + g^{1/2}\xi^{c}_{,c}J^{k}\phi_k + \tau^{j}_{,j} \\ &\quad = 0. \end{aligned} \tag{2.16}$$

If (2.13) is contracted with $g_{ia}g_{jb}$ we obtain

$$\xi^{i}_{,\phi_m} = 0, \tag{2.17}$$

which by inspection is seen to be necessary and sufficient to satisfy (2.13).

By use of the identities[8]

$$\pounds_\xi g^{ja} \equiv g^{ja}{}_{,c}\xi^{c} - g^{ca}\xi^{j}_{,c} - g^{jc}\xi^{a}_{,c}, \tag{2.18}$$

$$\xi^{c}_{;c} \equiv g^{-1/2}(g^{1/2}\xi^{c})_{,c} \equiv g^{-1/2}(g^{1/2})_{,c}\xi^{c} + \xi^{c}_{,c}, \tag{2.19}$$

Eq. (2.14) may be rewritten in the form

$$\begin{aligned}&[(\eta_{c,\phi_a} + \xi^{a}_{,c})g^{cj} + (\pounds_\xi g^{aj} + \tfrac{1}{2}\xi^{c}_{;c}g^{aj})]g^{ib} \\ &\quad - [(\eta_{c,\phi_a} + \xi^{a}_{,c})g^{ci} + (\pounds_\xi g^{ia} + \tfrac{1}{2}\xi^{c}_{;c}g^{ia})]g^{jb} \\ &\quad + [(\eta_{c,\phi_i} + \xi^{i}_{,c})g^{cb} + (\pounds_\xi g^{ib} + \tfrac{1}{2}\xi^{c}_{;c}g^{ib})]g^{ja} \\ &\quad - (\eta_{c,\phi_i} + \xi^{i}_{,c})g^{ca}g^{jb} - g^{ia}(\pounds_\xi g^{jb} + \tfrac{1}{2}\xi^{c}_{;c}g^{jb}) = 0. \end{aligned} \tag{2.20}$$

Contraction of (2.20) with g_{ab} gives

$$(\eta_{c,\phi_i} + \xi^{i}_{,c})g^{cj} - (\eta_{c,\phi_j} + \xi^{j}_{,c})g^{ci} = 0. \tag{2.21}$$

Contraction of (2.20) with g_{aj} yields

$$\begin{aligned}&2(\pounds_\xi g^{ij} + \tfrac{1}{2}\xi^{c}_{;c}g^{ij}) + (\eta_{c,\phi_i} + \xi^{c}_{,c})g^{ij} + 3(\eta_{c,\phi_i} + \xi^{i}_{,c})g^{jc} \\ &\quad - (\eta_{c}\phi_j + \xi^{j}_{,c})g^{ci} = 0. \end{aligned} \tag{2.22}$$

The difference (2.22) − (2.21) gives

$$2(\pounds_\xi g^{ij} + \tfrac{1}{2}\xi^{c}_{;c}g^{ij}) + (\eta_{c,\phi_i} + \xi^{c}_{,c})g^{ij} + 2(\eta_{c,\phi_i} + \xi^{i}_{,c})g^{jc} = 0, \tag{2.23}$$

which upon contraction with g_{ij} gives

$$\eta_{c,\phi_i} + \xi^{c}_{,c} = 0. \tag{2.24}$$

Use of (2.24) in (2.23) yields

$$\pounds_\xi g^{ij} + \tfrac{1}{2}\xi^{c}_{;c}g^{ij} + (\eta_{c,\phi_i} + \xi^{i}_{,c})g^{cj} = 0. \tag{2.25}$$

Substitution of (2.25) into (2.20) leads to

$$\pounds_\xi g_{ij} = 2\sigma g_{ij}, \quad \sigma \equiv \tfrac{1}{4}\xi^{i}_{;i}, \tag{2.26}$$

which implies that the vector ξ^i defines a conformal motion.

From (2.26) and (2.25) we obtain

$$\eta_{k,\phi_i} + \xi^{i}_{,k} = 0. \tag{2.27}$$

Hence (2.26) and (2.27) are necessary conditions that (2.20) be satisfied. It is easily seen they are also sufficient.

Integration of (2.27) gives

$$\eta_k = -\xi^{m}_{,k\phi_m} + N_k(x), \tag{2.28}$$

where the $N_k(x)$ are arbitrary functions of integration.

We now consider Eq. (2.15). By means of (2.17) and (2.28) Eq. (2.15) reduces to

$$f^{ji} = \frac{1}{\epsilon_0}t^{j}_{,\phi_i}, \tag{2.29}$$

in which

$$t^j \equiv g^{-1/2}\tau^j, \tag{2.30}$$

$$f^{ij} \equiv g^{ia}g^{jb}f_{ab}, \tag{2.31}$$

and

$$f_{ab} \equiv N_{b,a} - N_{a,b}. \tag{2.32}$$

Integration of (2.29) with use of (2.30) gives

$$\tau^j = \epsilon_0 g^{1/2}f^{jm}\phi_m + g^{1/2}G^j(x), \tag{2.33}$$

where $G^j(x)$ are functions of integration.

It is easily verified that (2.17), (2.28), and (2.33) are sufficient to satisfy (2.15).

We next consider (2.16) in which we eliminate η_c and τ^j by means of (2.28) and (2.33). If in the resulting equation we make use of (2.19), the definition of σ given by (2.26), and the relations

$$\pounds_\xi J^k \equiv J^{k}_{;c}\xi^{c} - J^{c}\xi^{k}_{;c} \equiv J^{k}_{,c}\xi^{c} - J^{c}\xi^{k}_{,c}, \tag{2.34}$$

$$g^{1/2}f^{jk}{}_{;j} \equiv (g^{1/2}f^{jk})_{,j}, \quad f^{jk} = -f^{kj}, \tag{2.35}$$

$$g^{1/2}G^{j}_{;j} \equiv (g^{1/2}G^{j})_{,j}, \tag{2.36}$$

we obtain

$$[\pounds_\xi J^k + 4\sigma J^k - \epsilon_0 f^{kj}{}_{;j}]\phi_k + [J^jN_j + G^{j}_{;j}] = 0. \tag{2.37}$$

It follows from (2.37) that

$$f^{kj}{}_{;j} = \frac{1}{\epsilon_0}[\pounds_\xi J^k + 4\sigma J^k], \tag{2.38}$$

$$J^jN_j + G^{j}_{;j} = 0. \tag{2.39}$$

Equations (2.28), (2.33), (2.38), and (2.39) [along with the definition of σ from (2.26)] are sufficient to satisfy (2.16).

We have now obtained solutions to the symmetry Eqs. (2.13)–(2.16). These solutions are collected together in the theorem to follow which summarizes the above work.

Theorem 2.1.

For the Lagrangian (1.4) of an electromagnetic field with sources [referred to a Riemannian space-time V_4 with metric $g_{ij}(x)$] an infinitesimal transformation of the type (1.6) and (1.7) will define a Noether symmetry mapping [in that (2.5) holds] if and only if ξ^i, η_i, τ^i are of the form

$$\xi^i = \xi^i(x), \tag{2.17'}$$

$$\eta_j = -\xi^{i}_{,j}\phi_i + N_j(x), \tag{2.28'}$$

$$\tau^j = g^{1/2}[\epsilon_0 f^{jm}\phi_m + G^j(x)], \tag{2.33'}$$

where ξ^i, N_j and G^j satisfy the conditions

$$\pounds_\xi g_{ij} = 2\sigma(x) g_{ij}, \quad \sigma = \tfrac{1}{4}\xi^i_{;i}, \tag{2.26'}$$

$$f^{mj}{}_{;j} = \frac{1}{\epsilon_0}(\pounds_\xi J^m + 4\sigma J^m),$$

$$(f^{mj} = g^{am} g^{jb} f_{ab}, \quad f_{ab} \equiv N_{b,a} - N_{a,b}), \tag{2.38'}$$

$$N_m J^m + G^m_{;m} = 0. \tag{2.39'}$$

Remark 2.1. We refer to the above set of equations as $\mathscr{F}$.[9]

Remark 2.2. It is of interest to note that if the definition of σ as given by (2.26) is used, then (2.38) may be expressed in the equivalent forms

$$f^{kj}{}_{;j} = \frac{1}{\epsilon_0}[(J^i\xi^j - J^j\xi^k)_{;j} + J^j_{;j}\xi^k], \tag{2.40}$$

$$(g^{1/2} f^{kj})_{;j} = \frac{1}{\epsilon_0}\pounds_\xi (g^{1/2} J^k). \tag{2.41}$$

From (1.7), (2.28), (2.31), and (2.32) the left-hand side of (2.41) describes the dynamics of the vector field $N_j(x)$, which may be interpreted as the "pure gauge" part of the symmetry deformation $\delta\phi_i$ of the four-vector potential $\phi_i(x)$. From the meaning of Lie differentiation we may interpret the right-hand side of (2.41) as the deformation in the four-current J^k associated with the point mapping (1.6) of the symmetry transformation (1.6) and (1.7). Hence in the Maxwell-like equation (2.41) we may consider the point deformation in the four-current to be the "source" of the "pure gauge deformation" in the four-vector potential.

Before proceeding to examine the possibilities of inferring charge conservation from the set of equations $\mathscr{F}$ in Theorem 2.1 we first examine these equations with regard to their consistency. By inspection the choice $\xi^i(x) = 0$, $G^i(x) = 0$, $N_i = N_{,i}$ where $N(x)$ satisfies $N_{,m}J^m = 0$ will define a solution of $\mathscr{F}$. This particular choice will be discussed in Sec. III.

If we assume $\xi^i \neq 0$, then it is evident from (2.26) that the space-time V_4 must admit a conformal motion defined by ξ^i. The problem of finding a solution to the set $\mathscr{F}$ (consistent with a given J^m) is essentially that of solving (2.26), (2.38), and (2.39) for ξ^i, N_i and G^i, because once these three functions are known η_j is determined by (2.28) and τ^j by (2.33). We shall consider this further in Sec. IV.

III. CHARGE CONSERVATION

We now examine the Noether symmetry mappings of Theorem 2.1 with regard to their relationship to the existence of charge conservation. The Noether symmetry condition (2.5) was based on the Lagrangian (1.4), not on Lagrange's equations, $\Lambda^a(\mathscr{L}) = 0$ [refer to (2.4)]; and in our determination of the Noether symmetry mappings no use was made of Lagrange's equations. Hence there is no *a priori* built-in assumption of charge conservation, i.e., $J^m_{;m}$ is not identically zero.

We first note that the choice $\xi^i(x) = 0$, $N_j = N(x)_{,j}$, $G^i(x) = 0$, reduces the set $\mathscr{F}$ of Noether symmetry equations of Theorem 2.1 to essentially

$$\eta_j = N_{,j}, \tag{3.1}$$

$$N_{,m}J^m = 0. \tag{3.2}$$

Equation (3.1) limits the Noether mappings to the familiar gauge transformation of the second kind. Equation (3.2) implies $\delta\mathscr{L} = N_{,m}J^m = 0$, which in turn implies $\delta\int\mathscr{L}d_4x = 0$. This latter condition is sufficient to obtain charge conservation.[3] Hence gauge transformations of the second kind which in addition satisfy (3.2) will be Noether mappings which lead to charge conservation.

We now consider the Noether symmetry mappings as described in Theorem 2.1 and investigate under what circumstances charge conservation is a concomitant of these general symmetry transformations which couple point mappings in a Riemannian space–time with "generalized gauge" deformations of the four-vector potential ϕ_i. We stress that in this investigation none of the restrictions of the preceding paragraph are assumed.

From the skew-symmetry of f^{kj} [refer to (2.31) and (2.32)] and of the expression $(J^i\xi^j - J^j\xi^k)$ it follows by covariant differentiation of (2.40) with x^k [and use of the definition of σ as given by (2.26)] that

$$\mu_{,i}\xi^i + 4\sigma\mu = 0, \tag{3.3}$$

where

$$\mu \equiv J^m_{;m}. \tag{3.4}$$

We regard (3.3) as a differential equation for μ with the coefficients ξ^i and σ determined by (2.26).

Equation (2.26) implies that if the Noether symmetries are to contain nontrivial point mappings, that is $\xi^i \neq 0$, then the Riemannian space–time under consideration must admit an r-parameter group of conformal motions defined by vectors $\xi^i_{(\alpha)}$, $\alpha = 1,\ldots, r$. In general these conformal motions will consist of proper conformal motions corresponding to σ's $\neq 0$ and improper conformal motions corresponding to σ's $= 0$. Those which are improper as is well-known are called motions.

We first examine (3.3) for those ξ^i which correspond to motion vectors. For each such motion vector [indicated by $\xi^i_{(\alpha')}$] (3.3) reduces to

$$\mu_{,i}\xi^i_{(\alpha')} = 0. \tag{3.5}$$

From (3.5) it cannot be shown that μ must be zero, since these equations are always satisfied when μ is an arbitrary constant. *Hence we conclude that in order to infer charge conservation by use of* (3.3) *the above-mentioned group of conformal motions must contain at least one proper conformal motion.*

Since all conformally flat space–times C_4 are known to admit a 15-parameter group G_{15} of conformal motions *we shall now restrict the V_4 to be conformally flat.* Conformally flat space–times include space–times of constant curvature which include Minkowski space–time M_4.

We shall now show that with respect to any conformally flat space–time that charge conservation (in the form $\mu = 0$) follows from (3.3) as a consequence of the existence of certain proper conformal motions.

Let

$$\bar{g}_{ij}(x) = e_i\delta_{ij}, \quad e_i = \pm 1, \tag{3.6}$$

define a Minkowski space–time M_4 and

$$g_{ij}(x) = e^{-2\theta(x)}\,\bar{g}_{ij}(x) = e^{-2\theta(x)}e_i\delta_{ij} \tag{3.7}$$

define a conformally flat space–time C_4. It is known that C_4 and M_4 admit the same 15-parameter conformal motion group defined by[10,11]

$$\xi^i = \bar{\xi}^i = a_j x^j x^i - \tfrac{1}{2}e_i a_i e_j x^j x^j + a_0 x^i + e_i\omega^i_j x^j + c^i, \quad \omega^i_j = -\omega^j_i, \tag{3.8}$$

where in M_4

$$\bar{\sigma}(x) = a_j x^j + a_0, \tag{3.9}$$

and in C_4

$$\sigma(x) = \bar{\sigma} - \bar{\xi}^m\theta_{,m}. \tag{3.10}$$

By use of (3.8), (3.9), and (3.10) the generators $X_\alpha \equiv \bar{\xi}^i_{(\alpha)}\partial_i = \xi^i_{(\alpha)}\partial_i$ of this 15-parameter conformal motion group along with their associated vectors $\bar{\xi}^i_{(\alpha)} = \xi^i_{(\alpha)}$ and conformal motion scalars $\bar{\sigma}_{(\alpha)}$, $\sigma_{(\alpha)}$ are expressible as follows. (In the remainder of this section $\alpha, \beta = 1,\ldots, 4.$)

$$P_\alpha = \partial_\alpha, \quad \bar{\xi}^i_{(\alpha)} = \delta^i_\alpha \quad \bar{\sigma}(P_\alpha) = 0, \quad \sigma(P_\alpha) = -\theta_{,\alpha}, \tag{3.11}$$

$$S_{\alpha\beta} = e_\alpha x^\alpha\partial_\beta - e_\beta x^\beta\partial_\alpha, \quad \bar{\xi}^i_{(\alpha\beta)} = e_\alpha\delta^i_\beta x^\alpha - e_\beta\delta^i_\alpha x^\beta,$$
$$\bar{\sigma}(S_{\alpha\beta}) = 0, \tag{3.12}$$
$$\sigma(S_{\alpha\beta}) = -e_\alpha x^\alpha\theta_{,\beta} + e_\beta x^\beta\theta_{,\alpha},$$

$$U = x^i\partial_i, \quad \bar{\xi}^i_{(0)} = x^i, \quad \bar{\sigma}(U) = 1, \quad \sigma(U) = 1 - x^m\theta_{,m}, \tag{3.13}$$

$$V_\alpha = x^\alpha x^j\partial_j - \tfrac{1}{2}\rho e_\alpha\partial_\alpha, \quad \bar{\xi}^i_{(\alpha)} = x^i x^\alpha - \tfrac{1}{2}\rho e_\alpha\delta^i_\alpha,$$
$$\bar{\sigma}(V_\alpha) = x^\alpha, \tag{3.14}$$
$$\sigma(V_\alpha) = x^\alpha - (x^m x^\alpha - \tfrac{1}{2}\rho e_\alpha\delta^m_\alpha)\theta_{,m}, \quad \rho \equiv \sum e_j x^j x^j.$$

For the C_4 space–time we now proceed to show that $\mu \equiv J^m_{;m} = 0$ is a consequence of the requirement that (3.3) hold for each of the five symmetry vectors which define the 5-parameter conformal motion subgroup (V_1, V_2, V_3, V_4, U). By means of (3.13) and (3.14) we obtain from (3.3)

$$\mu_{,m}x^m + 4(1 - x^m\theta_{,m}) = 0, \quad (U), \tag{3.15}$$

$$\mu_{,m}(x^m x^\alpha - \tfrac{1}{2}\rho e_\alpha\delta^m_\alpha) + 4\,[x^\alpha - (x^m x^\alpha - \tfrac{1}{2}\rho e_\alpha\delta^m_\alpha)\theta_{,m}]\mu = 0, \quad (V_\alpha). \tag{3.16}$$

By use of (3.15) we eliminate $\mu_{,m}x^m$ from (3.16) to obtain

$$\mu_{,\alpha} = 4\mu\theta_{,\alpha}. \tag{3.17}$$

Use of (3.17) in (3.15) implies $\mu = 0$.

Similarly, if we consider the five symmetry vectors which define the 5-parameter proper conformal motion subgroup (P_1, P_2, P_3, P_4, U), [refer to (3.11) and (3.13)] we obtain from (3.3) the set the equations consisting of

$$\mu_{,\alpha} - 4\mu\theta_{,\alpha} = 0, \quad (P_\alpha), \tag{3.18}$$

and Eq. (3.15). From (3.18) and (3.15) it also follows that $\mu = 0$.

We summarize the results of this section in the theorems and corollary to follow.

Theorem 3.1.

If the Noether symmetry conditions of Theorem 2.1 are satisfied then the four-current J^m and the conformal motion vector ξ^m must satisfy the condition

$$J^m_{;mi}\xi^i + 4\sigma J^m_{;m} = 0. \tag{3.3'}$$

Corollary 3.1.

Charge conservation in the form $J^m_{;m} = 0$ cannot be deduced from (3.3) based only on those symmetry vectors $\xi^i_{(\alpha')}$ which define motions $[\sigma_{(\alpha')} = 0]$.

Theorem 3.2.

In a conformally flat space–time with metric defined by (3.6) and (3.7) a necessary condition that the 5-parameter proper conformal motion group (V_1, V_2, V_3, V_4, U) defined by (3.13) and (3.14) (a subgroup of the 15-parameter conformal motion group) satisfy the Noether symmetry conditions given in Theorem 2.1 is that charge be conserved, i.e., $J^m_{;m} = 0$. A similar statement can be made with respect to the subgroup (P_1, P_2, P_3, P_4, U) defined by (3.11), (3.13).

Remark 3.1. The subgroups (P_α, U) and (V_α, U) referred to in Theorem 3.2 are isomorphic in that the commutators satisfy the conditions

$$[P_\alpha, P_\beta] = 0, \quad [P_\alpha, U] = P_\alpha, \tag{3.19}$$
$$[V_\alpha, V_\beta] = 0, \quad [V_\alpha, U] = V_\alpha. \tag{3.20}$$

It is easy to show by means of the coordinate transformation

$$y^i = -\frac{e_i x^i}{\rho(x)}, \quad \rho(x) \equiv \sum e_k x^k x^k, \tag{3.21}$$

with inverse

$$x^i = -[e_i y^i/\rho(y)], \tag{3.22}$$

that the generators P_α, V_α, U transform as follows

$$\begin{aligned} &P_\alpha(x) \to V_\alpha(y), \\ &V_\alpha(x) \to P_\alpha(y), \\ &U(x) \to -U(y). \end{aligned} \tag{3.23}$$

Hence by the coordinate transformation (3.21) plus a simple change in basis we have that the subgroups are related by

$$\begin{aligned} &(P_\alpha(x), U(x)) \to (V_\alpha(y), U^*(y)), \\ &(V_\alpha(x), U(x)) \to (P_\alpha(y), U^*(y)), \end{aligned} \tag{3.24}$$

where $U^* = -U$.

Remark 3.2. It can be shown that if *any* Riemannian space–time (not necessarily conformally flat) admits the 5-parameter group (P_α, U) of conformal motions defined by

$$\xi^i_{(\alpha)} = \delta^i_\alpha, \quad (P_\alpha), \tag{3.25}$$

$$\xi^i_{(0)} = x^i, \quad (U), \tag{3.26}$$

then, Eq. (3.3) (refer to Theorem 3.1) implies $J^m_{;m} = 0$. If however we substitute (3.25) into the conformal motion condition (2.26) we obtain as a condition on the metric of the space–time

$$g_{ij,\alpha} = (1/4g)g_{,\alpha}g_{ij}, \tag{3.27}$$

from which it follows that

$$g_{ij} = g^{1/4}g^0_{ij}, \quad g^0_{ij} = \text{consts}, \quad |\det g^0_{ij}| = 1. \tag{3.28}$$

Hence g^0_{ij} may be taken to be the metric of a flat space–time, and it therefore follows from (3.28) that the Riemannian space–time is conformally flat.

A similar result follows if any V_4 space–time admits the group (V_α, U) of conformal motions [refer to (3.13) and

(3.14)] in that (3.3) implies $J^m_{;m} = 0$, and the $\xi^i_{(\alpha)}$ associated with the generators V_α imply by use of (2.26) and the coordinate transformation (3.21) that the space–time is conformally flat.

Remark 3.3. It should be noted that if in (3.7) $\theta = \text{const}$ and hence the C_4 reduces to a flat space–time referred to Minkowski coordinates, then the above-stated Theorems 3.1 and 3.2 and Corollary 3.1 still hold. In this Minkowski space–time the generators of the 15-parameter conformal motion group (P_i, S_{ij}, U, V_i) defined respectivley by (3.11)–(3.14) are recognized to be the familiar translations, rotations, dilations, and inversions, respectively.

IV. EXISTENCE OF NOETHER SYMMETRIES IN C_4

In this section we continue with the assumption that the space–time is conformally flat, i.e., the metric is of the type (3.7). We shall show the set of equations (2.38) and (2.39) possesses a solution $N^i_{(\alpha)}$ and $G^i_{(\alpha)}$ for each of the $\xi^i_{(\alpha)}$ which define the two 5-parameter subgroups of conformal motions referred to in theorem 3.2. In doing so we make use of necessary condition that $J^m_{;m} = 0$.

With $J^m_{;m} = 0$, it follows that (2.40) may be written in the form[12]

$$(g^{1/2} f^{jm}_{(\alpha)})_{,j} = (1/\epsilon_0)[g^{1/2}(J^m \xi^j_{(\alpha)} - J^j \xi^m_{(\alpha)})]_{,j}. \tag{4.1}$$

By use of (2.31) and (3.7) it can be shown that (4.1) can be written as

$$\bar{g}^{ma}\bar{g}^{jb} f_{ab(\alpha),j} = (1/\epsilon_0)[e^{-4\theta}(J^m \xi^j_{(\alpha)} - J^j \xi^m_{(\alpha)})]_{,j}. \tag{4.2}$$

By means of (2.32) and (3.6) Eq. (4.2) is expressible as

$$e_m \bar{N}^j_{(\alpha),jm} - \Box^2 \bar{N}^m_{(\alpha)} = S^m_{(\alpha)}, \tag{4.3}$$

where

$$\bar{N}^j_{(\alpha)} \equiv \bar{g}^{jk} N_{(\alpha)k}, \tag{4.4}$$

$$S^m_{(\alpha)} \equiv (1/\epsilon_0)[e^{-4\theta}(J^m \xi^j_{(\alpha)} - J^j \xi^m_{(\alpha)})]_{,j}. \tag{4.5}$$

$$\Box^2 \bar{N}^m_{(\alpha)} \equiv \bar{g}^{jb} \bar{N}^m_{(\alpha),jb} = e_j \delta^{jb} \bar{N}^m_{(\alpha),jb}. \tag{4.6}$$

Note from the skew-symmetry of the term in brackets in (4.5) that

$$S^m_{(\alpha),m} = 0. \tag{4.7}$$

We assume the four current J^m has been established in the finite past and is contained in some spatially bounded region and vanishes outside the region. We also assume that $\theta(x)$ and its first derivatives are bounded functions of the coordinates. It then follows from (4.5) that the functions $S^m_{(\alpha)}(x)$ vanish outside the above-mentioned bounded region. For such $S^m_{(\alpha)}$, solutions to (4.3) will exist for $\bar{N}^j_{(\alpha)}$.[13] The functions $N_{(\alpha)k}$ can then be obtained by use of (4.4) and (3.6) and then expressed in the form

$$N_{(\alpha)j} = e_j \bar{N}^j_{(\alpha)}. \tag{4.8}$$

Corresponding to each $N_{(\alpha)j}$ of (4.8) there remains to be shown that (2.39) has a solution for $G^m_{(\alpha)}$. For the C_4 with metric (3.7), Eq. (2.39) takes the form

$$H^m_{(\alpha),m} = e^{-4\theta} N_{(\alpha)m} J^m, \tag{4.9}$$

where

$$H^m_{(\alpha)} \equiv e^{-4\theta} G^m_{(\alpha)}. \tag{4.10}$$

In general a solution to (4.9) will exist, and hence solutions $G^m_{(\alpha)}$ will exist for each $N_{(\alpha)m}$.[14]

We have thus established with respect to a conformally flat space–time that for each of the conformal motion vectors $\xi^i_{(\alpha)}$ referred to in Theorem 3.2 there exist associated functions $N_{(\alpha)i}$ and $G^i_{(\alpha)}$ which satisfy (2.38) and (2.39). Hence for each of these conformal motion vectors $\xi^i_{(\alpha)}$ there will exist an $\eta_{(\alpha)j}$ and $\tau^j_{(\alpha)}$ given by (2.28) and (2.33). This set $\{\xi^i_{(\alpha)}, \eta_{(\alpha)i}, \tau^i_{(\alpha)}\}$ defines a Noether symmetry mapping in the C_4 (see the last paragraph of Sec. II).[15]

V. CONCLUSION

Employing the Lagrangian (1.4) for an electromagnetic field with sources [in which the four-current $J^i(x)$ is assumed to be specifiable without consideration of any dependence upon a fundamental charge carrying field] we have established that in any conformally flat space–time (which includes Minskowski space–time) charge conservation in the form $J^i_{;i} = 0$ may be obtained as a concomitant of infinitesimal Noether symmetry mappings that are not gauge transformations of the second kind. These Noether mappings which lead to charge conservation are a combination of point mappings which define conformal motions in space–time and associated generalized gauge transformations of the four-vector potential. The existence of charge conservation as a concomitant of such Noether mappings has several interesting features.

(i) The *single* law of charge conservation is a consequence of the concerted action of *five* Noether symmetry transformations.

(ii) Either one of two isomorphically related 5-parameter groups of conformal motions may be used to define the point mapping part of these Noether symmetry transformations.

(iii) The conformal motion part of these Noether mappings associates geometrical symmetries with charge conservation. In Minkowski space–time such symmetries take the form of a scale change combined with either four translations or four inversions.

ACKNOWLEDGMENTS

The authors wish to thank W.R. Davis, P. Sommers, D.R. Oliver, and L.K. Norris for their helpful discussions.

[1] The metric is assumed to be given in that we neglect any coupling between this Riemannian space–time and the electromagnetic field as would be determined by the field equations of general relativity (Einstein–Maxwell equations).

[2] Unless otherwise indicated the Einstein summation notation is employed with lower case italic indices having the range 1–4. Since the functions $\phi_i(x)$ are arbitrary, at times we shall regard them as independent variables and at other times as functions of the coordinates x^i. To indicate coordinate differentiation in which ϕ_i is regarded as a function of the cordinates we use the notation $\phi_{i,j} \equiv d\phi_i/dx^j$, $\phi_{i,jk} = d(d\phi_i/dx^j)/dx^k$. With respect to a function $F[x^i, \phi_i(x), \phi_{i,j}(x)]$ in which x^i, ϕ_i, and $\phi_{i,j}$ are regarded as independent variables we use the notation $F_{,i} \equiv \partial F/\partial x^i$ to mean the change in F with respect to x^i with ϕ and $\phi_{i,k}$ held constant. Consistent with this nota-

tion we have $\phi_{i,j} \equiv 0$, $(\phi_{i,j})_{,k} \equiv 0$, etc. On the other hand the total change in F with respect to x^i is indicated by $F_{:i} \equiv dF/dx^i = \partial F/\partial x^i + (\partial F/\partial \phi_j)\phi_{j,i} + (\partial F/\partial \phi_{j,k})\phi_{j,ki}$. When we regard ϕ_i as a function of the coordinates x^i, we indicate covariant differentiation by a colon (:). For example, $F^i_{:j}[x^k, \phi_l(x)] \equiv dF^i/dx^j + F^k \left\{ {i \atop j\ k} \right\}$. In the case F^i is an explicit function of the coordinates only, then covariant differentiation may be indicated in an unambiguous manner by the usual semicolon (;) notation $F^i_{;j}$. The notation $F_{,\phi_i} = \partial F/\partial \phi_i$, etc.

[3]See, for example, F. Rohrlich, *Classical Charged Particles* (Addison–Wesley, Reading, 1965, p. 102. Although Rohrlich's derivation assumes a Minkowski space–time a similar procedure may be used in a Riemannian space–time.

[4]See, for example, S.S. Schweber, H.A. Bethe, and F. de Hoffmann, *Mesons and Fields, Volume* 1 *Fields* (Row, Peterson, and Company, Evanston, IL. 1954).

[5]Our preliminary calculations based on the Lagrangian for a charge carrying field (complex scalar field) in a *Riemannian space–time* indicate that it may be possible to obtain Noether symmetries which couple point mappings that define *conformal motions* in a Riemannian space–time to generalized gauge transformations of the first kind even though the complex scalar field is *not* massless.

[6]This identity holds for arbitrary $\phi_j(x)$ as may be verified by direct expansion. No assumption has been made that $\phi_i(x)$ satisfies Lagrange's equation $\Lambda^a(\mathscr{L}) = 0$, where $\Lambda^a(\mathscr{L})$ is defined by (2.4). This identity may be derived by a simple procedure which does not involve variation of the action integral. The derivation of this identity for this field theoretical case is similar to that for a classical particle dynamical system. See G.H. Katzin and J. Levine, J. Math. Phys. **15**, 1460 (1974).

[7]The notation ($\overset{\circ}{=}$) indicates that $\phi_i(x)$ is a solution of Lagrange's equations.

[8]Lie differentiation with respect to the vector ξ^i is denoted by $\pounds_\xi$.

[9]By means of the standard Noether procedure discussed at the beginning of Sec. II the Noether conservation laws (2.6) may be formulated and verified by use of (1.1), the set $\mathscr{F}$ of Noether symmetries, and Lagrange's equations.

[10]See G.H. Katzin and J. Levine, Tensor N.S. **17**, 249 (1966); **21**, 319 (1970).

[11]Barred quantities refer to the Minkowski space–time M_4.

[12]L.P. Eisenhart, *Riemannian Geometry* (Princeton U. P. Princeton, New Jersey, 1960), p. 32.

[13]J.A. Stratton, *Electromagnetic Theory* (McGraw–Hill, New York, 1941), Chap. VIII.

[14]The fact that such solutions exist was pointed out to the authors by Professor S. Schecter and Professor R.O. Fulp.

[15]If the condition $J^m_{;m} = 0$ which is required for the existence of either of the two 5-parameter conformal motion groups referred to in Theorem 3.2 is assumed to be satisfied, then the necesary condition (3.3) is satisfied identically for any of the 15 symmetry vectors $\xi^i_{(\alpha)}$ of the complete group of conformal motions. It then follows that the above analysis also shows the existence of a Noether symmetry mapping $\{\xi^i_{(\alpha)}, \eta_{(\alpha)i}, \tau_{(\alpha)}\}^*$ for each of the 15 vectors $\xi^i_{(\alpha)}$ which define the *complete* group of conformal motions.

Invariance properties of Yang equations and applications

Y. Brihaye and J. Nuyts
Université de l'Etat, Faculté des Sciences, Mons, Belgium

(Received 27 September 1979; accepted for publication 8 November 1979)

Exploiting the different invariances of the Yang equations we construct an infinite sequence of solutions. These depend on a series of polynomials given by recurrence relations

I. INTRODUCTION

Since the discovery of the instanton solution[1] the interest for self-dual Yang–Mills gauge fields defined by

$$F_{\mu\nu} = \tfrac{1}{2}\epsilon_{\mu\nu\rho\sigma}F^{\rho\sigma} \equiv \tilde{F}_{\mu\nu}, \tag{1.1}$$

has increased more and more.

Many authors have written the gauge potentials with different parametrizations and have obtained different solutions either explicitly (cf. Refs. 1–5) or implicitly (cf. Refs. 6 and 7).

The purpose of this note is to explicitly exhibit an infinite series of solutions to the SU (2) and SU (1,1) self-dual equations using the mathematical tool described in Ref. 6 and to study some of their properties. Let us first recall the basic content of Refs. 4 and 6.

Following Yang[4] it is convenient to work in a complexified version of space–time described by four coordinates Y, $\bar{Y}$, Z, $\bar{Z}$,

$$\sqrt{2}\begin{pmatrix} y & -z \\ \bar{z} & \bar{y} \end{pmatrix} = \begin{pmatrix} x^0 - ix^3 & -(x^2 + ix') \\ x^2 - ix^1 & x^0 + ix^3 \end{pmatrix}. \tag{1.2}$$

The gauge group is taken to be SU (2) or SU (1,1). The metric M is respectively the unit or the τ_3 Pauli matrix. Hence $m = \det M$ is one for SU (2) and -1 for SU (1,1). In terms of the new coordinates, (1.1) is written

$$F_{Y\bar{Y}} + F_{Z\bar{Z}} = 0, \tag{1.3a}$$
$$F_{\bar{Y}\bar{Z}} = 0,\ F_{YZ} = 0. \tag{1.3b}$$

From (1.3b) one can deduce that in the two anti-self-dual planes $(Y - Z)$ and $(\bar{Y} - \bar{Z})$ the gauge potentials are restricted to pure gauges, i.e.,

$$A_Y = -iD^{-1}D_{,Y},\ A_{\bar{Y}} = -iE^{-1}E_{,\bar{Y}},$$
$$A_Z = -iD^{-1}D_{,Z},\ A_{\bar{Z}} = -iE^{-1}E_{,\bar{Z}}, \tag{1.4}$$

where $D_{,Y}$ is the Y derivative of D.

The matrices D and E can be chosen to be triangular up to a gauge transformation, i.e.,

$$D = \frac{1}{\sqrt{f}}\begin{pmatrix} 1 & 0 \\ e & f \end{pmatrix},\ E^{-1} = \frac{1}{\sqrt{f}}\begin{pmatrix} 1 & g \\ 0 & f \end{pmatrix}. \tag{1.5}$$

The hermiticity conditions on the potentials (1.4) require that

$$e^* = -mg,\quad f = f^*. \tag{1.6}$$

The triple of functions (f, e, g) constitutes the basic object of the theory. It must obey the remaining conditions (1.3a) known as the Yang equations

$$\left(\frac{f_{,Y}}{f}\right)_{,\bar{Y}} + \left(\frac{f_{,Z}}{f}\right)_{,\bar{Z}} - \frac{e_{,Y}g_{,\bar{Y}}}{f^2} - \frac{e_{,Z}g_{,\bar{Z}}}{f^2} = 0, \tag{1.7}$$

$$f\left[\left(\frac{g_{,\bar{Y}}}{f^2}\right)_{,Y} + \left(\frac{g_{,\bar{Z}}}{f^2}\right)_{,Z}\right] = f\left[\left(\frac{e_{,Y}}{f^2}\right)_{,\bar{Y}} + \left(\frac{e_{,Z}}{f^2}\right)_{,\bar{Z}}\right] = 0. \tag{1.8}$$

II. MISCELLANEOUS TRANSFORMATIONS

In this section we recall the meaning of Bäcklund transformations,[6] of special algebraic transformations, of scale transformations and translations.[7] The last three transformations are particular cases of the most general D transformations

$$D' = VDU, \tag{2.1}$$

where V belongs to SL $(2,C)$ and depends only on $\bar{Y}$ and $\bar{Z}$ while U belongs to SU (2) or SU (1,1). The E transformation follows in such a way that the hermiticity relations (1.6) are still satisfied. Given any triplet $S = (f, e, g)$ the special algebraic transformation γ is defined by

$$\gamma\cdot S = \left(\frac{f}{\Delta}, \frac{-g}{\Delta}, \frac{-e}{\Delta}\right),\ \Delta = f^2 - eg, \tag{2.2}$$

while the scale transformation δ and the translation θ are defined by

$$\delta\cdot S = (kk^*f, k^{*2}e, k^2 g), \tag{2.3}$$
$$\theta\cdot S = (f, e + k^*, g - mk), \tag{2.4}$$

where k is a function of Y and Z alone.

Finally the Bäcklund transformation $\tilde{S} = \beta\cdot S$ is given by a set of differential equations

$$\tilde{f} = \frac{1}{f},\ \frac{\tilde{e}_{,Z}}{\tilde{f}} = -\frac{g_{,\bar{Y}}}{f},\ \frac{\tilde{g}_{,\bar{Z}}}{\tilde{f}} = \frac{e_{,Y}}{f},$$
$$\frac{\tilde{e}_{,Y}}{\tilde{f}} = \frac{g_{,\bar{Z}}}{f},\ \frac{\tilde{g}_{,\bar{Y}}}{\tilde{f}} = -\frac{e_{,Z}}{f}. \tag{2.5}$$

This defines the Bäcklund transformations up to integration constants which are arbitrary translations.

Under the four transformations above solutions of the Yang equations are transformed into solutions.

Let us note that γ^2 is the identity, while, β^2 is an arbitrary translation and that β and δ commute.

Refering to (1.6) it is easy to show that γ, δ, and θ respect the hermiticity of the solution while the Bäcklund transformation maps SU (2) soultions in SU (1,1) solutions and vice versa.

 0022-2488/80/040909-04$1.00

In what follows we shall use the composite transformation

$$\alpha = \beta\gamma, \tag{2.6}$$

to generate explicitely an infinite sequence of solutions S_n from a given one. This is the main result of our paper.

Note also that the Yang equations are invariant under the trivial sign change $f \to -f$ and independently $(e, g) \to (-e, -g)$.

III. TOPOLOGICAL NUMBER AND ACTION

Thanks to Eq. (1.1), the topological number T and the action A are equal up to a normalization factor. They are written below in terms of an integral over space–time (described by the usual and Yang variables),

$$T = \frac{1}{16\pi^2}\int d^4x\ \mathrm{tr}(\tilde{F}_{\mu\nu}F^{\mu\nu}) = \frac{1}{4\pi^2}\int Q(Y, \bar{Y}, Z, \bar{Z})\, dY d\bar{Y} dZ d\bar{Z}, \tag{3.1}$$

$$A = \int d^4x\ \mathrm{tr}(F_{\mu\nu}F^{\mu\nu}) = 16\pi^2 T. \tag{3.2}$$

In (3.1) we have put

$$Q = F_{Y\bar{Y}}F_{Z\bar{Z}} + F_{Y\bar{Z}}F_{\bar{Y}Z}, \tag{3.3}$$

which defining the gauge invariant matrix P introduced in Ref. 7,

$$P = D\,E^{-1} = \frac{1}{f}\begin{pmatrix} 1 & -g \\ e & f^2 - eg \end{pmatrix}, \tag{3.4}$$

can be written

$$Q = -\,\mathrm{tr}[(P_{,\bar{Y}}P^{-1})_{,Y}(P_{,\bar{Z}}P^{-1})_{,Z} - (P_{,\bar{Z}}P^{-1})_{,Y}(P_{,\bar{Y}}P^{-1})_{,Z}]. \tag{3.5}$$

Straightforward computations give the following results:

$$(P_K P^{-1})_{K'} = \left(\frac{g_K}{f^2}\right)_{K'} E_1 + \left[\frac{g_K e_{K'}}{f^2} - \left(\frac{f_K}{f}\right)_{K'}\right]E_2 + \left(\frac{e_{K'}}{f^2}\right)_K f^2 E_3, \tag{3.6}$$

$$E_1 = \begin{pmatrix} e & -1 \\ e^2 & -e \end{pmatrix},\quad E_2 = \begin{pmatrix} 1 & 0 \\ 2e & -1 \end{pmatrix},\quad E_3 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

The only nonzero traces of products of E's are

$$\mathrm{tr}(E_2)^2 = -2\,\mathrm{tr}\,(E_1E_3) = 2. \tag{3.7}$$

Hence one obtains

$$Q = f^2\left\{\left(\frac{g_{\bar{Y}}}{f^2}\right)_Y\left(\frac{e_Z}{f^2}\right)_{\bar{Z}} + \left(\frac{g_{\bar{Z}}}{f^2}\right)_Z\left(\frac{e_Y}{f^2}\right)_{\bar{Y}} - 2f^2\left[\frac{g_{\bar{Y}}}{f^2}\frac{e_Y}{f^2} - \frac{1}{f^2}\left(\frac{f_Y}{f}\right)_{\bar{Y}}\right] \times\left[\frac{g_{\bar{Z}}}{f^2}\frac{e_Z}{f^2} - \frac{1}{f^2}\left(\frac{f_{\bar{Z}}}{f}\right)_Z\right]\right\} - [(Y \longleftrightarrow Z), (\bar{Y} \longleftrightarrow \bar{Y}), (\bar{Z} \longleftrightarrow \bar{Z})]. \tag{3.8}$$

By using (1.7) and the hermiticity relations (1.6), Q can be reduced to

$$Q_m = 2\left[\left|\frac{e_Y g_{\bar{Y}}}{f^2} - \left(\frac{f_Y}{f}\right)_{\bar{Y}}\right|^2 + \left|\frac{g_{\bar{Y}}e_Z}{f^2} - \left(\frac{f_{\bar{Y}}}{f}\right)_Z\right|^2\right] + mf^2\left[2\left|\left(\frac{g_{\bar{Y}}}{f^2}\right)_Y\right|^2 + \left|\left(\frac{g_{\bar{Y}}}{f^2}\right)_Z\right|^2 + \left|\left(\frac{g_{\bar{Z}}}{f^2}\right)_Y\right|^2\right]. \tag{3.9}$$

Note that the modulus bar denotes the complex norm and that Q_+ [SU (2)] is as expected a positive quantity.

It is easy to see that the transformations γ, δ, and θ do not change Q and hence the topological quantum number.

IV. CONSTRUCTION OF SOLUTIONS

In this section, the main part of the paper, we explicitly write an infinite series of solutions of the Yang equations starting from the known one,

$$S = (f, e = f, g = f). \tag{4.1}$$

The equations then reduce to

$$f\,\Box\,1/f = 0, \tag{4.2}$$

and have a large class of solutions

$$\frac{1}{f} = \alpha_0 + \sum_{i=1}^{N}\frac{\alpha_i}{(Y - Y_i)(\bar{Y} - \bar{Y}_i) + (Z - Z_i)(\bar{Z} - \bar{Z}_i)}, \tag{4.3}$$

where α_i are real numbers and Y_i, Z_i are complex numbers, the locations of the instantons.

Let us start with a particular case of (4.1), (4.3), up to a translation, namely

$$S_0 = [R^2/(1 + \lambda R^2)](1, 1, 1) + (0, k^*, k), \tag{4.4}$$

where

$$R^2 = r + s, \tag{4.5}$$

$$r = Y\bar{Y}, \tag{4.6}$$

$$s = Z\bar{Z}. \tag{4.7}$$

Under the transformation α (2.6) one obtains

$$S_1 = \alpha \cdot S_0, \tag{4.8}$$

$$S_1 = \frac{1}{R^2}\left(kk^*, (k^*)^2\frac{Y}{\bar{Z}}, -(k)^2\frac{\bar{Y}}{Z}\right) + (\lambda kk^* + k + k^*, k^*{}_{,Z}Y - k^*{}_{,Y}Z, k_{,Y}\bar{Z} - k_{,Z}\bar{Y}) + (0, l^*, -l). \tag{4.9}$$

The integrability condition on the topological SU (2) density $Q^{(1)}$ requires that f_1 cannot be zero,

$$f_1 = \frac{1}{R^2}[(1 + \lambda R^2)kk^* + R^2(k + k^*)]. \tag{4.10}$$

If we want k to be analytic, we see (thanks to the Liouville theorem) that only the constant k is acceptable. Writing $k = a + ib$ we deduce that a and b must be outside (resp. inside) the circle defined in (4.11) for $\lambda > 0$ (resp. $\lambda < 0$),

$$(a + 1/\lambda)^2 + b^2 = (1/\lambda)^2. \tag{4.11}$$

The condition for the $\lambda = 0$ case is simply that a must be positive. The $Q_+^{(1)}$ function is then

$$Q_+^{(1)} = \frac{6(kk^*)^2(k + k^* + \lambda\, kk^*)^2}{[kk^* + R^2(k + k^* + \lambda\, kk^*)]^4}, \tag{4.12}$$

and its integral over space–time (3.1) takes the following values if λ is positive (resp. negative):

$T = 0$, if k is on the circle. (4.11),

$T = 1$, if k is outside (resp. inside) the circle (4.11),

$T = \infty$, if k is inside (resp. outside) the circle (4.11). (4.13)

The solutions with $T = 1$ are the 't Hooft–Polyakov instanton solutions.

We now define a particular sequence of solutions obtained from a dilated (δ) form of solution (4.4) with $\lambda = -1, k = 1$,

$$S_0 = \frac{s^2}{r+s-1}\left(\frac{r+s}{s}, Z^{-2}, \bar{Z}^{-2}\right), \tag{4.14}$$

by applying successive powers of α to it and by choosing the integration functions (translations) in a simple way.

Let

$$S_n = (f_n, e_n, g_n), \tag{4.15}$$

the sequence obtained by

$$S_{n+1} = \alpha \cdot S_n = \alpha^{n+1} \cdot S_0, \tag{4.16}$$

which takes the general form

$$S_n = \frac{1}{K_{n-1}}(K_n, Y^n Z^{n-2}, (-)^n \bar{Y}^n \bar{Z}^{n-2}). \tag{4.17}$$

Using the notation (4.6) and (4.7) the K_n ($n > 0$) polynomials in r and s satisfy the following recurrence relations:

$$K_n^2 = K_{n-1} K_{n+1} + (-)^n r^n s^{n-2}, \tag{4.18}$$

$$(K_{n+1})_{,s} = (n+1) K_n - r(K_n)_{,r}, \tag{4.19}$$

$$(K_{n+1})_{,r} = -(n-1) K_n + s(K_n)_{,s}. \tag{4.20}$$

The particular form (4.17) and the three conditons above on K_n guarantee that the two equations (1.8) are automatically satisfied.

We now list a few properties of these functions. First the integration condition on K_{n+1} following from (4.19) and (4.20) reads

$$R_n = r(K_n)_{,rr} + s(K_n)_{,ss} - n(K_n)_{,r} - (n-2)(K_n)_{,s} = 0. \tag{4.21}$$

Once it is satisfied for R_n it is automatically satisfied for R_{n+1}. The polynomials K_{2n+1} ($n \geqslant 0$) can be factorized in polynomials as

$$K_{2n+1} = L_{2n+1} M_{2n+1}, \tag{4.22}$$

while those of the form K_{2n} are

$$\begin{aligned} K_{2n} &= L_{2n-1} M_{2n+1} + (-)^{n+1} r^n s^{n-1} \\ &= M_{2n-1} L_{2n+1} + (-)^n r^n s^{n-1}. \end{aligned} \tag{4.23}$$

The factors themselves satisfy equations which using

$$A = r^2 + 2r(s-1) + (1+s)^2, \quad B = rs, \tag{4.24}$$

read

$$\begin{aligned} M_{2n+5} &= A L_{2n+3} - B M_{2n+1}, \\ L_{2n+5} &= M_{2n+3} - B L_{2n+1}, \end{aligned} \tag{4.25}$$

and

$$\begin{aligned} M_{2n+3}^2 - M_{2n+1} M_{2n+5} &= (-)^{n+1} r A B^n, \\ L_{2n+3}^2 - L_{2n+1} L_{2n+5} &= (-)^n r B^n. \end{aligned} \tag{4.26}$$

For the polynomials

$$m_i = M_{2i+1}, \quad l_i = L_{2i+1}, \tag{4.27}$$

the following generating functions

$$\mathcal{M} = \sum_{i=0}^{\infty} m_i x^i, \quad \mathcal{L} = \sum_{i=0}^{\infty} l_i x^i, \tag{4.28}$$

can be written

$$\begin{aligned} \mathcal{M} = [&(1 + x^2 B)(x m_1 + m_0 - A x l_0) \\ &+ A(x^2 l_1 + x l_0 - x^2 m_0)(1 + x^2 B)^2 - A x^2]^{-1}, \end{aligned} \tag{4.29}$$

$$\mathcal{L}(m_0, m_1, l_0, l_1) = \mathcal{M}(l_0, l_1, m_0, m_1).$$

The K_n for negative index are related to the K_n for positive index by

$$K_{-n}(r, s) s^{n+1} r^{n-1} = (-)^n K_n(-s, -r). \tag{4.30}$$

We close the paragraph by giving the first polynomials of the sequence:

$$\begin{aligned} K_1 &= 1 + s + r, \\ K_2 &= (1+s)^2 + rs, \\ K_3 &= (1+s)^2 + r(s+1)(s-1), \\ K_4 &= (1+s)^4 + r(s+1)^2(s-2) + r^2, \\ K_5 &= (1+s)^5 + r(s+1)^3(s-3) + 3r^2(s+1) - r^3, \end{aligned} \tag{4.31}$$

$$\begin{aligned} L_1 &= 1, \\ L_3 &= 1 + s, \\ L_5 &= (1+s)^2 - r, \\ M_1 &= 1 + s + r, \\ M_3 &= (1+s)^2 + r(s-1), \\ M_5 &= (1+s)^3 + r(1+s)(s-2) + r^2. \end{aligned} \tag{4.32}$$

The following ones are easily constructed by using (4.25).

V. DISCUSSION OF THE SOLUTIONS

In the preceding section we obtained, starting from the 't Hooft–Polyakov solution, an explicit infinite series of solutions. Unfortunately these solutions (except for S_1) do not correspond to a finite action. Indeed to the SU (1,1) solution S_2 corresponds to $Q^{(2)}$ which can be computed explicitly,

$$\begin{aligned} Q^{(2)}(r, s) &= \frac{(1+s)^4 + 4(1+s)^2 r + r^2 - rs}{[(1+s)^2 + rs]^4} \\ &\quad \times [2r^2 + 3r(3s+4) + 18(1+s)^2]. \end{aligned}$$

If one tries to integrate this function over the $r - s$ domain with measure $drds$ one sees two quadratic divergencies of opposite sign. As a consequence the value of the integral is ill defined. On the other hand, it is easy to see that for the SU (2) odd index S_{2n+1}, the f_{2n+1} function has a line of zero through the domain. Hence the action is truly infinite due to the positiveness of the action density.

Contrary to the hope expressed in Ref. 6 we have shown that a naïve application of the Bäcklund transformation generates a series of solutions which do not seem to be related to those of Atiyah. In other words, Atiyah's solutions keep

their secret far from the world of naive and pedestrian mathematics.

[1]A.A. Beliavin, A.M. Polyakoff, A.S. Schwartz, and Y.S. Tiupkin, Phys. Lett. B **59**, 85 (1975).
[2]R. Jackiw, C. Nohl, and C. Rebbi, Phys. Rev. D **15**, 1642 (1977).
[3]E. Witten, Phys. Rev. Lett. **38**, 121 (1977).
[4]C.N. Yang, Phys. Rev. Lett. **38**, 1377 (1977).
[5]Y. Brihaye and J. Nuyts, Phys. Lett. B **66**, 346 (1977).
[6]E. Corrigan, D.B. Fairlie, P. Goddard, and R.G. Yates, Commun. Math. Phys. **58**, 223 (1978).
[7]Y. Brihaye, D.B. Fairlie, J. Nuyts, and R.G. Yates, J. Math. Phys. **19**, 2528 (1978).

Linearization of relativistic nonlinear wave equations

Moshé Flato and Jacques Simon
Physique Mathématique, Faculté des Sciences Mirande, 21000-Dijon, France

(Received 15 February 1979)

Some linearization theorems for relativistic nonlinear wave equations are formulated and proved. Examples are discussed and the physical meaning of such a linearization is sketched. Finally, some remarks concerning gauge theories are presented.

I. INTRODUCTION

The evolution of many physical systems is described by a solution of a partial differential equation which is covariant under some transformation group. Famous examples of such systems are given by relativistic classical fields, the evolution of which is governed by Poincaré covariant equations. In these examples the situation can be usually described heuristically as follows. We are given a vector space E of initial conditions, namely a space of functions on $\mathbb{R}^3$ taking values in a finite-dimensional vector space, and an evolution equation of the type $(d/dt)\varphi_t = H\varphi_t$, where $\varphi_t \in E$ and H is some (linear or nonlinear) differential operator on E (in the variables of $\mathbb{R}^3$) with $H(0) = 0$. The evolution can thus be written in the form $\varphi_t = T_t(\varphi_0)$ and we have $T_{t+t'} = T_t T_{t'}$ and $T_t(0) = 0$.

Now if $g = (\Lambda, a)$ is an element of the Poincaré group $\mathscr{P} = \mathrm{SL}(2,\mathbb{C})\cdot\mathbb{R}^4$ we can define an operator on E by

$$(S_g\varphi_0)(x) = \psi((\Lambda, -a)^{-1}(0,x)),$$

where $x \in \mathbb{R}^3$, and ψ is defined by $\psi(t,x) = \varphi_t(x)$. Denoting as usual by P_0 the generator of time translations of $\mathscr{P}$, when $(0,a) = \exp(aP_0)$ we have $T_a = S_{\exp aP_0}$, $S_g(0) = 0$ and the Poincaré covariance of the evolution equation implies that $S_{gg'} = S_g S_{g'}$.

Conversely, if we are given a map $g \to S_g$ from $\mathscr{P}$ to the group of invertible maps in E, vanishing at the origin, satisfying $S_{gg'} = S_g S_{g'}$, we can define heuristically the generator

$$H = \frac{d}{dt}(S_{\exp tP_0})\Big|_{t=0},$$

and we have $H(0) = 0$ and $(d/dt)\varphi_t = H\varphi_t$ if we define $\varphi_t = S_{\exp tP_0}(\varphi_0)$.

A general study of nonlinear representations of Lie groups in infinite-dimensional vector spaces was presented in Ref. 1. For the sake of completeness we shall briefly recall some notation and definitions. Given a topological vector space E, we denote by $\mathscr{L}_n(E)$ the space of continuous n-linear symmetric maps from E^n to E, and by $\mathscr{F}(E)$ the space of formal series on E, of the form $f = \Sigma_{n\geqslant 1} f_n$ with $f_n \in \mathscr{L}_n(E)$.

Definition 1: A formal representation (S,E) of a real Lie group G in a topological vector space E is a homomorphism S from G to the group of invertible elements in $\mathscr{F}(E)$ (for the usual composition law of formal series) such that, if $S_g = \Sigma_{n\geqslant 1} S_g^n$, the maps $g \to S_g^n(\varphi_1,\ldots,\varphi_n)$ are measurable for every $n \geqslant 1$ and $\varphi_i \in E$ $(1 \leqslant i \leqslant n)$.

Obviously, S^1 is then a measurable linear representation of G on E, which we call the *free part* of S.

Definition 2: Two formal representations (S,E) and (S',E) of G are said to be equivalent if there exists an invertible $A \in \mathscr{F}(E)$ such that $S'_g = AS_gA^{-1}$ for all $g \in G$. A formal representation is said to be *linearizable* if it is equivalent to a linear representation.

The advantage one gains by introducing nonlinear representations of the Poincaré group to study covariant equations is that the Poincaré group structure brings in more constraints than that of the real line (the time evolution). We shall give two examples.

(1) The mere knowledge of the free part S^1 may give useful information on the formal representation S. For instance, if S^1 is the restriction of a unitary irreducible representation of $\mathscr{P}$ with real positive mass $m^2 > 0$ to its Fréchet space E of differentiable vectors, and if $g \to S_g^n(\varphi,\ldots,\varphi)$ is C^∞ from $\mathscr{P}$ to E for every φ in E, then the formal representation S of $\mathscr{P}$ on E is linearizable (Ref. 2, Prop. 3).

(2) Even more surprising: Suppose that the free part S^1 is the irreducible representation of $\mathscr{P}$ on $\mathbb{R}^4$ acting as $\mathrm{SO}_0(3,1)$. Then it can be proved that the formal representation $S = \Sigma_{n\geqslant 1} S^n$, restricted to $\mathrm{SL}(2,\mathbb{C})$, is equivalent to a linear representation. If $x,y \in \mathbb{R}^4$ (identified with the translations subgroup) we define $dS_x = (d/dt)S_{\exp tx}|_{t=0}$ and $\langle x,y\rangle = x^0y^0 - \Sigma_{i=1}^3 x^i y^i$. Then (up to equivalence) the only nonlinearizable representations of $\mathscr{P}$ on $\mathbb{R}^4$ are indexed by an integer $p \in \mathbb{N}$ and given, on the translations subroup, by the generators

$$dS_x(\varphi) = \langle\varphi,\varphi\rangle^p\langle x,\varphi\rangle\varphi,$$

or

$$dS'_x(\varphi) = \langle\varphi,\varphi\rangle^p\langle x,\varphi\rangle\varphi - (2p+1)(2p+2)^{-1}\langle\varphi,\varphi\rangle^{p+1}x,$$

the restriction to $\mathrm{SL}(2,\mathbb{C})$ being linear (Ref. 3, Prop. 7).

Therefore, in $\mathbb{R}^4$, if one wants to perturb a nonlinearizable representation of $\mathscr{P}$ so as to get another (nonequivalent) nonlinearizable representation of $\mathscr{P}$, *we have no choice for the value of the "coupling constant"*; it must be of the form $-(2p+1)(2p+2)^{-1}$. The situation here is quite different from the usual perturbation of a linear representation S^1 into $g \to S_g^{(\lambda)} = \Sigma_{n\geqslant 1}\lambda^{n-1}S_g^n$ in which every value $\lambda \in \mathbb{C}$ is admissible, and S^λ is equivalent to $S^{\lambda'}$ for $\lambda\lambda' \neq 0$.

In the following we shall study further the linearization of Poincaré group representations. Under some technical as-

 0022-2488/80/040913-05$1.00

sumptions we shall prove that if S^1 has a scalar mass $m^2 \geqslant 0$ and a given energy sign, the restriction of $S = \Sigma_{n \geqslant 1} S^n$ to the Euclidean group is linearizable (Prop. 2). We then prove that when S^1 is a direct sum of an irreducible representation with $m^2 > 0$ and of its contragredient representation (namely we admit both energy signs, which is the situation when the free part comes from the free Dirac equation, for instance) then, up to equivalence, the restrictions of S to the space of positive energy, or to the space of negative energy, is equal to S^1 (which of course does *not* imply that $S = S^1$ everywhere since S is nonlinear).

II. RESULTS

We shall first introduce some notation and definitions. We assume a finite dimensional real vector space T, a real Lie group $\mathscr{L}$ and a semidirect product $G = \mathscr{L} \cdot T$, where T is an Abelian invariant subgroup of G. We denote by T^* the dual group of T, identified with the dual vector space of T. We denote by $\langle , \rangle$ the bilinear form from $T^* \times T$ to $\mathbb{R}$ which defines the duality. We assume an orbit M of G in T^* and we denote by H the subgroup of G which stabilizes a given point $\mu \in M$. We assume a linear representation (L,F) of H in a complex finite-dimensional vector space F. We denote by $\mathscr{E} = C^\infty(M,F)$ the Fréchet space of C^∞ functions from M to F.

Suppose that there exists a differentiable section s of the canonical projection π: $G \to G/H$ which is normalized [i.e., $s(\pi(e)) = e, e$ identity of G]. We put $xs(\pi(y)) = s(\pi(xy)) \times w(x,\pi(y))(x,y \in G)$ [hence $w(x,\pi(y)) \in H$]. One then defines the induced representation $(U^L,\mathscr{E})$ of G in $\mathscr{E}$ by $(U^L_g f) = L(w(x^{-1},\pi(y))^{-1}) f(\pi(x^{-1}y))$ (we do not include here a ρ_H function since we do not care about unitarity and, moreover, the example we have in view is the Poincaré group for which $\rho_H = 1$).

Suppose that there exists a basis $\{e_1^*,...,e_n^*\}$ of T^* such that M is defined by the relations $x^r = h_r(x^1,...,x^m)$ $(r = m+1,...,n)$ where h_r are C^∞ functions on an open set Q of $B^* = \mathbb{R} e_1^* + \cdots + \mathbb{R} e_m^*$. We assume two coverings (Q_α) and (Q'_α) of B^* by open relatively compact sets such that $Q_\alpha \subset \overline{Q}_\alpha \subset Q'_\alpha$. We denote by $\{e_1,...,e_n\}$ the dual basis of $\{e_1^*,...,e_n^*\}$ and we put $B = \mathbb{R} e_1 + \cdots + \mathbb{R} e_m$.

We assume a formal representation $(S,\mathscr{E})$ of G in $\mathscr{E}$ such that

(a) $S^1_g = U^L_g$,

(b) $g \to S^n_g(\varphi,...,\varphi)$ is C^∞ from G to $\mathscr{E}$, for every $\varphi \in \mathscr{E}$.

We denote by $\hat{\otimes}^n \mathscr{E}$ the projective tensor product $\mathscr{E} \hat{\otimes} \mathscr{E} \hat{\otimes} \cdots \hat{\otimes} \mathscr{E}$ (n times). We have $\mathscr{L}_n(\mathscr{E}) \subset \mathscr{L}(\hat{\otimes}^n \mathscr{E}, \mathscr{E})$ and $\hat{\otimes}^n \mathscr{E} \simeq C^\infty(M^n, \otimes^n F)$. We denote by $C^\infty_c(M^n, \otimes^n F)$ the space of C^∞ functions with compact support from M^n to $\otimes^n F$. If $T \in \mathscr{L}(C^\infty_c(M^n, \otimes^n F), C^\infty(M,F))$ and $a \in M$, we denote by T_a the distribution $f \to (T(f))_a$ $(f \in C^\infty_c(M^n, \otimes^n F))$.

The Lie algebra $\underline{B}$ of B is represented in $\mathscr{L}_n(\mathscr{E})$ in the following way:

$$b\tau = dU^L_b \tau - \sum_{p=0}^{n-1} \tau((\otimes^p I) \otimes dU^L_b \otimes (\otimes^{n-p-1} I)),$$

where $b \in \underline{B}$, $\tau \in \mathscr{L}_n(\mathscr{E})$ and I is the identity in $\mathscr{E}$. We denote by $Z^1(\underline{B}, \mathscr{L}_n(\mathscr{E}))$ the space of 1-cocycles of the $\underline{B}$ module $\mathscr{L}_n(\mathscr{E})$ and by $B^1(\underline{B}, \mathscr{L}_n(\mathscr{E}))$ the corresponding space of 1-coboundaries.

Proposition 1: There exists $A \in \mathscr{F}(\mathscr{E})$, invertible, such that, if $S'_g = AS_gA^{-1}$,

$$S'^n_g = 0, \quad \text{when } g \in B, \quad \text{and } n \geqslant 2.$$

Proof: From Ref. 2, Prop. 1 it is sufficient to prove that $Z^1(\underline{B}, \mathscr{L}_n(\mathscr{E})) = B^1(\underline{B}, \mathscr{L}_n(E))$ for every $n \geqslant 2$.

To prove that equality we shall follow a method given by Guichardet[4] and Rideau[5] in a similar situation. Suppose that (ξ_α) is a partition of unity subordinate to the covering Q_α and ξ'_α is a C^∞ function, with compact support contained in Q'_α, and equal to 1 on Q_α. Choose $Z \in Z^1(\underline{B}, \mathscr{L}_n(\mathscr{E}))$ and define $|\alpha| = (\alpha_1,...,\alpha_n)$ and $Z^{|\alpha|}_x = (\otimes^n_{i=1} \xi_{\alpha_i}) Z_x$ $(x \in \underline{B})$. Then $Z^{|\alpha|}_x \in \mathscr{L}(C^\infty_c(M^n, \otimes^n F), C^\infty(M,F))$. If we endow the spaces of linear mappings with the topology of uniform convergence on bounded sets, we have a canonical topological isomorphism

$$\mathscr{L}(C^\infty_c(M^n, \otimes^n F),\ C^\infty(M,F)) \simeq C^\infty(M, \mathscr{L}(C^\infty_c(M^n, \otimes^n F), F)).$$

The 1-cocycle condition writes, if $a, a_1,...,a_n \in M$,

$$\left(a^i - \sum_{k=1}^n a^i_k\right) Z^{|\alpha|}_{e_i} = \left(a^j - \sum_{k=1}^n a^j_k\right) Z^{|\alpha|}_{e_i}.$$

Given $f \in C^\infty_c(M^n, \otimes^n F) \simeq C^\infty_c(Q^n, \otimes^n F)$, we extend it to B^n by zero in $B^n - Q^n$ and get a function $\tilde{f} \in C^\infty_c(B^n, \otimes^n F)$.

If $x_1,...,x_n \in B$ and $x = (x_1,...,x_{n-1})$, we define $\tilde{f}_x(x_n) = \tilde{f}(x_1,...,x_n)$. Take $y \in B$ and put $d = y - \Sigma_{k=1}^{n-1} x_k$. We have

$$\tilde{f}_x(x^1_n,...,x^m_n) = \tilde{f}_x(d^1, x^2_n,...,x^m_n) - \left(y^1 - \sum_{k=1}^n x^1_k\right) g^1_{x,y}(x_n),$$

where we have defined (D_1 being the derivation in the first variable)

$$g^1_{x,y}(x_n) = \int_0^1 D_1 \tilde{f}_x[d^1 + (x^1_n - d^1)s, x^2_n,...,x^m_n]\, ds.$$

We then write inductively:

$$\tilde{f}_x(d^1,...,d^{l-1}, x^l_n,...,x^m_n) = \tilde{f}_x(d^1,...,d^l, x^{l+1}_n,...,x^m_n) - \left(y^l - \sum_{k=1}^n x^l_k\right) g^l_{x,y}(x_n),$$

where

$$g^l_{x,y}(x_n) = \int_0^1 D_l \tilde{f}_x(d^1,...,d^{l-1}, d^l + (x^l_n - d^l)s, x^{l+1}_n,...,x^m_n) ds. \tag{1}$$

Therefore

$$\tilde{f}_x(x_n) = \tilde{f}_x(d) - \sum_{l=1}^m \left(y^l - \sum_{k=1}^n x^l_k\right) g^l_{x,y}(x_n).$$

We put $I^l_{yl}(x_1,...,x_n) = (\Sigma_{k=1}^n x^l_k - y^l)$. We then have

$$\tilde{f}(x_1,...,x_n) = \tilde{f}(x_1,...,x_{n-1}, d) + \sum_{l=1}^n I^l_{yl}(x_1,...,x_n) g^l_y(x_1,...,x_n),$$

where $g^l_y(x_1,...,x_n) = g^l_{x,y}(x_n)$. If $f \in C^\infty_c(B^n, \otimes^n F)$, define $T^{|\alpha|}_y(f) = \Sigma^n_{l=1}(Z^{|\alpha|}_{e_l}(g^l_y)_y)$. From (1) this defines a distribution $T \in \mathscr{L}(C^\infty_c(M^n, \otimes^n F), C^\infty(M,F))$. We are going to prove that $Z^{|\alpha|}_{e_l,y} = [(\Sigma^n_{k=1} x^l_k) - y^l]T^{|\alpha|}_y$ [i.e., for every $f \in C^\infty_c(B^n, \otimes^n F)$ we have $Z^{|\alpha|}_{e_l,y}(f) = T^{|\alpha|}_y(I^l_{yl} f)$. Define $F = I^l_{bl} f$ and, as before, the corresponding functions $G^{l'}_y$. Since $F(x_1,...,x_n) = 0$ if $(\Sigma^n_{k=1} x^l_k) - y^l = 0$ we have

$$G^{l'}_y = 0, \quad \text{if } l' > l, \tag{2}$$

$$G^{l'}_y(x_1,...,x_n) = 0, \quad \text{if } \sum_{k=1}^{n} x^l_k = y^l, \quad \text{when } l' \leqslant l.$$

Therefore, $G^{l'}_y = I^l_{yl} H^{l'}_y$ if $l' \leqslant l$, with $H^{l'}_y \in C^\infty(B^n, \otimes^n F)$. It results then from (2) that

$$F = \sum_{l'=1}^{l-1} I^{l'}_{yl'} I^l_{yl} H^l_y + I^l_y G^l_y, \tag{3}$$

$$f = \sum_{l'=s}^{l-1} I^{l'}_{yl'} H^{l'}_y + G^l_y.$$

We have

$$\begin{aligned} T^{|\alpha|}_y(I^l_{bl} f) = T^{|\alpha|}_y(F) &= \sum_{l'=1}^{l} Z^{|\alpha|}_{e_{l'},y}(G^{l'}_y) \\ &= \sum_{l'=1}^{l-1} Z^{|\alpha|}_{e_{l'},y}(I^l_{yl} H^{l'}_y) + Z^{|\alpha|}_{e_l,y}(G^l_y) \\ &= \sum_{l'=1}^{l-1} Z^{|\alpha|}_{e_l,y}(I^{l'}_{yl'} H^{l'}_y) + Z^{|\alpha|}_{e_l,y}(G^l_y) \\ &= Z^{|\alpha|}_{e_l,y}\left(\sum_{l'=1}^{l-1} I^{l'}_{y_{l'}} H^{l'}_y + G^l_y\right) \\ &= Z^{|\alpha|}_{e_l,y}(f), \text{ by (3).} \end{aligned}$$

Define

$$T_y = \sum_{|\alpha|} \left(\overset{n}{\underset{i=1}{\otimes}} \xi'_{\alpha_i}\right) T^{|\alpha|}_y;$$

$$T \in \mathscr{L}(C^\infty_c(B^n, \otimes^n F),\ C^\infty(B,F)),$$

and

$$\left[\left(\sum_{k=1}^{n} x^l_k\right) - y^l\right] T_y = I^l_{yl} T_y = \sum_{|\alpha|} \left(\overset{n}{\underset{i=1}{\otimes}} \xi'_{\alpha_i}\right) I^l_{bl} T^{|\alpha|}_y$$

$$= \sum_{|\alpha|} \left(\overset{n}{\underset{i=1}{\otimes}} \xi'_{\alpha_i}\right) Z^{|\alpha|}_{e_l,y} = \sum_{|\alpha|} \overset{n}{\underset{i=1}{\otimes}} \xi'_\alpha \xi_{\alpha_i} Z_{e_l,y}$$

$$= \sum_{|\alpha|} \overset{n}{\underset{i=1}{\otimes}} \xi_{\alpha_i} Z_{e_l,y} = Z_{e_l,y}.$$

Define $\sigma_n : C^\infty(B^n, \otimes^n F) \to C^\infty(B^n, \otimes^n F)$ by $(\sigma_n f)(x_1,...,x_n) = (n!)^{-1} \Sigma_{\sigma \in \mathfrak{S}_n} f(x_{\sigma(1)},...,x_{\sigma(n)})$, where $\mathfrak{S}_n$ is the permutation group of n elements. Define $(\sigma_n T)(f) = T(\sigma_n f)$. Since $Z_x \in \mathscr{L}_n(\mathscr{E})$, we have

$$Z_{e_l,y} = (\Sigma^n_{k=1} x^l_k - y^l)\sigma_n T_y.$$

Therefore,

$$Z^1(\underline{B}, \mathscr{L}_n(\mathscr{E})) = B^1(\underline{B}, \mathscr{L}_n(\mathscr{E})). \qquad \text{Q.E.D.}$$

Proposition 2: Let $\mathscr{P}$ denote the Poincaré group and M be a connected component of the two-sheeted hyperboloid $(x^0)^2 - \Sigma^3_{i=1}(x^i)^2 = m^2$ with $m^2 > 0$, or, for $m = 0$, of the vertexless light cone. Then, if H is a subgroup of $\mathscr{P}$ which is isomorphic to the Euclidean group $\tilde{E}(3) = SU(2) \cdot \mathbb{R}^3$ there exists $A \in \mathscr{F}(\mathscr{E})$, invertible, such that, if $S'_g = AS_g A^{-1}$, we have $S'^n_g = 0$ for all $g \in H$ and $n \geqslant 2$.

Proof: Denote here by B the radical of H. It results from Prop. 1 that there exists $U \in \mathscr{F}(\mathscr{E})$, invertible such that, if $T_g = US_g U^{-1}$, we have $T^n_g = 0$ for $g \in B$ and $n \geqslant 2$. Denote by Y a Levi factor of H. Being isomorphic to SU(2), Y is compact, and we can define $V = \int_Y T^1_{y^{-1}} T_y \, dy$. Then $T^1_y V = VT_y$ for all $y \in Y$. If $b \in B$, we have $T_y T^1_b T_{y^{-1}} = T^1_{yby^{-1}}$. Hence $VT^1_b = T^1_b V$. We then take $A = VU$. Q.E.D.

Given a unitary irreducible representation $(U, \mathscr{H})$ of the Poincaré group with $m^2 > 0$, denote by $(U^\infty, \mathscr{H}^\infty)$ the restriction of U to the Fréchet space of its differentiable vectors $\mathscr{H}^\infty$. We denote by $(\tilde{U}, \tilde{\mathscr{H}})$ the contragredient representation of $(U, \mathscr{H})$, by $(\tilde{U}^\infty, \tilde{\mathscr{H}}^\infty)$ its restriction to the space of differentiable vectors, and consider the representation $V = U^\infty \oplus \tilde{U}^\infty$ on $E = \mathscr{H}^\infty \oplus \tilde{\mathscr{H}}^\infty$.

Proposition 3: Let (S,E) be a formal representation of the Poincaré group $\mathscr{P}$, $S_g = \Sigma_{n \geqslant 1} S^n_g$ with $S^n_g \in \mathscr{L}_n(E)$, such that $g \to S^n_g(\varphi,...,\varphi)$ is C^∞ from $\mathscr{P}$ to E for every $\varphi \in E$ and that $S^1 = V$. Then there exists $A \in \mathscr{F}(E)$, invertible, such that, if $S'_g = AS_g A^{-1}$, we have $S'_g \varphi = V\varphi$ whenever $\varphi \in \mathscr{H}^\infty \cup \tilde{\mathscr{H}}^\infty$.

Proof: Let us write $E_1 = \mathscr{H}^\infty$ and $E_2 = \tilde{\mathscr{H}}^\infty$. We denote by $\mathscr{L}_n(E_i,E)$ the space of continuous n-linear symmetric mappings from E^n_i to E. By a proof similar to that of (Ref. 2, Prop. 2) one sees, in the notation of Ref. 2, that $Z^1(\mathscr{P}, \mathscr{L}_n(E_i,E)) = B^1(\mathscr{P}, \mathscr{L}_n(E_i,E))$ $(i,j = 1,2)$.

We have $S^2(\hat{\otimes}^2 S^1)^{-1} \in Z^1_\infty(\mathscr{P}, \mathscr{L}_2(E))$. Therefore, there exists $B^2_i \in \mathscr{L}_2(E_i,E)$ such that for $\varphi, \psi \in E_i$

$$S^2_g(\varphi \otimes \psi) = (S^1_g B^2_i - B^2_i \hat{\otimes}^2 S^1_g)(\varphi \otimes \psi).$$

Choose $B^2 \in \mathscr{L}_2(E)$ such that $B^2(\varphi^i \otimes \varphi^i) = B^2_i(\varphi^i \otimes \varphi^i)$ when $\varphi^i \in E^i$ $(i = 1,2)$; take for instance $B^2|_{E_i \otimes E_j} = 0$ if $i \neq j$.

Define $S^{(1)}_g = S_g$ and

$$S^{(2)}_g = (1 - B^2)^{-1} S^{(1)}_g (1 - B^2).$$

Then $S^{(2)2}_g|_{E_i} = 0$. Suppose that we have defined a formal representation $(S^{(p)}, E)$ of $\mathscr{P}$ such that $S^{(p)k}_g|_{E_i} = 0$ if $2 \leqslant k \leqslant p$. It results from the relation $S^{(p)}_{gg'} = S^{(p)}_g S^{(p)}_{g'}$ that $S^{(p)p+1} \times (\hat{\otimes}^{p+1} S^1)^{-1}|_{E_i} \in Z^1(\mathscr{P}, \mathscr{L}_n(E_i,E))$. Therefore, there exist $B^{p+1}_i \in \mathscr{L}_{p+1}(E_i,E)$ such that

$$S^{(p)p+1}_g \left(\overset{p+1}{\underset{k=1}{\otimes}} \varphi_k\right) = (S'_g B^{p+1}_i - B^{p+1}_i \hat{\otimes}^{p+1} S^1_g) \times \left(\overset{p+1}{\underset{k=1}{\otimes}} \varphi^i_k\right), \quad \varphi^i_k \in E_i.$$

Choose $B^{p+1} \in \mathscr{L}_{p+1}(E)$ such that $B^{p+1}(\otimes^{p+1}_{k=1} \varphi^i_k) = B^{p+1}_i(\otimes^{p+1}_{k=1} \varphi^i_k)$ if $\varphi^i_k \in E_i$; take for instance $B^{p+1}|_{E_{i_1} \hat{\otimes} \cdots \hat{\otimes} E_{i_{p+1}}} = 0$ if all $i_1,...,i_{p+1}$ are not equal. Define $S^{(p+1)}_g = (1 - B^{p+1})^{-1} S^{(p)}_g (1 - B^{p+1})$. We have $S^{(p+1)p+1}_g|_{E_i} = 0$. Consider the polynomial $P_n = (1 - B^1) \cdots (1 - B^n)$. We have $S^{(n)}_g = P_n^{-1} S_g P_n$. The

sequence $(P_n)_{n\geqslant 2}$ is such that, if $m\geqslant n$, $P_m - P_n$ contains terms of degree strictly greater than n. Therefore, this sequence has a limit $A\in\mathcal{F}(E)$ in the topology of the filtration by the degree. A has an inverse since its term of degree one is the identity. We thus have $A^{-1}S_g A\varphi = V\varphi$ if $\varphi\in E_1\cup E_2$. Q.E.D.

Remark: Similar results can be obtained for the linearization of the time translation *alone*.

III. DISCUSSION

We should start our discussion by exhibiting some examples of wave equations related to previous sections.

Example 1: As was mentioned before, if φ is a function of t, the "wave equation"

$$\frac{d}{dt}\varphi = \langle\varphi,\varphi\rangle^p\varphi^0\varphi - (2p+1)(2p+2)^{-1}\langle\varphi,\varphi\rangle^{p+1}e_0,$$

where $p\in\mathbb{N}$, $\varphi\in\mathbb{R}^4$ and $e_0 = (1,0,0,0)$ is *nonlinearizable* and yet covariant in our generalized sense under the "Poincaré group."

Example 2: In the physical four-dimensional space–time the wave equations treated by many authors, of the form $(\Box + m^2)\varphi = F(\varphi)$ with F analytic, are *linearizable* on the subspace of positive energy initial conditions as well as on the subspace of negative energy initial conditions. Since in classical field theories these are the subspaces of physical interest (corresponding respectively to particles and antiparticles) we believe that this result is of some interest.

Example 3: Once more in the physical four-dimensional space–time consider the *first order* (in time) evolution equation

$$\frac{\partial\varphi}{dt} = (-\Delta + m^2)^{1/2}\varphi + F(\varphi).$$

If we take as the space of initial conditions the Fourier transform of the usual Wigner space of square-integrable functions on the positive energy hyperboloid, $F(\varphi)$ will leave this space invariant and therefore the equation will be *linearizable*!

What is the physical motivation and the physical meaning of our linearization program? Here are some points:

(1) In case of linearizability of a given family of wave equations we have an algorithm for solving explicitly such equations. Such a way of solving the equation (if in addition convergence is proved) will necessarily be exhaustive for the *given* space of initial conditions.

(2) What we are doing (in case of linearizability) is a kind of a generalized time-independent Hamilton–Jacobi transformation in classical field theory, where by "change of coordinates" in the initial fields space, we bring the system into a linear (and thus completely integrable) form.

(3) In any time-evolution system (first order system in time), if we denote the nonlinear evolution by $S(t)$ and the linear (free) part by $U(t)$, one can check that formally the operator $\lim_{T\to\infty}(2T)^{-1}\int_{-T}^{T}U(-t)S(t)\,dt$ [or any suitable invariant mean on $\mathbb{R}$ applied on $t\to U(-t)S(t)$] intertwines between the nonlinear evolution and the corresponding linear one.

If one develops $S(t)$ into a power series, there exist mainly three obstacles for giving a mathematically precise meaning to such an operator:

(a) being able to interchange the integral with the summation which defines the power series for $S(t)$,

(b) convergence of each term in the summation after the interchange is performed, and

(c) convergence of the whole series in a suitable topology.

The formal linearization procedure ensures at least point (b), which seems to be the most crucial. In "soliton physics" examples are known of convergence of such an intertwining integral in the soliton sector.

(4) There exists a certain connection between what we are doing and S-matrix formalism of *classical* field theory. In a sense we are performing a type of abstract scattering theory. To be more precise it is quite evident that the wave operator (often called the Möller operator) is an intertwining operator between the free motion and the nonlinear motion. Therefore, if one wants to demonstrate the existence of a (not necessarily trivial) S matrix in classical field theory via the existence of the Möller operator (the most usual technique applied for such an aim), a necessary condition would be linearizability of our nonlinear equation in our sense discussed above.

(5) In addition to all that was said above, the main physical motivation to our mathematical study is the utmost aim of classifying "all possible interactions" (in a well-defined mathematical sense) which can be generated by a given family of fields. It is rather clear that the problem of equivalence classes of interactions (in a nonlinear sense) of nonlinear covariant field equations is equivalent to the characterization of equivalence classes of nonlinear representations of the covariance group having a given free part, a problem with which we have been dealing precisely.

(6) In Example 1, an interesting phenomenon occurs: A numerical coefficient $-(2p+1)/(2p+2)$ appears in our equation in an *essential* way. If we *change* this coefficient by another one (different from zero) the resulting equation will *no more* be covariant under our covariance group. In other words, this numerical coefficient is *necessary* for the commutation relations of the nonlinear representation of the Lie algebra to hold.

One can therefore speculate in an evident way: Suppose that in infinite-dimensional nonlinear representations of the Poincaré group similar phenomena to that of Example 1 occur (a thing which is pretty reasonable!). Is it not possible that dimensionless coupling constants in physics (and especially α in electrodynamics) might occur in such a manner, and therefore might be calculable from the nonlinear group theoretical formalism itself?

Though nonlinear field theories form a subject of great interest by itself, one cannot deny the fact that the most popular representatives of such theories have been in the last decade the so-called gauge theories. Most of the modern theories of weak as well as of strong interactions are in fact gauge theories. Therefore, though our article does not deal directly with them, we shall end our paper with some remarks concerning gauge theories.

Evidently, one canot avoid being impressed by the simplicity and mathematical elegance of gauge theories. Yet one can still have some doubts about their interpretation as well as about the belief that among nonlinear theories *only* gauge theories have direct connection with physical reality.

Here are some remarks concerning such questions, most of which apply to both Abelian and non-Abelian gauge theories.

Suppose that we have a classical field theory of spinor and vector fields. To be more precise, let us consider a system of wave equations of the type $L(A,\psi) = N(A,\psi)$, where A stands for the vector fields, ψ for the Fermi fields, L the free (linear) part, and N for the interacting (nonlinear) part.

Suppose that $L(A + a,\psi) = L(A,\psi)$ but that $N(A + a,\psi) \neq N(A,\psi)$ (note that intentionally we do not touch here the spinor fields), where a is a translation in the solution space of the linear equation in the vector field part. (In electrodynamics a would for instance be the famous pure gauge $\partial_\mu \varphi$. Of course the whole point of the usual gauge invariance there would be to transform also the ψ field in such a way that

$$L(A_\mu,\psi) - N(A_\mu,\psi) \equiv L(A_\mu + \partial_\mu\varphi, e^{i\varphi}\psi) - N(A_\mu + \partial_\mu\varphi, e^{i\varphi}\psi)$$

holds).

Now, we can easily convince ourselves that the set of solutions of the system $L(A,\psi) = N(A,\psi)$ and that of the system $L(A,\psi) = N(A + a,\psi)$ are at least mathematically equivalent: There is a formal one-to-one mapping which maps the solutions of the first system bijectively onto the solutions of the second system.

In quantum field theory the situation seems in a way even better: Suppose (as is usually done) that our quantization procedure is based on functional-integration techniques. Then, since in such techniques we integrate on the classical free fields, the two theories can be expected to be even physically equivalent, at least formally.

In other words, even without touching the spinor fields (and thus without having necessarily the usual gauge invariance for our theory) the situation can still be physically interesting.

The by now classical argument of Yang and Mills is physically simple and appealing: If one believes in gauge invariance as a fundamental principle, massless vector fields obeying Yang–Mills equations "must" exist. If (for physical applications) they require mass, give them mass by the Higgs–Kibble mechanism. If the fields $F^i_{\mu\nu}$ are not gauge *invariant* (only $F^i_{\mu\nu}\,F^i_{\mu\nu}$ is!) still they are observables in a certain sense as their quanta should represent the vector particles. If one cannot always determine even up to a gauge transformation in the non-Abelian case (Gribov's difficulty) the potentials from the fields, this fact might turn to be rather an advantage for the famous confinement problem.

In other words, gauge theories have become something like a dogma!

One should note that actually in gauge theories the only term we have to compensate is a pure gauge. In other words, vector fields are not needed to ensure gauge invariance, and fictitious scalar fields can do. As a matter of fact all that is needed is to replace all derivatives of the Fermi fields by $\partial_\mu - ie\partial_\mu A$ and postulate that the scalar fields A transform like $A \to A + \Lambda$ under local gauge transformations.

The scalar fields A do not have a free (kinetic) Lagrangian and the variation of the total Lagrangian with respect to A gives nothing more than the conservation equations for the currents.

The last but not least remark is the following: Look at the Hilbert space of the classical fields and note that the Poincaré group acts in this space and in particular on the Fermi fields, on which we concentrate. Let $U(x)$ be matrices over a ring of functions on $\mathbb{R}^4$ which are unitary on the spinor fields and identity on the vector fields. Then the operation $\psi \to U\psi$, $A_\mu \to A_\mu$, $P_\mu \to UP_\mu U^+$ and $M_{\mu\nu} \to UM_{\mu\nu}U^+$ is a trivial symmetry of the nonlinear Lagrangian which is a different way of interpreting gauge invariance of spinor fields without touching the vector fields.

Evidently in such an interpretation, which is a change of basis in Hilbert space (similarity transformation), the differential operator which defined the field equation does not stay invariant but transforms by similarity.

Therefore, our operation U does not send solutions of the equation to themselves, but defines a weak covariance exactly like conformal transformations in some massive theories (as there, this notion turns out to be useful). It is also quite clear that in quantized Yang–Mills theories more general fields of operators $U(x)$ on $\mathbb{R}^4$ are admissible, and that these might be of some physical interest.

To sum up, we end with two interesting questions for future research which arise from our discussion:

(1) One should study situations in which usual gauge invariance is not satisfied (nonminimal coupling) and see whether they can apply in particle physics.

(2) Since contrary to Yang and Mills philosophy we do not think that the photon exists thanks to the usual gauge invariance, what will then be the true mechanism which forces the existence of the photon?

ACKNOWLEDGMENTS

We thank Christian Fronsdal, Bengt Nagel, and Daniel Sternheimer for interesting discussions.

[1] M. Flato, G. Pinczon, and J. Simon, Ann. Sci. Ec. Norm. Sup. 4e sér. **10**, 405 (1977).
[2] M. Flato and J. Simon, Lett. Math. Phys. **2**, 155 (1977).
[3] G. Pinczon and J. Simon, Lett. Math. Phys. **2**, 499 (1978).
[4] A. Guichardet, "Extensions des Représentations Induites des Produits Semi-Directs," Preprint M357-06-78, Centre Mathématique Ecole Polytechnique, Palaiseau, France (1978).
[5] G. Rideau, "On the Extensions of Mass-Zero Representations of the Poincaré Group," Preprint, Université de Paris VII (France).

Operator ordering and functional integrals for Fermi systems

T. Suzuki[a)]
International Centre for Theoretical Physics, Trieste, Italy
and Istituto di Fisica Teorica dell 'Università, Trieste, Italy

(Received 5 January 1979)

Operator ordering for Fermi systems is formulated generally with the aid of the Grassmann numbers. The ordering schemes are restricted to a special class called s-ordering in contrast with those for Bose systems. A relation, similar to that for Bose systems, is shown to exist between different values of the equal-time contractions in the Dyson–Wick expansions of the S matrix and different operator orderings in the interaction Hamiltonian. The S matrix is independent of the ordering scheme chosen. In the functional integral context, different orderings correspond to different discretization prescriptions for the definition of the functional integrals in terms of coherent states for Fermi operators.

I. INTRODUCTION

In general, interaction Hamiltonians contain products of noncommuting operators at equal times in quantum theory. Since Dyson's time-ordering operation does not prescribe the ordering of operators when two or more times coincide, the Dyson–Wick expansion cannot be applied to perturbation calculations in its simple form.

In previous papers (Refs. 1, 2, and 3, to be referred to as I, II, and III, respectively) the author (in collaboration with Leschke and Hirshfeld) has investigated in detail the operator-ordering problem for Bose systems in quantum mechanics and in quantum field theory. They pointed out the necessity of prescribing definite (regularized) values for the equal-time contractions in order to be able to apply Dyson–Wick perturbation methods. The values chosen are intimately related to the ordering of the operators in the interaction Hamiltonian. Different orderings lead to different Feynman rules, although the S matrix is independent of the ordering chosen.

Different operator orderings correspond to different definitions of functional integration in the functional integral method.[4] In Ref. 5, the author has studied the general correspondence rules (general ordering schemes) in the framework of the functional integrals for Bose systems. It was found that the functional integration in the lattice approximation and the phase space function should be defined in an ordering dependent way.[4] Among general ordering schemes, special ordering schemes, called u ordering and s ordering, are shown to have the simplest expressions in the phase space and the coherent state representations, respectively.

In this paper we discuss generally operator ordering for Fermi systems and develop generalized Dyson–Wick expansions for the S matrix. Then we study the functional integral expressions for the general ordering schemes based on the recent work by Ohnuki and Kashiwa,[6] who discussed the coherent-state representation of the functional integral for Fermi systems. We shall find that similar conclusions to those for the Bose case in all respects are obtained for the Fermi case.

II. OPERATOR ORDERING FOR FERMI SYSTEMS

We start by defining operator ordering generally for Fermi systems. This may easily be done in much the same way as was done for Bose systems in I–III. The only change is that one must introduce anticommuting c numbers, i.e., the Grassmann numbers.[7]

For simplicity, let us discuss first a system with one degree of freedom described by creation and annihilation operators $\hat{a}$ and $\hat{a}^+$, which satisfy the usual anticommutation relations. As for the Bose case, we define a generalized Weyl operator

$$D_s(\alpha,\beta) \equiv F_s(\alpha,\beta) e^{(\alpha\hat{a} + \beta\hat{a}^+)}, \tag{2.1}$$

where α and β are the Grassmann numbers assumed to anticommute with $\hat{a}$ and $\hat{a}^+$, respectively, also. Due to the property of the Grassmann numbers, the function $F_s(\alpha,\beta)$ is generally expressed as [under the condition $F_s(0,0) = 1$]

$$F_s(\alpha,\beta) = 1 + c_0\alpha + c_1\beta + \tfrac{1}{2}s\alpha\beta, \tag{2.2}$$

where c_0, c_1, and s are arbitrary c numbers. That F_s is limited to the above simple form is characteristic of the Fermi case, contrary to the Bose case.

Let us now assume that $c_0 = c_1 = 0$, the meaning of which will be explained soon. Then one sees from the form of Eq. (2.2) that the general ordering scheme in the Fermi case is restricted to only that called s ordering in the Bose case.[1,2] The s-ordered product $\{\ \}_s$ of an operator $G(\hat{a},\hat{a}^+)$ is defined by the relation

$$\{G(\hat{a},\hat{a}^+)\}_s = G\left(\frac{\partial}{\partial\alpha}, \frac{\partial}{\partial\beta}\right) D_s(\alpha,\beta)|_{\alpha=\beta=0}. \tag{2.3}$$

Note that, inside the ordering bracket $\{\ \}_s$, all operators must be treated as if they were the Grassmann numbers. We shall use left derivatives for the definition of the derivatives with respect to the Grassmann numbers.[7] The assumption $c_0 = c_1 = 0$ about the form of $F_s(\alpha,\beta)$ corresponds to

[a)] On leave of absence from the Department of Physics, Kanazawa University, 920 Kanazawa, Japan.

 0022-2488/80/040918-06$1.00

requiring

$$\{\hat{a}\}_s = \hat{a} \quad \text{and} \quad \{\hat{a}^+\}_s = \hat{a}^+ . \tag{2.4}$$

Some of the orderings covered by the s ordering are as follows: $s = 1,0$, and -1 correspond to the normal, Weyl, and antinormal ordering, respectively.

Now we introduce a Grassmann number function $g_s(a,a^+)$, called an s function, associated with the operator $G(\hat{a},\hat{a}^+)$ for a given s. This function is uniquely defined by

$$G(\hat{a},\hat{a}^+) \equiv \{g_s(\hat{a},\hat{a}^+)\}_s . \tag{2.5}$$

The meaning of the right-hand side is that one must replace the Grassmann numbers a and a^+ in the s function by the operators $\hat{a}$ and $\hat{a}^+$, and s order the result.

Differently ordered products and the corresponding functions of an operator G are related to each other in the following way:

$$\{G(\hat{a},\hat{a}^+)\}_s = \left\{\left(1 + \frac{s-s'}{2}\frac{\partial^2}{\partial\hat{a}\partial\hat{a}^+}\right)G(\hat{a},\hat{a}^+)\right\} = \left\{\exp\left(\frac{s-s'}{2}\frac{\partial^2}{\partial\hat{a}\partial\hat{a}^+}\right)G(\hat{a},\hat{a}^+)\right\} , \tag{2.6}$$

and

$$g_{s'}(a,a^+) = \left[1 + \frac{s-s'}{2}\frac{\partial^2}{\partial a\partial a^+}\right]g_s(a,a^+) = \exp\left(\frac{s-s'}{2}\frac{\partial^2}{\partial a\partial a^+}\right)g_s(a,a^+) . \tag{2.7}$$

The proof is easily obtained by observing that $G(\hat{a},\hat{a}^+)$ is expressed in general as $A_1 + A_2\hat{a} + A_3\hat{a}^+ + A_4\hat{a}\hat{a}^+$. (See also the Appendix in II.)

The above procedure may be generalized simply to many degrees of freedom. Since one may consider different orderings for different degrees of freedom, Eq. (2.3) is rewritten as

$$\{G(\hat{a},\hat{a}^+)\}_{s=\{s_i\}} = G\left(\frac{\partial}{\partial\alpha},\frac{\partial}{\partial\beta}\right)D_s(\alpha,\beta)|_{\alpha=\beta=0} , \tag{2.8}$$

where

$$D_s(\alpha,\beta) = \exp\left(\sum_i \frac{s_i}{2}\alpha_i\beta_i\right)\exp\left[\sum_i(\alpha_i\hat{a}_i + \beta_i\hat{a}_i^+)\right] . \tag{2.9}$$

Also, Eqs. (2.6) and (2.7) are re-expressed as

$$\{G(\hat{a},\hat{a}^+)\}_{s'} = \left\{\exp\left(\sum_i \frac{s_i - s_i'}{2}\frac{\partial^2}{\partial\hat{a}_i\partial\hat{a}_i^+}\right)G(\hat{a},\hat{a}^+)\right\}_{s'} , \tag{2.10}$$

$$g_{s'}(a,a^+) = \exp\left(\sum_i \frac{s_i - s_i'}{2}\frac{\partial^2}{\partial a_i\partial a_i^+}\right)g_{s'}(a,a^+) . \tag{2.11}$$

(Note that the product of two Grassmann numbers commutes with any Grassmann number and operator. The proof is therefore given in the same way as in the Appendix of II.)

In quantum field theory, the ordering terms are divergent. It is therefore necessary to introduce some regularization method to treat the divergent integral. In III, we have discussed quantum field theory for the Bose systems on the basis of dimensional regularization and the usual cutoff method. Similar discussions may be carried out also about the Fermi case. In Appendix A we shall show, as an example, operator ordering of a free Dirac theory.

III. GENERALIZED WICK THEOREM AND DYSON-WICK EXPANSIONS FOR THE S MATRIX

We now generalize the usual Wick theorem to the case when two or more times coincide. It is done analogously to the procedure for the Bose systems. We introduce a time-ordering operation T_s, which extends the usual definition to the case of equal times through

$$T_s\{G(\hat{a}(t),\hat{a}^+(t),t)\} \equiv e^{itH^0}\{G(\hat{a},\hat{a}^+,t)\}_s e^{-itH^0} , \tag{3.1}$$

where H^0 is a free Hamiltonian. $\hat{a}(t)$ and $\hat{a}^+(t)$ are operators in the interaction representation. We easily find that

$$T_s\{g_{s'}(\hat{a}(t),\hat{a}^+(t),t)\} = G(\hat{a}(t),\hat{a}^+(t),t) . \tag{3.2}$$

It is convenient to introduce the following notation:

$$\hat{a}^\nu(t) = \begin{cases}\hat{a}(t), & \text{for } \nu = 1,\\ \hat{a}^+(t), & \text{for } \nu = 2.\end{cases} \tag{3.3}$$

For an arbitrary operator function, the generalized Wick theorem may be given by

$$T_s\{G(\hat{a}^\nu)\} = \left\{Z^0_{s,s'}\left[\frac{\delta}{\delta\hat{a}^\nu}\right]G(\hat{a}^\nu)\right\} . \tag{3.4}$$

The functional differential operator on the right-hand side is obtained from the Gaussian functional

$$Z^0_{s,s'}[\alpha^\nu] = \exp\left[-\tfrac{1}{2}\sum_{\nu,\nu'}\int dt \times \int dt'\,\Delta^{\nu\nu'}_{s,s'}(t-t')\alpha^\nu(t)\alpha^{\nu'}(t')\right], \tag{3.5}$$

by substituting the Grassmann number argument $\alpha^\nu(t)$ with the differential operator $\delta/\delta\hat{a}^\nu(t)$. The contractions appearing in Eqs. (3.5) are

$$\Delta^{\nu\nu'}_{s,s'}(t-t') \equiv T_s\{\hat{a}^\nu(t)\hat{a}^{\nu'}(t')\} - \{\hat{a}^\nu(t)\hat{a}^{\nu'}(t')\}_{s,s'} . \tag{3.6}$$

The proof of the generalized Wick theorem (3.4) is given by the combination of the usual Wick theorem[8] (for different times) with the formula (2.10) (for equal times), as shown in Appendix B.

Now let us apply the theorem to the scattering operator

$$S = T\left\{\exp\left[-i\int dt H^I(\hat{a}^\nu(t),t)\right]\right\} , \tag{3.7}$$

where H^I is an interaction Hamiltonian. First we find

$$S = T_s\left\{\exp\left[-i\int dt h^I_s(\hat{a}^\nu(t),t)\right]\right\} , \tag{3.8}$$

where we have used Eq. (3.2). Here, h^I_s is the s function of H^I. Then the generalized Wick theorem yields

$$S = \left\{Z^0_{s,N}\left[\frac{\delta}{\delta\hat{a}^\nu}\right]\exp\left[-i\int dt h^I_s(\hat{a}^\nu(t),t)\right]\right\}_N , \tag{3.9}$$

where, for simplicity, we have chosen $\mathscr{S}^I$ to represent normal ordering. The contractions in $Z^0_{s,N}$ are

$$\Delta^{\nu\nu'}_{s,N}(t-t') = \langle 0|T_s\{\hat{a}^\nu(t)\hat{a}^{\nu'}(t')\}|0\rangle . \tag{3.10}$$

Just as in the Bose case, Eqs. (3.8) and (3.9) show how the s dependence of the function h^I_s is compensated for by the corresponding s dependence of the equal-time prescription of the time ordering operation in such a way that the S operator is independent of s.

Let us show a simple example with the following

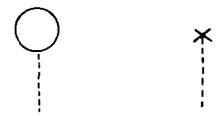


FIG. 1. Tadpole diagrams.

Lagrangian:

$$\mathscr{L} = \bar{\psi}(i\gamma^\mu\partial_\mu - m_0)\psi + \tfrac{1}{2}[\partial_\mu\phi\partial^\mu\phi - \mu_0^2\phi^2] - g\{\bar{\psi},\psi\}_N\phi\,, \tag{3.11}$$

where ψ is a Dirac field and ϕ is a scalar field. Since a normal-ordered form of interaction is assumed, there is no contribution from tadpole diagrams. Let us evaluate tadpole diagrams in the s-ordering scheme. Since

$$h_N^I(\psi,\phi) = g\bar{\psi}\psi\phi\,, \tag{3.12}$$

we obtain [using Eq. (A6)]

$$h_{\mathscr{S}}^I(\psi,\phi) = g\bar{\psi}\psi\phi + \frac{(1-s)}{2}\,\mathrm{Tr}[S_{\alpha\beta}^{+}(0) - S_{\alpha\beta}^{-}(0)]\phi\,, \tag{3.13}$$

where, for simplicity, we have used the same s for every degree of freedom. We are assuming some regularization method [e.g., Eq. (A9)]. Since by definition

$$\langle 0|T_{\mathscr{S}}(\bar{\psi}(x)\psi(x))|0\rangle = \frac{(s-1)}{2}\,\mathrm{Tr}[S_{\alpha\beta}^{+}(0) - S_{\alpha\beta}^{-}(0)]\,, \tag{3.14}$$

we easily find that both contributions from Figs. 1(a) and 1(b) cancel each other as expected.

IV. FUNCTIONAL INTEGRAL

In this section we present the functional integral expressions for the general ordering scheme. First let us follow the recent work by Ohnuki and Kashiwa[6]. Coherent states are introduced as follows:

$$\hat{a}_j|(\xi)_n\rangle = \xi_j|(\xi)_n\rangle, \tag{4.1}$$

and

$$\langle(\xi)_n|\hat{a}_j = \langle(\xi)_n|\xi_j\,, \tag{4.2}$$

where

$$|(\xi)_n\rangle \equiv \exp\left[-\sum_{j=1}^{n}\xi_j\hat{a}_j^{+}\right]|0\rangle,$$

and

$$\langle(\xi)_n| \equiv \langle 0|\delta(\xi_1,\hat{a}_1)\cdots\delta(\xi_n,\hat{a}_n)\,, \tag{4.4}$$

with

$$\delta(\xi_j,\hat{a}_j) = \frac{1}{i}(\xi_j - \hat{a}_j)\,. \tag{4.5}$$

Here, ξ_j are the Grassmann number. Likewise,

$$|(\xi)_n^*\rangle \equiv \delta(\hat{a}_n^{+},\xi_n^*)\cdots\delta(\hat{a}_1^{+},\xi_1^*)|0\rangle, \tag{4.6}$$

and

$$\langle(\xi)_n^*| \equiv \langle 0|\exp\left(-\sum_{j=1}^{n}\hat{a}_j\xi_j^*\right), \tag{4.7}$$

turn out to be coherent states for $\hat{a}_j^{+}$ with the Grassmann number eigenvalue ξ_j^*. The completeness conditions are given, e.g., by

$$\int\int e^{-\sum_j \xi_j\xi_j^*}|(\xi)_n^*\rangle\,\langle(\xi)_n|(d\xi)_n^*(d\xi)_n = 1, \tag{4.8}$$

and

$$\int\int e^{-\sum_j \xi_j^*\xi_j}|(\xi)_n\rangle\,\langle(\xi)_n^*|(d\xi)_n(d\xi)_n^* = 1\,, \tag{4.9}$$

where

$$(d\xi)_n = d\xi_n\cdots d\xi_1\,, \tag{4.10}$$

and

$$(d\xi)_n^* = d\xi_1^*\cdots d\xi_n^*\,. \tag{4.11}$$

For any state vector, one can define the Grassmann representatives by

$$\phi_A((\xi)_n) \equiv \langle(\xi)_n|A\rangle,\ \phi_A((\xi)_n^*) \equiv \langle(\xi)_n^*|A\rangle \tag{4.12}$$

and their adjoints

$$\tilde{\phi}_A((\xi)_n^*) \equiv \langle A|(\xi)_n^*\rangle,\ \tilde{\phi}_A((\xi)_n) \equiv \langle A|(\xi)_n\rangle\,. \tag{4.13}$$

For the vacuum, one finds

$$\phi_0((\xi)_n^*) = \tilde{\phi}_0((\xi)_n) = 1\,. \tag{4.14}$$

As shown in Ref. 5, let us define a generating functional:

$$W_{\mathscr{S}}[\eta] = \left\langle T_{\mathscr{S}}\left\{\exp\left[i\int_{t_I}^{t_F}dt\,(\tilde{a}^{+}(t)\eta^*(t) - \tilde{a}(t)\eta(t))\right]\right\}\right\rangle_{ba}, \tag{4.15}$$

where $\eta(t)$ and $\eta^*(t)$ are external Grassmann number source functions and $\tilde{a}(t)$ and $\tilde{a}^{+}(t)$ are N-component operators in the Heisenberg representation. Here $\mathscr{S}$ means an N-component set $\{s_j\}$. The time-ordering operation $T_{\mathscr{S}}$ for the case of equal times is defined, in this case, by

$$T_{\mathscr{S}}\{A(\tilde{a}(t),\tilde{a}^{+}(t),t)\} \equiv \{A(\tilde{a}(t),\tilde{a}^{+}(t),t)\}_{\mathscr{S}}\,, \tag{4.16}$$

where

$$\{A(\tilde{a}(t),\tilde{a}^{+}(t),t)\}_{\mathscr{S}} \equiv U_{tt_I}^{-1}\{A(\hat{a},\hat{a}^{+},t)\}_{\mathscr{S}}U_{tt_I}\,, \tag{4.17}$$

and

$$U_{tt_I} \equiv \exp[-i(t-t_I)H]\,. \tag{4.18}$$

The bracket $\langle\cdot\rangle_{ba}$ means

$$\langle\cdot\rangle_{ba} = \langle B|U_{t_F t_I}(\cdot)|A\rangle\,. \tag{4.19}$$

It should be noted that, since $W_{\mathscr{S}}[\eta]$ is defined by using the $T_{\mathscr{S}}$ product, all operators, even in the products at equal times, behave as if they were the Grassmann numbers in $W_{\mathscr{S}}[\eta]$. This is essential when one tries to get the functional integral expression of a quantum theory, because the functional integral approach is a Grassmann-number method for the Fermi systems just as it is a c-number method for the Bose systems.

The lattice approximation to Eq. (4.15) is given by

$$W_{\mathscr{S}}^{(N)}[\eta] = \left\langle T_{\mathscr{S}}\left[\exp(i\sum_{k=1}^{N}\Delta_t\tilde{a}_k^{\alpha}\eta_k^{\alpha})\right]\right\rangle_{ba}, \tag{4.20}$$

where $\eta_k^{\alpha} = (\eta^*(t_k), -\eta(t_k))$, $\tilde{a}_k^{\alpha} = (\tilde{a}^{+}(t_k), \tilde{a}(t_k))$, $t_k = t_I + k\Delta_t$, and $\Delta_t = (t_F - t_I)/N$.

For an arbitrary function $f(\tilde{a})$ we have

$$f\left(\frac{-\partial}{i\Delta_t \partial\eta_k}\right) W_{\mathscr{S}}^{(N)}[\eta] = \langle \{e^{i\Delta_t \hat{a}_N^{\alpha} \eta_N^{\alpha}}\}_{\mathscr{S}} \cdots \{f(\tilde{a}_k) e^{i\Delta_t \hat{a}_k^{\alpha} \eta_k^{\alpha}}\}_{\mathscr{S}} \cdots \{e^{i\Delta_t \hat{a}_1^{\alpha} \eta_1^{\alpha}}\}_{\mathscr{S}} \rangle_{ba} , \tag{4.21}$$

which reduces to

$$\begin{aligned}
&\langle B | \{e^{i\Delta_t \hat{a}^{\alpha} \eta_N^{\alpha}}\}_{\mathscr{S}} e^{-\Delta_t H} \cdots \{f(\hat{a}) e^{i\Delta_t \hat{a}^{\alpha} \eta_k^{\alpha}}\}_{\mathscr{S}} e^{-i\Delta_t H} \cdots \{e^{i\Delta_t \hat{a}^{\alpha} \eta_1^{\alpha}}\}_{\mathscr{S}} e^{-i\Delta_t H} | A \rangle \\
&\quad = \int\int \tilde{\phi}_B((\xi_N)_n) e^{-\sum_{k=0}^{N} \sum_{j=1}^{n} \xi_{k,j}^* \xi_{k,j}} \langle (\xi_N)_n^* | \{e^{i\Delta_t \hat{a}^{\alpha} \eta_N^{\alpha}}\}_{\mathscr{S}} e^{-i\Delta_t H} | (\xi_{N-1})_n \rangle \cdots (\xi_k)_n^* | \{f(\hat{a}) e^{i\Delta_t \hat{a}^{\alpha} \eta_k^{\alpha}}\}_{\mathscr{S}} \\
&\qquad \times e^{-i\Delta_t H} | (\xi_{k-1})_n \rangle \cdots \langle (\xi_1)_n^* | \{e^{i\Delta_t \hat{a}^{\alpha} \eta_1^{\alpha}}\}_{\mathscr{S}} e^{-i\Delta_t H} | (\xi_0)_n \rangle \phi_A((\xi_0)_n^*) \prod_{l=0}^{N} (d\xi_l)_n (d\xi_l)_n^* ,
\end{aligned} \tag{4.22}$$

where we have used the completeness relation (4.9).

Now, from Eqs. (4.3), (4.7), (2.5), and (2.11), we observe that the following relations hold:

$$\langle (\xi_k)_n^* | G(\hat{a}^+, \hat{a}) | (\xi_{k-1})_n \rangle = \exp\left(\sum_{j=1}^{n} \xi_{k,j}^* \xi_{k-1,j}\right) g_N(\xi_k^*, \xi_{k-1}) \tag{4.23}$$

$$= \exp\left(\sum_{j=1}^{n} \xi_{k,j}^* \xi_{k-1,j}\right) \exp\left(\sum_{j=1}^{n} \frac{s_j - 1}{2} \frac{\partial^2}{\partial\xi_{k-1,j} \partial\xi_{k,j}^*}\right) g_{\mathscr{S}}(\xi_k^*, \xi_{k-1}) , \tag{4.24}$$

where g_N and $g_{\mathscr{S}}$ are N and $\mathscr{S}$ functions of G, respectively. From Eq. (2.10) we also get

$$\langle (\xi_k)_n^* | \{G(\hat{a}^+, \hat{a})\}_{\mathscr{S}} | (\xi_{k-1})_n \rangle = \exp\left(\sum_{j=1}^{n} \xi_{k,j}^* \xi_{k-1,j}\right) \exp\left[\sum_{j=1}^{n} \frac{s_j - 1}{2} \frac{\partial^2}{\partial\xi_{k-1,j} \partial\xi_{k,j}^*}\right] G(\xi_k^*, \xi_{k-1}) . \tag{4.25}$$

When use is made of the partial integration,[7] we obtain

$$\begin{aligned}
&\int\int e^{-\sum_{j=1}^{n} \xi_{k,j}^* (\xi_{k,j} - \xi_{k-1,j})} \exp\left(\sum_{j=1}^{n} \frac{s_j - 1}{2} \frac{\partial^2}{\partial\xi_{n-1,j} \partial\xi_{k,j}^*}\right) G(\xi_k^*, \xi_{k-1}) (d\xi_k)_n^* (d\xi_{k-1})_n \\
&\quad = \int\int e^{-\sum_{j=1}^{n} \xi_{k,j}^* (\xi_{k,j} - \xi_{k-1,j})} \exp\left[-\sum_{j=1}^{n} \frac{s_j - 1}{2} (\xi_{k,j} - \xi_{k-1,j}) \frac{\partial}{\partial\xi_{k-1,j}}\right] G(\xi_k^*, \xi_{k-1}) (d\xi_k)_n^* (d\xi_{k-1})_n \\
&\quad = \int\int e^{-\sum_{j=1}^{n} \xi_{k,j}^* (\xi_{k,j} - \xi_{k-1,j})} G(\xi_k^*, \xi_{k-1}^{\mathscr{S}}) (d\xi_k)_n^* (d\xi_{k-1})_n ,
\end{aligned} \tag{4.26}$$

where

$$\xi_{k-1,j}^{\mathscr{S}} \equiv \frac{(1+s_j)}{2} \xi_{k-1,j} + \frac{(1-s_j)}{2} \xi_{k,j} . \tag{4.27}$$

Consequently,

$$f\left(-\frac{\partial}{i\Delta_t \partial\eta_k}\right) W_{\mathscr{S}}^{(N)}[\eta] = \int\int \tilde{\phi}_B((\xi_N)_n) f(\xi_k^*, \xi_{k-1}^{\mathscr{S}}) \exp\left(\sum_{l=0}^{N} i\Delta_t \mathscr{L}_l\right) \phi_A((\xi_0)_n^*) \prod_{l=0}^{N} (d\xi_l)_n (d\xi_l)_n^* , \tag{4.28}$$

where

$$\begin{aligned}
&\mathscr{L}_l = i \sum_{j=1}^{n} \xi_{l,j}^* (\xi_{l,j} - \xi_{l-1,j})/\Delta_t - h_{\mathscr{S}}(\xi_l^*, \xi_{l-1}^{\mathscr{S}}) + \sum_{j=1}^{n} (\xi_{l,j}^* \eta_{l,j}^* - \xi_{l-1,j}^{\mathscr{S}} \eta_{l,j}), \quad \text{for } l = 1 \sim N , \\
&\mathscr{L}_0 = i \sum_{j=1}^{n} \xi_{0,j}^* \xi_{0,j} \Big/ \Delta t .
\end{aligned} \tag{4.29}$$

Here we have assumed that $h_{\mathscr{S}}$ commutes with all Grassmann numbers. The result (4.28) may be interpreted as a generalization of the expression obtained in Ref. 6 for normal ordering. Note that, for general s ordering, the s function $h_{\mathscr{S}}$ of the Hamiltonian is used in addition to the "discretization" rule expressed by $(\xi_{k-1}^{\mathscr{S}})$. This is analogous to the s-ordering case for Bose systems[5].

As a simple example, let us take the following model:

$$H = m(\hat{a}^+ \hat{a} + \hat{b}^+ \hat{b}) + g\hat{a}^+ \hat{b}^+ \hat{b}\hat{a} . \tag{4.30}$$

Then

$$\begin{aligned}
h_{\mathscr{S}=(s_1,s_2)}(\xi^*, \xi) = {} & m(\xi^{(1)*} \xi^{(1)} + \xi^{(2)*} \xi^{(2)}) + \frac{m}{2}(2 - s_1 - s_2) + g\Big[\xi^{(1)*} \xi^{(2)*} \xi^{(2)} \xi^{(1)} + \frac{(1-s_1)}{2} \xi^{(2)*} \xi^{(2)} \\
& + \frac{(1-s_2)}{2} \xi^{(1)*} \xi^{(1)} + \frac{(1-s_1)(1-s_2)}{4}\Big] ,
\end{aligned} \tag{4.31}$$

where $\xi^{(1)}$ and $\xi^{(2)}$ correspond to the operators $\hat{a}$ and $\hat{b}$, respectively. Therefore, the generating functional between the vacuum states is given as follows:

$$W_{\mathscr{F}}[\eta] = \lim_{N\to\infty} \int\int \exp\left[i \sum_{l=0}^{N} \Delta_t \mathscr{L}_l \right] \prod_{l=0}^{N} (d\xi_l^{(2)} d\xi_l^{(1)} d\xi_l^{(1)*} d\xi_l^{(2)*}), \tag{4.32}$$

where

$$\mathscr{L}_l = i \sum_{j=1,2} \xi_l^{(j)*} (\xi_l^{(j)} - \xi_{l-1}^{(j)})/\Delta_t - h_{\mathscr{F}}(\xi_l^{(1)*}, \xi_{l-1}^{s_1,(1)}, \xi_l^{(2)*}, \xi_{l-1}^{s_2,(2)}) + \sum_{j=1,2} (\xi_l^{(j)*}\eta_l^{(j)*} - \xi_{l-1}^{s_j,(j)}\eta_l^{(j)}), \quad \text{for } l = 1\sim N,$$

$$\mathscr{L}_0 = i \sum_{j=1,2} \xi_0^{(j)*} \xi_0^{(j)} / \Delta t. \tag{4.33}$$

Here

$$\xi_{l-1}^{s_i(i)} \equiv \frac{(1+s_i)}{2} \xi_{l-1}^{(i)} + \frac{(1-s_i)}{2} \xi_l^{(i)}. \tag{4.34}$$

V. SUMMARY

We have studied operator ordering for Fermi systems. Once use is made of the Grassmann numbers, the procedure is completely analogous to that for Bose systems. The general ordering for Fermi systems is restricted to the s ordering only. The generalized Wick theorem and the Dyson–Wick expansions for the S matrix have been given analogously. The S matrix is independent of the ordering chosen. Using coherent states for Fermi operators, we have obtained the functional integral expression of the generating functional for the s ordering, which is again similar to the coherent state expression for the Bose system. As far as the functional integral expressions are concerned, two features are different from the naively expected form. One must use the s function $h_{\mathscr{F}}$ in place of the classical Hamiltonian. In addition, one must take the discretization prescription corresponding to the s ordering.

ACKNOWLEDGMENTS

The author would like to thank Professor Abdus Salam, the International Atomic Energy Agency, and UNESCO for hospitality at the International Center for Theoretical Physics, Trieste, and Professor G. Furlan for hospitality at the Instituto di Fisica Teorica dell'Università di Trieste. He is also grateful to the National Research Council of Italy (CNR) for financial support.

APPENDIX A

Since it is useful for application, we shall give the ordering scheme for a free Dirac theory. The Dirac field is expanded as

$$\psi_\alpha(x) = \psi_\alpha^{(+)}(x) + \psi_\alpha^{(-)}(x),$$

$$\psi_\alpha^{(+)}(x) = (2\pi)^{-3/2} \int d^3p \sqrt{\frac{m}{E_p}} \times \sum_{\sigma=\pm} \hat{a}_\sigma(p) u_\sigma^\alpha(p) e^{-ipx},$$

$$\psi_\alpha^{(-)}(x) = (2\pi)^{-3/2} \int d^3p \sqrt{\frac{m}{E_p}} \times \sum_{\sigma=\pm} \hat{b}_\sigma^+(p) v_\sigma^\alpha(p) e^{ipx}. \tag{A1}$$

If we assume (s_σ^1, s_σ^2) ordering for $\hat{a}_\sigma$ and $\hat{b}_\sigma$, the generalized Weyl operator becomes

$$D_{\mathscr{F}}(\xi,\eta) = F_{\mathscr{F}}(\xi,\eta) \exp\left\{ \int d^3x \times \sum_\alpha [\xi_\alpha^1(x)\psi_\alpha^{(+)}(x) + \xi_\alpha^2(x)\bar{\psi}_\alpha^{(+)}(x) + \eta_\alpha^1(x) \times \bar{\psi}_\alpha^{(-)}(x) + \eta_\alpha^2(x)\psi_\alpha^{(-)}(x)] \right\}, \tag{A2}$$

with

$$F_{\mathscr{F}}(\xi,\eta) = \exp\left[\sum_{\alpha,\beta,i=1,2} \frac{1}{2} \times \int I_{s,\alpha\beta}^i(x-y) \xi_\alpha^i(x) \eta_\beta^i(y) d^3x d^3y \right], \tag{A3}$$

where operators in Eqs. (A2) are in the Schrödinger representation. Here, $I_{s,\alpha\beta}^i$ are defined as follows:

$$I_{s,\alpha\beta}^1(x-y) = \frac{m}{(2\pi)^3} \int \frac{d^3p}{E_p} \sum_\sigma s_\sigma^1 u_\sigma^\alpha(p) \bar{u}_\sigma^\beta(p) e^{-ip(x-y)}, \tag{A4}$$

$$I_{s,\alpha\beta}^2(x-y) = \frac{m}{(2\pi)^3} \int \frac{d^3p}{E_p} \sum_\sigma s_\sigma^2 \bar{v}_\sigma^\alpha(p) v_\sigma^\beta(p) e^{-ip(x-y)}. \tag{A5}$$

For example,

$$\{\psi_\alpha^{(+)}(x)\bar{\psi}_\beta^{(-)}(y)\}_{\mathscr{F}} = \psi_\alpha^{(+)}(x)\bar{\psi}_\beta^{(-)}(y) - \frac{m}{2(2\pi)^3} \times \int \frac{d^3p}{E_p} \sum_\sigma (s_\sigma^1 + 1) u_\sigma^\alpha(p) \bar{u}_\sigma^\beta(p) e^{-ip(x-y)}. \tag{A6}$$

In the limit of equal space–time, the extra terms become divergent. If we assume $s_\sigma^i = s^i$, there are two types of divergences. In dimensional regularization[3] they are replaced by

$$\int \frac{d^3p}{(2\pi)^3} \to \int \frac{d^{n-1}p}{(2\pi)^{n-1}} = 0, \tag{A7}$$

$$\int \frac{d^3p}{(2\pi)^3 E_p} \to \int \frac{d^{n-1}p}{(2\pi)^{n-1} E_p} = \frac{2m^{n-2}}{(4\pi)^{n/2}} \Gamma\left(1 - \frac{n}{2}\right). \tag{A8}$$

For example,

$$\mathrm{Tr}(S_{\alpha\beta}^{(+)}(0) - S_{\alpha\beta}^{(-)}(0))$$

$$= m \int \frac{d^3p}{(2\pi)^3 E_p} \to \frac{2m^{n-1}}{(4\pi)^{n/2}} \Gamma\left(1 - \frac{n}{2}\right), \tag{A9}$$

where $S_{\alpha\beta}^{+(-)}$ is a usual positive (negative) energy part of the free Green function.

APPENDIX B

Let us prove the generalized Wick theorem (3.2). First we note that for different times the usual Wick theorem is expressed as

$$T\{G(\hat{a}^\nu(t))\} = \left\{ Z_N^0 \left[\frac{\delta}{\delta \hat{a}^\nu(t)} \right] G(\hat{a}^\nu(t)) \right\}_N , \tag{B1}$$

where Z_N^0 is given by Eq. (3.5) with the definition of the contraction

$$\begin{aligned} \Delta_N^{\nu\nu'}(t-t') &\equiv T\{\hat{a}^\nu(t)\hat{a}^{\nu'}(t')\} - \{\hat{a}^\nu(t)\hat{a}^{\nu'}(t')\}_N \\ &= \langle 0|T\{\hat{a}^\nu(t)\hat{a}^{\nu'}(t')\}|0\rangle . \end{aligned} \tag{B2}$$

If we combine Eq. (B1) with the formula (2.10) for equal times, we obtain the generalized Wick theorem for the case $\mathscr{S}' = N$. Now, since in the interaction representation, in general,

$$\hat{a}_i(t) = \hat{a}_i e^{-i\omega_i t}, \quad \hat{a}_i^+(t) = \hat{a}_i^+ e^{i\omega_i t}, \tag{B3}$$

We can generalize the formula (2.10) to the case of unequal times as follows:

$$\{G(\hat{a}^\nu(t))\}_{\mathscr{S}} = \left\{ \tilde{Z}^0_{\mathscr{S}\mathscr{S}'} \left[\frac{\delta}{\delta \hat{a}^\nu(t)} \right] G(\hat{a}^\nu(t)) \right\}_{\mathscr{S}'} , \tag{B4}$$

where $\tilde{Z}^0_{\mathscr{S}\mathscr{S}'}$ is given by Eq. (3.5) with the following replacement of the definition of the contraction:

$$\begin{aligned} \Delta^{\nu\nu'}_{\mathscr{S}\mathscr{S}'}(t-t') \to &\{\hat{a}^\nu(t)\hat{a}^{\nu'}(t')\}_{\mathscr{S}} \\ &- \{\hat{a}^\nu(t)\hat{a}^{\nu'}(t')\}_{\mathscr{S}'} . \end{aligned} \tag{B5}$$

Combination of the theorem (3.2) for $\mathscr{S}' = N$ and Eq. (B4) leads to the generalized Wick theorem for general $\mathscr{S}'$.

[1] H. Leschke, A.C. Hirshfeld, and T. Suzuki, Phys. Lett. A **67**, 87 (1978).
[2] H. Leschke, A.C. Hirshfeld, and T. Suzuki, Phys. Rev. D **18**, 2834 (1978), and references therein.
[3] T. Suzuki, A.C. Hirshfeld, and H. Leschke, to appear in Prog. Theor. Phys. **63**, (1980), No. 1.
[4] H. Leschke and M. Schmutz, Z. Phys. B **27**, 85 (1977).
[5] T. Suzuki, "Functional integrals and general correspondence rules", Kanazawa Univ. preprint DPKU-7905.
[6] Y. Ohnuki and T. Kashiwa, Prog. Theor. Phys. **60**, 548 (1978).
[7] F.A. Berezin, *The Method of Second Quantization* (Academic, New York, 1966).
[8] G.S. Agarwal and E. Wolf, Phys. Rev. D **2**, 2161, 2187, 2206 (1970).

A recursion formula for sp(3,R) matrix elements

G. Rosensteel [a)]
Department of Physics, Tulane University, New Orleans, Louisiana 70118

(Received 11 September 1979; accepted for publication 26 November 1979)

A recursion formula is derived for the reduced matrix elements of the sp(3,R) discrete series with respect to a basis symmetry—adapted to its u(3) subalgebra. These matrix elements are needed in the symplectic nuclear model of collective motion.

1. DEFINITION OF sp(3,R)

The noncompact symplectic algebra has been identified as the dynamical algebra describing monopole and quadrupole collective excitations in the nuclear harmonic oscillator shell model.[1-8] The physically relevant discrete series irreducible representations of sp(n,R) were constructed by Godement[9] and their reduction into irreps of its maximally compact subgroup U(n) was determined by the author.[10] Although required for applications, the sp(n,R) matrix elements are known only for the cases of one- ($n = 1$)[1] and two- ($n = 2$)[2,8] dimensional collective motion in which use was made of the low dimensional Cartan isomorphisms $C_1 \simeq A_1$, $C_2 \simeq B_2$. In this paper, a recursion formula is derived for the sp(3,R) matrix elements with respect to a basis symmetry-adapted to the u(3) subalgebra.

A basis for the complexification C_3 of sp(3,R) is given by the one-body bilinear products in the harmonic oscillator bosons,

$$\begin{aligned} A_{ij} &= \tfrac{1}{2}\sum_{\alpha=1}^{A} c_{\alpha i}^{+} c_{\alpha j}^{+} , \\ B_{ij} &= \tfrac{1}{2}\sum_{\alpha=1}^{A} c_{\alpha i} c_{\alpha j} , \\ C_{ij} &= \tfrac{1}{2}\sum_{\alpha=1}^{A} (c_{\alpha i}^{+} c_{\alpha j} + c_{\alpha j} c_{\alpha i}^{+}) , \end{aligned} \tag{1}$$

where $c_{\alpha j}$ denotes the oscillator boson for the α particle in the A-particle space. This basis is especially useful because of its relation with the unitary subalgebra.

The u(3) subalgebra is spanned by the number-conserving C_{ij}. Thus the number operator is $H = \Sigma_{i=1}^{3} C_{ii}$, and the su(3) subalgebra is spanned by the eight independent traceless operators $C_{ij} - (1/3)\delta_{ij} H_0$. The commutation relations of the C_{ij} with themselves and with the A_{ij} and B_{ij} are particularly simple.

Firstly, the traceless operators may be viewed as an su(3) irreducible tensor operator, via the adjoint representation, transforming according to the $(\lambda\mu) = (11)$ su(3) irrep. The tensor operator $C^{(11)}_{lm}$ contains angular momentum $l = 2$ and $l = 1$ components normalized by $C^{(11)}_{l=1,m=0} = L_0$ and $C^{(11)}_{l=2,m=0} = 3^{-1/2}(2C_{33} - C_{11} - C_{22})$. In addition, the operators A_{ij} and B_{ij} are also su(3) tensor operators transforming according to the (20) and (02) irreps, respectively. Both $A^{(20)}_{lm}$ and $B^{(02)}_{lm}$ contain $l = 2$ and $l = 0$ components normalized by

$$\begin{aligned} A^{(20)}_{l=0,m=0} &= (\tfrac{2}{3})^{1/2}(A_{11} + A_{22} + A_{33}), \\ A^{(20)}_{l=2,m=0} &= 3^{-1/2}(2A_{33} - A_{11} - A_{22}) , \end{aligned} \tag{2a}$$

$$\begin{aligned} B^{(02)}_{l=0,m=0} &= (\tfrac{2}{3})^{1/2}(B_{11} + B_{22} + B_{33}), \\ B^{(02)}_{l=2,m=0} &= 3^{-1/2}(2B_{33} - B_{11} - B_{22}) . \end{aligned} \tag{2b}$$

Moreover, $A^{(20)}$ and $B^{(02)}$ are $2\hbar\omega$ raising and lowering operators,

$$[\hbar\omega H_0, A^{(20)}_{lm}] = 2\hbar\omega A^{(20)}_{lm} , \tag{3a}$$

$$[\hbar\omega H_0, B^{(02)}_{lm}] = -2\hbar\omega B^{(02)}_{lm} , \tag{3b}$$

and $B^{(02)}$ is the adjoint of $A^{(20)}$,

$$B^{(02)}_{lm} = (-1)^{l-m}(A^{(20)}_{l-m})^{\dagger} . \tag{4}$$

The remaining defining commutation relations are given by

$$[A^{(20)}_{l_1 m_1}, A^{(20)}_{l_2 m_2}] = 0, \tag{5}$$

and

$$\begin{aligned} [B^{(02)}_{l_1 m_1}, A^{(20)}_{l_2 m_2}] = {} & \tfrac{1}{2}\sqrt{10}\sum_{lm} ((02) l_1 m_1; (20) l_2 m_2 | (11) lm) C^{(11)}_{lm} \\ & + 2(\tfrac{2}{3})^{1/2} ((02) l_1 m_1; (20) l_2 m_2 | (00)) H_0 , \end{aligned} \tag{6}$$

where the coefficients are su(3) Clebsch–Gordons.

Suppose now that $|iN(\lambda\mu)\alpha\rangle$ are orthonormal basis states for a sp(3,R) irrep, which are symmetry-adapted to the u(3) subalgebra. Thus, N is the eigenvalue of H_0 and $(\lambda\mu)$ fixes the su(3) content. The index i is used, when necessary, to distinguish multiply-occurring u(3) irreps; states with different i are orthogonal. The quantum numbers $\alpha \equiv KLM$ range over the basis states corresponding to the $(\lambda\mu)$ irrep. The K quantum number is similar to that defined by Vergados.[11] Thus, states with different K quantum numbers are orthogonal and the range of K is the same as the range of the Elliott K quantum number. The LM denote the so(3) $\supset$ so(2) content which, if $\lambda \geqslant \mu$, are projected from the su(3) highest weight state and, if $\lambda < \mu$, from the su(3) lowest weight state.

The matrix elements of the $A^{(20)}$, $B^{(02)}$, and $C^{(11)}$ operators may be calculated from their reduced matrix elements via the Wigner–Eckart theorem using the Wigner coefficients of su(3). These coefficients are readily available due to the work of Draayer and Akiyama.[12] With these authors' phase conventions and the convention for the so(3) projection, one finds

$$\begin{aligned} &\langle iN(\lambda\mu)\alpha' | C^{(11)}_{\beta} | iN(\lambda\mu)\alpha\rangle \\ &\quad = ((\lambda\mu)\alpha; (11)\beta | (\lambda\mu)\alpha')_{\rho=1} \langle (\lambda\mu) \| C^{(11)} \| (\lambda\mu)\rangle , \end{aligned} \tag{7}$$

where

 0022-2488/80/040924-04$1.00

$$\langle(\lambda\mu)\|C^{(11)}\|(\lambda\mu)\rangle = \begin{cases} g_{\lambda\mu}, & \mu = 0 \\ -g_{\lambda\mu}, & \mu \neq 0 \end{cases},$$

$$g_{\lambda\mu} = 2\left(\frac{\lambda^2+\mu^2+\lambda\mu+3\lambda+3\mu}{3}\right)^{1/2}, \tag{8}$$

and $\alpha' = K'L'M'$, $\alpha = KLM$, and $\beta = lm$.

Similarly,

$$\langle i'N'(\lambda'\mu')\alpha'|A^{(20)}_{\beta}|iN(\lambda\mu)\alpha\rangle$$
$$= ((\lambda\mu)\alpha;(20)\beta|(\lambda'\mu')\alpha')\langle i'N'(\lambda'\mu')\|A^{(20)}\|iN(\lambda\mu)\rangle\,, \tag{9}$$

$$\langle i'N'(\lambda'\mu')\alpha'|B^{(02)}_{\beta}|iN(\lambda\mu)\alpha\rangle$$
$$= ((\lambda\mu)\alpha;(02)\beta|(\lambda'\mu')\alpha')\langle i'N'(\lambda'\mu')\|B^{(02)}\|iN(\lambda\mu)\rangle\,. \tag{10}$$

But, since $B^{(02)}$ is the adjoint of $A^{(20)}$,

$$\langle i'N'(\lambda'\mu')\|B\|iN(\lambda\mu)\rangle$$
$$= (-1)^{\lambda'+\mu'+\lambda+\mu}\left(\frac{\dim(\lambda\mu)}{\dim(\lambda'\mu')}\right)^{1/2}$$
$$\times\langle iN(\lambda\mu)\|A^{(20)}\|iN'(\lambda'\mu')\rangle^{*}\,. \tag{11}$$

Note that from the commutation relations (3),

$$\langle i'N'(\lambda'\mu')\|A\|iN(\lambda\mu)\rangle = 0, \quad \text{if } N' \neq N+2\,. \tag{12}$$

Therefore, it has been found, due to the u(3) symmetry adaptation of both the basis states and the sp(3,R) operators, that the only undetermined matrix elements are the u(3) reduced matrix elements of $A^{(20)}$. Moreover, these matrix elements must be solutions to the commutation relations, Eqs. (5 and 6).

2. Sp(3,R) DISCRETE SERIES

Before deriving a recursion relation for the $A^{(20)}$ reduced matrix elements, it is necessary to review some basic properties of the sp(3,R) discrete series.[4,9,10]

For every irrep of u(3), $N_0(\lambda_0\mu_0)$, there is associated an irreducible unitary representation of sp(3,R). The representation space for the $N_0(\lambda_0\mu_0)$ symplectic irrep is given by

$$\mathscr{H}_{N_0(\lambda_0\mu_0)} = \bigoplus_{n=0,2,4\cdots} \mathscr{H}^{(N_0+n)}_{N_0(\lambda_0\mu_0)}\,, \tag{13}$$

where $\mathscr{H}^{(N_0+n)}_{N_0(\lambda_0\mu_0)}$ is the eigenspace of H_0 belonging to the eigenvalue N_0+n and $\mathscr{H}^{(N_0)}_{N_0(\lambda_0\mu_0)}$ carries the $N_0(\lambda_0\mu_0)$ irrep of u(3).

The subspace $\mathscr{H}^{(N_0+n)}_{N_0(\lambda_0\mu_0)}$ is generated by the polynomials of degree $n/2$ in the raising operators $A^{(20)}$ acting on $\mathscr{H}^{(N_0)}_{N_0(\lambda_0\mu_0)}$. However, because the raising operators commute, Eq. (5), only symmetric polynomials are nonzero. It is known that the su(3) content $(\hat\lambda\hat\mu)$ of the symmetric coupling of $n/2$ copies of (2,0) tensors is enumerated by the set[13]

$$\Omega^{(n)} = \{(\hat\lambda\hat\mu)|n_1+n_2+n_3 = n,\quad \hat\lambda = n_1-n_2,\quad \hat\mu = n_2-n_3\,,$$
where n_1,n_2,n_3 range over the even nonnegative integers with $n_1 \geqslant n_2 \geqslant n_3\}$. (14)

Hence,

$$\mathscr{H}^{(N_0+n)}_{N_0(\lambda_0\mu_0)} = \bigoplus_{(\hat\lambda\hat\mu)\in\Omega^{(n)}} A^{n(\hat\lambda\hat\mu)}\mathscr{H}^{(N_0)}_{N_0(\lambda_0\mu_0)}\,, \tag{15}$$

where there are $n/2$ copies of $A^{(20)}$ in the symmetric coupling forming $A^{n(\hat\lambda\hat\mu)}$.

TABLE 1. Symmetric couplings and their parents.

n	$\Omega^{(n)}$ $(\hat\lambda,\hat\mu)$	Possible parents $(\hat\lambda',\hat\mu')$
$n=0$	(0,0)	–
$n=2$	(2,0)	–
$n=4$	(4,0)	(2,0)
	(0,2)	(2,0)
$n=6$	(6,0)	(4,0)
	(2,2)	(4,0),(0,2)
	(0,0)	(0,2)
$n=8$	(8,0)	(6,0)
	(4,2)	(6,0),(2,2)
	(0,4)	(2,2)
	(2,0)	(2,2),(0,0)
$n=10$	(10,0)	(8,0)
	(6,2)	(8,0),(4,2)
	(2,4)	(4,2),(0,4)
	(4,0)	(4,2),(2,0)
	(0,2)	(0,4),(2,0)

The tensor operators $A^{n(\hat\lambda\hat\mu)}$ may be computed recursively,

$$A^{n(\hat\lambda\hat\mu)}_{\gamma} = \sum_{\beta,\gamma'}((20)\beta;(\hat\lambda'\hat\mu')\gamma'|(\hat\lambda\hat\mu)\gamma)A^{(20)}_{\beta}A^{n-2(\hat\lambda'\hat\mu')}_{\gamma'}\,, \tag{16}$$

where the "parent" $(\hat\lambda'\hat\mu')$ of $(\hat\lambda\hat\mu)$ is chosen by convention once and for all from the set

$$\Omega^{(n-2)}\cap\{(\hat\lambda-2,\hat\mu),(\hat\lambda,\hat\mu+2),(\hat\lambda+2,\hat\mu-2)\}\,. \tag{17}$$

In the accompanying table, $(\hat\lambda\hat\mu)\in\Omega^{(n)}$ and their possible parents $(\hat\lambda'\hat\mu')$ are listed up to $n=10$. Different choices of parents define different phases for the tensors $A^{n(\hat\lambda\hat\mu)}$, and, as will become apparent, different phases for the matrix elements of $A^{(20)}$.

A basis for the subspace $\mathscr{H}^{(N_0+n)}_{N_0(\lambda_0\mu_0)}$ is given by

$$\Phi[(\hat\lambda\hat\mu)\rho N(\lambda\mu)\alpha]$$
$$\equiv \sum_{\gamma,\alpha_0}((\hat\lambda\hat\mu)\gamma;(\lambda_0\mu_0)\alpha_0|(\lambda\mu)\alpha)_{\rho}A^{n(\hat\lambda\hat\mu)}_{\gamma}|N_0(\lambda_0\mu_0)\alpha_0\rangle\,, \tag{18}$$

where $|N_0(\lambda_0\mu_0)\alpha_0\rangle$ spans $\mathscr{H}^{(N_0)}_{N_0(\lambda_0\mu_0)}$, ρ denotes the multiplicity of $(\lambda\mu)$ in the coupling $(\hat\lambda\hat\mu)\otimes(\lambda_0\mu_0)$, and $N = N_0+n$ is the eigenvalue of H_0.

Two basis states with different u(3) content $N(\lambda\mu)$ are necessarily orthogonal. On the other hand, two such states with identical u(3) symmetry $N(\lambda\mu)$, yet distinct $(\hat\lambda\hat\mu)\rho$ quantum numbers, are not in general orthogonal. Therefore, we diagonalize the metric matrix in order to define the orthonormal basis $|iN(\lambda\mu)\alpha\rangle$. Thus,

$$\langle\Phi[(\hat\lambda_2\hat\mu_2)\rho_2 N(\lambda\mu)\alpha]|\Phi[(\hat\lambda_1\hat\mu_1)\rho_1 N(\lambda\mu)\alpha]\rangle$$
$$= \sum_i R^{N(\lambda\mu)}(i,(\hat\lambda_2\hat\mu_2)\rho_2)\omega_i^{N(\lambda\mu)}R^{N(\lambda\mu)}(i,(\hat\lambda_1\hat\mu_1)\rho_1)\,, \tag{19}$$

where $R^{N(\lambda\mu)}$ is the orthogonal diagonalization matrix and $\omega_i^{N(\lambda\mu)}$ are the eigenvalues of the metric matrix. Note that the metric matrix and, hence, its eigenvalues and eigenvectors forming $R^{N(\lambda\mu)}$ are independent of $\alpha = KLM$. Moreover, as

will become apparent from the recursion relation, the metric matrix is real. Therefore, the orthonormal basis is given by

$$|iN(\lambda\mu)\alpha\rangle = (\omega_i^{N(\lambda\mu)})^{-1/2} \sum_{(\hat{\lambda}\hat{\mu}),\rho} R^{N(\lambda\mu)}(i,(\hat{\lambda}\hat{\mu})\rho)\Phi[(\hat{\lambda}\hat{\mu})\rho N(\lambda\mu)\alpha]. \quad (20)$$

3. RECURSION FORMULAS

We are now in a position to derive a recursion formula for the reduced matrix elements of $A^{(20)}$ at the Nth oscillator level, $\langle iN(\lambda\mu)\|A^{(20)}\|i'N-2(\lambda'\mu')\rangle$, in terms of the data at the $N-2$ oscillator level, viz. $\langle i'N-2(\lambda'\mu')\|A^{(20)}\|i''N-4(\lambda''\mu'')\rangle$, as well as $R^{N-2(\lambda'\mu')}$ and $\omega_{i'}^{N-2(\lambda'\mu')}$. This is achieved in two steps by first deriving a formula for the metric matrix at the Nth level in terms of the data at the $N-2$ level (Formula A), and then determining the reduced matrix elements at the Nth level in terms of the metric matrix at the Nth level in addition to the $N-2$ level data (Formula B).

These two formulas are given as follows:

Formula A:

$$\begin{aligned}
&\langle\Phi[(\hat{\lambda}_2\hat{\mu}_2)\rho_2 N(\lambda\mu)\alpha]|\Phi[(\hat{\lambda}_1\hat{\mu}_1)\rho_1 N(\lambda\mu)\alpha]\rangle = \sum_{\substack{(\lambda'_2\mu'_2)\rho'_2\\(\lambda'_1\mu'_1)\rho'_1}} U((20)(\hat{\lambda}'_2\hat{\mu}'_2)(\lambda\mu)(\lambda_0\mu_0);(\hat{\lambda}_2\hat{\mu}_2)\rho_2(\lambda'_2\mu'_2)\rho'_2)\\
&\times U((20)(\hat{\lambda}'_1\hat{\mu}'_1)(\lambda\mu)(\lambda_0\mu_0);(\hat{\lambda}_1\hat{\mu}_1)\rho_1(\lambda'_1\mu'_1)\rho'_1)\cdot\sum_{i'_2,i'_1}(\omega_{i'_2}^{N-2(\lambda'_2\mu'_2)})^{1/2}R^{N-2(\lambda'_2\mu'_2)}(i'_2,(\hat{\lambda}'_2\hat{\mu}'_2)\rho'_2)\\
&\cdot(\omega_{i'_1}^{N-2(\lambda'_1\mu'_1)})^{1/2}R^{N-2(\lambda'_1\mu'_1)}(i'_1,(\hat{\lambda}'_1\hat{\mu}'_1)\rho'_1)\cdot\Big\{\sum_{i''(\lambda''\mu'')}(-1)^{\lambda''-\mu''+\lambda-\mu}U((20)(\lambda''\mu'')(\lambda\mu)(20);(\lambda'_1\mu'_1)(\lambda'_2\mu'_2))\\
&\cdot\langle i'_2N-2(\lambda'_2\mu'_2)\|A\|i''N-4(\lambda''\mu'')\rangle\langle i'_1N-2(\lambda'_1\mu'_1)\|A\|i''N-4(\lambda''\mu'')\rangle+\delta((\lambda'_2\mu'_2),(\lambda'_1\mu'_1))\delta(i'_2i'_1)\\
&\cdot[(\tfrac{10}{3})^{1/2}U((\lambda'_1\mu'_1)(11)(\lambda\mu)(20);(\lambda'_1\mu'_1)(20))\langle(\lambda'_1\mu'_1)\|C\|(\lambda'_1\mu'_1)\rangle+\tfrac{2}{3}(N-2)]\Big\}, \qquad (21)
\end{aligned}$$

where $(\hat{\lambda}'_k\hat{\mu}'_k)$ is the parent of $(\hat{\lambda}_k\hat{\mu}_k)$, $(\lambda'_k\mu'_k)\rho'_k$ runs over the coupling $(\hat{\lambda}'_k\hat{\mu}'_k)\otimes(\lambda_0\mu_0)$, i'_k runs over the normalized states with u(3) symmetry $N-2(\lambda'_k\mu'_k)$, $i''(\lambda''\mu'')$ runs over all normalized states at the $N-4$ level, and U denotes the U-coefficient or 6-$(\lambda\mu)$ coefficients of su(3).[12]

Formula B:

$$\begin{aligned}
&\langle i_2N(\lambda_2\mu_2)\|A^{(20)}\|i'_1N-2(\lambda'_1\mu'_1)\rangle\\
&=(\omega_{i_2}^{N(\lambda_2\mu_2)})^{-1/2}\sum_{(\hat{\lambda}_2\hat{\mu}_2)\rho_2}R^{N(\lambda_2\mu_2)}(i_2,(\hat{\lambda}_2\hat{\mu}_2)\rho_2)\cdot\sum_{(\lambda'_2\mu'_2)\rho'_2}U((20)(\hat{\lambda}'_2\hat{\mu}'_2)(\lambda_2\mu_2)(\lambda_0\mu_0);(\hat{\lambda}_2\hat{\mu}_2)\rho_2(\lambda'_2\mu'_2)\rho'_2)\\
&\cdot\sum_{i'_2}(\omega_{i'_2}^{N-2(\lambda'_2\mu'_2)})^{1/2}R^{N-2(\lambda'_2\mu'_2)}(i'_2,(\hat{\lambda}'_2\hat{\mu}'_2)\rho'_2)\cdot\Big\{\sum_{i''(\lambda''\mu'')}(-1)^{\lambda''-\mu''+\lambda'_1-\mu'_1}U((20)(\lambda''\mu'')(\lambda_2\mu_2)(20);(\lambda'_2\mu'_2)(\lambda'_1\mu'_1))\\
&\cdot\langle i'_2N-2(\lambda'_2\mu'_2)\|A\|i''N-4(\lambda''\mu'')\rangle\langle i'_1N-2(\lambda'_1\mu'_1)\|A\|i''N-4(\lambda''\mu'')\rangle\\
&+(-1)^{\lambda_2-\mu_2+\lambda'_2+\mu'_2}\delta((\lambda'_2\mu'_2),(\lambda'_1\mu'_1))\delta(i'_2,i'_1)\\
&\cdot[(\tfrac{10}{3})^{1/2}U((\lambda'_1\mu'_1)(11)(\lambda_2\mu_2)(20);(\lambda'_1\mu'_1)(20))\langle(\lambda'_1\mu'_1)\|C\|(\lambda'_1\mu'_1)\rangle+\tfrac{2}{3}(N-2)]\Big\}, \qquad (22)
\end{aligned}$$

where $(\hat{\lambda}_2\hat{\mu}_2)$ runs over the elements of $\Omega^{(N-N_0)}$, which couple with $(\lambda_0\mu_0)$ to form $(\lambda_2\mu_2)$ with nonzero multiplicity ρ_2. $(\hat{\lambda}'_2\hat{\mu}'_2)$, the parent of $(\hat{\lambda}_2\hat{\mu}_2)$, $(\lambda'_2\mu'_2)\rho'_2$, which runs over the couplings $(\hat{\lambda}'_2\hat{\mu}'_2)\otimes(\lambda_0\mu_0)$, i'_2, which runs over the normalized states with u(3) symmetry $N-2(\lambda'_2\mu'_2)$, and $i''(\lambda''\mu'')$, which runs over all normalized states at the $N-4$ level.

In order to prove formula A, first use the definition of the Φ, Eq. (18), the recursion relation for the tensors $A^{n(\hat{\lambda}\hat{\mu})}$, Eq. (16), and the definition of the U-coefficients to obtain

$$\begin{aligned}
&\langle\Phi[(\hat{\lambda}_2\hat{\mu}_2)\rho_2 N(\lambda\mu)\alpha]|\Phi[(\hat{\lambda}_1\hat{\mu}_1)\rho_1 N(\lambda\mu)\alpha]\rangle\\
&=\sum_{\substack{(\lambda'_2\mu'_2)\rho'_2\\(\lambda'_1\mu'_1)\rho'_1}}U((20)(\hat{\lambda}'_2\hat{\mu}'_2)(\lambda\mu)(\lambda_0\mu_0);(\hat{\lambda}_2\hat{\mu}_2)\rho_2(\lambda'_2\mu'_2)\rho'_2)U((20)(\hat{\lambda}'_1\hat{\mu}'_1)(\lambda\mu)(\lambda_0\mu_0);(\hat{\lambda}_1\hat{\mu}_1)\rho_1(\lambda'_1\mu'_1)\rho'_1)\\
&\cdot\sum_{\alpha'_2,\alpha'_1}\sum_{\beta_2,\beta_1}((20)\beta_2;(\lambda'_2\mu'_2)\alpha'_2|(\lambda\mu)\alpha)((20)\beta_1;(\lambda'_1\mu'_1)\alpha'_1|(\lambda\mu)\alpha)\\
&\cdot\langle\Phi[(\hat{\lambda}'_2\hat{\mu}'_2)\rho'_2N-2(\lambda'_2\mu'_2)\alpha'_2]|(A_{\beta_2}^{(20)})^{\dagger}A_{\beta_1}^{(20)}\Phi[(\hat{\lambda}'_1\hat{\mu}'_1)\rho'_1N-2(\lambda'_1\mu'_1)\alpha'_1]\rangle. \qquad (23)
\end{aligned}$$

Secondly, expand the Φ on the right side of (23) in terms of the orthonormal basis via Eq. (20). Then, formula A will be proven if the term in braces in that formula is equal to

$$\sum_{\alpha_2'\alpha_1'}\sum_{\beta_2,\beta_1}((20)\beta_2;(\lambda_2'\mu_2')\alpha_2'|(\lambda\mu)\alpha)((20)\beta_1;(\lambda_1'\mu_1')\alpha_1'|(\lambda\mu)\alpha)$$
$$\cdot\langle(i_2'N-2(\lambda_2'\mu_2')\alpha_2'|(A^{(20)}_{\beta_2})^{\dagger}A^{(20)}_{\beta_1}|i_1'N-2(\lambda_1'\mu_1')\alpha_1'\rangle \quad (24)$$

Next, if we write

$$(A^{(20)}_{\beta_2})^{\dagger}A^{(20)}_{\beta_1}=A^{(20)}_{\beta_1}(A^{(20)}_{\beta_2})^{\dagger}+[(A^{(20)}_{\beta_2})^{\dagger},A^{(20)}_{\beta_1}], \quad (25)$$

then the matrix elements of the first term may be computed by inserting a complete set of states at the $N-4$ level, while the commutator may be evaluated using the adjoint property, Eq. (4), and the commutation relation, Eq. (6). Finally, formula A is obtained if we employ Eq. (9) and reexpress the result using the U-coefficients.

Formula B is derived with a similar strategy. First, express $|i_2N(\lambda_2\mu_2)\alpha_2\rangle$ in terms of $\Phi[(\hat{\lambda}_2\hat{\mu}_2)\rho_2N(\lambda_2\mu_2)\alpha_2]$, Eq. (20), employ the definition of Φ, Eq. (18), and the recursion relation, Eq. (16), and then simplify with the U-coefficients to obtain

$$\langle i_2N(\lambda_2\mu_2)\alpha_2|A^{(20)}_{\beta}|i_1'N-2(\lambda_1'\mu_1')\alpha_1'\rangle$$
$$=(\omega^{N(\lambda_2\mu_2)}_{i_2})^{-1/2}\sum_{(\hat{\lambda}_2\hat{\mu}_2)\rho_2}R^{N(\lambda_2\mu_2)}(i_2,(\hat{\lambda}_2\hat{\mu}_2)\rho_2)$$
$$\sum_{(\lambda_2'\mu_2')\rho_2'}U((20)(\hat{\lambda}_2'\hat{\mu}_2')(\lambda_2\mu_2)(\lambda_0\mu_0);$$
$$(\hat{\lambda}_2\hat{\mu}_2)\rho_2(\lambda_2'\mu_2')\rho_2')$$
$$\sum_{\alpha_2'}\sum_{\beta_2}((20)\beta_2;(\lambda_2'\mu_2')\alpha_2'|(\lambda_2\mu_2)\alpha_2)$$
$$\cdot\langle\Phi[(\hat{\lambda}_2'\hat{\mu}_2')\rho_2'N-2(\lambda_2'\mu_2')\alpha_2']|(A^{(20)}_{\beta_2})^{\dagger}$$
$$\times A^{(20)}_{\beta}|i_1'N-2(\lambda_1'\mu_1')\alpha_1'\rangle. \quad (26)$$

Use Eq. (20) again in order to express Φ on the right side in terms of the orthonormal basis. Formula B will be established if the term in braces in that formula is equal to

$$\sum_{\alpha_2'}\sum_{\beta_2}((20)\beta_2;(\lambda_2'\mu_2')\alpha_2'|(\lambda_2\mu_2)\alpha_2)$$
$$\cdot\langle i_2'N-2(\lambda_2'\mu_2')\alpha_2'|(A^{(20)}_{\beta_2})^{\dagger}A^{(20)}_{\beta}|i_1'N-2(\lambda_1'\mu_1')\alpha_1'\rangle. \quad (27)$$

This is shown by using Eq. (25) as before.

4. DISCUSSION

Since the index i defining the orthonormal state $|iN(\lambda\mu)\alpha\rangle$ has no special significance, it may be desirable to replace it with some other quantum number, which also serves to distinguish states with identical u(3) content. This can be achieved after recursion formulas A and B have been used to compute matrix elements with respect to the indexed basis $|iN(\lambda\mu)\alpha\rangle$ by diagonalizing a u(3) invariant operator, say $[A^{(20)}\times C^{(11)}\times B^{(02)}]^{(00)}$. Its eigenvalues Y serve as the new quantum number and its eigenvectors define the change of basis from $|iN(\lambda\mu)\alpha\rangle$ to $|YN(\lambda\mu)\alpha\rangle$.

It should be noted that a definite choice of phase has been made here. It can be altered by either a different phase convention for the su(3) Wigner coefficients and/or a different choice of parents.

In one special case, the reduced matrix elements can be computed directly. Consider the infinite set of highest weight states $|N_0+n(\lambda_0+n,\mu_0)HW\rangle, n=0,2,4,\cdots$. These states carry an irrep of the A_1 subalgebra spanned by A_{33}, B_{33} and C_{33}. Since the irreps of the A_1 subalgebra are well known, we have immediately

$$\langle N_0+n+2(\lambda_0+n+2,\mu_0)\|A^{(20)}\|N_0+n(\lambda_0+n,\mu_0)\rangle$$
$$=\left(\frac{(N_0+2\lambda_0+\mu_0+3n/2)(2+n)}{3}\right)^{1/2}. \quad (28)$$

ACKNOWLEDGMENT

I would like to thank Dr. R.T. Sharp, Dr. J. Patera, and Dr. D.J. Rowe for stimulating discussions.

[1]S. Goshen and H.J. Lipkin, Ann. Phys. (N.Y.) **6**, 301 (1959).
[2]S. Goshen and H.J. Lipkin, in "Spectroscopic and Group Theoretical Methods in Physics," edited by F. Block (North-Holland, Amsterdam, 1968).
[3]G. Rosensteel, Ph.D. Thesis, U. of Toronto, 1975; G. Rosensteel and D.J. Rowe, *Phys. Rev. Lett.* **38**, 10 (1977); "The Sp(3,R) Model of Nuclear Collective Motion," in "Proceedings of the V International Colloquium on Group Theoretical Methods in Physics," edited by R.T. Sharp and B. Kolman (Academic, New York 1977).
[4]G. Rosensteel and D.J. Rowe, "On the Algebraic Formulation of Collective Models III. The Symplectic Shell Model of Collective Motion," Ann. Phys. (N.Y.) (1980), preprint.
[5]F. Arickx, Nucl. Phys. A **284**, 264 (1977); Nucl. Phys. A **268**, 347 (1976).
[6]F. Arickx, J. Broeckhove, and E. Deumens, Nucl. Phys. A **318**, 269 (1979).
[7]K.T. Hecht and D. Braunschweig, Nucl. Phys. A **295**, 34 (1978).
[8]D. Peterson, Ph.D. thesis, U. of Michigan, Ann Arbor, 1979.
[9]R. Godement, *Seminaire Cartan* (Ecole Normale Superieure, Paris, 1958).
[10]G. Rosensteel and D. J. Rowe, Int. J. Theor. Phys. **16**, 63 (1977).
[11]J.D. Vergados, Nucl. Phys. A **111**, 681 (1968).
[12]J.P. Draayer and Yoshimi Akiyama, J. Math. Phys. **14**, 1904 (1973); Comput. Phys. Commun. **5**, 405 (1973).
[13]R.M. Thrall, Am. J. Math. **64**, 371 (1942).

Transverse instability of breathers in resonant media

Yuji Kodama and Mark J. Ablowitz
Department of Mathematics, Clarkson College, Potsdam, New York 13676

(Received 19 July 1979; accepted for publication 26 November 1979)

It is shown analytically that the breather (0π pulse) of transparent optical pulse propagation in a resonant medium is unstable with respect to long transverse perturbations.

1. INTRODUCTION

In a previous paper,[1] we have shown that the 2π pulse of self-induced transparency (SIT) is unstable to transverse perturbations. Other work in this direction has also been done in Refs. 2 and 3. In such a transparent medium, there are mainly two kinds of one-dimensional ultrashort optical pulses, solitons, (i.e., the "0π pulse" and "2π pulse") which propagate coherently without attenuation.[4] We discuss in this paper the transverse instability of 0π pulses (breathers) and show that 0π pulse is unstable. We follow and extend the method which was developed by Zakharov and Rubenchik[5] for studing the transverse instability of the soliton in the nonlinear Schrödinger equation. This work is entirely analytical in character. For SIT, it complements the work in Ref. 6 which was essentially numerical, and did not specialize to solitons. The work in Ref. 6 demonstrates the importance of transverse effects.

As a simple, and illuminating example, we shall use this method to study the transverse stability of the breather solution of the sine-Gordon equation.

We take the two dimensional sine-Gordon equation to be of the form:

$$\partial_t^2 u - \partial_x^2 u + \sin u = \alpha \partial_y^2 u, \tag{1.1}$$

in which the breather solution is written by

$$u_0 = 4\tan^{-1}(\mu\,\mathrm{sech} p\cos q), \tag{1.2}$$

where $p = \mu x/(1+\mu^2)^{1/2} + p_0$, $q = t/(1+\mu^2)^{1/2} + q_0$ and, p_0, q_0, and μ are arbitrary constants. Linearizing (1.1) by

$$u = u_0 + u_1, \quad |u_1| \ll |u_0|, \tag{1.3}$$

we have

$$\partial_t^2 u_1 - \partial_x^2 u_1 + \cos u_0 \cdot u_1 = \alpha \partial_y^2 u_1. \tag{1.4}$$

For the short wave perturbation, if we take α to be positive then (1.4) is stable (well-posed). On the other hand, for the long wave perturbation, since the transverse effects ($\partial_y^2 u$) becomes small in this limit, it has an effect after a long time $t \simeq 1/|k_y|$, where k_y is a wavenumber of the transverse direction. It is convenient to use the long time scale $T = \epsilon t$ ($\epsilon \simeq |k_y|$) in order to investigate long wave perturbations. Then (1.4) becomes

$$\begin{aligned}\partial_t^2 u_1 - \partial_x^2 u_1 + \cos u_0 \cdot u_1 \\ = -\epsilon 2\partial_t \partial_T u_1 + \epsilon^2(\alpha \partial_Y^2 u_1 - \partial_T^2 u_1),\end{aligned} \tag{1.5}$$

where $Y = \epsilon y$. For u_1, we take the form,

$$u_1 = \hat{u}_1(x,t) e^{\Omega T} \sin K_Y Y, \tag{1.6}$$

then we have

$$\begin{aligned} &L\hat{u}_1 \equiv (\partial_t^2 - \partial_x^2 + \cos u_0)\hat{u}_1 = G, \\ &G \equiv -\epsilon 2\Omega \partial_t \hat{u}_1 - \epsilon^2(\Omega^2 + \alpha K_Y^2)\hat{u}_1. \end{aligned} \tag{1.7}$$

The solvability condition of (1.7) may be given by the orthogonality condition

$$\iint_D wG\,dx\,dt = 0, \tag{1.8}$$

where w is a solution to $Lw = 0$, which satisfies the boundary conditions,

$$w \to 0 \quad \text{as} \quad |x| \to \infty, \quad w\text{: periodic with respect to } t. \tag{BC}$$

Here the domain of integration D is given by

$$D = \{(x,t) \mid -\infty < x < \infty, \quad t\text{-periodic}\}.$$

This implies that in (1.7) one should find the solution u_1 which satisfies the boundary conditions (BC). To solve (1.7), we use the following perturbation method.

Expanding $\hat{u}_1 = v_0 + \epsilon v_1 + \cdots$, and $\Omega = \Omega_1 + \epsilon\Omega_2 + \cdots$, we obtain, at leading order,

$$Lv_0 = 0, \tag{1.9}$$

from which we have the solution, subject to (BC),

$$v_0 = A^- u_{0x} + A^+ u_{0t}. \tag{1.10}$$

Here the subscripts $+$ and $-$ express the parity of the function with respect to the space coordinate x, and A^+ and A^- are arbitrary constants. At first order in ϵ, we have

$$Lv_1 = -2\Omega_1 \partial_t v_0, \tag{1.11}$$

from which we have the solution

$$\begin{aligned} v_1 = &-\Omega_1 t u_{0t} A^+ + \Omega_1 x u_{0t} A^- \\ &+ (\text{homogeneous solutions}). \end{aligned} \tag{1.12}$$

Using the fact that $\partial_\mu u_0$ (derivative with respect to μ) is a homogeneous solution (i.e., $L\partial_\mu u_0 = 0$), one can find a solution satisfying the (BC),

$$v_1 = -\Omega_1\left(\frac{1}{\mu^2} x u_{0x} + 2\,\frac{1+\mu^2}{\mu^2}\sin\frac{u_0}{2}\right)A^+ + \Omega_1 x u_{0t} A^-. \tag{1.13}$$

At order ϵ^2, we have

$$Lv_2 = -2\Omega_1 v_{1t} - (\Omega_1^2 + \alpha K_Y^2)v_0 - 2\Omega_2 v_{0t} \equiv G_2. \tag{1.14}$$

Taking $Lu_{0x} = 0$ and $Lu_{0t} = 0$ into account, the secularity conditions (1.8) are given by

$$\iint_D u_{0x} G_2\,dx\,dt = 0, \quad \iint_D u_{0t} G_2\,dx\,dt = 0, \tag{1.15}$$

 0022-2488/80/040928-04$1.00

where $D = \{(x,t) | -\infty < x < \infty, -\pi(1+\mu^2)^{1/2} \leqslant t < \pi(1+\mu^2)^{1/2}\}$. From (1.15), we obtain, respectively,

$$\left[\Omega_1^2 \int_D\int (u_{0x}^2 + u_{0t}^2)\, dx\, dt + \alpha K_Y^2 \int_D\int u_{0x}^2\, dx\, dt\right] A^- = 0, \tag{1.16}$$

$$\left[\Omega_1^2 \int_D\int \left(2\cos\frac{u_0}{2} - 1\right) u_{0t}^2\, dx\, dt - \frac{\alpha\mu^2 K_Y^2}{1+\mu^2} \times \int_D\int u_{0t}^2\, dx\, dt\right] A^+ = 0.$$

By the direct calculations, the integrations in Eq. (1.16) are given by

$$\int\int u_{0x}^2\, dx\, dt = 32\pi\left(|\mu| - \sin^{-1}\left(\frac{\mu^2}{1+\mu^2}\right)^{1/2}\right), \tag{1.17a}$$

$$\int\int u_{0t}^2\, dx\, dt = 32\pi \sin^{-1}\left(\frac{\mu^2}{1+\mu^2}\right)^{1/2}, \tag{1.17b}$$

$$\int\int \left(2\cos\frac{u_0}{2} - 1\right) u_{0t}^2\, dx\, dt = 32\pi\frac{|\mu|}{1+\mu^2}. \tag{1.17c}$$

We obtain, for the antisymmetric mode (A^-)

$$\Omega_1^2 = -\alpha\left(1 - \frac{1}{|\mu|}\sin^{-1}\left(\frac{\mu^2}{1+\mu^2}\right)^{1/2}\right) K_Y^2, \tag{1.18a}$$

and for the symmetric mode (A^+)

$$\Omega_1^2 = \left(\alpha|\mu| \sin^{-1}\left(\frac{\mu^2}{1+\mu^2}\right)^{1/2}\right) K_Y^2. \tag{1.18b}$$

Note that $(1/|\mu|)\sin^{-1}[\mu^2/(1+\mu^2)]^{1/2} < 1$ for all μ. Thus the breather of the sine-Gordon equation is unstable to long wave perturbations. It is interesting to see the weak limit ($\mu \to 0$) of (1.18). In this limit (1.1) can be reduced to the nonlinear Schrödinger equation, and (1.18a,b) become, respectively,

$$\Omega_1^2 = -\frac{\alpha}{3}\mu^2 K_Y^2, \quad \Omega_1^2 = \alpha\mu^2 K_Y^2, \tag{1.19}$$

which agrees with the result in Ref. 5.

2. 0π PULSE OF SIT

We consider the two-dimensional SIT equations given by Mattar and Newstein,[6]

$$E_\eta = P + iF\nabla_\rho^2 E, \quad P_\tau = E(W-1),$$
$$W_\tau = -\tfrac{1}{2}(E^*P + P^*E). \tag{2.1}$$

Equation (2.1) are obtained from the Maxwell–Bloch equations by using the standard slowly varying envelope approximations in a two level medium. Here E,P are the nondimensional complex amplitudes of the electric field, and polarization, respectively, $W-1$ is the energy stored/atom (i.e., the normalized population inversion, where $|W-1| \leqslant 1$, and $W=0$ is the rest state), and F, the Fresnel number, is a dimensionless constant. Furthermore we note that in Eq. (2.1) we assume long relaxation times, and on-line resonance. (Modifications to those assumptions can be readily incorporated into our analysis). In (2.1) E, P, and W are functions of the transverse coordinate ρ, longitudinal coordinate η, and retarded time τ. Equation (2.1) has a breather solution,[4]

$$E_0 = u_{0\tau}, \quad u_0 \equiv 4\tan^{-1}\left(\frac{C_1}{C_2}\,\mathrm{sech}p\cos q\right), \tag{2.2}$$

where

$$p = C_1\{\tau - \eta/(C_1^2 + C_2^2)\} + p_0,$$
$$q = C_2\{\tau + \eta/(C_1^2 + C_2^2)\} + q_0, \tag{2.3}$$

and, $C_1 > 0$, C_2, p_0, and q_0 are arbitrary constants. Here we note that u_0 satisfies the sine-Gordon equation in the light-cone coordinates,

$$u_{0\tau\eta} + \sin u_0 = 0. \tag{2.4}$$

Linearizing (2.1) and eliminating P and W, we obtain

$$E_{1\tau\eta} + E_1 + \tfrac{1}{2}E_1\frac{\partial}{\partial\eta}\int^\tau E_0 E_0^*\, d\tau' + \tfrac{1}{2}E_0\frac{\partial}{\partial\eta}\int^\tau (E_0^* E_1 + E_0 E_1^*)\, d\tau' - \tfrac{1}{2}iFE_0\int^\tau (E_0^*\nabla_\rho^2 E_1 - E_0\nabla_\rho^2 E_1^*)\, d\tau' - iF\nabla_\rho^2 E_{1\tau} = 0 \tag{2.5}$$

where E_0 is the breather solution (2.2), and E_1 the small fluctuation due to perturbation, $|E_1| \ll |E_0|$. By writing $E_1 = \phi + i\psi$ and noting $\partial_\eta \int^\tau E_0^2\, d\tau' = 2(\cos u_0 - 1)$, we have the system of equations,

$$L_0\psi \equiv \psi_{\tau\eta} + \cos u_0\psi \equiv G_0,$$
$$L_1\phi \equiv \phi_{\tau\eta} + \cos u_0\phi + u_{0\tau}\frac{\partial}{\partial\eta}\int^\tau u_0\phi\, d\tau' \equiv G_1, \tag{2.6}$$

where

$$G_0 = F\nabla_\rho^2\phi_\tau,$$
$$G_1 = -F\nabla_\rho^2\psi_\tau - Fu_{0\tau}\int^\tau u_{0\tau}\nabla_\rho^2\psi\, d\tau'. \tag{2.7}$$

For short wave perturbations (i.e., $\nabla_\rho^2\psi$, $\nabla_\rho^2\phi$ are large), one can show that (2.6) is stable (well-posed). On the other hand, for the long wave perturbation (i.e., $\nabla_\rho^2\psi$, $\nabla_\rho^2\phi$ are small), we use the scaling coordinates $T = \epsilon\eta$, $\mathbf{Y} = \epsilon\boldsymbol{\rho}$ ($\epsilon \simeq |\mathbf{k}\rho|$, where $\mathbf{k}\rho$ is the wavenumber of transverse perturbation). Assuming that ψ and ϕ take the forms,

$$\phi = \hat{\phi}(\tau,\eta)e^{\Omega T}\sin\mathbf{K}_Y\cdot\mathbf{Y}, \quad \psi = \hat{\psi}(\tau,\eta)e^{\Omega T}\sin\mathbf{K}_Y\cdot\mathbf{Y}, \tag{2.8}$$

we have

$$L_0\hat{\psi} = -\epsilon\Omega\hat{\psi}_\tau - \epsilon^2 FK_Y^2\hat{\phi}_\tau,$$
$$L_1\hat{\phi} = -\epsilon\Omega N\cdot\hat{\phi} + \epsilon^2 FK_Y^2 N\cdot\hat{\psi}, \tag{2.9}$$

where $K_Y \equiv |\mathbf{K}_Y|$, and N is an operator given by

$$N\cdot = \partial_\tau + u_{0\tau}\int^\tau d\tau'\, u_{0\tau'}. \tag{2.10}$$

In this problem, the proper boundary conditions are given by

$$\hat{\phi},\hat{\psi} \to 0 \text{ as } |p| \to \infty, \quad \hat{\phi},\hat{\psi} \text{ are } q\text{-periodic.} \tag{BC}$$

Expanding $\hat{\phi}$, $\hat{\psi}$, and Ω by

$$\hat{\phi} = \phi_0 + \epsilon\phi_1 + \cdots, \quad \hat{\psi} = \psi_0 + \epsilon\psi_1 + \cdots,$$
$$\Omega = \Omega_1 + \epsilon\Omega_2 + \cdots, \tag{2.11}$$

at leading order, we have

$$L_0\psi_0 = 0, \quad L_1\phi_0 = 0, \tag{2.12}$$

from which solutions subject to (BC) are

$$\psi_0 = A_0^1 u_{0\tau} + A_1^0 u_{0\eta}, \quad \phi_0 = A_0^2 u_{0\tau\tau} + A_1^1 u_{0\tau\eta}. \tag{2.13}$$

Here the superscripts i ($i = 0,1,2$) denote the order of differentiation with respect to τ, the subscripts j ($j = 0,1$) denote the order of differentiation with respect to η, and A_j^i are arbitrary constants. At order ϵ, we have

$$L_0\psi_1 = -\Omega_1\psi_{0\tau}, \quad L_1\phi_1 = -\Omega_1 N\cdot\phi_{0\tau}. \tag{2.14}$$

Using $L_0\partial_{C_i}u_0 = 0$, $L_1\partial_{C_i}u_0 = 0$ ($i = 1,2$), we obtain the solutions with (BC),

$$\begin{aligned}\psi_1 = &-\Omega_1\frac{C_1^2+C_2^2}{C_1^2}\left(\frac{1}{4C_1}pu_{0\tau} - \frac{C_1^2+C_2^2}{4C_1}pu_{0\eta}\right.\\ &\left.+\sin\frac{u_0}{2}\right)A_1^0 + \Omega_1\frac{C_1^2+C_2^2}{C_1^2}\left(\frac{3C_1^2-C_2^2}{4C_1}pu_{0\tau}\right.\\ &\left.+\frac{C_1^2+C_2^2}{4C_1}pu_{0\eta} - \sin\frac{u_0}{2}\right)A_0^1,\end{aligned} \tag{2.15a}$$

$$\begin{aligned}\phi_1 = &-\Omega_1\frac{C_1^2+C_2^2}{C_1^2}\left(\frac{1}{4C_1}pu_{0\tau\tau} - \frac{C_1^2+C_2^2}{4C_1}pu_{0\tau\eta}\right.\\ &\left.+\frac{\partial}{\partial\tau}\sin\frac{u_0}{2} + u_{0\tau}\right)A_1^1 + \Omega_1\frac{C_1^2+C_2^2}{C_1^2}\\ &\times\left(\frac{3C_1^2-C_2^2}{4C_1}pu_{0\tau\tau} + \frac{C_1^2+C_2^2}{4C_1}pu_{0\tau\eta}\right.\\ &\left.-\frac{\partial}{\partial\tau}\sin\frac{u_0}{2}\right)A_0^2.\end{aligned} \tag{2.15b}$$

At order ϵ^2, we have

$$\begin{aligned}&L_0\psi_2 = -\Omega_1\psi_{1\tau} - FK_Y^2\phi_{0\tau} - \Omega_2\psi_{0\tau} \equiv G_{02},\\ &L_1\phi_2 = -\Omega_1 N\cdot\phi_1 + FK_Y^2 N\cdot\psi_0 - \Omega_2 N\cdot\phi_0 \equiv G_{12}.\end{aligned} \tag{2.16}$$

The solvability conditions of those equations are given by

$$\int_D\int u_{0\tau}G_{02}\,d\tau\,d\eta = 0, \quad \int_D\int u_{0\eta}G_{02}\,d\tau\,d\eta = 0, \tag{2.17a}$$

$$\int_D\int u_{0\tau\tau}G_{12}\,d\tau\,d\eta = 0, \quad \int_D\int u_{0\tau\eta}G_{12}\,d\tau\,d\eta = 0. \tag{2.17b}$$

where the domain of integration D is given by

$$D = \{(\tau,\eta)|\ -\infty < p(\tau,\eta) < \infty,\ -\pi \leqslant q(\tau,\eta) < +\pi\}.$$

On the domain D, one can show that L_0 and L_1 are self adjoint operators. From (2.17) we obtain

$$\tilde{I}(\Omega_1^2, FK_Y^2)\cdot A = 0, \tag{2.18}$$

where $A = (A_0^2, A_1^1, A_0^1, A_1^0)^T$ and $\tilde{I}$ is a matrix (function of Ω_1^2 and FK_y^2) given by

$$\begin{pmatrix} \Omega_1^2 I_{11} & \Omega_1^2 I_{12} & FK_Y^2 I_{13} & FK_Y^2 I_{14}\\ \Omega_1^2 I_{21} & \Omega_1^2 I_{22} & FK_Y^2 I_{23} & FK_Y^2 I_{24}\\ FK_Y^2 I_{31} & FK_Y^2 I_{32} & \Omega_1^2 I_{33} & \Omega_1^2 I_{34}\\ FK_Y^2 I_{41} & FK_Y^2 I_{42} & \Omega_1^2 I_{43} & \Omega_1^2 I_{44}\end{pmatrix}. \tag{2.19}$$

Here the components I_{ij} ($i = 1,\ldots,4$ and $j = 1,\ldots,4$) are given in the Appendix. From (2.18), we obtain

$$J_0 Z^2 + J_1 Z + J_2 = 0, \tag{2.20}$$

where $Z = \{\Omega_1^2/(FK_Y^2)\}^2$ and J_0, J_1, and J_2 are given in terms of the I_{ij}. The J_i ($i = 0,1,2$) are not zero. From (2.20) it is obvious that Ω_1 has a positive real part (since Z is proportional to Ω_1^4), i.e., the 0π pulse is unstable with respect to transverse perturbations.

If C_1 is small in comparison with $|C_2|$ fixed (i.e., weak limit), then Eq. (2.1) can be reduced to a system of nonlinear Schrödinger equations,

$$\begin{aligned}&ik_0\frac{\partial A}{\partial T_2} - \frac{1}{k_0^2}\frac{\partial^2 A}{\partial X^2} - \frac{1}{2k_0^2}A^2A^* + k_0F\nabla_Y^2 A = 0,\\ &ik_0\frac{\partial B}{\partial T_2} + \frac{1}{k_0^2}\frac{\partial^2 B}{\partial X^2} + \frac{1}{2k_0^2}B^2B^* + k_0F\nabla_Y^2 B = 0,\end{aligned} \tag{2.21}$$

where $X = \epsilon(\tau - \eta/k_0^2)$, $T_2 = \epsilon^2\eta$, and A,B are given by the expansions,

$$\begin{aligned}E = &\epsilon\left(A(X,T_2,\mathbf{Y})\exp\left\{i\left(k_0\tau + \frac{1}{k_0}\eta\right)\right\} + B(X,T_2,\mathbf{Y})\right.\\ &\left.\times\exp\left\{-i\left(k_0\tau + \frac{1}{k_0}\eta\right)\right\}\right) + \epsilon^2 E_2 + \cdots,\end{aligned} \tag{2.22}$$

$$\begin{aligned}W = &\epsilon^2\left(\frac{1}{2k_0^2}(AA^* + BB^*) - \frac{1}{2k_0^2}AB^*\right.\\ &\times\exp\left\{2i\left(k_0\tau + \frac{1}{k_0}\eta\right)\right\}\\ &\left.+ (\text{complex conjugate})\right) + \epsilon^3 W_2 + \cdots.\end{aligned} \tag{2.23}$$

In this limit, the 0π pulse (2.2) can be approximated by

$$E_0 \sim -4C_1\ \text{sech}p\ \sin q - 4\frac{C_1^2}{C_2}\text{sech}p\tanh p\cos q,$$

and (2.20) gives, at the lowest order of C_1 (i.e., for symmetrical mode),

$$\Omega_1^2 \simeq \pm 2\frac{C_1^2}{C_2^3}FK_Y^2. \tag{2.24}$$

This also demonstrates, in the weak limit, that the 0π pulse is unstable to the transverse variations. Equation (2.24) agrees with the results in Ref. 5, if we take $k_0 = C_2$.

We note that as $\mu\to\infty$ the growth rate of the instability (1.18) is proportional to $\mu^{1/2}$. However the method considered here breaks down as $\mu\to\infty$ since the periodicity in t is assumed order 1 [i.e., $2\pi(1+\mu^2)^{1/2} \sim O(1)$]. In the limit $\mu\to\infty$ the breather tends to a double zero solution (which is also given by the limit of zero relative velocity of kink-antikink solution). Since both kink and antikink solutions are stable, this suggests that the double zero solution might be neutrally stable. Although our method (without extension) does not apply, it may be that a method based upon the conservation laws[7] is applicable.

ACKNOWLEDGMENT

This research is sponsored by Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant No. AFSOR 78 3674. We also appreciate comments made to us by A.C. Newell.

APPENDIX: COMPONENTS OF $\tilde{I}(\Omega_1^2, FK_Y^2)$

Here we list up the components I_{ij} of the matrix (2.19).

$$I_{11} = \frac{C_1^2 + C_2^2}{C_1^2}\left[\frac{3C_1^2 - C_2^2}{4C_1}\iint u_{0\tau\tau} N\cdot(pu_{0\tau\tau}) + \frac{(C_1^2 + C_2^2)^2}{4C_1}\iint u_{0\tau\tau} N\cdot(pu_{0\tau\eta}) - (C_1^2 + C_2^2)\iint u_{0\tau\tau} N\cdot\left(\frac{\partial}{\partial\tau}\sin\frac{u_0}{2}\right)\right],$$

$$I_{12} = \frac{1}{C_1^2 + C_2^2} I_{11} - \frac{1}{4C_1}\iint u_{0\tau\tau} N\cdot(pu_{0\tau\tau}) - \iint u_{0\tau\tau} N\cdot u_{0\tau},$$

$$I_{13} = -\iint u_{0\tau\tau} N\cdot u_{0\tau},$$

$$I_{14} = -\iint u_{0\tau\tau} N\cdot u_{0\eta},$$

$$I_{21} = \frac{C_1^2 + C_2^2}{C_1^2}\left[\frac{3C_1^2 - C_2^2}{4C_1}\iint u_{0\tau\eta} N\cdot(pu_{0\tau\tau}) + \frac{(C_1^2 + C_2^2)^2}{4C_1}\iint u_{0\tau\eta} N\cdot(pu_{0\tau\eta}) - (C_1^2 + C_2^2)\iint u_{0\tau\eta} N\cdot\left(\frac{\partial}{\partial\tau}\sin\frac{u_0}{2}\right)\right],$$

$$I_{22} = \frac{1}{C_1^2 + C_2^2} I_{21} - \frac{1}{4C_1}\iint u_{0\tau\eta} N\cdot(pu_{0\tau\tau}) - \iint u_{0\tau\eta} N\cdot u_{0\tau},$$

$$I_{23} = -\iint u_{0\tau\eta} N\cdot u_{0\tau},$$

$$I_{24} = -\iint u_{0\tau\eta} N\cdot u_{0\eta},$$

$$I_{31} = -\iint u_{0\tau}^2,$$

$$I_{32} = -\iint u_{0\tau\tau} u_{0\tau\eta},$$

$$I_{33} = \frac{C_1^2 + C_2^2}{C_1^2}\left[\frac{3C_1^2 - C_2^2}{4C_1}\iint u_{0\tau}\frac{\partial}{\partial\tau}(pu_{0\tau}) + \frac{(C_1^2 + C_2^2)^2}{4C_1}\iint u_{0\tau}\frac{\partial}{\partial\tau}(pu_{0\eta}) - (C_1^2 + C_2^2)\iint u_{0\tau}\frac{\partial}{\partial\tau}\sin\frac{u_0}{2}\right],$$

$$I_{34} = \frac{1}{C_1^2 + C_2^2} I_{33} - \frac{1}{4C_1}\iint u_{0\tau}\frac{\partial}{\partial\tau}(pu_{0\tau}),$$

$$I_{41} = I_{32},$$

$$I_{42} = -\iint u_{0\tau\eta}^2,$$

$$I_{43} = \frac{C_1^2 + C_2^2}{C_1^2}\left[\frac{3C_1^2 - C_2^2}{4C_1}\iint u_{0\eta}\frac{\partial}{\partial\tau}(pu_{0\tau}) + \frac{(C_1^2 + C_2^2)^2}{4C_1}\iint u_{0\eta}\frac{\partial}{\partial\tau}(pu_{0\eta}) - (C_1^2 + C_2^2)\iint u_{0\eta}\frac{\partial}{\partial\tau}\sin\frac{u_0}{2}\right],$$

$$I_{44} = \frac{1}{C_1^2 + C_2^2} I_{43} - \frac{1}{4C_1}\iint u_{0\eta}\frac{\partial}{\partial\tau}(pu_{0\tau}).$$

[1] M.J. Ablowitz and Y. Kodama, Phys. Lett. A **70**, 83 (1979).
[2] L.A. Bolshov, V.V. Likhansky, and A.P. Napartovich, Sov. Phys. JETP **45**, 928 (1978).
[3] L.A. Bolshov and V.V. Likhansky, Zh. Eksp. Teor. Fiz. **75**, 2047 (1978).
[4] For example, G.L. Lamb Jr., Rev. Mod. Phys. **43**, 99 (1971).
[5] V.E. Zakharov and A.M. Rubenchik, Sov. Phys. JETP **38**, 494 (1974).
[6] F.P. Matter and M.C. Newstein, IEEE J. Quantum Electron **QE-13**, 507 (1977).
[7] A.C. Newell, private communication.

Weakly nonlinear theory of coherent Langmuir waves. I. Wave wave interaction

M. Trocheris
Association Euratom-CEA Sur la Fusion, Département de Physique du Plasma et de la Fusion Contrôlée, Centre d'Etudes Nucléaires, Boîte Postale n°6, 92260 Fontenay-Aux-Roses, France

(Received 8 August 1979; accepted for publication 30 November 1979)

A weakly nonlinear analysis of the Vlasov equation is made in the case of small amplitude Langmuir waves. The nonlinear terms are treated as a small perturbation in the framework of the asymptotic theory of Krylov and Bogoliubov. In order to apply this method, the Vlasov system of equations is transformed by taking as new unknown functions a complete system of constants of the motions of the linearized equations. Such a system can be found with the help of essentially Van Kampen's and Case's expansion in normal modes. Resonant wave particle interaction is avoided by cutting off the distribution functions in velocity space and by considering waves of phase velocities larger than the cutoff. In this case, the main physical effect is mode coupling and Davidson's nonlinear system of equations on the wave amplitudes is recovered. The main point is that the derivation is made without neglecting the free streaming portions of the distribution functions. A discussion of the validity of the approximation and of the relevant time scales is presented.

1. INTRODUCTION

This paper deals with the weakly nonlinear theory of Langmuir waves in a Vlasov plasma. The plasma is supposed to be fully ionized with a sufficiently high temperature and a small enough density so that collisions may be neglected. Then only collective effects due to the average field come into play and the plasma can be described by the usual Vlasov–Maxwell system of equations. A wave in this plasma is considered as a small perturbation of a uniform equilibrium state. In the first order or linear approximation the waves propagate independently of one another. However, for larger amplitudes or for long enough time scales a variety of nonlinear effects appear, which are classified according to an elaborate terminology. They include in particular wave–wave and wave–particle interactions and they have been extensively studied in the framework of the weak turbulence theory (see for instance Tsytovich[1]) in which the waves are essentially assumed to be incoherent. In this paper the opposite case of coherent waves is considered, that is to say no random character is assumed for the waves, which are thus supposed to evolve according to the Vlasov–Maxwell equations without any loss of information. Many nonlinear problems have been considered in this context and in particular wave–wave interaction (see Davidson,[2] Chaps. 2–6).

In this latter work a system of mode coupling equations is established, but the derivation involves neglecting the free streaming parts of the perturbed distribution functions. The aim of the present paper is to show that the result still holds when the free streaming parts are no longer neglected, which is not trivial. There is no *a priori* reason why there should be a closed system of equations on the wave amplitudes, in which the perturbed distribution functions are not involved except through the wave amplitudes. However, it will be shown that this is in fact the case. Yet the proof can be made only when the density of resonant particles is negligible, i.e., when the phase velocities of the waves are large enough so that the distribution functions are negligible for such velocities.

A weakly nonlinear analysis will be used and the orderings will be made according to the powers of a small parameter ϵ, which is of the order of the amplitude of the waves or of the density perturbation δn associated with the wave. Denoting by n_0 the unperturbed density, we assume

$$\epsilon \sim \delta n / n_0 . \tag{1.1}$$

We may choose an initial state which differs slightly from a stable equilibrium, and use the free energy argument of Gardner[3] to ensure that ϵ remains small for all times.

With a view to arriving at a description of nonlinear evolution over long times, approximations will be used which are uniformly valid in time as much as possible. This can be achieved in particular by means of the multiple time scale method as in Freeman *et al.*[4] and in Davidson[2] for the treatment of the wave–wave interaction. In this work an equivalent form of this approximation namely the original asymptotic method of Krylov and Bogoliubov[5] and Bogoliubov and Mitropolsky,[6] will be used. This method yields approximate solutions which are valid in principle over the interval

$$t_0 \leqslant t \leqslant \frac{1}{\epsilon \omega_p}$$

and contributions of order ϵ^2 are neglected in the evolution of quantities like the wave amplitudes which are of order ϵ initially. In the following, the asymptotic method will be applied in its simplest form often called the "method of averaging" (Bogoliubov and Mitropolsky[6]; Minorsky[7]). This method is applicable to a system of linear equations perturbed by small quadratic terms provided the unknown functions are the constants of the motions of the linear system. A prerequisite is therefore to cast the system of the Vlasov equations into such a form. It will be seen that this

 0022-2488/80/040932-09$1.00

can be achieved by expanding the distribution fuctions in a complete system of normal modes, which are essentially the same as Van Kampen's[8] and Case's.[9]

As the only effect considered in this paper is the wave–wave interaction, we choose the problem of Langmuir waves in an infinite magnetic field which affords the simplest example of resonant mode coupling (see Aamodt and Drummond[10]). As usual, the ions are replaced by a uniform background of charge. The distribution function of the electrons $f(\mathbf{X}, v, t)$ depends on three space coordinates $\mathbf{X}$ and on one velocity component v along the direction of the magnetic field and the system of equations is the following:

$$f(\mathbf{X}, v, t) = f_0(v, t) + \sum_{\mathbf{K}\neq 0} e^{i\mathbf{K}\cdot\mathbf{X}} f_{\mathbf{K}}(v, t) \left(\int_{-\infty}^{+\infty} f_0(v)\, dv = 1 \right), \tag{1.2}$$

$$\frac{\delta f_{\mathbf{K}}}{\delta t} + ik_{\parallel} v f_{\mathbf{K}} + \frac{e}{m} (E_{\mathbf{K}})_{\parallel} \frac{\delta f_0}{\delta v} = - \sum_{\mathbf{q}\neq 0} \frac{e}{m} (E_{\mathbf{K}-\mathbf{q}})_{\parallel} \frac{\delta f_{\mathbf{q}}}{\delta v}, \tag{1.3}$$

$$\frac{\delta f_0}{\delta t} = - \sum_{\mathbf{K}\neq 0} \frac{e}{m} (E_{\mathbf{K}})_{\parallel} \frac{\delta f_{-\mathbf{k}}}{\delta v}, \tag{1.4}$$

$$(E_{\mathbf{K}})_{\parallel} = -i \frac{k_{\parallel}}{k^2} 4\pi e n_0 \int_{-\infty}^{+\infty} f_{\mathbf{K}}\, dv, \tag{1.5}$$

where $k_{\parallel}$ and $E_{\parallel}$ denote the components of $\mathbf{K}$ and $\mathbf{E}$ along the direction of the magnetic field, respectively.

In order to avoid wave particle resonant interaction we will assume that the distribution functions f_0 and $f_{\mathbf{K}}$ vanish for all t outside a finite interval of the velocity $-V < v < V$, and we will choose cases where the phase velocities of the waves lie outside this interval.

Following this Introduction, the asymptotic method is recalled in Sec. 2. The relevant expansion is established in Sec. 3. The final equations on the wave amplitudes are derived in Sec. 4 and the conditions of validity are discussed in Sec. 5.

2. THE ASYMPTOTIC METHOD

In order to recall the asymptotic method in its classical form (see Bogoliubov and Mitropolsky[6]), let us consider the following system of n differential equations on n unknown functions $x_j(t)$:

$$\frac{dx_j}{dt} = \epsilon F_j(t, x) \qquad (j = 1,2,...,n), \tag{2.1}$$

where x stands for $x_1, x_2, ..., x_n$ in the right-hand side and where ϵ is the small parameter. The method yields an approximate solution which differs from the exact solution by an error of order ϵ and which is valid for

$$t_0 \leqslant t < \tau/\epsilon, \tag{2.2}$$

where τ is a constant independent of ϵ.

The most classical procedure to find this solution is the "method of averaging"[6,7] whereby the right-hand side (rhs) of Eq. (2.1) is replaced by its average over the explicit dependence on t, i.e., over the dependence on t for constant x. The method can be generalized to the case where the average over t is well defined only for part of the rhs of Eq. (2.1). Suppose that Eq. (2.1) can be written in the following form:

$$\frac{dx_j}{dt} = \epsilon G_j(t, x) + \epsilon H_j(t, x), \tag{2.3}$$

where G_j has no well defined average over t whereas H_j has a vanishing average over the explicit dependence on t. Then the approximate solution satisfies a simplified system of equations, which is obtained by striking out the term ϵH_j in the rhs of Eq. (2.3).

Let us now outline a proof of this result by starting from a simple but rather strong condition on $H_j(t, x)$, namely that its integral over t for constant x,

$$K_j(t, x) = \int_{t_0}^{t} H_j(t', x)\, dt', \tag{2.4}$$

remains bounded for all $t > t_0$. The derivatives $\delta H_j/\delta x_K$ are also assumed to satisfy the same condition as H_j. Defining the new unknown functions $\bar{x}_j$ by

$$\bar{x}_j = x_j - \epsilon K_j(t, x), \tag{2.5}$$

we have the following by differentiating with respect to t:

$$\frac{d\bar{x}_j}{dt} = \frac{dx_j}{dt} - \epsilon H_j(t, x) - \epsilon \frac{\delta K_j}{\delta x_k} \cdot \frac{dx_k}{dt} = \epsilon G_j(t, x) + O(\epsilon^2) = \epsilon G_j(t, \bar{x}) + O(\epsilon^2). \tag{2.6}$$

Terms $O(\epsilon^2)$ on the rhs can be neglected as they lead to an error of order ϵ on $\bar{x}$ for t satisfying Eq. (2.2).

The result can easily be extended to the case where G_j and H_j depend on ϵ provided they remain bounded in the limit $\epsilon \to 0$. The condition on H_j is then that K_j remain bounded for all t satisfying Eq. (2.2).

The equations to be considered in the following are not simply differential equations like Eq. (2.1). In addition to a discrete set of unknown functions of t like $x_j(t)$, we shall have to deal with unknown functions depending on a continuous parameter λ. If $Y(t,\lambda)$ is such a function, the corresponding type of equation will be the following:

$$\delta Y/\delta t = \epsilon \mathscr{F}(t,\lambda;Y), \tag{2.7}$$

where the dependence on Y in the rhs is not simply algebraic, but may contain for instance derivatives of Y with respect to λ or integrals of Y over λ. However, this dependence will turn out to be linear or quadratic in Y. Thanks to this simplifying feature, the above results can easily be generalized to equations of the type (2.7) or to equations of the following type analogous to Eq. (2.3):

$$\frac{\delta Y}{\delta t} = \epsilon \mathscr{G}(t,\lambda;Y) + \epsilon \mathscr{H}(t,\lambda;Y). \tag{2.8}$$

In the same way as above the term $\mathscr{H}$ in Eq. (2.8) can be neglected provided its integral over the explicit dependence on t remains bounded in the interval (2.2). In addition, the same condition has to hold for the expression analogous to $(\delta K_j/\delta x_K)(dx_K/dt)$ which is obtained simply by differentiating $\mathscr{H}$ with respect to its implicit dependence on t through Y.

In Sec. 4 the asymptotic method will be applied to a

system of equations which is a combination of the types (2.3) and (2.8), requiring another straightforward extension of the theory. A difficulty arises from the fact that the system of equations considered in Sec. 4 is infinite. This is connected with the usual expansion in Fourier series of the spatial dependence of the distribution function. No attempt at a rigorous treatment of this point will be made. We will therefore simply retain the rule which enables one to neglect terms in the rhs of equations of the type (2.3) and (2.8) and we will admit that the solution of the simplified equations is an approximation of the solution of the original equations in the same way as in the above discussion.

3. THE EXPANSION IN NORMAL MODES

Before applying the asymptotic method to the Vlasov equations (1.2)–(1.5), it is necessary to cast them into a form analogous to Eq. (2.1), which means writing them in terms of new unknown functions which are constants of the motion in the linear approximation. This can be achieved by means of a complete system of normal modes of the linearized equations. Such a system was proposed by Van Kampen[8] and Case[9] and it will be our starting point. However, it will have to be slightly modified before it is exactly suited to our problem, as we have assumed that the distribution functions vanish outside the interval $(-V,V)$.

In the linear approximation, the Vlasov equation (1.2)–(1.5) can be written in the form

$$\frac{\delta f_{\mathbf{K}}}{\delta t} + i\,k_{\|} L_k f_{\mathbf{K}} = 0, \quad \frac{\delta f_0}{\delta t} = 0, \tag{3.1}$$

with $k = |\mathbf{K}|$ and

$$L_k f_{\mathbf{K}} = v f_{\mathbf{K}} - \frac{\omega_p^2}{k^2}\frac{\delta f_0}{\delta v}\int_{-V}^{V} f_{\mathbf{K}}(v')\,dv'. \tag{3.2}$$

We want to construct a complete system of eigenfunctions of the linear operator L_k, in terms of which an arbitrary function of v can be expanded. This can be achieved by simple modifications of Case's derivation. A different proof is possible by the method of the resolvent (Kato,[11] Titchmarsh[12]) and it is outlined in Appendix A. This latter derivation[13] requires very simple conditions of validity, namely that the functions $f_0(v)$ and $f_{\mathbf{K}}(v)$ have, respectively, three and two continuous derivatives and vanish outside $(-V,V)$.

Before giving the eigenfunctions and stating the expansion formula, it is necessary to recall some properties of the dispersion function. For λ complex or real outside $(-V,V)$, the dispersion relation is obtained by equating to zero the function

$$\epsilon(k,\lambda) = 1 - \frac{\omega_p^2}{k^2}\int_{-V}^{V}\frac{f_0'(v)}{v-\lambda}\,dv \quad \left[f_0'(v) = \frac{\delta f_0}{\delta v}\right]. \tag{3.3}$$

We will assume

$$f_0'(v) > 0, \text{ for } -V < v < 0, \text{ and } f_0'(v) < 0, \text{ for } 0 < v < V,$$

in which case it can be shown that $\epsilon(k,\lambda)$ has no complex zeros in λ and has at most two real zeros outside $(-V,V)$:

$$\nu_k^1 > V \quad \text{and} \quad \nu_k^2 < -V,$$

which both exist for small enough k. With $-V \leqslant \lambda \leqslant V$, expression (3.3) for $\epsilon(k,\lambda)$ is no longer defined and the following functions appear in the theory:

$$\epsilon_0(k,\lambda) = 1 - \frac{\omega_p^2}{k^2} P\int_{-V}^{V}\frac{f_0'(v)}{v-\lambda}\,dv, \tag{3.4}$$

$$\epsilon_{\pm}(k,\lambda) = \epsilon_0(k,\lambda) \mp i\pi\frac{\omega_p^2}{k^2} f_0'(\lambda). \tag{3.5}$$

The notation P in Eq. (3.4) denotes the Cauchy principal value of the integral. Expression (3.5) is the limit of $\epsilon(k,\lambda)$ as λ tends to a real value in the interval $(-V,V)$ either from above or from below the real axis. With the assumptions made on $f_0(v)$, it can be shown that

$$\epsilon_{\pm}(k,\lambda) \neq 0, \quad \text{for } -V \leqslant \lambda \leqslant V. \tag{3.6}$$

The discrete spectrum consists of the two zeros and ν_k^1 of ν_k^2 and the corresponding eigenfunctions are simply

$$\varphi_k^{\alpha}(v) = \frac{f_0'(v)}{v - \nu_k^{\alpha}} \quad (\alpha = 1,2), \tag{3.7}$$

$$L_k \varphi_k^{\alpha} = \nu_k^{\alpha}\varphi_k^{\alpha}. \tag{3.8}$$

The same eigenvalues ν_k^1 and ν_k^2 also belong to the eigenfunctions of the adjoint operator L_k^{+} defined by

$$(f, L_k g) = (L_k^{+} f, g), \tag{3.9}$$

with the following notation for the scalar product of two functions:

$$(f,g) = \int_{-V}^{V} f(v)g(v)\,dv. \tag{3.10}$$

Thus, we have

$$\bar{\varphi}_k^{\alpha}(v) = \frac{1}{v - \nu_k^{\alpha}}, \tag{3.11}$$

$$L_k^{+}\bar{\varphi}_k^{\alpha} = \nu_k^{\alpha}\bar{\varphi}_k^{\alpha}. \tag{3.12}$$

Further, the following orthogonality relations hold:

$$(\bar{\varphi}_k^1, \varphi_k^2) = (\bar{\varphi}_k^2, \varphi_k^1) = 0. \tag{3.13}$$

The continuous spectrum consists of all real numbers of the interval $(-V,V)$. The corresponding "eigenfunctions" are singular and can be written with the following notations which prove convenient:

$$\begin{aligned} {}_v\phi_{\lambda}^k &= \frac{\omega_p^2}{k^2} f_0'(v) P\frac{1}{v-\lambda} + \epsilon_0(k,\lambda)\delta(v-\lambda), \\ {}_v\bar{\phi}_{\lambda}^k &= \frac{\omega_p^2}{k^2} f_0'(\lambda) P\frac{1}{v-\lambda} + \epsilon_0(k,\lambda)\delta(v-\lambda), \\ L_k\phi_{\lambda}^k &= \lambda\phi_{\lambda}^k, \quad L_k^{+}\bar{\phi}_{\lambda}^k = \lambda\bar{\phi}_{\lambda}^k. \end{aligned} \tag{3.14}$$

These expressions can be considered as two sided linear functionals or distributions. Their values for given functions $f(v)$ or $g(\lambda)$ will be denoted in the following way:

$$\begin{aligned} \langle f, \phi_{\lambda}^k\rangle &= \langle f(v), {}_v\phi_{\lambda}^k\rangle \\ &= \frac{\omega_p^2}{k^2} P\int_{-V}^{V} f(v) f_0'(v)\frac{dv}{v-\lambda} + \epsilon_0(k,\lambda) f(\lambda), \end{aligned} \tag{3.15}$$

$$\begin{aligned} \langle {}_v\phi^k, g\rangle &= \langle {}_v\phi_{\lambda}^k, g(\lambda)\rangle \\ &= \frac{\omega_p^2}{k^2} f_0'(v) P\int_{-V}^{V} g(\lambda)\frac{d\lambda}{v-\lambda} + \epsilon_0(k,v) g(v), \end{aligned} \tag{3.16}$$

with similar expressions for ${}_v\bar{\phi}^k_\lambda$.

The linear functionals ϕ^k and $\bar{\phi}^k$ are then "eigenfunctionals" of L_k and L_k^+, respectively, in the following sense:

$$\langle L_k^+ f, \phi^k_\lambda \rangle = \lambda \langle f, \phi^k_\lambda \rangle, \quad \langle L_k f, \bar{\phi}^k_\lambda \rangle = \lambda \langle f, \bar{\phi}^k_\lambda \rangle.$$

The following orthogonality relations are also easily established:

$$\langle \bar{\varphi}^\alpha_k, \phi^k_\lambda \rangle = 0 \quad (\lambda \neq \nu^\alpha_k), \quad \langle \varphi^\alpha_k, \bar{\phi}^k_\lambda \rangle = 0 \quad (\lambda \neq \nu^\alpha_k).$$

With these definitions and notations an arbitrary distribution function $f_{\mathbf{K}}(v)$ can be expanded in eigenfunctions of L_k:

$$f_{\mathbf{K}}(v) = \sum_{\alpha=1,2} a^\alpha_{\mathbf{K}} \varphi^\alpha_k(v) + \langle {}_v\phi^k_\lambda, c_{\mathbf{K}}(\lambda) \rangle . \tag{3.17}$$

Conversely, the "coefficients" $a_{\mathbf{K}}$ and $c_{\mathbf{K}}$ of this expansion can be expressed in terms of the function $f_{\mathbf{K}}$ in the following way:

$$a^\alpha_{\mathbf{K}} = \frac{(\bar{\varphi}^\alpha_k, f_{\mathbf{K}})}{(\bar{\varphi}^\alpha_k, \varphi^\alpha_k)}, \tag{3.18}$$

$$c_{\mathbf{K}}(\lambda) = \frac{\langle f_{\mathbf{K}}, \bar{\phi}^k_\lambda \rangle}{\epsilon_r(k,\lambda)\epsilon_{-r}(k,\lambda)}. \tag{3.19}$$

The correspondence with the notations of Van Kampen[8] and Case[9] is given in Appendix E.

These results are readily applied to the solution of the linearized Vlasov equations. In that case $f_0(v)$, as well as the eigenvalues and the eigenfunctions, are independent of the time t. Therefore, if the solution $f_{\mathbf{K}}(v,t)$ is expanded according to Eq. (3.17), the time dependence will appear only in the coefficient $a^\alpha_{\mathbf{K}}(t)$ and $c_{\mathbf{K}}(\lambda,t)$. Further, this time dependence is very simple. By differentiating Eq. (3.18) with respect to t and using Eqs. (3.1), (3.9), and (3.12) one finds easily

$$\frac{da^\alpha_{\mathbf{K}}}{dt} + i k_\parallel \nu^\alpha_k a^\alpha_{\mathbf{K}} = 0 \tag{3.20}$$

and similarly

$$\frac{\delta}{\delta t} c_{\mathbf{K}} + i k_\parallel \lambda c_{\mathbf{K}} = 0. \tag{3.21}$$

We therefore have the following constants of the motion for the linearized Vlasov equations:

$$A^\alpha_{\mathbf{K}} = e^{i k_\parallel \nu^\alpha_k t} a_{\mathbf{K}}, \quad \Gamma_{\mathbf{K}}(\lambda) = e^{i k_\parallel \lambda t} c_{\mathbf{K}}(\lambda, t), \tag{3.22}$$

where $a^\alpha_{\mathbf{K}}$ and $c_{\mathbf{K}}$ can be expressed in terms of $f_{\mathbf{K}}$ by means of Eqs. (3.18) and (3.19).

Equations (3.17)–(3.19) together with Eq. (3.22) provide the change of unknown functions which is required to write the Vlasov equations in the form suitable for the application of the asymptotic method. In the general case the space averaged distribution function $f_0(v)$ is time dependent, but we will first assume that this time dependence is negligible. In that case, the change of unknown functions is straightforward[13] and Eq. (1.3) on the functions $f_{\mathbf{K}}(v,t)$ for $\mathbf{K} \neq 0$ are found to be equivalent to a system of equations on $A^\alpha_{\mathbf{K}}(t)$ and $\Gamma_{\mathbf{K}}(\lambda,t)$ of which a first part can be written in the following way:

$$(\bar{\varphi}^\alpha_k, \varphi^\alpha_k) \frac{d}{dt} A^\alpha_{\mathbf{K}} = \sum_{\substack{\mathbf{K}', \mathbf{K}'' \\ (\mathbf{K}' + \mathbf{K}'' = \mathbf{K})}} i k'_\parallel \{\mathrm{I} + \mathrm{II}\} \{\mathrm{I}' + \mathrm{II}''\}, \tag{3.23}$$

where I, II, I′, II′ denote the following expressions:

$$\mathrm{I} = \sum_\beta A^\beta_{\mathbf{K}'} e^{i(\omega - \omega')t},$$

$$\mathrm{II} = \frac{\omega_p^2}{k'^2} \int_{-V}^{V} \Gamma_{\mathbf{K}'}(\lambda) e^{i(\omega - k_\parallel \lambda)t} d\lambda, \tag{3.24}$$

$$\mathrm{I}' = \sum_\gamma A^\gamma_{\mathbf{K}''} e^{-i\omega'' t} \left(\bar{\varphi}^\alpha_k, \frac{\delta \varphi^\gamma_{k''}}{\delta v} \right),$$

$$\mathrm{II}' = -\left(\frac{\delta \bar{\varphi}^\alpha_k}{\delta v}, \langle {}_v\phi^{k''}_\lambda, \quad \Gamma_{\mathbf{K}''}(\lambda) e^{-i k''_\parallel \lambda t} \rangle \right),$$

with the notation

$$\omega = k_\parallel \nu^\alpha_k, \quad \omega' = k'_\parallel \nu^\beta_{k'}, \quad \omega'' = k''_\parallel \nu^\gamma_{k''} .$$

Expression II has been transformed by means of the following identity:

$$\int_{-V}^{V} \langle {}_v\phi^{k'}, c_{\mathbf{K}''} \rangle dv = \int_{-V}^{V} c_{\mathbf{K}'}(\lambda) d\lambda,$$

which can be easily established from symmetry property of the bilinear functional ${}_v\phi^k_\lambda$ (see Appendix B):

$$(\langle f, \phi \rangle, g) = (f, \langle \phi, g \rangle). \tag{3.25}$$

The second part of the system of equations on the $A^\alpha_{\mathbf{K}}$ and $\Gamma_{\mathbf{K}}$ has the form

$$\frac{\delta}{\delta t} \Gamma_{\mathbf{K}} = \text{quadratic expression in the } A\text{'s and } \Gamma\text{'s}, \tag{3.26}$$

and it needs not be written explicitly.

4. APPLICATION OF THE ASYMPTOTIC METHOD

By definition of the small parameter ϵ, the following ordering holds in the weakly nonlinear problem considered:

$$f_{\mathbf{K}} \sim E_{\mathbf{K}} \sim A^\alpha_{\mathbf{K}} \sim \Gamma_{\mathbf{K}} \sim \epsilon \tag{4.1}$$

and this is supposed to hold for all times as stated in the Introduction. Then the asymptotic method enables one to find approximate solutions which are valid for a time interval of the form (2.2), i.e., in our case

$$t_0 \leqslant t \lesssim \frac{1}{\epsilon \omega_p}. \tag{4.2}$$

In Eq. (3.23) and (3.26) the quadratic forms are of order ϵ^2 so that they lead to a finite relative variation of $A^\alpha_{\mathbf{K}}$ and $\Gamma_{\mathbf{K}}$ over the time interval (4.2). A term of order ϵ^3 in the same equations only contributes a change of $A^\alpha_{\mathbf{K}}$ and $\Gamma_{\mathbf{K}}$ by an amount of order ϵ^2 which can be neglected, in agreement with the asymptotic method.

It follows in particular that the time dependence of $f_0(v)$ can indeed by neglected in deriving Eq. (3.23) and (3.26). In fact, by differentiating Eq. (3.18) with respect to t and taking into account that $\delta f_0/\delta t$ is of order ϵ^2 according to Eq. (1.4), one easily obtains

$$(\bar{\varphi}^\alpha_k, \varphi^\alpha_k) \frac{da^\alpha_{\mathbf{K}}}{dt} = \left(\bar{\varphi}^\alpha_k, \frac{\delta f_{\mathbf{K}}}{\delta t} \right) + O(\epsilon^3).$$

The terms $O(\epsilon^3)$ can be neglected and from there on the derivation proceeds rigourously up to Eq. (3.23). A similar justification holds for the derivation of Eq. (3.26).

It will be seen later on that the ordering in ϵ is not quite so simple, due to the presence of some secular terms in Eq. (3.26). This difficulty will be discussed in Sec. 5 and the result will be that the range of validity of the final equations is generally limited to $t = O(\epsilon^{-1/2})$ rather than to $t = O(\epsilon^{-1})$. However, for the sake of simplicity we will first apply the asymptotic method to the system of equations (3.23) by pretending that all quadratic terms in the equations for $A^{\alpha}_{\mathbf{K}}$ and $\Gamma_{\mathbf{K}}$ are of order ϵ^2 for all t. The validity of the derivation will be reexamined in Sec. 5 after establishing a more realistic ordering.

We will now see how Eq. (3.23) can be simplified by applying the rule defined in Sec. 2. The products of expressions (3.24) in the right hand side of Eq. (3.23) give rise to four terms of which only the term $\mathrm{I}\times\mathrm{I}'$ will eventually remain. We will examine the three other terms one by one.

Let us start with the term

$$\mathrm{II}\times\mathrm{I}' = \frac{\omega_p^2}{k'^2}\sum_{\gamma}\left(\bar{\varphi}^{\alpha}_{k}, \frac{\delta\varphi^{\gamma}_{k''}}{\delta v}\right)e^{i(\omega-\omega'')t} \times A^{\gamma}_{\mathbf{K}''}\int_{-V}^{V}\Gamma_{\mathbf{K}'}(\lambda)e^{-k'_{\parallel}\lambda t}\,d\lambda.$$

The explicit dependence on t of this expression, in the sense of Sec. 2, includes all dependence except through $A^{\gamma}_{\mathbf{K}''}$ and $\Gamma_{\mathbf{K}}$, and we have to study the integral of $\mathrm{II}\times\mathrm{I}'$ over this explicit dependence on t. Consider first the integral over λ. It is a Fourier integral and it decreases as $t\to+\infty$, the faster the more regular the function $\Gamma_{\mathbf{K}'}(\lambda)$ is. By assuming that $f_0(v)$ and $f_{\mathbf{K}'}(v)$ have a sufficient number of continuous derivatives, one can make sure that $\Gamma_{\mathbf{K}'}(\lambda)$ has at least two continuous derivatives in which case the Fourier integral can be shown to decrease as $1/t^2$ through a double integration by parts. The other factors in $\mathrm{II}\times\mathrm{I}'$ are bounded so that the integral of $\mathrm{II}\times\mathrm{I}'$ over the explicit time dependence is in fact convergent for $t\to+\infty$. This result can also be shown when $\Gamma_{\mathbf{K}'}$ is replaced by $\delta\Gamma_{\mathbf{K}'}/\delta t$. Therefore, the term $\mathrm{II}\times\mathrm{I}'$ in Eq. (3.23) can be neglected.

Consider now the term $\mathrm{I}\times\mathrm{II}'$. By using the symmetry relation (3.25), it can be brought back to a form analogous to $\mathrm{II}\times\mathrm{I}'$

$$\begin{aligned}\mathrm{I}\times\mathrm{II}' &= -\sum_{\beta}e^{i(\omega-\omega')t}A^{\beta}_{\mathbf{K}'}\left(\frac{\delta\bar{\varphi}^{\alpha}_{k}}{\delta v}, \left\langle {}_{v}\phi^{k''}_{\lambda}, \Gamma_{\mathbf{K}''}(\lambda)e^{-ik''_{\parallel}\lambda t}\right\rangle\right)\\ &= -\sum_{\beta}e^{i(\omega-\omega')t}A^{\beta}_{\mathbf{K}'}\int_{-V}^{V}F(\lambda)\Gamma_{\mathbf{K}''}(\lambda)e^{-ik''_{\parallel}\lambda t}\,d\lambda,\end{aligned}$$

where the function

$$F(\lambda) = \left\langle \frac{\delta\bar{\varphi}^{\alpha}_{k}}{\delta v}, {}_{v}\phi^{k''}_{\lambda}\right\rangle$$

can have as many continuous derivatives as is required provided $f_0(v)$ has sufficiently many. The term $\mathrm{I}\times\mathrm{II}'$ can therefore be neglected in exactly the same way as $\mathrm{II}\times\mathrm{I}'$.

Finally, it is easily seen that the term $\mathrm{II}\times\mathrm{II}'$ can also be neglected for the same reasons as $\mathrm{II}\times\mathrm{I}'$ and $\mathrm{I}\times\mathrm{II}'$.

We therefore arrive at a simplified system of equations in which only the term $\mathrm{I}\times\mathrm{I}'$ is retained in the right-hand side of Eq. (3.23). This simplified system can be written in the following form:

$$\frac{dA^{\alpha}_{\mathbf{K}}}{dt} = \sum_{\substack{\mathbf{K}',\mathbf{K}''\\(\mathbf{K}'+\mathbf{K}''=\mathbf{K})}}\sum_{\beta,\gamma}e^{i(\omega-\omega'-\omega'')t}M(\mathbf{K},\alpha;\mathbf{K}',\beta;\mathbf{K}'',\gamma)A^{\beta}_{\mathbf{K}'}A^{\gamma}_{\mathbf{K}''} \tag{4.3}$$

or equivalently

$$\frac{da^{\alpha}_{\mathbf{K}}}{dt} + ik_{\parallel}v^{\alpha}_{k}a^{\alpha}_{\mathbf{K}} = \sum_{\substack{\mathbf{K}',\mathbf{K}''\\(\mathbf{K}'+\mathbf{K}''=\mathbf{K})}}\sum_{\beta,\gamma}M(\mathbf{K},\alpha;\mathbf{K}',\beta;\mathbf{K}'',\gamma)a^{\beta}_{\mathbf{K}'}a^{\gamma}_{\mathbf{K}''}, \tag{4.4}$$

with the notation

$$M(\mathbf{K},\alpha;\mathbf{K}',\beta;\mathbf{K}'',\gamma) = ik'_{\parallel}\frac{(\bar{\varphi}^{\alpha}_{k},\delta\varphi^{\gamma}_{k''}/\delta v)}{(\bar{\varphi}^{\alpha}_{k},\varphi^{\alpha}_{k})}. \tag{4.5}$$

Thus, the coefficients of the continuous spectrum $\Gamma_{\mathbf{K}}(\lambda,t)$ have disappeared from the equations so that a closed system of equations on the wave amplitudes $A^{\alpha}_{\mathbf{K}}$ is obtained as a direct consequence of the asymptotic approximation, and this is the main result of this paper.

The various mode coupling terms in Eqs. (4.3) will contribute significantly to the evolution of the wave amplitude, i.e., by an amount of order ϵ for $t\sim 1/\epsilon\omega_p$, only insofar as the resonance condition is satisfied:

$$\omega-\omega'-\omega''=0 \quad\text{or}\quad \omega-\omega'-\omega''=O(\epsilon). \tag{4.6}$$

In the latter case the factor $\exp i(\omega-\omega'-\omega'')t$ varies by a finite amount over a time interval of order $1/\epsilon\omega_p$ and the corresponding mode coupling term plays a significant part in the equations.

It is shown in Appendix C that Eq. (4.3) is the same as Davidson's mode coupling equation.

5. CONDITIONS OF VALIDITY

In order to assess the conditions of validity of the theory more precisely, it is necessary to consider also Eq. (3.26). The right-hand side of this equation is formed of terms which are quadratic in the following expressions:

$$\begin{aligned}&\text{(a) } a^{\beta}_{\mathbf{K}'},\quad \int_{-V}^{V}c_{\mathbf{K}'}(\lambda,t)\,d\lambda,\\ &\text{(b) } a^{\gamma}_{\mathbf{K}''},\quad c_{\mathbf{K}''}(\lambda,t),\quad \frac{\delta}{\delta\lambda}c_{\mathbf{K}''}(\lambda,t),\\ &P\int_{-V}^{V}\frac{c_{\mathbf{K}''}(v,t)}{v-\lambda}\,dv,\quad \frac{\delta}{\delta\lambda}P\int_{-V}^{V}\frac{c_{\mathbf{K}''}(v,t)}{v-\lambda}\,dv,\end{aligned} \tag{5.1}$$

each term being formed by the product of one expression (a) by one expression (b), with a coefficient function of λ and t. As will be checked later, the dominant term for large t corresponds to the product

$$a^{\beta}_{\mathbf{K}'}\frac{\delta}{\delta\lambda}c_{\mathbf{K}''}(\lambda,t)$$

and appears in Eq. (3.26) in the following way:

$$\frac{\delta\Gamma_{\mathbf{K}}}{\delta t} = F(\lambda,t)A^{\beta}_{\mathbf{K}'}\exp\{ik'_{\parallel}(\lambda-v^{\beta}_{k'})t\}$$
$$\times\left[-ik''_{\parallel}t\Gamma_{\mathbf{K}'}+\frac{\delta\Gamma_{\mathbf{K}'}}{\delta\lambda}\right]+\cdots, \tag{5.2}$$

where the coefficient $F(\lambda,t)$ is slowly varying in t. Such a term is never resonant as

$$|\lambda|\leqslant V \quad \text{and} \quad |v^{\beta}_{k'}|>V.$$

The corresponding contribution to the evolution of $\Gamma_{\mathbf{K}}$ therefore remains of the same order as the term itself, namely of order $\epsilon^2 t$, that is to say of the same order as the contribution from a normal nonsecular and resonant term. However, the situation becomes worse if we consider the derivative of $\Gamma_{\mathbf{K}}$ with respect to λ. From Eq. (5.2) we then get

$$\frac{\delta^2\Gamma_{\mathbf{K}}}{\delta t\delta\lambda}\sim A^{\beta}_{\mathbf{K}'}\left[k'_{\parallel}k''_{\parallel}t^2\Gamma_{\mathbf{K}'}\right.$$
$$\left.+i(k'_{\parallel}-k''_{\parallel})t\frac{\delta\Gamma_{\mathbf{K}'}}{\delta\lambda}+\frac{\delta^2\Gamma_{\mathbf{K}'}}{\delta\lambda^2}\right]. \tag{5.3}$$

The corresponding term being nonresonant, we may conclude

$$\frac{\delta\Gamma_{\mathbf{K}}(\lambda,t)}{\delta\lambda}-\frac{\delta\Gamma_{\mathbf{K}}(\lambda,0)}{\delta\lambda}\sim\epsilon^2t^2$$

and

$$\frac{\delta\Gamma_{\mathbf{K}}}{\delta\lambda}\sim\max(\epsilon,\epsilon^2t^2).$$

If we consider higher order derivatives, we are led to assume

$$\frac{\delta^n\Gamma_{\mathbf{K}}}{\delta\lambda^n}\sim\max(\epsilon,\epsilon^2t^{n+1}). \tag{5.4}$$

Accepting this estimate, we can now come back to the discussion of Sec. 4. A key point was the fact that the Fourier integral in II×I′ decreased as $1/t^2$ for large t because the second derivative $\delta^2\Gamma_{\mathbf{K}}/\delta\lambda^2$ remained of order ϵ, which apparently is not the case. By using Eq. (5.4) instead and by making a careful evaluation of upper bounds it is possible to show that

$$\int_0^T(\mathrm{II}\times\mathrm{I}')\,dt\sim\epsilon(\epsilon T)^2, \tag{5.5}$$

where the integral is taken over the explicit dependence on t. The same result holds also for the term I×II′, and the term II×II′ remains negligible compared to the others. The order of magnitude (5.5) must be compared to the contribution of the term I×I′ to $A^{\alpha}_{\mathbf{K}}$ which is of order ϵ^2T, so that the condition of validity of Eq. (4.3) is

$$\epsilon(\epsilon T)^2\ll\epsilon^2T \quad \text{or} \quad \epsilon T\ll 1, \tag{5.6}$$

which limits the range of validity for t to an interval much smaller than anticipated initially.

Now we must make sure that the term shown in Eq. (5.2) is dominant for large t in the right-hand side of Eq. (3.26), when the order of the derivatives of $\Gamma_{\mathbf{K}}$ is that given by Eq. (5.4). As a result of the discussion of the various terms given in Appendix D one can reasonably assume that this is the case if

$$\omega_p t\leqslant\omega_p T\sim\epsilon^{-1/2}. \tag{5.7}$$

Then condition (5.6) is also satisfied. Therefore, the simplified equations (4.3) and (4.4) appear to be justified for a time interval limited by Eq. (5.7) but not for $\omega_p t\sim\epsilon^{-1}$.

6. CONCLUSION

Summarizing the results, a closed nonlinear system of equations on the wave amplitudes has been derived and is given by Eqs. (4.4) and (4.5). The main starting assumption is that the nonlinear terms in the Vlasov equation (1.3) are small, and the asymptotic method of Krylov and Bogoliubov or Bogoliubov and Mitropolsky has been used. The result is the same as in Davidson's treatment of coherent wave–wave interaction, but it has been derived in this paper without neglecting the free streaming portions of the distribution functions. A detailed discussion shows that the validity of the approximation does not extend to the full time scale (4.2), but is rather restricted to $\omega_p t\sim\epsilon^{-1/2}$. Finally, it should be recalled that the distribution functions $f_0(v)$ and $f_{\mathbf{K}}(v)$ were assumed to vanish outside a finite interval $-V<v<V$ and that the waves considered had phase velocities outside this interval. Therefore, new features may appear in the theory when resonant wave particle interaction is included and this will be the subject of a subsequent paper.

ACKNOWLEDGMENT

The author wishes to thank Dr. J.L. Soulé for fruitful discussions and pertinent remarks.

APPENDIX A. PROOF OF THE EXPANSION FORMULA

For functions $f(v)$ which vanish outside the interval $(-V,V)$, the operator L_k defined by Eq. (3.2) is seen to be bounded in a certain sense and therefore it is to be expected that the method of the resolvent[11,12] will be well suited to derive an expansion in eigenfunctions of L_k. This derivation is available in a previous report[13] and it will be only outlined below.

The assumptions on the function $f(v)$ to be expanded and on $f'_0(v)$ occurring in L_k have been given in Sec. 3.

The first step is to solve the following equation for the function $h(v,\lambda)$ for real v and complex λ:

$$(L_k-\lambda)h(v,\lambda)=f(v), \tag{A1}$$

with the result

$$h(v,\lambda)=\frac{f(v)}{v-\lambda}+\frac{\omega_p^2}{k^2}\frac{f'_0(v)}{v-\lambda}\frac{K(\lambda)}{\epsilon(\lambda)}, \tag{A2}$$

where $\epsilon(\lambda)$ is given by Eq. (3.3) and where

$$K(\lambda)=\int_{-V}^{V}\frac{f(v)}{v-\lambda}\,dv. \tag{A3}$$

It can be shown that $h(v,\lambda)$ is holomorphic in the complex λ plane outside the zeros of $\epsilon(\lambda)$ and outside the interval $[-V,V]$ of the real axis.

The next step is to show by using lemmas, which are recalled in an earlier paper,[14] that for $\lambda\to\infty$:

$$h(v,\lambda)\sim-\frac{1}{\lambda}f(v)$$

or more precisely that

$$\lim_{|\lambda| \to +\infty} \lambda h(v,\lambda) = -f(v) \tag{A4}$$

uniformly in θ if $\lambda = |\lambda| \exp i\theta$.

Then $h(v,\lambda)$ is integrated in the complex λ plane along a circle $C(R)$ of radius R having its center at the origin. From Eq. (A4) it is easily seen that

$$\lim_{R \to +\infty} \int_{C(R)} h(v,\lambda)\, d\lambda = -2\pi i f(v)$$

and even that

$$\int_{C(R)} h(v,\lambda)\, d\lambda = -2\pi i f(v), \tag{A5}$$

for R large enough so that no zeros of $\epsilon(\lambda)$ are left outside $C(R)$. In fact with the assumptions made in Sec. 3 there are only two such zeroes ν_1 and ν_2.

The method then consists in changing the path of integration in Eq. (A5) by letting it shrink towards the segment $[-V,V]$ of the real axis and by collecting the contributions from the poles ν_1 and ν_2 of $h(v,\lambda)$. In this way the following relation obtains:

$$f(v) = -\frac{1}{2\pi i}\int_{\gamma_1} h\, d\lambda - \frac{1}{2\pi i}\int_{\gamma_2} h\, d\lambda - \frac{1}{2\pi i}\int_{\Gamma} h\, d\lambda, \tag{A6}$$

where γ_1 and γ_2 are small circles around ν_1 and ν_2 and Γ is a rectangle around the segment $[-V,V]$ of the real axis, which can be defined by the four points

$$V+\delta \pm i\eta \quad \text{and} \quad -V-\delta \pm i\eta,$$

$\eta > 0$, $\delta > 0$ being small enough so that ν_1 and ν_2 are outside Γ.

The integrals along γ_1 and γ_2 are easily shown to yield the terms of the discrete spectrum in the expansion formula

$$-\frac{1}{2\pi i}\int_{\gamma_\alpha} h(v,\lambda)\, d\lambda = \frac{(f,\bar{\varphi}_\alpha)}{(\varphi_\alpha,\bar{\varphi}_\alpha)}\varphi_\alpha(v) \quad (\alpha = 1,2).$$

Then the real labor is to go over from the integral along Γ in Eq. (A6) to the contribution of the continuous spectrum in the expansion formula.[13] Suffice it to state here that one can show the following relation:

$$-\frac{1}{2\pi i}\int_{\Gamma} h\, d\lambda = -\frac{1}{2\pi i}\lim_{\delta\to 0}\lim_{\eta\to 0}\int_{\Gamma} h\, d\lambda = \langle {}_v\phi^k_\mu, c(\mu)\rangle,$$

with the notations of Sec. 3 and with expression (3.19) for $c(\lambda)$.

APPENDIX B. SYMMETRY PROPERTY OF THE FUNCTIONAL ϕ

Equation (3.25) expresses a useful symmetry property of the bilinear functional ϕ:

$$(\langle f,\phi\rangle, g) = (f,\langle \phi, g\rangle).$$

Replacing $\langle f,\phi\rangle$ and $\langle \phi,g\rangle$ by their expressions (3.15) and (3.16), respectively, this relation becomes

$$\int_{-V}^{V} d\lambda\, g(\lambda) P\int_{-V}^{V} f(v) f_0'(v)\frac{dv}{v-\lambda} = \int_{-V}^{V} dv\, f(v) f_0'(v) P\int_{-V}^{V} g(\lambda)\frac{d\lambda}{v-\lambda}. \tag{B1}$$

This equality can be proved with the following assumptions on $f_0'(v)$, $f(v)$, and $g(v)$:

(i) f_0' and g vanish outside the interval $[-V,V]$,
(ii) f_0' has two continuous derivatives,
(iii) f and g have one continuous derivative.

Consider the integral

$$K(\epsilon) = \int_{D(\epsilon)} \frac{g(\lambda) f_0'(v) f(v)}{v-\lambda}\, dv\, d\lambda,$$

where the domain of integration $D(\epsilon)$ is obtained from the square $[-V,V]\times[-V,V]$ by removing from it the diagonal strip $|v-\lambda| < \epsilon$.

Let us denote by I and J, respectively, the left-hand and right-hand side of Eq. (B1). The difference $I - K(\epsilon)$ can be written as a sum of integrals which can all be proved to tend to zero as $\epsilon \to 0$, so that

$$I = \lim_{\epsilon\to 0} K(\epsilon).$$

In the same way, it can be shown that

$$J = \lim_{\epsilon\to 0} K(\epsilon),$$

so that

$$I = J.$$

The derivation of the relation

$$\int_{-V}^{V} \langle {}_v\phi^k, c_{\mathbf{K}'}\rangle\, dv = \int_{-V}^{V} c_{\mathbf{K}'}(\lambda)\, d\lambda,$$

then goes as follows:

$$\int_{-V}^{V} \langle {}_v\phi, c\rangle\, dv = (1,\langle \phi, c\rangle) = (\langle 1,\phi\rangle, c) = (1,c) = \int_{-V}^{V} c(\lambda)\, d\lambda.$$

APPENDIX C. COMPARISON WITH R.C. DAVIDSON'S MODE COUPLING EQUATION

The mode coupling equation (4.3) can be shown to agree with R.C. Davidson's result in Chap. 6 of his book.

In order to make the comparison, one must first write Davidson's equation for the case of an infinite magnetic field. The result is fairly easy to guess from the case of no magnetic field [Eqs. (58), (59), and (60) of Chap. 6] and it can be derived unambiguously from the case of a uniform magnetic field [Eqs. (65) and (66)].

Integrating f_0 over $\mathbf{V}_\perp$, writing $v_\parallel = v$ and assuming

$$f_0(v) = 0, \quad \text{for } |v| > V,$$

one finds Davidson's relevant equations in the following form:

$$\frac{\delta}{\delta t}\hat{\phi}^{\alpha}_{\mathbf{K}} = \sum_{\mathbf{K}',\mathbf{K}''}\ \sum_{\beta,\gamma}^{(\mathbf{K}'+\mathbf{K}''=\mathbf{K})} \left[\bar{M}(\mathbf{K},\omega;\mathbf{K}',\omega';\mathbf{K}'',\omega'') \Big/ 2k^2 \frac{\delta\epsilon(\mathbf{K},s)}{\delta s}\Big|_{s=-i\omega} \right] \hat{\phi}^{\beta}_{\mathbf{K}'}\hat{\phi}^{\gamma}_{\mathbf{K}''},$$

$$\left.\frac{\delta\epsilon}{\delta s}\right|_{s=-i\omega} = -i\frac{4\pi e^2}{mk^2}\int_{-V}^{V}\frac{k_\parallel f_0'(v)}{(\omega-k_\parallel v)^2}\,dv,$$

$$\bar{M}(\mathbf{K},\omega;\mathbf{K}',\omega';\mathbf{K}'',\omega'') = \frac{4\pi e^3}{m^2}\int_{-V}^{V}\frac{dv}{\omega-k_\parallel v}\times\left[k_\parallel'\frac{\delta}{\delta v}\frac{k_\parallel'' f_0'(v)}{\omega''-k_\parallel'' v}+\begin{pmatrix}k'\leftrightarrow k''\\ \omega'\leftrightarrow\omega''\end{pmatrix}\right],$$

with the shorthand notation

$$\omega=\omega_{\mathbf{K}}^{\alpha},\quad \omega'=\omega_{\mathbf{K}'}^{\beta},\quad \omega''=\omega_{\mathbf{K}''}^{\gamma}.$$

The correspondence with the notation of this paper is easily established by writing

$$\phi=\sum_{\mathbf{K}}\phi_{\mathbf{K}}e^{i\mathbf{K}\cdot\mathbf{X}},\quad \phi_{\mathbf{K}}=\sum_{\alpha}\hat{\phi}_{\mathbf{K}}^{\alpha}\exp(-i\omega_{\mathbf{K}}^{\alpha}t),$$

$$\mathbf{E}_{\mathbf{K}}=-i\mathbf{K}\phi_{\mathbf{K}}=-i\mathbf{K}\frac{m}{e}A_{\mathbf{K}}^{\alpha}\exp(-\omega_{\mathbf{K}}^{\alpha}t),$$

and thus

$$\hat{\phi}_{\mathbf{K}}^{\alpha}=\frac{m}{e}A_{\mathbf{K}}^{\alpha}.$$

Further, it is easily seen that

$$\left.\frac{\delta\epsilon}{\delta s}\right|_{s=-i\omega_{\mathbf{K}}^{\alpha}} = -i\frac{\omega_p^2}{k_\parallel k^2}(\bar{\varphi}_{\mathbf{K}}^{\alpha},\varphi_{\mathbf{K}}^{\alpha}),$$

taking into account that $f_0(v)$ is normalized to 1 in this paper.

In this way Davidson's equation can be translated into the notations of this paper with the following result:

$$\frac{\delta}{\delta t}A_{\mathbf{K}}^{\alpha}=\sum_{\mathbf{K}',\mathbf{K}''}\sum_{\beta,\gamma}\frac{m}{e}\frac{ik_\parallel}{\omega_p^2}\times\frac{(1/2)\bar{M}(\mathbf{K},\omega;\mathbf{K}',\omega';\mathbf{K}'',\omega'')}{(\bar{\varphi}_{\mathbf{K}}^{\alpha},\varphi_{\mathbf{K}}^{\alpha})}A_{\mathbf{K}'}^{\beta}A_{\mathbf{K}''}^{\gamma},$$

where $(1/2)\bar{M}$ can be replaced by the unsymmetrized matrix

$$M_1(\mathbf{K},\omega;\mathbf{K}',\omega';\mathbf{K}'',\omega'')=\frac{e}{m}\omega_p^2\frac{k_\parallel'}{k_\parallel}(\bar{\varphi}_{\mathbf{K}}^{\alpha},\frac{\delta}{\delta v}\varphi_{\mathbf{K}''}^{\gamma})$$

so that the following final form is obtained:

$$\frac{\delta}{\delta t}A_{\mathbf{K}}^{\alpha}=\sum_{\mathbf{K}',\mathbf{K}''}\sum_{\beta,\gamma}ik_\parallel'\frac{(\bar{\varphi}_{\mathbf{K}}^{\alpha},(\delta/\delta v)\varphi_{\mathbf{K}''}^{\gamma})}{(\bar{\varphi}_{\mathbf{K}}^{\alpha},\varphi_{\mathbf{K}}^{\alpha})}A_{\mathbf{K}'}^{\beta}A_{\mathbf{K}''}^{\gamma},$$

which is identical to Eq. (4.3).

APPENDIX D

We consider the various possible terms in the rhs of the equation of evolution of $\Gamma_{\mathbf{K}}$ as given by Eq. (5.1), and we wish to check that the term shown in Eq. (5.2) is dominant when the derivatives of $\Gamma_{\mathbf{K}}$ are of the order of magnitude (5.4). According to the asymptotic method described in Sec. 2, we should compare the various terms resulting from Eq. (5.1) after integration over the "explicit time dependence" from 0 to T with $T=O(\epsilon^{-1/2})$. Considering first the term shown in Eq. (5.2), which was assumed to be dominant, we find the following expression:

$$A_{\mathbf{K}'}^{\beta}\Gamma_{\mathbf{K}''}\int_0^T F(\lambda,t)\exp ik'(\lambda-\nu_{k'}^{\beta})t\,dt\sim\epsilon^2T\sim\epsilon^{3/2},$$

taking into account that $\lambda-\nu_{k'}^{\beta}\neq 0$, as $|\lambda|<V$ and $|\nu_{k'}^{\beta}|>V$. (We write k' and k'' instead of $k_\parallel'$ and $k_\parallel''$.)

Considering now one by one all the other products of an expression (a) by an expression (b) of Eq. (5.1), one can show that they all yield a contribution of order higher than $\epsilon^{3/2}$, with the following two exceptions:

$$\left[\int_{-V}^{V}c_{\mathbf{K}'}(u,t)\,du\right]\frac{\delta}{\delta\lambda}c_{\mathbf{K}''}(\lambda,t), \tag{D1}$$

$$a_{\mathbf{K}'}\frac{\delta}{\delta\lambda}P\int_{-V}^{V}\frac{c_{\mathbf{K}''}(u,t)}{u-\lambda}\,du. \tag{D2}$$

The latter expression turns out to be especially difficult to deal with. It can be first decomposed in the following way:

$$a_{\mathbf{K}'}\frac{\delta}{\delta\lambda}\left\{\int_{-V}^{V}\frac{\Gamma_{\mathbf{K}''}(u,t)-\Gamma_{\mathbf{K}''}(\lambda,t)}{u-\lambda}e^{-ik''ut}+\Gamma_{\mathbf{K}''}(\lambda,t)P\int_{-V}^{V}\frac{e^{-ik''ut}}{u-\lambda}\,du\right\}, \tag{D3}$$

where the last integral can be written

$$P\int_{-V}^{V}\frac{e^{-ik''ut}}{u-\lambda}\,du=\int_C\frac{e^{-ik''ut}}{u-\lambda}\,du-\frac{k''}{|k''|}i\pi e^{-ik''\lambda t}, \tag{D4}$$

where the contour C avoids the pole λ with a small semicircle below the real axis if $k''t>0$ and above the real axis if $k''t<0$. The integral along C can then be shown to decrease fast enough with t as $t\to+\infty$ so that it causes no problem. The last term in Eq. (D4) gives rise to a term of the same form as Eq. (5.2) in the equation of evolution of $\Gamma_{\mathbf{K}}$. We are thus left with the first part of Eq. (D3), namely,

$$a_{\mathbf{K}'}\frac{\delta}{\delta\lambda}\int_{-V}^{V}\frac{\Gamma_{\mathbf{K}''}(u,t)-\Gamma_{\mathbf{K}''}(\lambda,t)}{u-\lambda}e^{-ik''ut}\,du. \tag{D5}$$

From the estimates (5.4) of the derivatives of $\Gamma_{\mathbf{K}}$, it is possible to show that the integral of Eq. (D5) over t from 0 to T is of order higher than ϵ for $T\sim\epsilon^{-1/2}$, but it does not seem to be possible to show that it is of order higher than $\epsilon^{3/2}$ without additional knowledge on $\Gamma_{\mathbf{K}''}(u,t)$. However, it become possible by using a "model dependence" of $\Gamma_{\mathbf{K}''}(u,t)$ on u and t.

Let us consider a function $\Gamma(u,t)$ satisfying the simple equation

$$\frac{\delta\Gamma}{\delta t}=\alpha(u,t)\Gamma,\quad\text{with}\quad\alpha(u,t)=-i\epsilon lte^{i(\omega+mu)t}, \tag{D6}$$

where ω, k, l, and m are real and $k=1+m$. Thus,

$$\Gamma(u,t)=\Gamma_0(u)e^{\int_0^t\alpha(u,s)\,ds}. \tag{D7}$$

Equation (D6) has the essential features of Eq. (5.2) and the derivatives of $\Gamma(u,t)$ given by Eq. (D7) can be shown to satisfy condition (5.4) if $\Gamma_0(u)$ is itself of order ϵ.

Then we wish to study the expression

$$H_V(\lambda,t)=\frac{\delta}{\delta\lambda}\int_{-V}^{V}\frac{\Gamma(u,t)-\Gamma(\lambda,t)}{u-\lambda}e^{-iku t}\,du. \tag{D8}$$

Considering the limiting case $V\to\infty$ for simplicity, it is possible to show the following:

This paper	Van Kampen	Case
$f_0'(v)$	$-F(v)$	$-\frac{k^2}{\omega_p^2}\eta(v)$
$f_k(v)$	$f(k,v)$	$g(v)$
${}_v\phi_\lambda^k$	$\bar{g}_{k,\lambda}(v)$	$g_\lambda(v)$
$c_k(u)$	$C(k,u)$	$A(u)$
$\epsilon_0(k,u)$	$\frac{\omega_p^2}{k^2}F(u)\lambda(k,u)$	$\lambda(u)$
$\epsilon_\pm(k,u)$	$\frac{\omega_p^2}{k^2}F(u)[\lambda(k,u)\pm i\pi]$	$\lambda(u)\pm i\pi\eta(u)$

$$K(\lambda,t)=\int_0^T H_\infty(\lambda,t)\,dt=O\left(\epsilon^{3/4}\log\frac{1}{\epsilon}\right)$$

if

$$T=O(\epsilon^{-1/2}) \text{ and } \Gamma_0(u)=O(\epsilon).$$

This result holds uniformly in λ outside fixed intervals around the "resonances" λ_0, λ_1:

$$|\lambda-\lambda_0|>\delta, \quad |\lambda-\lambda_1|>\delta',$$

where λ_0 and λ_1 are defined by

$$\omega+m\lambda_0=0 \quad \text{and } \omega-l\lambda_1=0,$$

respectively.

The corresponding order of magnitude for the contribution of Eq. (D2) to the evolution of $\Gamma_{\mathbf{K}}$ would then be of order $\epsilon^{7/4}\log(1/\epsilon)$ which tends to zero faster than $\epsilon^{3/2}$, although admittedly by a small margin.

APPENDIX E

The correspondence between the notations used in this paper, by Van Kampen[8] and by Case,[9] is given in the table.

However, it should be noted that the problem treated by Van Kampen and Case is that of one-dimensional Langmuir waves in a plasma without magnetic field. Then the wave number is a scalar k instead of the vector $\mathbf{K}$ of this paper.

[1]V.N. Tsytovich, *Nonlinear Effects in Plasma* (Plenum, New York, 1970).
[2]R.C. Davidson, *Methods of Nonlinear Plasma Theory* (Academic, New York, 1972).
[3]C.S. Gardner, Phys. Fluids **6**, 839 (1963).
[4]E.A. Frieman, *et al.*, Phys. Fluids **6**, 1298 (1963). E.A. Frieman and P.R. Rutherford, Ann. Phys. **28**, 134 (1964).
[5]N. Krylov and N.N. Bogoliubov, *Introduction to Nonlinear Mechanics* (Princeton University, Princeton, 1947).
[6]N.N. Bogoliubov and I. Mitropolsky, *Les Méthodes Asymptotiques en Théorie des Oscillations Nonlinéaires* (Gauthier-Villars, Paris, 1962).
[7]N. Minorsky, *Nonlinear Oscillations* (Van Nostrand, New York, 1962).
[8]N.G. Van Kampen, Physica (Utrecht) **21**, 949 (1955).
[9]K.M. Case, Ann. Phys. N.Y. **7**, 349 (1959).
[10]R.E. Aamodt and W.E. Drummond, Phys. Fluids **7**, 1816 (1964).
[11]T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1966), Chap. 1.
[12]E.C. Titchmarsh, *Eigenfunction Expansion Associated with Second Order Differential Equations* (Oxford University, London, 1958).
[13]M. Trocheris, "Méthodes asymptotiques dans la théorie des ondes de Langmuir cohérentes," Rapport EURATOM-CEA-n° 864 (1976).
[14]M. Trocheris, Nucl. Fusion **5**, 299 (1965); M. Trocheris, Symp. Appl. Math. Am. Math. Soc. **28**, 100 (1967).

Weakly nonlinear theory of coherent Langmuir waves. II. Wave particle interaction

M. Trocheris

Association Euratom-CEA Sur La Fusion, Département de Physique du Plasma et de la Fusion Contrôlée, Centre d'Etudes Nucléaires-Boîte Postale No. 6, 92260 FONTENAY-AUX-ROSES, FRANCE

(Received 8 August 1979; accepted for publication 30 November 1979)

A weakly nonlinear analysis was carried out in Part 1 in the case of Langmuir waves, by neglecting the tail of the distribution function where particle velocities can be equal to phase velocities of waves. The treatment is extended in Part 2 by including particles of all velocities, so that resonant wave particle interaction becomes possible. The asymptotic theory of Krylov and Bogolioubov is applied to the Vlasov system of equations after writing it in terms of unknown functions which are constants of the motion of the linearized equations. A system of such constants can be found with the help of a suitable expansion in normal modes, which differs from Van Kampen's expansion by the fact that weakly damped waves are treated on the same footing as weakly growing waves. It is then found that the free streaming portions of the distribution functions can no longer be left out, and the final result is a system of equations on both the wave amplitudes and the free streaming perturbations, which are coupled through resonant wave particle interaction.

1. INTRODUCTION

The treatment presented in Part 1[1] will be extended to include resonant wave–particle interaction and as a result the wave amplitudes will no longer be found to evolve independently of the perturbed distribution functions. A system of equations will be derived for the coupled evolution of wave amplitudes and distribution functions corresponding to free-streaming perturbations. This system has already been derived by a heuristic method.[2] It contains a new type of interaction between a wave and a free-streaming perturbation which has recently been studied experimentally.[3] This new interaction is also closely connected with a previous result found by R.W.B. Best[4] in his analysis of nonlinear plasma oscillations and recently confirmed by M. Gros *et al.*[5] with the help of the multiple water-bag model. In fact the latter authors mention in the introduction of their paper that a problem remains because this new type of interaction is not found in the usual multiple time-scale perturbation treatment of the Vlasov equation. This may refer in particular to R.C. Davidson's work[6] in which wave–wave interaction is treated in this way but by neglecting the free-streaming portions of the distribution function. As was mentioned previously[2] the equations derived in the present paper provide the required complement and the method used is equivalent to the multiple time-scale method in the lowest significant order.

As compared to Part 1, the new feature in this paper will be that the distribution functions will no longer be assumed to vanish outside a finite interval of the velocity v so that first of all Landau damping will come into play. Essentially the same scheme as in Part 1 will be followed, but in order to have a satisfactory physical description, slightly damped waves and weakly unstable waves will be treated on the same footing. This is possible, but it requires a much deeper modification of Van Kampen's expansion[7,8] which can be made in the appropriate way by using earlier results.[9]

The functions of the velocities which will enter the final system of equations are suggested by the asymptotic form for large times of the solution of the linearized equations. This asymptotic form can be written in the following way[2,10]:

$$f_k(v,t) \simeq \sum_\alpha a_k^\alpha(t)\,\frac{f_0'(v)}{v - v_k^\alpha} + \psi_k(v,t), \tag{1.1}$$

with

$$a_k^\alpha(t) = a_k^\alpha(0)e^{-ikv_k^\alpha t}, \quad \psi_k(v,t) = \psi_k(v,0)e^{-ikvt}. \tag{1.2}$$

The sum over α extends to the roots of the dispersion relation and $\psi_k(v,t)$ represents a free streaming perturbation. It will be seen that the asymptotic method already used in Part 1 leads naturally to a description in terms of a slow evolution of the asymptotic form (1.1). In fact this is to be expected if the wave amplitudes are small enough, since the solution of the linearized equation will always find the time to relax to its asymptotic form (1.1), before it is substantially altered by the effect of the nonlinear terms. This remark is the basis of the heuristic treatment referred to above.[2]

As we are not especially interested in mode coupling, which was treated in Part 1, we will consider in the following the simple problem of one dimensional Langmuir waves in a plasma without magnetic field. The system of the Vlasov–Maxwell equations can then be written in the following form:

$$f(x,v,t) = f_0(v,t) + \sum_{k\neq 0} e^{ikx} f_k(v,t), \quad \left(\int_{-\infty}^{+\infty} f_0\,dv = 1\right) \tag{1.3}$$

$$\frac{\partial f_k}{\partial t} + ikvf_k + \frac{e}{m}E_k\frac{\partial f_0}{\partial v} = -\frac{e}{m}\sum_{q\neq 0} E_{k-q}\frac{\partial f_q}{\partial v}, \tag{1.4}$$

$$\frac{\partial f_0}{\partial t} = -\frac{e}{m}\sum_{k\neq 0} E_{-k}\frac{\partial f_k}{\partial v}, \tag{1.5}$$

$$E_k = -i4\pi e n_0/k \int_{-\infty}^{+\infty} f_k\,dv. \tag{1.6}$$

The theory will be developed for this system and the

 0022-2488/80/040941-08$1.00

result will also be given for the case of an infinite magnetic field considered in Part 1.

The asymptotic method used in the following will be the same as in Part 1, to which the reader can refer. The appropriate expansion in normal modes is presented in Sec. 2. Then in Sec. 3, the initial system of equations (1.4)(1.5)(1.6). is transformed into a new system of equations on the new unknown functions $a_k^\alpha(t)$ and $c_k(\lambda,t)$, which are the coefficients of the discrete and continuous spectrum in the expansion of $f_k(v,t)$. In Sec. 4, the asymptotic method is applied to this new system and the final equations are obtained in terms of $a_k^\alpha(t)$ and $\psi_k(v,t)$. A number of properties of these equations are discussed in Sec. 5 and the result for the case of an infinite magnetic field is given in Sec. 6.

2. EXPANSION IN NORMAL MODES

It will be seen that it is possible to describe weakly damped waves as normal modes in the same way as weakly unstable waves by using a suitable expansion of the distribution function $f_k(v,t)$. This expansion is easily obtained from previous work[9] through a slight modification which is given in Appendix A.

The functions f_0 and f_k are assumed to be defined in the complex v plane and holomorphic in v on both sides of the real axis, for values of the imaginary part of v which may be at most of the order of ω/k_0 if k_0 is the smallest value of k in the Fourier series (1.3). In addition, conditions have to be imposed on the behaviour of these functions as the real part of v tends to infinity.[9]

The starting point is the linear operator L_k defined by writing the linearized Vlasov equation in the following form:

$$\frac{\partial f_k}{\partial t} + ik\,L_k f_k = 0\,. \tag{2.1}$$

With the assumptions made on $f_0(v)$, this operator can be written:

$$L_k f(v) = vf(v) - \frac{\omega_p^2}{k^2}\frac{\partial f_0}{\partial v}\int_\Delta f(u)\,du\,, \tag{2.2}$$

where the path of integration over u is a straight line Δ parallel to the real axis in the complex v plane, which will be suitably chosen for each value of k. It is also useful to consider the adjoint operator L_k^+ defined by

$$(L_k f,g)_k = (f,L_k^+ g)_k, \tag{2.3}$$

with the scalar product

$$(f,g)_k = \int_\Delta f(v)g(v)\,dv\,. \tag{2.4}$$

The main properties of these operators and the corresponding expansion formula are given below. Further details and precise conditions of validity can be found in a previous paper.[9]

A. Discrete spectrum

The discrete spectrum is composed of zeros of the usual dispersion functions

$$\epsilon_\pm(k,\lambda) = 1 - \frac{\omega_p^2}{k^2}\int_{C_\pm}\frac{f_0'(v)}{v-\lambda}\,dv, \quad \left(f_0' = \frac{\partial f_0}{\partial v}\right) \tag{2.5}$$

with the usual path of integration consisting of the real axis plus possibly a loop around the pole, so as to leave the pole above C_+ and below C_-.

It is easily seen that a zero ν_k^α of $\epsilon_+(k,\lambda)$ lying above Δ as well as a zero of $\epsilon_-(k,\lambda)$ lying below Δ is an eigenvalue of L_k and the following relations hold

$$L_k\varphi_k^\alpha(v) = \nu_k^\alpha\varphi_k^\alpha(v), \quad \varphi_k^\alpha(v) = \frac{f_0'(v)}{v-\nu_k^\alpha}, \tag{2.6}$$

$$L_k^+\bar\varphi_k^\alpha(v) = \nu_k^\alpha\bar\varphi_k^\alpha(v), \quad \bar\varphi_k^\alpha(v) = \frac{1}{v-\nu_k^\alpha}. \tag{2.7}$$

For a positive value of k, zeros of $\epsilon_+(k,\lambda)$ lying above the real axis correspond to unstable modes and zeros lying below the real axis correspond to damped modes. By choosing Δ a suitable distance below the real axis it will be possible to include slightly damped modes in the discrete spectrum as well as unstable modes. Similarily, for a negative value of k, Δ will be chosen a suitable distance above the real axis.

B. Continuous spectrum

The continuous spectrum could be thought of as consisting of all the points of Δ in the complex plane. In fact one must rather consider a strip $B(k)$ parallel to the real axis, extending on both sides of Δ, and containing no zeros of $\epsilon_+(k,\lambda)$ nor $\epsilon_-(k,\lambda)$. The role of eigenfunctions is played by linear functionals operating on functions of v or λ which are defined in $B(k)$. These eigenfunctionals can be defined by the following relations in a way similar to (3.15)(3.16) of Part 1.

$$\langle f,\phi_\lambda^k\rangle = \frac{\omega_p^2}{k^2}\int_{C_+} f_0'(v)f(v)\frac{dv}{v-\lambda} + \epsilon_+(k,\lambda)f(\lambda)\,, \tag{2.8}$$

$$\langle f,\bar\phi_\lambda^k\rangle = \frac{\omega_p^2}{k^2}f_0'(\lambda)\int_{C_+} f(v)\frac{dv}{v-\lambda} + \epsilon_+(k,\lambda)f(\lambda)\,, \tag{2.9}$$

$$\langle {}_v\phi^k,g\rangle = \frac{\omega_p^2}{k^2}f_0'(v)\int_{\Gamma_+} g(\lambda)\frac{d\lambda}{v-\lambda} + \epsilon_+(k,v)g(v)\,, \tag{2.10}$$

$$\langle {}_v\bar\phi^k,g\rangle = \frac{\omega_p^2}{k^2}\int_{\Gamma_+} f_0'(\lambda)g(\lambda)\frac{d\lambda}{v-\lambda} + \epsilon_+(k,v)g(v)\,. \tag{2.11}$$

The notation for the contours is the same as in (2.5): C_+ and Γ_+ pass below the pole, C_- and Γ_- pass above the pole.

The result of L_k operating on ϕ_λ^k can be defined by

$$\langle f,L_k\phi_\lambda^k\rangle = \langle L_k^+ f,\phi_\lambda^k\rangle, \tag{2.12}$$

and then one may write

$$L_k\phi_\lambda^k = \lambda\phi_\lambda^k, \tag{2.13}$$

and in a similar way

$$L_k^+\bar\phi_\lambda^k = \lambda\bar\phi_\lambda^k\,. \tag{2.14}$$

The following symmetry relations hold and prove useful[9]:

$$(\langle f,\phi^k\rangle,g)_k = (f,\langle\phi^k,g\rangle)_k, \tag{2.15}$$

$$(\langle f,\bar\phi^k\rangle,g)_k = (f,\langle\bar\phi^k,g\rangle)_k\,. \tag{2.16}$$

C. Orthogonality relations

The eigenfunctions and functionals satisfy the following orthogonality relations:

$$(\bar{\varphi}^{\alpha}_{k},\varphi^{\beta}_{k})_{k}=0, \quad (\nu^{\alpha}_{k}\neq\nu^{\beta}_{k}) \tag{2.17}$$

$$\langle\bar{\varphi}^{\alpha}_{k},\phi^{k}_{\lambda}\rangle=0, \quad \langle\varphi^{\alpha}_{k},\bar{\phi}^{k}_{\lambda}\rangle=0, \quad (\lambda\in B(k)). \tag{2.18}$$

D. Expansion formula

Let $f(v)$ have the same properties as we have assumed for $f_0(v)$ and $f_k(v)$ in the same domain of the complex v plane. Then for all v in this domain, the following expansion formula holds

$$f(v)=\sum_{\alpha}a^{\alpha}_{k}\varphi^{\alpha}_{k}(v)+\langle{}_{v}\phi^{k}_{\lambda},c(\lambda)\rangle_{B(k)}. \tag{2.19}$$

The sum over α extends to the eigenvalues of the discrete spectrum, i.e., to the zeros of $\epsilon_{+}(k,\lambda)$ lying above $B(k)$ and to the zeros of $\epsilon_{-}(k,\lambda)$ lying below $B(k)$. The function $c(\lambda)$ which plays the part of the coefficient for the continuous spectrum is defined for λ in $B(k)$. The eigenfunctional ${}_{v}\phi^{k}_{\lambda}$ is defined by (2.10) when v is in $B(k)$. For v outside $B(k)$ the expression used above in (2.19) is defined by analytic continuation, which is straight forward by suitably choosing one of the two expressions given in (2.10). This definition is recalled by the index $B(k)$ in (2.19).

The expansion formula is completed by the expressions for the coefficients a^{α}_{k} and $c(\lambda)$ in terms of $f(v)$

$$a^{\alpha}_{k}=(\bar{\varphi}^{\alpha}_{k},f)_{k}/(\bar{\varphi}^{\alpha}_{k},\varphi^{\alpha}_{k})_{k}, \tag{2.20}$$

$$c(\lambda)=\langle f,\bar{\phi}^{k}_{\lambda}\rangle/\epsilon_{+}(k,\lambda)\epsilon_{-}(k,\lambda). \tag{2.21}$$

In principle there is an expansion for every choice of $B(k)$. In practice $B(k)$ will be chosen so as to include all weakly damped modes in the discrete spectrum, which implies that $B(k)$ lies below the real axis for $k>0$ and above the real axis for $k<0$.

The choice of $B(k)$ implies that a distinction can be made between weakly damped modes which are included in the discrete spectrum and heavily damped modes which are absorbed in the continuous spectrum. The same distinction is usually made in the treatment of the linearized equation through Laplace transformation, so that for instance all weakly damped modes appear in the time asymptotic form (1.1) of the solution.

When comparing our result (2.19) with Van Kampen's expansion the discrete spectrum of eigenvalues appears to be extended into the complex half plane corresponding to damped modes by virtue of the definition (2.2) of the operator L_k.

3. TRANSFORMATION OF THE EQUATIONS

The Vlasov system of equations (1.4)(1.5)(1.6) can be transformed by replacing f_k for each k by its expansion (2.19), provided that $f_k(v)$ and $f_0(v)$ are assumed to have the required properties of analyticity. We will carry out this transformation and thus obtain a system of equations on the coefficients a^{α}_{k} and $c_k(\lambda)$ which occur in the expansion of $f_k(v)$.

The Vlasov system of equations can be written in the following form

$$\frac{\partial f_k}{\partial t}+ikL_kf_k=Q_k, \tag{3.1}$$

$$Q_k=i\omega_p^2\sum_{\substack{k'+k''=k\\ k'k''\neq 0}}\frac{1}{k'}\left(\int_{-\infty}^{+\infty}f_{k'}\,dv\right)\frac{\partial f_{k''}}{\partial v}, \tag{3.2}$$

together with the equation of evolution for f_0 (1.5). In the same way as in Part 1 we will make the calculations by neglecting the time dependence in f_0 and we will justify this approximation later.

We start from the expressions (2.20) and (2.21) for the coefficients a^{α}_{k} and $c_k(\lambda)$ of the expansion of $f_k(v)$. By differentiating with respect to t and by using (3.1)(2.3)(2.7)(2.14), we obtain as in Part 1

$$(\bar{\varphi}^{\alpha}_{k},\varphi^{\alpha}_{k})_{k}\left(\frac{da^{\alpha}_{k}}{dt}+ik\nu^{\alpha}_{k}a^{\alpha}_{k}\right)=(\bar{\varphi}^{\alpha}_{k},Q_k)_{k}, \tag{3.3}$$

$$\epsilon_{+}(k,\lambda)\epsilon_{-}(k,\lambda)\left[\frac{\partial}{\partial t}c_k(\lambda,t)+ik\lambda c_k(\lambda,t)\right]=\langle Q_k,\bar{\phi}^{k}_{\lambda}\rangle. \tag{3.4}$$

In the expression (3.2) for Q_k the distribution functions $f_{k'}$ and $f_{k''}$ have to be replaced by their expansions according to (2.19). For $f_{k'}$ the calculation can be made as in Part 1, by using the symmetry relation (2.15) with the result:

$$n_{k'}=\int_{-\infty}^{+\infty}f_{k'}(v)\,dv=\int_{\Delta}f_{k'}(v)\,dv$$

$$=\sum_{\beta}\frac{k'^2}{\omega_p^2}a^{\beta}_{k'}+\int_{\Delta}c_{k'}(\lambda)d\lambda \qquad (\Delta\subset B(k)). \tag{3.5}$$

In replacing $f_{k''}$ by its expansion, a problem arises with the expression:

$$(\bar{\varphi}^{\alpha}_{k},Q_k)_{k}=i\omega_p^2\sum_{(k'+k''=k)}\frac{1}{k'}n_{k'}\left(\bar{\varphi}^{\alpha}_{k},\frac{\partial f_{k''}}{\partial v}\right)_{k}. \tag{3.6}$$

By definition of the scalar product, the integration over v is taken along a parallel to the real axis in $B(k)$ and thanks to the properties of $f_{k''}$ it can also be taken along any parallel to the real axis with a suitable loop around ν^{α}_{k}. But a more precise choice of contours must be made when integrating the various terms arising from the expansion of $f_{k''}$ as they contain poles for $v=\nu^{\gamma}_{k''}$. We will take contours which pass these poles on the same side as a parallel to the real axis in $B(k'')$ does. This choice will be recalled by adding a subscript k'' to the scalar products in addition to the subscript k.

A similar problem is encountered with $\langle Q_k,\bar{\phi}^{k}_{\lambda}\rangle$, namely the functional $\bar{\phi}^{k}_{\lambda}$ must be defined for functions which have poles at the $\nu^{\gamma}_{k''}$. We will then take the definition (2.9) of $\bar{\phi}^{k}_{\lambda}$ inside $B(k'')$ and extend it outside by analytic continuation. This will be indicated by $B(k'')$ added as a subscript to the functional.

Making use of these notations, the calculation is straightforward and the system of equations on a^{α}_{k} and $c_k(\lambda)$ is easily obtained. It is convenient to write it in terms of A^{α}_{k} and $\Gamma_k(\lambda)$ which are the constants of the motion in the linear approximation and which are defined by:

$$a^{\alpha}_{k}(t)=\exp(-ik\nu^{\alpha}_{k}t)A^{\alpha}_{k}(t), \tag{3.7}$$

$$c_k(\lambda,t)=\exp(-ik\lambda t)\Gamma_k(\lambda,t). \tag{3.8}$$

The resulting system of equations is composed of two

parts which are coupled through the quadratic terms. The first part gives the time evolution of A_k^α and can be written in the following way:

$$(\bar{\varphi}_k^\alpha,\varphi_k^\alpha)_k \frac{dA_k^\alpha}{dt} = \sum_{(k'+k''=k)} ik'[\mathrm{I}+\mathrm{II}][\mathrm{I}'+\mathrm{II}'], \quad (3.9)$$

$$\mathrm{I} = \sum_\beta e^{i(\omega-\omega')t} A_{k'}^\beta, \quad \mathrm{II} = \frac{\omega_p^2}{k'^2} e^{i\omega t} \int_\Delta e^{-ik'\lambda t} \Gamma_{k'}(\lambda,t)\, d\lambda,$$

$$\mathrm{I}' = \sum_\gamma e^{-i\omega'' t} A_{k''}^\gamma \left(\bar{\varphi}_k^\alpha, \frac{\partial \varphi_{k''}^\gamma}{\partial v}\right)_{k,k''} \quad (\Delta \subset B(k')),$$

$$\mathrm{II}' = -\left(\frac{\partial \bar{\varphi}_k^\alpha}{\partial v}, \langle {}_v\phi_\lambda^{k''}, e^{-ik''\lambda t}\Gamma_{k''}(\lambda,t)\rangle_{B(k'')}\right)_{k,k''},$$

with the notation:

$$\omega = kv_k^\alpha, \quad \omega' = k'v_{k'}^\beta, \quad \omega'' = k''v_{k''}^\gamma. \quad (3.10)$$

The second part gives the time evolution of $\Gamma_k(\lambda,t)$ and can be written in a similar way

$$\epsilon_+(k,\lambda)\epsilon_-(k,\lambda)\frac{\partial}{\partial t}\Gamma_k(\lambda,t) = \sum_{(k'+k''=k)} ik'[\mathrm{I}+\mathrm{II}][\mathrm{I}'+\mathrm{II}'], \quad (3.11)$$

$$\mathrm{I} = \sum_\beta e^{i(k\lambda-\omega')t} A_{k'}^\beta, \quad \mathrm{II} = \frac{\omega_p^2}{k'^2} e^{ik\lambda t} \int_\Delta e^{-ik'\mu t}\Gamma_{k'}(\mu,t)\, d\mu,$$

$$\mathrm{I}' = \sum_\gamma e^{-i\omega'' t} A_{k''}^\gamma \left\langle \frac{\partial \varphi_{k''}^\gamma}{\partial v}, {}_v\bar{\phi}_\lambda^k \right\rangle_{B(k'')} \quad (\Delta \subset B(k')),$$

$$\mathrm{II}' = \left\langle \frac{\partial}{\partial v}\langle {}_v\phi_\mu^{k''}, e^{-ik''\mu t}\Gamma_{k''}(\mu,t)\rangle_{B(k'')}, {}_v\bar{\phi}_\lambda^k \right\rangle_{B,(k'')}$$

4. APPLICATION OF THE ASYMPTOTIC METHOD

In this section, the asymptotic method will be applied to the transformed system of equations (3.9)–(3.11). But before doing so, it is necessary to rediscuss the choice of the small parameter ϵ and to study certain integrals which occur in the calculations.

A. The small parameter of the asymptotic theory

By definition of the small parameter ϵ:

$$f_k \sim E_k \sim a_k^\alpha \sim c_k \sim \epsilon \quad (4.1)$$

so that the right-hand sides of Eq. (3.9)–(3.11) appear to be of order ϵ^2. However, as was already apparent in Part 1, the expansion of $\partial f_{k''}/\partial v$ introduces expressions of the form:

$$\frac{\partial}{\partial v}(e^{-ik''vt}F(v)) \sim \epsilon t, \text{ with } F(v) \sim \epsilon. \quad (4.2)$$

The corresponding terms in (3.9)–(3.11) are of order $\epsilon^2 t$ and can bring contributions of order $\epsilon^2 t^2$ to the evolution of A_k^α and Γ_k. The validity of the approximation will therefore be limited to

$$\epsilon^2 t^2 = O(\epsilon) \text{ and therefore } t = O(\epsilon^{-1/2}).$$

On the same time scale, the mode coupling terms which appear in Eq. (3.9) do not contain expressions of the form (4.2) and will therefore bring a contribution of order $\epsilon^{3/2}$ to the evolution of A_k. Such terms are interesting physically and should therefore be retained. Therefore in the right-hand side of (3.9), terms will be neglected only if they bring contributions which are of higher order in ϵ than $\epsilon^{3/2}$. A similar remark can be made on Eq. (3.11) and the same rule will be followed in neglecting terms in the right hand side.

We now come back to the time dependence of f_0 which was neglected in deriving Eqs. (3.9)–(3.11). This approximation can be justified in the same way as in Part 1. Thus:

$$\frac{\partial f_0}{\partial t} \sim \epsilon^{3/2},$$

$$(\bar{\varphi}_k^\alpha,\varphi_k^\alpha)_k \frac{da_k^\alpha}{dt} = \left(\bar{\varphi}_k^\alpha, \frac{\partial f_k}{\partial t}\right)_k + O(\epsilon^{5/2}).$$

The terms omitted in Eq. (3.3) and similarly in Eq. (3.4), are therefore of order $\epsilon^{5/2}$ and can be neglected, as they would give contributions of order ϵ^2 to A_k^α and Γ_k for $t = O(\epsilon^{-1/2})$.

B. Study of certain integrals

We shall have to study the behaviour of Fourier integrals of the form:

$$J(kt) = \int_{-\infty}^{+\infty} e^{-ikvt} f(v)\, dv, \quad (4.3)$$

as $t \to +\infty$. It was seen in Part 1 that this integral is $O(t^{-2})$ if $f(v)$ has two derivatives. But we shall have to deal with functions $f(v)$ which are quite regular for real v but which have a pole close enough to the real axis to have a large influence on the value of the integral. The integral can then be written

$$J(kt) = \int_{-\infty}^{+\infty} \frac{g(v)}{v-\mu} e^{-ikvt}\, dv, \quad (4.4)$$

where μ is complex with a small imaginary part and where $g(v)$ is analytic in v and has no singularity close to the real axis. In that case one can write:

$$J(kt) = \int_{-\infty}^{+\infty} \frac{g(v)-g(\mu)}{v-\mu} e^{-ikvt}\, dv + g(\mu)\int_{-\infty}^{+\infty} \frac{e^{-ikvt}}{v-\mu}\, dv. \quad (4.5)$$

The first integral has the same properties as (4.3) when $f(v)$ has no singularity and the second integral has a simple value:

$$\int_{-\infty}^{+\infty} \frac{e^{-ikvt}}{v-\mu}\, dv = e^{-ik\mu t} \times \begin{cases} 0, & \text{if } k\,\mathrm{Im}\mu > 0 \\ -2\pi i \dfrac{k}{|k|}, & \text{if } k\,\mathrm{Im}\mu < 0 \end{cases} \quad (t>0). \quad (4.6)$$

This result is easily generalized to the case where the integral in (4.4) is taken along a contour (C) consisting of the real axis plus a possible loop around μ in the complex v plane. Then the second integral in (4.5) vanishes when (C) lies below μ for $k>0$, and when (C) lies above μ for $k<0$: in both cases we shall say that (C) lies on the "proper side" of μ for the value of k considered.

C. First system of equations

We are now going to apply the asymptotic method to

the first system of transformed equations (3.9). We consider each of the terms I×I′, II×I′, I×II′ and II×II′ successively and we try and simplify them by using the rule established in Part 1. According to this rule a term may be neglected in the right-hand side of (3.9) and (3.11) when its integral over the explicit dependence on t remains of order $\epsilon^{3/2}$ for $t = O(\epsilon^{-1/2})$.

The term I×I′ must be retained and it is significant every time the resonance condition on the frequencies is satisfied:

$$\omega = \omega' + \omega'' . \tag{4.7}$$

The term II×I′ can be written:

$$\mathrm{II}\times\mathrm{I}' = \sum_{\gamma} \frac{\omega_p^2}{k'^2} \left(\bar{\varphi}^{\alpha}_{k}, \frac{\partial \varphi^{\alpha}_{k''}}{\partial v} \right)_{k,k''} A^{\gamma}_{k''} e^{i(\omega-\omega'')t} \times \int_{\Delta} e^{-ik'\lambda t} \Gamma_{k'}(\lambda,t)\, d\lambda ,$$

where Δ is a parallel to the real axis in $B(k')$. The explicit time dependence of this expression in the sense of the asymptotic theory is obtained by ignoring the time dependence of $A^{\gamma}_{k''}$ and $\Gamma_{k'}$. The proof is then the same as for the corresponding term II×I′ in Part 1, with the only difference that $\Gamma_{k'}$ may have poles close to the real axis at the zeros of $\epsilon_{\pm}(k',\lambda)$. But the contour of integration Δ lies in $B(k')$ and therefore is automatically on the "proper side" of these poles for k'. Therefore the integral over λ along Δ has the same properties as if these poles were absent and the proof given in Part 1 still holds, so that the term II×I′ may be neglected.

Let us consider now the term I×II′:

$$\mathrm{I}\times\mathrm{II}' = -\sum_{\beta} e^{i(\omega-\omega')t} A^{\beta}_{k'} \left(\frac{\partial \bar{\varphi}^{\alpha}_{k}}{\partial v}, \times \langle {}_{v}\phi^{k''}_{\lambda}, e^{-ik''\lambda t} \Gamma_{k''}(\lambda,t) \rangle_{B(k'')} \right)_{k,k''}$$

This term is not always negligible but it can be simplified considerably and namely by using the symmetry relation (2.15) to write

$$\left(\frac{\partial \bar{\varphi}^{\alpha}_{k}}{\partial v}, \langle {}_{v}\phi^{k''}, c_{k''} \rangle \right) = \left(\left\langle \frac{\partial \bar{\varphi}^{\alpha}_{k}}{\partial v}, {}_{v}\phi^{k''} \right\rangle, c_{k''} \right).$$

The right-hand side of this equation is seen to have the following form:

$$\int_{\Delta} G(\lambda,t)\Gamma_{k''}(\lambda,t) e^{-ik''\lambda t}\, d\lambda. \tag{4.8}$$

Where $G(\lambda,t)$ is slowly dependent on t so that the integral can be discussed as (4.3) and (4.4). A number of different cases must be considered according to the signs of k and k'', the relative position of $B(k)$ and $B(k'')$, and the position of ν^{α}_{k} with respect to $B(k)$ and $B(k'')$. More details on the discussion can be found in an internal report.[11] The final result is the following:

$$\mathrm{I}\times\mathrm{II}' = (\eta-\eta'')i\pi \sum_{\beta} e^{i(\omega-\omega')t} A^{\beta}_{k'} \times \left[\frac{\partial}{\partial v} \epsilon_{-\eta''}(k'',v) e^{-ik''vt} \Gamma_{k''}(v,t) \right]_{v=\nu^{\alpha}_{k}} \tag{4.9}$$

$$(\eta = k/|k|, \eta'' = k''/|k''|).$$

The term II×II′ is easily shown to be negligible.

Thus by simplifying the system of equations (3.9), the following final form is obtained:

$$(\bar{\varphi}^{\alpha}_{k}, \varphi^{\alpha}_{k})_{k} \frac{dA^{\alpha}_{k}}{dt} = \sum_{(k'+k''=k)} ik' \sum_{\beta} A^{\beta}_{k'} \left\{ \sum_{\gamma} e^{i(\omega-\omega'-\omega'')t} \times \left(\bar{\varphi}^{\alpha}_{k}, \frac{\partial \varphi^{\gamma}_{k''}}{\partial v} \right)_{k,k''} A^{\gamma}_{k''} + (\eta-\eta'')i\pi e^{i(\omega-\omega')t} \times \left[\frac{\partial}{\partial v} \epsilon_{-\eta''}(k'',v) e^{-ik''vt} \Gamma_{k''}(v,t) \right]_{v=\nu^{\alpha}_{k}} \right\}. \tag{4.10}$$

D. Second system of equations

We now have to apply the same method to system (3.11). In the same way as above the terms II×I′ and II×II′ of this system can be shown to be negligible. The term I×II′ requires a more detailed but relatively simple discussion. Integrals of the type (4.8) are still encountered and the corresponding terms are neglected when the path of integration Δ lies in $B(k'')$ and therefore on the proper side of the poles for k''. In this way, the following expressions are obtained by discarding the negligible terms:

$$\mathrm{II}' = \epsilon_{\eta''}(k,\lambda) \frac{\partial}{\partial \lambda} \langle {}_{\lambda}\phi^{k''}, c_{k''} \rangle_{B(k'')} + \cdots = \epsilon_{\eta''}(k,\lambda) \frac{\partial}{\partial \lambda} \{ \epsilon_{-\eta''}(k'',\lambda) c_{k''}(\lambda,t) \} + \cdots . \tag{4.11}$$

Thus the final simplified form of (3.11) is the following:

$$\epsilon_{\eta}(k,\lambda)\epsilon_{-\eta}(k,\lambda) \frac{\partial}{\partial t} \Gamma_{k}(\lambda,t) = \sum_{(k'+k''=k)} ik' \sum_{\beta} A^{\beta}_{k'} \left\{ \sum_{\gamma} e^{i(k\lambda-\omega'-\omega'')t} A^{\gamma}_{k''} \times \left\langle \frac{\partial \varphi^{\gamma}_{k''}}{\partial v}, {}_{v}\bar{\phi}^{k}_{\lambda} \right\rangle_{B(k'')} + e^{i(k\lambda-\omega')t} \epsilon_{\eta''}(k,\lambda) \times \frac{\partial}{\partial \lambda} [\epsilon_{-\eta''}(k'',\lambda) e^{-ik''\lambda t} \Gamma_{k''}(\lambda,t)] \right\}. \tag{4.12}$$

It is interesting to rewrite Eqs. (4.10) and (4.12) in terms of the coefficients a^{α}_{k} of the discrete spectrum and of the functions $\psi_{k}(v,t)$ considered in the introduction, which are related to $c_{k}(v,t)$ by:

$$\psi_{k}(v,t) = \epsilon_{-\eta}(k,v) c_{k}(v,t) . \tag{4.13}$$

Thus the final equations take the following form:

$$(\bar{\varphi}^{\alpha}_{k}, \varphi^{\alpha}_{k})_{k} \left(\frac{da^{\alpha}_{k}}{dt} + ik\nu^{\alpha}_{k} a^{\alpha}_{k} \right) = \sum_{(k'+k''=k)} \left\{ \sum_{\beta,\gamma} ik' \left(\frac{\partial \varphi^{\gamma}_{k''}}{\partial v}, \bar{\varphi}^{\alpha}_{k} \right)_{k,k''} a^{\beta}_{k'} a^{\gamma}_{k''} + \sum_{\beta} (\eta''-\eta)\pi k' a^{\beta}_{k'} \left(\frac{\partial \psi_{k''}}{\partial v} \right)_{v=\nu^{\alpha}_{k}} \right\}, \tag{4.14}$$

$$\epsilon_{\eta}(k,v) \left(\frac{\partial}{\partial t} + ikv \right) \psi_{k}(v,t) = \sum_{(k'+k''=k)} ik' \sum_{\beta} a^{\beta}_{k'} \left\{ \sum_{\gamma} a^{\gamma}_{k''} \left\langle \frac{\partial \varphi^{\gamma}_{k''}}{\partial u}, {}_{u}\bar{\phi}^{k}_{v} \right\rangle_{B(k'')} + \epsilon_{\eta''}(k,v) \frac{\partial \psi_{k''}}{\partial v} \right\}, \tag{4.15}$$

where

$\eta = k/|k|$ or sign of k, $\quad \eta'' = k''/|k''|$ or sign of k''.

E. Time evolution of f_0

A necessary complement of the above Eqs. (4.14) and (4.15) is Eq. (1.5) which gives the evolution of f_0 with time. Replacing f_k and f_{-k} by their expansion in the right-hand side of (1.5) we obtain:

$$\frac{\partial f_0}{\partial t} = -ik \sum_k (\mathrm{I} + \mathrm{II})(\mathrm{I}' + \mathrm{II}') \tag{4.16}$$

with

$$\mathrm{I} = \sum_\beta a^\beta_{-k}, \qquad \mathrm{II} = \frac{\omega_p^2}{k^2} \int_\Delta c_{-k}(\lambda)\, d\lambda,$$

$$\mathrm{I}' = \sum_\gamma a^\gamma_k \frac{\partial \varphi^\gamma_k}{\partial v}, \qquad \mathrm{II}' = \frac{\partial}{\partial v} \langle {}_v\phi^k, c_k \rangle_{B(k)}.$$

As in the previous cases, the terms containing expression II may be neglected. Consider now expression II' which can be written:

$$\mathrm{II}' = \frac{\partial}{\partial v} \left\{ \frac{\omega_p^2}{k^2} f_0'(v) \int_\Delta \frac{\Gamma_k(\lambda)}{v - \lambda} e^{-ik\lambda t}\, d\lambda + \epsilon_{-\eta}(k,v) c_k(v) \right\},$$

where the straight line Δ lies in $B(k)$ and therefore on the proper side of the poles v and ν^α_k. Thus the integral can be neglected and the following simplified equation obtains:

$$\frac{\partial f_0}{\partial t} = \sum_k \sum_\beta -ik\, a^\beta_{-k} \left\{ \sum_\gamma a^\gamma_k \frac{\partial \varphi^\gamma_k}{\partial v} + \frac{\partial \psi_k}{\partial v} \right\}. \tag{4.17}$$

5. DISCUSSION OF THE FINAL EQUATIONS

In this section, a number of important features of the final Eqs. (4.14) and (4.15) are discussed.

A. Coupling of a wave with a free streaming perturbation

Equations (4.14) which describe the evolution of the wave amplitudes a^α_k differ from the mode coupling equations found in Part 1 by the presence of an additional term which couples the wave amplitudes to the free streaming perturbations. This term is significant, subject to a resonance condition which was discussed previously and which is that the two waves involved (k,α) and (k',β) have the same velocity. In fact in the case of one dimensional Langmuir waves considered in this paper, the dispersion relation is such that this condition cannot be satisfied exactly. If the two waves have neighbouring values of k and ω, they can interact nearly resonantly with a free-streaming perturbation of small wave number $k'' = k - k'$. However it will be mentioned in Sec. 6 that in the case of an infinite magnetic field with oblique propagation the resonant condition can be satisfied exactly.

It is seen in Eq. (4.14) that the term in question contains a factor $(\eta - \eta'')$ and therefore vanishes unless $k\,k'' < 0$. It was shown previously[2] that, as a consequence of this condition, the following relation holds between the absolute values of the wave numbers:

$$|k| = |k'| - |k''| \tag{5.1}$$

which implies in particular $|k'| > |k''|$.

B. Regularity in v

A basic assumption of the theory is that the functions $f_k(v)$ are holomorphic in v near the real axis. As we expect $f_k(v)$ to differ only slightly from the asymptotic from (1.1) of the linear theory, we anticipate that expression (1.1) is also holomorphic near the real axis in the complex plane and this can indeed be shown by using expressions (4.13) and (2.21) of ψ_k and c_k in terms of f_k. Thus ψ_k must have poles at the zeros of $\epsilon_\eta(k,v)$ and no other singularities near the real axis. Assuming this property to hold, it is interesting to check that there is no contradiction with equation (4.15). The left hand side of this equation is then seen to have no singularity. In the right hand side, the only possible singularities are poles at the zeros $\nu^\gamma_{k''}$ of $\epsilon_{\eta''}(k'',v)$. In the neighborhood of each of these poles the following relation can be written:

$$\left\langle \frac{\partial \varphi^\gamma_{k''}}{\partial u}, {}_u\bar{\phi}^k_v \right\rangle_{B(k'')} = \epsilon_{\eta''}(k,v) \frac{\partial}{\partial v} \varphi^\gamma_{k''}(v) + h(v), \tag{5.2}$$

where $h(v)$ is holomorphic near the pole considered. Thus all possible singularities of the right hand side of Eq. (4.15) are contained in the expression:

$$\epsilon_{\eta''}(k,v) \frac{\partial}{\partial v} \left\{ \sum_\gamma a^\gamma_{k''} \varphi^\gamma_{k''} + \psi_{k''} \right\} \tag{5.3}$$

which has no singularities near the real axis by virtue of the regularity of (1.1).

C. Conditions of validity

The conditions of validity of the theory and the relevant time scales were discussed in Part 1, Sec. 5. This discussion must be extended to include the new features introduced by the resonant particles.

If we consider Eq. (4.12) for $\Gamma_k(\lambda,t)$, a first difference with Part 1 is that the terms involving a derivative with respect to λ can now be resonant for:

$$\lambda = \nu^\beta_{k'}. \tag{5.4}$$

For these resonant values of λ the estimates of the derivatives of Γ_k have to be reconsidered. In part 1 they were assumed to have the following order of magnitude

$$\frac{\partial^n \Gamma_k}{\partial \lambda^n} \sim \mathrm{Max}(\epsilon, \epsilon^2 t^{n+1}), \tag{5.5}$$

which remains justified for nonresonant values of λ. In the opposite case, when λ is close to a resonant value $\nu^\beta_{k'}$, the same reasoning on Eq. (4.12) leads to an increased power of t:

$$\frac{\partial^n \Gamma_k}{\partial \lambda^n} \sim \mathrm{Max}(\epsilon, \epsilon^2 t^{n+2}) \quad (\lambda \simeq \nu^\beta_{k'}). \tag{5.6}$$

This estimate is valid for λ close enough to $\nu^\beta_{k'}$ so that $(\lambda - \nu^\beta_{k'})t$ remains finite, i.e.,

$$\lambda - \nu^\beta_{k'} \sim \epsilon^{1/2} \text{ for } t \sim \epsilon^{-1/2}.$$

Thus in estimating integrals over λ, the contribution of the small resonant intervals of λ will turn out to be of the same order as the estimates made in Part 1 for a finite interval of λ. Therefore the results of Part 1 are essentially unchanged by the presence of the resonances.

Another new feature is the fact that integrals over the velocity of the form (4.3) now extend from $-\infty$ to $+\infty$, whereas they were taken over a finite interval $(-V,V)$ in Part 1. Therefore the orders of magnitude of Γ_k and its derivatives (5.5) can no longer be considered as defined uniformly in λ in the usual sense. A stronger definition should be used to ensure the convergence of the integrals taken over the upper bounds of Γ_k or its derivatives. This can be done in principle as in a previous paper.[9] For instance the order of magnitude of Γ_k can be understood in terms of an upper bound of $|\lambda^2\Gamma_k(\lambda)|$ in addition to an upper bound of $|\Gamma_k(\lambda)|$.

In discussing the order of magnitude of the neglected terms, one should also take care of the poles which may occur at the roots of the dispersion relation, at least when they are close to the real axis. Let us consider for instance the typical integral:

$$J(t) = \int_{(C)} \Gamma_k(\lambda,t)e^{-ik\lambda t}\,d\lambda,$$

where the path of integration (C) lies by the "proper side" of the poles. This integral can be treated by the procedure indicated in Eq. (4.5) and it can be decomposed into a sum of integrals of the type:

$$J_1(t) = \int_{-\infty}^{+\infty} \frac{F(\lambda)-F(\mu)}{\lambda-\mu} e^{-ik\lambda t}\,d\lambda,$$

where $F(\lambda)$ is a regular function of λ whose derivatives satisfy (5.5). This is not quite the same as assuming the order of magnitude (5.5) directly for the derivatives of Γ_k. In fact it can be shown[11] that:

$$\int_0^T J(t)\,dt \sim \epsilon^{3/4}, \quad \text{for } T\sim\epsilon^{-1/2},$$

whereas the corresponding expression in Part 1 was proved to be of order ϵ. However this result is sufficient to justify the derivation of Eq. (4.10) governing the evolution of A_k^α.

The discussion of Eq. (4.12) involves estimates of more complicated integrals. The most difficult to deal with has the following form:

$$K(\lambda,t) = \frac{\partial}{\partial\lambda}\int_{-\infty}^{+\infty} \frac{F(u,t)-F(\lambda,t)}{u-\lambda} e^{-iku t}\,du,$$

where F is again a regular function of u or λ whose derivatives satisfy (5.5) with $F\sim\epsilon$. In order that the corresponding terms of (3.11) may be neglected in the derivation of (4.12), one should have:

$$\int_0^T K(\lambda,t)\,dt \sim \epsilon^\alpha, \quad (\alpha > \tfrac{1}{2}), \quad \text{for } T\sim\epsilon^{-1/2}.$$

It has not been possible to derive this result directly from the orders of magnitude (5.5). However one can find good reasons to believe that it holds as a consequence of the actual dependence of $F(u,t)$ on u and t in the same way as in Part 1.

Thus Eqs. (4.10) and (4.12) appear to have the same range of validity as in Part 1, namely: $\omega_p t \lesssim \epsilon^{-1/2}$.

6. RESULTS FOR THE CASE OF AN INFINITE MAGNETIC FIELD

The above theory can easily be repeated for the case of Langmuir waves in an infinite magnetic field considered in Part 1. The only difference with Part 1 is then that the space averaged distribution function $f_0(v)$ extends to all values of v.

The operator L_k remains the same as in Sec. 2. The frequency of a normal mode is of the form $k_\parallel \nu_k^\alpha$, where ν_k^α is the root of the same dispersion relation as above. The choice of the strip B in the complex v plane and the expansion formula will depend on $k = |\underline{\kappa}|$ and $k_\parallel$, so that all quantities relevant to the expansion formula will bear the index $\underline{\kappa}$ instead of k.

The coefficients of the expansion $a_{\underline{\kappa}}^\alpha(t)$ and $c_{\underline{\kappa}}(\lambda,t)$ are defined in terms of $f_{\underline{\kappa}}(v,t)$ by:

$$a_{\underline{\kappa}}^\alpha(t) = \frac{(\bar{\varphi}_{\underline{\kappa}}^\alpha, f_{\underline{\kappa}})_{\underline{\kappa}}}{(\bar{\varphi}_{\underline{\kappa}}^\alpha, \varphi_{\underline{\kappa}}^\alpha)_{\underline{\kappa}}}, \tag{6.1}$$

$$c_{\underline{\kappa}}(\lambda,t) = \frac{\langle f_{\underline{\kappa}}, \bar{\phi}_\lambda^k\rangle}{\epsilon_+(k,\lambda)\epsilon_-(k,\lambda)} B(\underline{\kappa}), \tag{6.2}$$

where the index $\underline{\kappa}$ on the scalar product means that the integral is to be taken along a contour lying on the same side of $\nu_{\underline{\kappa}}^\alpha$ as $B(\underline{\kappa})$.

The system of Eqs. (4.14) and (4.15) is then replaced by the following:

$$\begin{aligned}(\bar{\varphi}_{\underline{\kappa}}^\alpha, \varphi_{\underline{\kappa}}^\alpha)_{\underline{\kappa}}\left(\frac{da_{\underline{\kappa}}^\alpha}{dt} + ik_\parallel \nu_{\underline{\kappa}}^\alpha a_{\underline{\kappa}}^\alpha\right) &= \sum_{(\underline{\kappa}'+\underline{\kappa}''=\underline{\kappa})}\sum_\beta \left\{ ik_\parallel' a_{\underline{\kappa}'}^\beta \sum_\gamma a_{\underline{\kappa}''}^\gamma \left(\frac{\partial\varphi_{\underline{\kappa}''}^\gamma}{\partial v}, \bar{\varphi}_{\underline{\kappa}}^\alpha\right)_{\underline{\kappa},\underline{\kappa}''} \right. \\ &\quad \left. + (\eta''-\eta)\pi k_\parallel' a_{\underline{\kappa}'}^\beta \left(\frac{\partial\psi_{\underline{\kappa}''}}{\partial v}\right)_{v=\nu_{\underline{\kappa}}^\alpha}\right\},\end{aligned} \tag{6.3}$$

$$\begin{aligned}\epsilon_\eta(k,v)\left(\frac{\partial}{\partial t} + ik_\parallel v\right)\psi_{\underline{\kappa}} &= \sum_{(\underline{\kappa}'+\underline{\kappa}''=\underline{\kappa})}\sum_\beta ik_\parallel' a_{\underline{\kappa}'}^\beta \left\{ \sum_\gamma a_{\underline{\kappa}''}^\gamma \left\langle \frac{\partial\varphi_{\underline{\kappa}''}^\gamma}{\partial u}, {}_u\bar{\phi}_v^k \right\rangle_{B(\underline{\kappa}'')} \right. \\ &\quad \left. + \epsilon_{\eta''}(k,v)\frac{\partial\psi_{\underline{\kappa}''}}{\partial v}\right\},\end{aligned} \tag{6.4}$$

where

$$\eta = \frac{k_\parallel}{|k_\parallel|}, \quad \eta'' = \frac{k_\parallel''}{|k_\parallel''|}.$$

The dispersion function $\epsilon_\pm(k,\lambda)$ is still given by (2.5), but now k is defined by:

$$k^2 = |\underline{\kappa}|^2 = k_\parallel^2 + k_\perp^2 \tag{6.5}$$

and the wave corresponding to a zero $\nu_{\underline{\kappa}}^\alpha$ of the dispersion function has a frequency:

$$\omega_{\underline{\kappa}}^\alpha = k_\parallel \nu_{\underline{\kappa}}^\alpha, \tag{6.6}$$

so that $\nu_{\underline{\kappa}}^\alpha$ is the parallel phase velocity.

The last term in the right-hand side of Eq. 6.3 corresponds to the interaction between a wave and a free-streaming perturbation and it is significant only if the following resonance condition is satisfied

$$ik_\parallel \nu_{\underline{\kappa}}^\alpha - ik_\parallel' \nu_{\underline{\kappa}'}^\beta - ik_\parallel'' \nu_{\underline{\kappa}}^\alpha = ik_\parallel'(\nu_{\underline{\kappa}}^\alpha - \nu_{\underline{\kappa}'}^\beta) = 0,$$

which means that the two waves must have the same parallel phase velocity.

With the dispersion relation (6.6) it is in fact possible to find two waves with the same phase velocity if they corre-

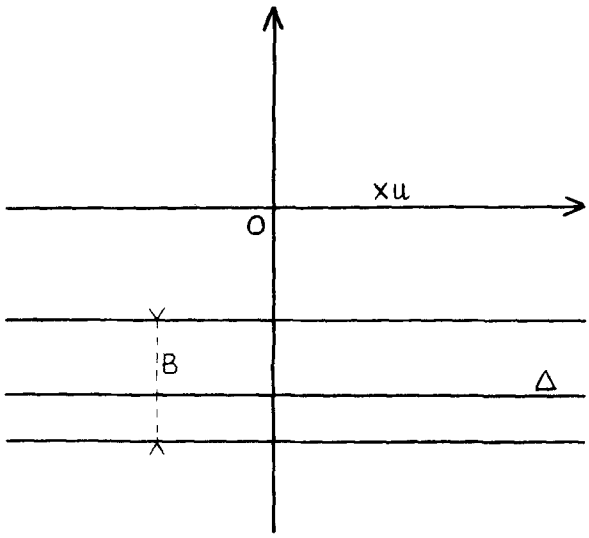


FIG. 1.

spond to different values of $k_\perp$. This possibility will be examined in detail in connexion with experimental work.[3]

7. CONCLUSION

The weakly nonlinear evolution of electron plasma waves has been treated with a small parameter proportional to the wave amplitudes and by using the ordering of the Klylov and Bogoliubov method, which is equivalent to the first order of the multiple time scale method. It is found that a description in terms of wave amplitudes alone is not possible in the general case and that the final result can be interpreted as a coupled evolution of wave amplitudes and free-streaming perturbations. In particular a new interesting phenomenon appears in the equations, namely the nonlinear interaction of a wave with a free-streaming perturbation. As was to be expected from Part 1 of this work[1] such an interaction occurs through resonant particles. It is obviously closely connected with the second order wave found earlier by Best[4] although by using a different ordering.

The final system of equations previously justified by a heuristic argument,[2] has been derived in a systematic way and its conditions of validity have been discussed. It is shown in particular that the equations are valid over a time scale $O(1/\sqrt{\epsilon})$ instead of $O(1/\epsilon)$ as might have been expected. Thus the validity should extend up to the time scale of particle trapping and presumably including the occurrence of particle trapping.

Eqs. (4.14) and (4.15) of this paper are identical with Eqs. (35) and (46) of Ref. 2 when a few trivial errors are corrected in these latter equations.

APPENDIX

In this Appendix, we use the notations of a previous paper[9] which omit the index k.

Let us consider the case where $k > 0$ (see Fig. 1). Let $f(u)$ and $f_0(u)$ be two given functions of class 2 in a wide strip S of the complex u plane extending on both sides of the real axis. The expansion of $f(u)$ derived in Ref. 9 is valid inside a strip B included in S and containing no zero of $R_+(u)$ nor of $R_-(u)$. This expansion can be written

$$f(u) = S_1^B(u) + S_2^B(u) + \langle {}_u\phi_\lambda, c(\lambda)\rangle\,, \tag{A1}$$

where $c(\lambda)$ is defined in B according to Eq. (2.21), and where

$$S_1^B(u) = \sum_\alpha a_\alpha \varphi_\alpha(u)$$

$$S_2^B(u) = \sum_\beta a_\beta \varphi_\beta(u)\,.$$

The sum over α extends to all zeros of $R_+(u)$ located above B and the sum over β to all zeros of $R_-(u)$ located below B.

For $u \in B$, expression (1) may be written:

$$f(u) = S_1^B(u) + S_2^B(u) + R_-(u)c(u) + \frac{\omega_p^2}{k^2} f_0'(u) \int_\Delta \frac{c(\lambda)}{u-\lambda}\, d\lambda\,, \tag{A2}$$

where the integral over λ is taken along a straight line Δ parallel to the real axis and lying in B below the point u. This integral is therefore defined and holomorphic in u not only in B but also above B, i.e., in a half plane which we denote by B_+. The sum $S_2^B(u)$ is holomorphic inside both B_+ and S, i.e., in $B_+ \cap S$. Consider now the two remaining terms in the right hand side of (2):

$$S_1^B(u) + R_-(u)c(u). \tag{A3}$$

$S_1^B(u)$ is meromorphic in $B_+ \cap S$ and has poles at the zeros of $R_+(u)$. According to (2.21), $R_-(u)c(u)$ is defined in S and its singularities are poles at the zeros of $R_+(u)$. It is easily verified that the two residues of these poles cancel out in expression (3). Therefore the right hand side of Eq. (A2) is holomorphic in $B_+ \cap S$. As a result, Eq. (A2) is valid at every point of $B_+ \cap S$ and in particular on the real axis and near the real axis. Expansion formula (2.19) is thus proved for $k > 0$. A similar derivation applies when $k < 0$.

[1] M. Trocheris, J. Math. Phys. **21**, 932 (1980).
[2] M. Trocheris, Plasma Phys. **21**, 75 (1979).
[3] F. Gervais, J. Olivain, A. Quemeneur and M. Trocheris, to be published.
[4] R.W.B. Best, Physica **74**, 183 (1974).
[5] M. Gros, P. Bertrand, G. Bauman, J. Plasma Phys. **20**, 465 (1978).
[6] R.C. Davidson, *Methods in Non-Linear Plasma Theory* (Academic, New York, 1972).
[7] N.G. Van Kampen, Physica **21**, 949 (1955).
[8] K.M. Case, Ann. Phys. (N.Y.) **7**, 349 (1959).
[9] M. Trocheris, Nucl. Fusion **5**, 299 (1965).
[10] L. Landau, J. Phys. USSR **10**, 25 (1946); and D.C. Montgomery and D.A. Tidman, *Plasma Kinetic Theory* (McGraw-Hill, New York, (1964).
[11] M. Trocheris, "Methode asymptotique dans la théorie non linéaire des ondes de Langmuir cohérentes," EUR-CEA-FC 364 (1976).

ERRATA

Erratum: Dense electron-gas response at any degeneracy [J. Math. Phys. 19, 32 (1978)]

C. Gouedard and C. Deutsch
Université Paris XI, 91405 Orsay, France

(Received 19 October 1979; accepted for publication 9 November 1979)

In Figs. 2 and 3, Z should read $\alpha = \beta\mu$.

 0022-2488/80/040949-01$0.00