Theory of ground state correlations of closed shell nuclei: A density-matrix formulation*†

Franz Krejs‡ and Abraham Klein

Department of Physics, University of Pennsylvania, Philadelphia, Pennsylvania 19104
(Received 30 November 1972; final revised manuscript received 23 January 1973)

A general theory of long range correlations in the ground state of a closed shell nucleus is outlined, using a number of ideas introduced or adumbrated, but not always successfully developed, in previous work of the authors and others. These include the following: (i) Any operator can be formally expressed as a series in particle-hole ($p-h$) excitation and destruction operators which should converge rapidly for a suitable choice of single-particle basis at the closed shells. By means of these relations and with the help of sum rules all observables can be computed without the explicit use of wave functions. (ii) For a fixed choice of single-particle basis, the ground state energy can be computed from a variational principle in which in lowest nontrivial approximation, in consequence of (i), the parameters are amplitudes of the same type as occur in the random phase approximation (RPA). The precise relation to RPA is fully detailed. (iii) The single particle basis chosen is the natural orbital basis of Lowdin, which presents itself as the obvious choice both for formulation of the dynamics and for the algorithm of solution. (iv) Though the theory deals with pairs of fermion operators, particular attention is paid that all Pauli principle restrictions are satisfied. (v) It is verified that in the perturbation limit, known results are reproduced. The method is generally applicable to any system for which the correlations can be viewed as particle-hole excitations with respect to a reference Slater determinant.

1. INTRODUCTION

The deviation of the ground state of doubly closed shell nuclei from the pure single-particle description is well-documented. For instance, in the O^{16} ground state we note the evidence from stripping that the p-shell is not completely full and from pickup that the d-shell is not completely empty.[1,2] The reasons for not leaving the interpretation of such experiments completely to the admirable, large and impressive shell model calculations,[3,4] but rather to try to develop simplified pictures and models, hardly needs documentation.

Since the paper which follows is in many ways quite formal, let us emphasize at the beginning that is represents a mathematical formulation of the picture of ground state correlations as arising from the virtual excitation of vibrations—the tradition relating to the RPA (quasi-boson) theory of ground state correlations.[5]

A few years ago attention was called to certain failures of this theory to yield correct results in the limit of perturbation theory,[6,7,8] a failure tied to improper treatment of the Pauli principle. Satisfactory methods of dealing with this difficulty have been proposed,[9–13] so that we may now focus on the development of a systematic method which is capable of going beyond the lowest approximation.

In this paper we have given final form to a variational algebraic method which for the problem at hand—the closed shell nucleus—was suggested by Michailovic and Rosina[14] (see also Ref. 15), and studied further by the present authors.[10] A number of points were however left obscure by the previous work. Mainly it was not clear how one could go systematically to a complete theory and also how the amplitudes which enter the variational principle—which look like RPA amplitudes—can differ from the latter without any inconsistency arising.

From the mathematical point of view our work can be regarded as a practical, but nonrigorous contribution to the problem of calculating variationally the two-body density matrix.[16,17] We emphasize here only the new results in this paper: (i) A systematic expansion—applicable at closed shells—of any operator in terms of polynomials in $p-h$ excitation and destruction operators which is the underlying mathematical tool of this paper (see also Ref. 12). This expansion should converge well when applied to the computation of matrix elements between low-lying states at the closed shell. (ii) The formulation of the variational principle so that the variational equations will have the RPA matrix structure, which can now be handled by standard eigenvalue means.[18,19] (iii) The use of the natural orbitals of Lowdin[20] (for a review of their applications to atomic physics, see Davidson[21]), which have been reconsidered recently for nuclear problems.[22,23] Our use of them is distinct from previous ones, but is most natural in that it is *directly* suggested by their definition. (iv) The description of what should prove to be a practical algorithm for the solution of the variational equations. (v) Clarification of the relationship to the RPA.

In exposing the improved results, we have gone over matters discussed in previous work only where deemed necessary.

The importance of the problem at hand has led to the study of other methods of dealing with the problem of ground state correlations and low-lying excitations, some of which have already proved themselves of value. We shall complete this introduction with a brief review of *other than* the straight shell model calculations and perturbation theory.

The most useful and versatile algebraic method so far developed is that of Rowe.[24,25] In application he has eschewed the rigid orthodoxy of which we may be accused by using shell-model wave functions in an essential way in his calculations, in order to simplify the application to open shell nuclei.

Elegant, but so far only limited diagrammatic treatments have been developed by Ellis[9] and by Padjen and Ripka.[26]

We note finally the variant of the shell model—the multiconfiguration Hartree–Fock method—championed by Faessler and coworkers[27,28] and by Ho-Kim,[29] which is a variation of the shell model. One mixes Slater determinants in a shell model type of calculation but hopes to keep the number of basis states tractable by

Copyright © 1973 by the American Institute of Physics

optimizing with respect to the choice of single particle states.

To avoid misunderstanding, we emphasize that we deal exclusively with the problem: Given the model Hamiltonian—how does one solve it to good approximation? We have not dealt with the problem of deriving the model Hamiltonian from first principles. This fundamental aspect has been studied recently by many authors,[30] particularly with respect to the 2-particle interaction (closed shell +2 nuclei). The method summarized by Kirson[30] promises also to be useful for the particle-hole interaction at closed shells.

2. STRUCTURE OF THE SPACE AND BASIC OPERATOR IDENTITIES

Let us consider a system of N particles interacting via short range forces as described by the Hamiltonian

$$H = \sum_{ab} h(a|b)\psi_a^\dagger \psi_b + \frac{1}{4}\sum_{abcd} V(abcd)\psi_a^\dagger \psi_b^\dagger \psi_d \psi_c, \qquad (2.1)$$

where a refers to some complete set of single particle functions. (We have specifically in mind the application to the nuclear shell model at "closed shells," but the method to be developed applies *a fortiori* to a problem such as the long range correlations of the electron gas at high densities.) Here $\psi_a^\dagger, \psi_a$ are fermion creation, annihilation operators for the modes designated by a, and the numerical matrices h and V satisfy the conditions of hermiticity,

$$\begin{aligned} h(a|b) &= h(b|a)^*, \\ V(ab|cd) &= V(cd|ab)^*, \end{aligned} \qquad (2.2)$$

whereas V is also taken in antisymmetrical form,

$$V(ab|cd) = -V(ba|cd) = -V(ab|dc). \qquad (2.3)$$

Thus direct and exchange interactions are treated on an equal footing, as seems most appropriate for a short range interaction.

In an individual particle picture the ground state, $|0\rangle$, of our system is approximated by a state $|\Phi_0\rangle$ which is a Slater determinant of N single-particle wave functions. We choose the set $\{a\}$ to contain the single-particle states out of which $|\Phi_0\rangle$ is made up. We then split the set $\{a\}$ into two subsets:

$$\{a\} = \{i\} U \{p\}, \qquad (2.4)$$

where $\{i\} = \{i, j, k \cdots\}$ is the set of N states occupied in $|\Phi_0\rangle$ and $\{p\} = \{p, q, r, \cdots\}$ is the set of unoccupied states, which together with $\{i\}$ forms a complete set. The state $|\Phi_0\rangle$ is then characterized as usual by the conditions

$$\psi_p|\Phi_0\rangle = \psi_i^\dagger|\Phi_0\rangle = 0. \qquad (2.5)$$

Though we shall not carry out any formal operator transformations, we shall nevertheless use the customary language of particle-hole (p–h) excitations suggested by (2.5). In addition to the state $|\Phi_0\rangle$ the complete space of states of our many-body system will then be characterized by various p–h subspaces, generated by the operation of $\psi_p^\dagger\psi_i, \psi_p^\dagger\psi_q^\dagger\psi_i\psi_j, \cdots$ on $|\Phi_0\rangle$. We are interested in a situation where the admixture of such configurations into the ground state is nonnegligible; but where we require nevertheless that an individual particle picture remain valid in the sense that a choice of the set $\{a\}$ can be made such that the total probability for the excitation of ν p–h pairs in the true ground state is a rapidly decreasing function of ν. For our purposes we shall express this requirement somewhat differently, namely we shall assume that individual particle and hole occupation probabilities, strictly zero in the limit of the single-particle picture, remain small compared to unity,

$$\begin{aligned} &\langle 0|\psi_p^\dagger\psi_p|0\rangle \ll 1, \\ &\langle 0|\psi_i\psi_i^\dagger|0\rangle \ll 1. \end{aligned} \qquad (2.6)$$

We shall develop a method designed to exploit these inequalities, based on the following theorem: In the space of states characterized by $|\Phi_0\rangle$ and p–h excitations built upon $|\Phi_0\rangle$, every operator can be formally expressed as an infinite series in the operators $(\psi_p^\dagger\psi_i)$ and $(\psi_i^\dagger\psi_p)$. If we consider a subspace defined by limiting the number of p labels, the series reduces to a finite polynomial.

For a concise proof and statement of this theorem we need a temporary change of notation. We write

$$\psi_p \to \psi_+(1), \qquad \psi_i \to \psi_-(1), \qquad (2.7)$$

and use the summation convention for repeated numerical labels, where in each case the summation is over the subspace indicated by $\pm$. Thus for the number operator we have

$$\hat{n} = \sum_a \psi_a^\dagger\psi_a = \psi_+^\dagger(1)\psi_+(1) + \psi_-^\dagger(1)\psi_-(1). \qquad (2.8)$$

If we deal with the space of N particles, this becomes

$$\hat{n} - N = \psi_+^\dagger(1)\psi_+(1) - \psi_-(1)\psi_-^\dagger(1). \qquad (2.9)$$

The proof of the theorem makes essential use of (2.9).

Thus by pre- and post-multiplication, (2.9) becomes

$$\begin{aligned} \psi_+^\dagger(1)(\hat{n} - N)\psi_+(2) = {}& \psi_+^\dagger(1)\psi_+^\dagger(3)\psi_+(3)\psi_+(2) \\ & - [\psi_+^\dagger(1)\psi_-(3)][\psi_-^\dagger(3)\psi_+(2)]. \end{aligned} \qquad (2.10)$$

If it is understood that (2.10) operates in the space of N-particle states, $\hat{n}$ may be replaced by $N-1$ and we thus find

$$\begin{aligned} \psi_+^\dagger(1)\psi_+(2) = {}& -[\psi_+^\dagger(1)\psi_-(3)][\psi_-^\dagger(3)\psi_+(2)] \\ & + \psi_+^\dagger(1)\psi_+^\dagger(3)\psi_+(3)\psi_+(2), \end{aligned} \qquad (2.11)$$

of which the first term is already in the desired form.

To obtain an expression for a $2p2p$-operator we have to insert the operator $(\hat{n} - N)$ twice and get

$$\begin{aligned} \psi_+^\dagger(1)\psi_+^\dagger(2)\psi_+(2')\psi_+(1') = {}& \tfrac{1}{2}\psi_+^\dagger(12)\psi_-(34)\psi_-^\dagger(43)\psi_+(2'1') \\ & - 2\psi_+^\dagger(123)\psi_+(32'1') - \tfrac{1}{2}\psi_+^\dagger(1234)\psi_+(432'1'), \end{aligned} \qquad (2.12)$$

where for conciseness we have introduced the notation

$$\psi_+^\dagger(123\cdots n) \equiv \psi_+^\dagger(1)\psi_+^\dagger(2)\cdots\psi_+^\dagger(n) \qquad (2.13)$$

and similarly for the other operators.

The h–h operators can be treated in the same way. This procedure can be carried further to higher and higher order up to Np–Nh operators where it terminates because any expression having more than N destruction operators on the right will give a vanishing result when

operating on any state of the N particle system. Following the route described, a general formula can be given to express an np–np (nh–nh) operator by np–nh operators and higher pp (hh) operators:

$$\begin{aligned}
&\psi_+^\dagger(12\cdots n)\psi_+(n'\cdots 1') \\
&\quad = \frac{1}{n!}\,\psi_+^\dagger(1\cdots n)\psi_-(n+1\cdots 2n) \\
&\qquad \times \psi_-^\dagger(2n\cdots n+1)\psi_+(n'\cdots 1') \\
&\qquad - \frac{1}{n!}\left\{\sum_{j_1=1}^{3\prime}\ \sum_{j_2=j_1+1}^{j_1+3\prime}\cdots\ \sum_{j_{n-1}=j_{n-2}+1}^{n+1\prime} j_1 j_2\cdots j_{n-1}\right\} \\
&\qquad \times \psi_+^\dagger(12\cdots n+1)\psi_+(n+1,n'\cdots 1') \\
&\qquad - \frac{1}{n!}\left\{\sum_{j_1=1}^{5\prime}\ \sum_{j_2=j_1+1}^{j_1+5\prime}\cdots\ \sum_{j_{n-2}=j_{n-3}+1}^{n+2\prime} j_1 j_2\cdots j_{n-2}\right\} \\
&\qquad \times \psi_+^\dagger(12\cdots n+2)\psi_+(n+2,n+1,n'\cdots 1') \\
&\qquad - \frac{1}{n!}\left\{\sum_{j_1=1}^{7\prime}\ \sum_{j_2=j_1+1}^{j_1+7\prime}\cdots\ \sum_{j_{n-3}=j_{n-4}+1}^{n+3\prime} j_1 j_2\cdots j_{n-3}\right\} \\
&\qquad \times \psi_+^\dagger(12\cdots n+3)\psi_+(n+3\cdots n'\cdots 1') \\
&\qquad \vdots \\
&\qquad - \frac{1}{n!}\left\{\sum_{j_1=1}^{2n-1\prime} j_1\right\}\psi_+^\dagger(1\cdots 2n-1)\psi_+(2n-1\cdots n'\cdots 1') \\
&\qquad - \frac{1}{n!}\,\psi_+^\dagger(1\cdots 2n)\psi_+(2n\cdots n'\cdots 1'). \qquad (2.14)
\end{aligned}$$

The prime on the sum is to indicate that it only runs over either even or odd integers. Similarly for h–h operators:

$$\begin{aligned}
&\psi_-(1\cdots n)\psi_-^\dagger(n'\cdots 1') \\
&\quad = \frac{1}{n!}\,\psi_+^\dagger(n+1\cdots 2n)\psi_-(1\cdots n) \\
&\qquad \times \psi_-^\dagger(n'\cdots 1')\psi_+(2n\cdots n+1) \\
&\qquad - \frac{1}{n!}\left\{\sum_{j_1=1}^{3\prime}\ \sum_{j_2=j_1+1}^{j_1+3\prime}\cdots\ \sum_{j_{n-1}=j_{n-2}+1}^{n+1\prime} j_1 j_2\cdots j_{n-1}\right\} \\
&\qquad \times \psi_-(1\cdots n,n+1)\psi_-^\dagger(n+1n'\cdots 1') \\
&\qquad - \frac{1}{n!}\left\{\sum_{j_1=1}^{5\prime}\ \sum_{j_2=j_1+1}^{j_1+5\prime}\cdots\ \sum_{j_{n-2}=j_{n-3}+1}^{n+2\prime} j_1 j_2\cdots j_{n-2}\right\} \\
&\qquad \times \psi_-(1\cdots n+2)\psi_-^\dagger(n+2,n+1,n'\cdots 1') \\
&\qquad \vdots \\
&\qquad - \frac{1}{n!}\left\{\sum_{j_1=1}^{2n-1\prime} j_1\right\}\psi_-(1\cdots 2n-1) \\
&\qquad \times \psi_-^\dagger(2n-1\cdots n+1,n'\cdots 1') \\
&\qquad - \frac{1}{n!}\,\psi_-(1\cdots 2n)\psi_-^\dagger(2n\cdots n+1,n'\cdots 1'). \qquad (2.15)
\end{aligned}$$

Successive applications of (2.14) and (2.15) allows expression of any pp or hh operator by a sum of products of p–h operators. This sum will involve up to Np–Nh operators. If the state $|\Phi_0\rangle$ is a good first approximation to the ground state, the residual interaction will primarily admix $2p$–$2h$ configurations as we shall see below and higher configurations will be of decreasing importance. This fact will render approximate forms of (2.14) and (2.15) practical tools of calculation.

3. APPLICATION TO COMPUTATION OF THE ENERGY: CHOICE OF ORBITALS

To utilize the results just found we bring H, Eq. (2.1), to normal form with respect to the state $|\Phi_0\rangle$. This standard procedure yields

$$\hat{H} = W_{N0} + \hat{H}_{11} + (\hat{H}_{20} + \text{h.c.}) + \hat{H}_{22} + (\hat{H}_{40} + \text{h.c.}) + \hat{H}'_{22} + (\hat{H}_{31} + \text{h.c.}), \qquad (3.1)$$

where

$$W_{NO} = \langle\Phi_0|H|\Phi_0\rangle = \sum_i h(i|i) + \frac{1}{2}\sum_{ij} V(ij|ij), \qquad (3.2)$$

$$H_{11} = -\sum_{ij}\left(h(i|j) + \sum_k V(ik|jk)\right)\psi_j\psi_i^\dagger + \sum_{pq}\left(h(p|q) + \sum_i V(pi|qi)\right)\psi_p^\dagger\psi_q, \qquad (3.3)$$

$$H_{20} = \sum_{pi}\left(h(pi) + \sum_j V(pj|ij)\right)\psi_p^\dagger\psi_i, \qquad (3.4)$$

$$H_{22} = \sum_{piqj}{}', V(pj|iq)\psi_p^\dagger\psi_i\psi_j\psi_q^\dagger, \qquad (3.5)$$

$$H_{40} = \frac{1}{4}\sum_{piqj} V(pq|ij)\psi_p^\dagger\psi_i\psi_q^\dagger\psi_j, \qquad (3.6)$$

$$H'_{22} = \frac{1}{4}\sum_{pqrs} V(pq|rs)\psi_p^\dagger\psi_q^\dagger\psi_s\psi_r + \frac{1}{4}\sum_{ijkl} V(ij|kl)\psi_l\psi_k\psi_i^\dagger\psi_j^\dagger, \qquad (3.7)$$

$$H_{31} = \frac{1}{2}\sum_{pqri} V(pq|ir)\psi_p^\dagger\psi_i\psi_q^\dagger\psi_r + \frac{1}{2}\sum_{pijk} V(pi|jk)\psi_p^\dagger\psi_k\psi_j\psi_i^\dagger. \qquad (3.8)$$

Let us at this point make a definite choice for the single particle basis $\{a\}$. By far the most commonly used single particle states are the Hartree–Fock (HF) states. The fact that the latter give the best expectation value for the total energy does not imply, however, that they will also give the best expectation values of other operators. Furthermore, if one wants to go beyond the independent particle description, the HF states do not provide any advantages over other possible choices of the single particle basis. Recently, alternative ways of choosing a single particle basis, which have some attractive and convenient properties, have therefore received attention.[22,23] A very useful and convenient choice for our purposes are the so-called natural orbitals introduced by Lowdin.[20,21] These orbitals are defined by the condition that they render the single particle density matrix diagonal, i.e.,

$$\rho_{ab} \equiv \langle 0|\psi_b^\dagger\psi_a|0\rangle = \delta_{ab}\rho_a \qquad (3.9)$$

for all $a,b \in \{a\}$. The natural orbitals yield a Slater determinant which satisfies the condition:

$$\text{Tr}(\langle 0|\psi_b^\dagger\psi_a|0\rangle - \langle\Phi_0|\psi_b^\dagger\psi_a|\Phi_0\rangle)^2 = \text{minimum}. \qquad (3.10)$$

This leads to the interpretation that in the true ground state the occupation numbers for the states $\{i\}$ are as

close to unity and for the states $\{p\}$ as close to zero as possible when $\{i\}$ and $\{p\}$ are natural orbitals.

Let us then choose natural orbitals as the single particle basis. From the defining relation (3.9) one sees that $\hat{H}_{20}$ will not contribute to the ground state expectation value of $\hat{H}$ and that in the case of $\hat{H}_{11}$ contributions will only come from the diagonal part.

The ground state expectation value of $\hat{H}$ is now evaluated. Using the expansions (2.14) and (2.15), we express all $p-p$ and $h-h$ operators in (3.1) in terms of $p-h$ operators up to products of four $p-h$ operators. We give a few details of that evaluation since it should aid in the clarification of our basic assumptions.

We utilize in (3.1) the following approximate forms[7,10] of (2.14) and (2.15):

$$\langle 0|\psi_j\psi_i^\dagger|0\rangle = \sum_p \langle 0|(\psi_j\psi_p^\dagger)(\psi_p\psi_i^\dagger)|0\rangle - \frac{1}{2}\sum_{k,pq}\langle 0|(\psi_j\psi_p^\dagger)(\psi_k\psi_q^\dagger)(\psi_q\psi_k^\dagger)(\psi_p\psi_i^\dagger)|0\rangle, \quad (3.11)$$

$$\langle 0|\psi_p^\dagger\psi_q|0\rangle = \sum_i \langle 0|(\psi_p^\dagger\psi_i)(\psi_i^\dagger\psi_q)|0\rangle - \frac{1}{2}\sum_{r,ij}\langle 0|(\psi_p^\dagger\psi_i)(\psi_r^\dagger\psi_j)(\psi_j^\dagger\psi_r)(\psi_i^\dagger\psi_q)|0\rangle, \quad (3.12)$$

$$\langle 0|\psi_p^\dagger\psi_q^\dagger\psi_s\psi_r|0\rangle = \frac{1}{2}\sum_{ij}\langle 0|(\psi_p^\dagger\psi_i)(\psi_q^\dagger\psi_j)(\psi_j^\dagger\psi_s)(\psi_i^\dagger\psi_r)|0\rangle, \quad (3.13)$$

$$\langle 0|\psi_i\psi_j\psi_k^\dagger\psi_l^\dagger|0\rangle = \frac{1}{2}\sum_{pq}\langle 0|(\psi_i\psi_p^\dagger)(\psi_j\psi_q^\dagger)(\psi_q\psi_k^\dagger)(\psi_p\psi_l^\dagger)|0\rangle. \quad (3.14)$$

These relations are exact if the ground state $|0\rangle$ contains at most $2p-2h$ admixtures.

To evaluate (3.11) to (3.14), we introduce the amplitudes

$$Y_{pi}^*(\alpha) = \langle\alpha|\psi_p^\dagger\psi_i|0\rangle, \quad (3.15)$$

$$Z_{pi}(\alpha) = \langle\alpha|\psi_{\bar{i}}^\dagger\psi_{\bar{p}}|0\rangle, \quad (3.16)$$

where $\bar{i},\bar{p}$ are the time-reversed orbits to i,p. (In the purely formal steps which follow, we shall be careless about such distinctions and drop the bars.) We shall refer to the set of states $|\alpha\rangle$ which occur in the definitions (3.15), (3.16) as $1p-1h$ states. This is only a rough and ready characterization, and it is one of the aims of this work to give a more precise dynamical characterization of these states. In the limit of a very weak residual interaction (see Sec. 8) the nonvanishing Y_{pi} are 0(1), whereas the Z_{pi} are proportional to matrix elements of the residual interaction. The Z_{pi} are the small quantities in our theory of the ground state energy, and our procedure is to expand all observables in a power series in these quantities.

For example, in the treatment of $\hat{H}_{40}$ we encounter the matrix element

$$\langle 0|(\psi_p^\dagger\psi_i)(\psi_q^\dagger\psi_j)|0\rangle = \sum_\alpha \langle 0|\psi_p^\dagger\psi_i|\alpha\rangle\langle\alpha|\psi_q^\dagger\psi_j|0\rangle. \quad (3.17)$$

A similar treatment is accorded all $2p-2h$ expectation values. However, in (3.13) and (3.14) we also have $4p-4h$ operators, and if our theory is to be kept simple, further compromise is necessary, at least initially.

Consider for instance the expression

$$\langle 0|(\psi_p^\dagger\psi_q^\dagger\psi_i\psi_j)(\psi_j^\dagger\psi_i^\dagger\psi_s\psi_r)|0\rangle = \langle 0|\psi_p^\dagger\psi_q^\dagger\psi_i\psi_j|0\rangle\langle 0|\psi_j^\dagger\psi_i^\dagger\psi_s\psi_r|0\rangle + \sum_I \langle 0|\psi_p^\dagger\psi_q^\dagger\psi_i\psi_j|I\rangle\langle I|\psi_j^\dagger\psi_i^\dagger\psi_s\psi_r|0\rangle, \quad (3.18)$$

where the states $|I\rangle$ which intervene in the second sum can be considered roughly to comprise the $2p-2h$ space. This second sum is composed of terms of order Z^2 $\langle I|\psi_j^\dagger\psi_r|\alpha\rangle^2 = 0(Z^4)$, i.e., the average amplitude for finding a $p-h$ pair in a state of the space $|\alpha\rangle$ should not differ substantially from the average value of Z. On the other hand, the first term of (3.18) is $0(Z^2)$ and if we confine ourselves for the moment to this approximation (see below for further discussion), we have

$$\langle 0|\psi_p^\dagger\psi_q^\dagger\psi_i\psi_j\psi_j^\dagger\psi_i^\dagger\psi_s\psi_r|0\rangle = \sum_{\alpha\alpha'} Z_{pi}^*(\alpha)Y_{qj}^*(\alpha)Y_{sj}(\alpha')Z_{ri}(\alpha'). \quad (3.19)$$

If we apply the results (3.17) and (3.19) to the evaluation of (3.11)–(3.14), we arrive at the following expression of the ground state energy:

$$\begin{aligned} W_0 = \langle 0|H|0\rangle = W_{NO} &+ \sum_{pi}(\epsilon_p - \epsilon_i)\Big\{\sum_\alpha Z_{pi}^*(\alpha)Z_{pi}(\alpha) \\ &- \frac{1}{2}\sum_{\alpha\beta}\sum_{qj} Z_{pi}^*(\alpha)Y_{qj}^*(\alpha)Y_{qj}(\beta)Z_{pi}(\beta)\Big\} \\ &+ \sum_{piqj}\sum_\alpha V(pj|iq)Z_{pi}^*(\alpha)Z_{qj}(\alpha) \\ &+ \frac{1}{4}\sum_{piqj}\sum_\alpha [V(pq|ij)Z_{pi}^*(\alpha)Y_{qj}^*(\alpha) \\ &+ V(ij|pq)Y_{pi}(\alpha)Z_{qj}(\alpha)] \\ &- \frac{1}{8}\sum_{\substack{pqrs\\ijkl}}\sum_{\alpha\beta}[V(pq|rs)\delta_{li}\delta_{kj} \\ &+ V(ij|kl)\delta_{qr}\delta_{ps}]Z_{pl}^*(\alpha)Y_{qk}^*(\alpha)Y_{rj}(\beta)Z_{si}(\beta), \end{aligned} \quad (3.20)$$

where

$$\mathcal{E}_a = h(a|a) + \sum_i V(ai|ai). \quad (3.21)$$

Equation (3.20) represents a simplification attendant also upon the assumption of natural orbitals.

In what follows, we shall seek to minimize W_0 treating the amplitudes Y and Z as variational parameters. We consider below the constraints imposed on these amplitudes. First we discuss briefly how one could go beyond the present approximation. As we have already noted below (3.18), the next level of approximation would bring in new amplitudes of the form $\langle I|\psi_j^\dagger\psi_r|\alpha\rangle$. It is idle to imagine that these could also be determined accurately from a ground state variational principle. Rather, by analogy, these should be calculated in the first instance from a corresponding variational principle for $W_\alpha = \langle\alpha|\hat{H}|\alpha\rangle$ which can be formulated in analogy with what follows below. In a more sophisticated treatment, insofar as some of the amplitudes Y and Z occur in W_α and the higher amplitudes occur in W_0, one is led to coupled variational principles, and to the extent that the coupling is important one has an expression of collective motion. These aspects are more clearly formulated in connection with specific applications, and we shall therefore not pursue these matters presently in any additional detail.

4. SUBSIDIARY CONDITIONS FOR THE VARIATIONAL PRINCIPLE

Subsidiary conditions which constrain the variation of the amplitudes Y and Z are implied by the algebra of the p–h operators and by the Pauli principle. Constraints arising from the algebra come from the ground state expectation value of the following commutators:

$$[\psi_j^\dagger \psi_q, \psi_p^\dagger \psi_i] = \delta_{pq}\delta_{ij} - \delta_{pq}\psi_i\psi_j^\dagger - \delta_{ij}\psi_p^\dagger\psi_q, \tag{4.1}$$

$$[\psi_j^\dagger \psi_q, \psi_i^\dagger \psi_p] = 0. \tag{4.2}$$

If we take the ground state expectation value, the fact that we choose natural orbitals will make the r.h.s. of (4.1) diagonal. We then obtain from this equation the condition

$$\sum_\alpha \{Y_{qj}(\alpha)Y_{pi}^*(\alpha) - Z_{qj}(\alpha)Z_{pi}^*(\alpha)\} = \delta_{ij}\delta_{pq}\{1 - \langle 0|\psi_i\psi_i^\dagger|0\rangle - \langle 0|\psi_p^\dagger\psi_p|0\rangle\} \equiv \delta_{ij}\delta_{pq}\{pi\}, \tag{4.3}$$

where

$$\{pi\} \equiv 1 - \langle 0|\psi_i\psi_i^\dagger|0\rangle - \langle 0|\psi_p^\dagger\psi_p|0\rangle$$
$$= 1 - \left(\sum_{r\alpha} Z_{ri}^*(\alpha)Z_{ri}(\alpha) - \frac{1}{2}\sum_{rsk}\sum_{\alpha\beta} Z_{ri}^*(\alpha)Y_{sk}^*(\alpha)Y_{sk}(\beta)Z_{ri}(\beta)\right)$$
$$- \left(\sum_{k\alpha} Z_{pk}^*(\alpha)Z_{pk}(\alpha) - \frac{1}{2}\sum_{klr}\sum_{\alpha\beta} Z_{pk}^*(\alpha)Y_{rl}^*(\alpha)Y_{rl}(\beta)Z_{pk}(\beta)\right). \tag{4.4}$$

It is now convenient to introduce renormalized amplitudes y, z. Using a vector notation standard in the treatments of the RPA, the relation between the renormalized amplitudes y, z and the amplitudes Y, Z can be written in the following form:

$$x_{pi}(\alpha) = \begin{pmatrix} y_{pi}(\alpha) \\ z_{pi}^*(\alpha) \end{pmatrix} = \begin{pmatrix} Y_{pi}(\alpha)\{pi\}^{-1/2} \\ Z_{pi}^*(\alpha)\{pi\}^{-1/2} \end{pmatrix} = \sum_{qj} C_{pi,qj}^{1/2} X_{qj}(\alpha), \tag{4.5}$$

where the diagonal matrix $C^{1/2}$ is given by

$$C_{pi,qj}^{1/2} = \{pi\}^{-1/2}\delta_{pq}\delta_{ij}. \tag{4.6}$$

In terms of the new amplitudes, condition (4.1) becomes

$$\sum_\alpha \{y_{qj}(\alpha)y_{pi}^*(\alpha) - z_{qj}(\alpha)z_{pi}^*(\alpha)\} = \delta_{pq}\delta_{ij}. \tag{4.7}$$

Evaluation of the commutator (4.2) gives in terms of y, z:

$$\sum_\alpha \{y_{qj}^*(\alpha)z_{pi}^*(\alpha) - y_{pi}^*(\alpha)z_{qj}^*(\alpha)\} = 0. \tag{4.8}$$

The commutators (4.1) and (4.2) for the p–h operators were of course derived by use of the anticommutation relations for the fermion field operators $\psi^\dagger$ and ψ. It is not difficult to verify that the only other constraint on the amplitudes Y, Z which does not also involve additional amplitudes, arises from the relation

$$\psi_p^\dagger\{\psi_i, \psi_j\}_+ \psi_q + (q \leftrightarrow p) = 0, \tag{4.9}$$

derived from the basic anticommutator relation $\{\psi_i, \psi_j\}_+ = 0$ by pre- and post-multiplication. The combination symmetric in p and q has been chosen in (4.9) because the constraints imposed by the antisymmetric combination are already implied by (4.8). Evaluation of (4.9) yields:

$$\sum_\alpha \{Z_{pi}^*(\alpha)Y_{qj}^*(\alpha) + Z_{pj}^*(\alpha)Y_{qi}^*(\alpha) + Z_{qi}^*(\alpha)Y_{pj}^*(\alpha) + Z_{qj}^*(\alpha)Y_{pi}^*(\alpha)\} = 0. \tag{4.10}$$

Expressed in terms of renormalization amplitudes, (4.10) becomes

$$\sum_\alpha \{pi\}^{-1/2}\{qj\}^{-1/2}(z_{pi}^*(\alpha)y_{qj}^*(\alpha) + z_{qj}^*(\alpha)y_{pi}^*(\alpha)) + \sum_\alpha \{pj\}^{-1/2}\{qi\}^{-1/2}(z_{pj}^*(\alpha)y_{qi}^*(\alpha) + z_{qi}^*(\alpha)y_{pj}^*(\alpha)) = 0. \tag{4.11}$$

Besides the conditions (4.7) and (4.8) arising from the algebra, (4.10) is the only further condition on the amplitudes as a consequence of the basic fermion anticommutation relations for $\psi^\dagger$ and ψ. Therefore, (4.10) may be viewed as a consequence of the Pauli principle and satisfaction of this equation assures that the Pauli principle is satisfied.

For the variation, these subsidiary conditions and their complex conjugates will be taken into account by Lagrange multipliers. For condition (4.7) we use Lagrange multipliers $\frac{1}{4}\lambda_{qj,pi}$, for (4.8) we use $\frac{1}{4}\mu_{qj,pi}$ and for (4.11) $\frac{1}{8}\nu_{pi,qj}$ will be used. From the structure of the subsidiary conditions the various Lagrange multipliers have the following symmetries,

$$\lambda_{pi,qj} = \lambda_{qj,pi}^*, \tag{4.12}$$

$$\mu_{pi,qj} = -\mu_{qj,pi}, \tag{4.13}$$

$$\nu_{pi,qj} = \nu_{qj,pi} = \nu_{pj,qi} = \nu_{qi,pj}. \tag{4.14}$$

Subtraction of the constraints, multiplied by the appropriate Lagrange multipliers, will allow variations of the amplitudes y and z as if they were unconstrained.

5. THE VARIATIONAL EQUATIONS

The renormalized amplitudes y^* and z will now be chosen as variational parameters. Before we actually can apply the variational principle we still have to express the ground state energy W_0, Eq. (3.20), in terms of the renormalized amplitudes. Subtraction of the subsidiary conditions by means of the Lagrange multipliers (4.12)–(4.14) then gives the following variational expression:

$$\delta\Big\{W_{NO} + \sum_{qj}(\epsilon_q - \epsilon_j)\frac{1}{\{qj\}}\sum_{\alpha\beta} z_{qj}^*(\alpha)z_{qj}(\beta) \times \left(\delta_{\alpha\beta} - \frac{1}{2}\sum_{rk}\frac{1}{\{rk\}}y_{rk}^*(\alpha)y_{rk}(\beta)\right) + \sum_{piqj}\sum_\alpha \frac{1}{\{pi\}^{1/2}}\frac{1}{\{pj\}^{1/2}}\Big\{V(pj|iq)z_{pi}^*(\alpha)z_{qj}(\alpha) + \tfrac{1}{4}V(pq|ij)z_{pi}^*(\alpha)y_{qj}^*(\alpha) + \tfrac{1}{4}V(ij|pq)z_{qj}(\alpha)y_{pi}(\alpha)\Big\} - \frac{1}{8}\sum_{ijkl}\sum_{pq}\sum_{\alpha\beta}\frac{V(ij|kl)}{\{pl\}^{1/2}\{qk\}^{1/2}\{qj\}^{1/2}\{pi\}^{1/2}} \times z_{pl}^*(\alpha)y_{qk}^*(\alpha)y_{qj}(\beta)z_{pi}(\beta) - \frac{1}{8}\sum_{pqrs}\sum_{ij}\sum_{\alpha\beta}\frac{V(pq|rs)}{\{pi\}^{1/2}\{qj\}^{1/2}\{rj\}^{1/2}\{si\}^{1/2}} \times z_{pi}^*(\alpha)y_{qj}^*(\alpha)y_{rj}(\beta)z_{si}(\beta)$$

$$-\frac{1}{4}\sum_{piqj}\sum_{\alpha}[\lambda_{pi,qj}(y^*_{pi}(\alpha)y_{qj}(\alpha)-z^*_{pi}(\alpha)z_{qj}(\alpha)$$
$$-\delta_{pq}\delta_{ij})+\text{C.C.}]$$
$$-\frac{1}{4}\sum_{piqj}\sum_{\alpha}[\mu_{pi,qj}(y^*_{pi}(\alpha)z^*_{qj}(\alpha)-z^*_{pi}(\alpha)y^*_{qj}(\alpha))+\text{C.C.}]$$
$$-\frac{1}{4}\sum_{piqj}\sum_{\alpha}\left[\nu_{pi,qj}\left(\frac{1}{\{pi\}^{1/2}\{qj\}^{1/2}}z_{pi}(\alpha)y_{qj}(\alpha)\right.\right.$$
$$\left.\left.+\frac{1}{\{pj\}^{1/2}\{qj\}^{1/2}}z^*_{pj}(\alpha)y^*_{qi}(\alpha)\right)+\text{C.C.}\right]\Bigg\}=0. \quad (5.1)$$

In carrying out the variation, there is a certain freedom in symmetrization of the main energy term in consequence of the differential forms of the constraints. With malice aforethought, we shall use the following prearrangement attendant upon (4.7):

$$\sum_{\alpha}z^*_{pi}(\alpha)\delta[z_{qj}(\alpha)]$$
$$=\frac{1}{2}\sum_{\alpha}z^*_{pi}(\alpha)\delta[z_{qj}(\alpha)]+\frac{1}{2}\sum_{\alpha}y_{qj}(\alpha)\delta[y_{pi}(\alpha)]. \quad (5.2)$$

The preference for this particular choice is that it is the only one which permits the transition from Eq. (5.3) to Eq. (5.4) below. The proof of this is given in Sec. 7.

Carrying out the variation in this way leads to the following equations in matrix notation:

$$\sum_{qj}\begin{pmatrix}A_{pi,qj} & B_{pi,qj}\\ B^*_{pi,qj} & A^*_{pi,qj}\end{pmatrix}\begin{pmatrix}y_{qj}(\alpha)\\ z^*_{qj}(\alpha)\end{pmatrix}$$
$$=\tau_3\sum_{qj}\begin{pmatrix}\lambda_{pi,qj} & \mu_{pi,qj}\\ \mu^*_{pi,qj} & \lambda^*_{pi,qj}\end{pmatrix}\begin{pmatrix}y_{qj}(\alpha)\\ z^*_{qj}(\alpha)\end{pmatrix}. \quad (5.3)$$

The component matrices A and B will be given below. As we shall prove in Sec. 7, we can introduce linear combinations of the solutions of (4.7) such that the Lagrange multiplier matrix becomes diagonal. With this the variational equations finally take the form

$$\sum_{qj}\begin{pmatrix}A_{pi,qj} & B_{pi,qj}\\ B^*_{pi,qj} & A^*_{pi,qj}\end{pmatrix}\begin{pmatrix}y_{qj}(I)\\ z^*_{qj}(I)\end{pmatrix}=\lambda_I\tau_3\begin{pmatrix}y_{pi}(I)\\ z^*_{pi}(I)\end{pmatrix}, \quad (5.4)$$

where

$$A_{pi,qj}=\frac{1}{\{pi\}^{1/2}\{qj\}^{1/2}}[(\epsilon_p-\epsilon_i)\delta_{pq}\delta_{ij}+V(pj\,|\,iq)]+(\Delta A)_{pi,qj}, \quad (5.5)$$

$$B_{pi,qj}=\frac{1}{\{pi\}^{1/2}\{qj\}^{1/2}}\Bigg\{\tfrac{1}{2}V(pq\,|\,ij)-\tfrac{1}{2}(\epsilon_q-\epsilon_j)$$
$$\times\left(\sum_{I'}z_{qj}(I')y_{pi}(I')\right)-\tfrac{1}{2}(\epsilon_p-\epsilon_i)\left(\sum_{I'}z_{pi}(I')y_{qj}(I')\right)$$
$$-\frac{1}{8}\sum_{kl}V(lk\,|\,ji)\left[\frac{1}{\{qk\}^{1/2}\{pl\}^{1/2}}\left(\sum_{I'}y_{pl}(I')z_{qk}(I')\right)\right.$$
$$\left.-\frac{1}{\{ql\}^{1/2}\{pk\}^{1/2}}\left(\sum_{I'}y_{ql}(I')z_{pk}(I')\right)\right]$$
$$-\frac{1}{8}\sum_{rs}V(pq\,|\,rs)\left[\frac{1}{\{si\}^{1/2}\{rj\}^{1/2}}\left(\sum_{I'}y_{si}(I')z_{rj}(I')\right)\right.$$
$$\left.-\frac{1}{\{sj\}^{1/2}\{ri\}^{1/2}}\left(\sum_{I'}y_{sj}(I')z_{ri}(I')\right)\right]+\nu_{pi,qj}\Bigg\}$$
$$+(\Delta B)_{pi,qj}. \quad (5.6)$$

The quantities (ΔA) and (ΔB) in (5.5) and (5.6) are the contributions which come from the variation of the quantities $\{pi\}^{1/2}$. The leading contributions have their origin in the term $\langle 0\,|\,H_{40}\,|\,0\rangle$ and its hermitian conjugate and are given by the equations

$$(\Delta A)_{pi,qj}=\tfrac{1}{16}\delta_{pq}\delta_{ij}\sum_{stkl}\left[V(st\,|\,kl)\left(\sum_{\gamma}z^*_{sk}(\gamma)y^*_{tl}(\gamma)\right)+\text{C.C.}\right]$$
$$\times\frac{1}{\{sk\}^{1/2}\{tl\}^{1/2}}\left\{\frac{1}{\{tl\}}(\delta_{li}+\delta_{tp})+\frac{1}{\{sk\}}(\delta_{ki}+\delta_{sp})\right\}, \quad (5.7)$$

$$(\Delta B)_{pi,qj}=-\frac{1}{64}\sum_{stkl}\left[V(st\,|\,kl)\left(\sum_{\gamma}z^*_{sk}(\gamma)y^*_{tl}(\gamma)\right)+\text{C.C.}\right]$$
$$\times\frac{1}{\{sk\}^{1/2}\{tl\}^{1/2}}\left(\sum_{\beta}(y_{pi}(\beta)z_{qj}(\beta)+y_{qj}(\beta)z_{pi}(\beta))\right)$$
$$\times\left\{\frac{1}{\{tl\}}(\delta_{lj}+\delta_{li}+\delta_{tq}+\delta_{tp})\right.$$
$$\left.+\frac{1}{\{sk\}}(\delta_{kj}+\delta_{ki}+\delta_{sq}+\delta_{sp})\right\}. \quad (5.8)$$

Omitted contributions are at least $0(z^3)$.

Equation (5.4) has the same matrix structure as the RPA and as we discuss in Sec. 7, the same orthonormality conditions apply, namely

$$x^\dagger(I')\tau_3x(I)=\delta_{II'}. \quad (5.9)$$

6. SOLUTION OF THE VARIATIONAL EQUATIONS

Let us now consider how one actually would solve the Eqs. (5.4). These are nonlinear equations for the amplitudes y and z^* and we consequently have recourse to an iterative procedure. As will be shown in Sec. 8, the Lagrange multiplier ν associated with the Pauli principle condition (4.11) vanishes in the limit of weak residual interaction. To start our calculation we shall set it equal to zero. We chose natural orbitals as our single particle basis, but as is clear from the defining relation (3.9), calculation of the natural orbitals requires knowledge of the true ground state $|0\rangle$. We thus do not actually have the natural orbital basis and have to start with some approximation to it. For this we choose a Hartree–Fock basis. Within our level of accuracy the residual interaction in a Hartree–Fock description does not admix $1p1h$ configurations to the HF ground state but at least $2p2h$ configurations. Therefore, part of the requirement for natural orbitals, namely

$$\langle 0\,|\,\psi^\dagger_p\psi_i\,|\,0\rangle=0, \quad (6.1)$$

will be well satisfied from the beginning, if we start with a HF basis as a first approximation to natural

orbitals. When one goes beyond the HF limit, however, it is well to keep in mind that in a general calculation, the use of natural orbitals no longer precludes the admixture of some one-particle–one-hole excitations in the ground state wave function. The reentry of such components does not appear to us to be a source of concern since low-lying excitations of this sort will mostly be ruled out by parity and angular momentum considerations and is in any case limited by the variational condition.

In the first iterative step we shall thus choose a HF basis and shall set all terms which are nonlinear in the amplitudes equal to zero. $C^{1/2}$ then becomes the unit matrix and

$$A^0_{pi,qj} = (\epsilon_p - \epsilon_i)\delta_{pq}\delta_{ij} + V(pj\,|\,iq), \tag{6.2}$$

$$B^0_{pi,qj} = \tfrac{1}{2} V(pq\,|\,ij). \tag{6.3}$$

With this the variational equations (5.4) are readily solved and gives amplitudes $y^0_{pi}(I)$ and $z^{0*}_{pi}(I)$ as a first approximation.

By means of a unitary transformation we can now go from the HF basis to a new basis which is a better approximation to the natural orbital basis. Namely the amplitudes y^0 and z^0 allow calculation of the density matrix elements

$$\langle 0\,|\,\psi^\dagger_p \psi_q\,|\,0\rangle \simeq \sum_{kI} z^{0*}_{pk}(I) z^0_{qk}(I) - \frac{1}{2} \sum_{klr} \sum_{II'} z^{0*}_{pk}(I) y^{0*}_{rl}(I) y_{rl}(I') z^0_{qk}(I') \tag{6.4}$$

and similarly for $\langle 0\,|\,\psi^\dagger_i \psi_j\,|\,0\rangle$. This density matrix can then be diagonalized by means of a unitary transformation of the single particle HF basis. The single particle energies $\mathcal{E}_a$ and the matrix elements $V(ab\,|\,cd)$ are then transformed to the new basis as are the amplitudes $y^0_{pi}(I)$ and $z^0_{pi}(I)$. Now new component matrices A^1 and B^1 are calculated where y^0, z^0 are taken for the computation of the nonlinear terms. The variational equations (5.4) are then solved with A^1 and B^1 to give new amplitudes $y^1_{pi}(I)$ and $z^{1*}_{pi}(I)$.

The procedure outlined above is repeated with diagonalization of the density matrix and calculation of new component matrices A, B. The iteration is then carried on until self-consistency is achieved, i.e., until the amplitudes y and z^* no longer change in successive iterations.

As we pointed out before, if the true ground state is described in terms of natural orbitals, the admixture of p–h configurations to the ground state Slater determinant is smallest.

This indicates that the residual interaction which effects this admixture, can usually be expected to be small too. For small residual interaction the Lagrange multiplier ν goes to zero and the treatment outlined above is adequate. If ν cannot be expected to be zero then one faces a more difficult situation. The homogeneous equations (5.4) involving the Lagrange multipliers ν and λ_I have then to be solved simultaneously with the subsidiary condition (4.11) arising from the Pauli principle, and the normalization condition (5.9) which in this case is equivalent to the conditions imposed by the algebra. A method for the solution of such a problem has been developed, but its exposition will be left to the context of specific applications.

7. SOME PROPERTIES OF THE VARIATIONAL EQUATIONS

We consider in particular the proof that (5.3) can be replaced by (5.4). We write the former as

$$\mathcal{K} x(\alpha) = \tau_3 \mathcal{N} x(\alpha). \tag{7.1}$$

Equation (7.1), which emerges most directly from the variation is to be solved in conjunction with the subsidiary conditions. We are thus seeking the null space of the matrix $\mathcal{K} - \tau_3 \mathcal{N}$. If we have found members of that space, then any linear combination of these is equally a member of the space. We may then ask if we can find a linear combination

$$x(I) = \sum C_{I\alpha} x(\alpha) \tag{7.2}$$

such that $x(I)$ is an eigenvector of the matrix $\mathcal{N}$, i.e.,

$$\mathcal{N} x(I) = \lambda_I x(I). \tag{7.3}$$

The vector $x(I)$ must then also satisfy (5.4).

It is easily shown that a necessary condition for the compatibility of (5.3) and (5.4) is that $\mathcal{K}$ have the RPA structure, namely

$$\tau_1 \mathcal{K}^* \tau_1 = \mathcal{K}, \tag{7.4}$$

and this condition, which is here satisfied, resulted from the symmetrization (5.2). Equation (7.4) leads, of course, to the famous doubling of solutions known from the RPA.

The essential part of the argument proceeds now as follows. Except under the special circumstances where we have chosen an unstable starting HF solution, the solutions of (5.4) will be real and fall into two sets $\pm\lambda_I$ with associated vectors $x(I)$ and $\tau_1 x^*(I)$. The orthogonality relation (5.9) follows from the equations and the normalization is imposed. The solutions then satisfy a matrix completeness relation,

$$\sum_I \{x(I)x^\dagger(I) - \tau_1 x^*(I)[\tau_1 x^*(I))^\dagger\} = \tau_3. \tag{7.5}$$

But this is just a matrix version of the subsidiary conditions (4.7), (4.8) and their complex conjugates. This then assures us of the validity of (7.2) and (7.3), i.e., the $x(I)$ are indeed a choice of the $x(\alpha)$, in fact the simplest possible choice.

In Sec. 9, we shall investigate the relationship of the present formulation to the RPA. In the latter method the states $|I\rangle$ are identified with eigenstates of the Hamiltonian. It is therefore well to emphasize that there is no such imputation in the present treatment. The states $|\alpha\rangle$ or $|I\rangle$ span a certain space. It is the same space as that of the RPA solutions, but no condition that the Hamiltonian be diagonal in the excited subspace has been imposed. Thus the λ_I are not in general the excitation energies of the system. These remarks resolve a nonexistent dilemma raised in our previous work.

8. THE PERTURBATION LIMIT

In this section we show that in the limit of weak residual interaction, the solutions of the variational equations (5.4) indeed give the correct perturbation theory result for the correlation energy. As we shall see, this also implies that the Lagrange multipliers ν associated with the Pauli principle condition (4.11) vanish.

In this limit ground state correlations, described by the amplitudes z, will be small. The renormalization matrix $C^{1/2}$ will approach unity and we have

$$z^*_{pi}(I) \cong Z^*_{pi}(I). \tag{8.1}$$

and

$$y_{pi}(I) \simeq Y_{pi}(I). \tag{8.2}$$

The excited states become uncoupled p–h excitations, i.e.,

$$|I\rangle \simeq |rk\rangle = \psi^\dagger_r \psi_k |0\rangle. \tag{8.3}$$

with the help of (3.15) we then find for the amplitudes $Y_{pi}(I)$

$$Y_{pi}(rk) = \langle 0|\psi^\dagger_i \psi_p |rk\rangle \sim \langle \Phi_0 |\psi^\dagger_i \psi_p |rk\rangle = \delta_{pr}\,\delta_{ik}. \tag{8.4}$$

Remembering that in this limit V is small compared to the single particle energies $\mathcal{E}$ and Z is small compared to Y, we find from the equation for Y in (5.4)

$$\lambda(rk) = \epsilon_r - \epsilon_k, \tag{8.5}$$

where we used the value (8.4) for Y. Neglecting the particle-particle and hole-hole scattering terms, the component matrix B^* becomes

$$B^*_{pi,qj} = \tfrac{1}{2} V(ij|pq) - \tfrac{1}{2}(\epsilon_p - \epsilon_q - \epsilon_i - \epsilon_j)\, Z^*_{pi}(qj), \tag{8.6}$$

if we assume that in this limit $\nu_{pi,qj} = 0$. With (8.4), (8.5), and (8.6) we find from the equation for $Z^*_{pi}(I)$ in (5.4)

$$Z_{pi}(qj) = -\frac{V(ij|pq)}{\epsilon_p + \epsilon_q - \epsilon_i - \epsilon_j}. \tag{8.7}$$

We can now verify the assumption that the Lagrange multiplier ν vanishes in the weak limit by actually calculating the correlation energy. Inserting the values (8.4) and (8.7) in the expression (3.20) for the ground state energy we find to second order in V for the correlation energy:

$$W_O - W_{NO} = -\frac{1}{4} \sum_{piqj} \frac{|V(pq|ij)|^2}{\epsilon_p + \epsilon_q - \epsilon_i - \epsilon_j}. \tag{8.8}$$

This is the correct result as given by second order perturbation theory. The vanishing of the Lagrange multipliers $\nu_{pi,qj}$ in this limit is also consistent with the fact that the values (8.4) and (8.7) for the amplitudes Y, Z satisfy the Pauli principle restriction (4.10) as they should.

9. RELATION TO THE RPA EXCITED STATES

Given the observations of the previous section, a relation of the eigenvalue problem Eq. (5.4) to the RPA is easily stated. First we must restrict ourselves to the approximation (8.6) for the matrix $B^*_{pi,qj}$ and next we insert in that expression the perturbation value (8.7) for $Z^*_{pi}(qj)$. We thus obtain

$$B^*_{pi,qj} = V(ij|pq), \tag{9.1}$$

which is the required form. This informs us that for weak coupling the eigenvalues, λ_I, will not differ much from the excitation energies and can for all intents and purposes be identified with them.

It is, however, unnecessary to be satisfied with this state of affairs. One can first construct the space of states $|I\rangle$ which are characterized by the variational problem and afterwards diagonalize the Hamiltonian in this space to find the actual excited states. This is so formulated that if the Hamiltonian is already diagonal, we are dealing with Rowe's version of the RPA.

To carry out these steps, we define a set of states $|I\rangle$ by the requirements

$$|I\rangle = 0^\dagger_I |0\rangle, \tag{9.2}$$

$$\langle I'|I\rangle = \delta_{II'}, \tag{9.3}$$

$$0_I|0\rangle = 0 = \langle I'|0_I|0\rangle, \tag{9.4}$$

$$0^\dagger_I = \sum_{pi} [\bar{y}_{pi}(I)\psi^\dagger_p \psi_i - \bar{z}^*_{pi}\psi^\dagger_i \psi_p]. \tag{9.5}$$

From (9.3) and (9.4), we obtain, using (9.5)

$$\delta_{II'} = x^\dagger(I')\bar{x}(I), \tag{9.6}$$

$$0 = (\tau_1 x^*(I'))^\dagger \bar{x}(I), \tag{9.7}$$

where

$$\bar{x}_{pi}(I) = \begin{pmatrix} \bar{y}_{pi}(I) \\ \bar{z}^*_{pi}(I) \end{pmatrix}. \tag{9.8}$$

If we choose

$$\bar{x}(I) = CX(I), \tag{9.9}$$

(9.6) and (9.7) reduce to the orthonormality requirements for the solutions of the variational problem, so that this choice allows us to identify the states with these solutions.

Next we construct the matrix

$$\langle I'|(H - W)|I\rangle = \tfrac{1}{2}\{\langle 0|[0_{I'}, [H, 0^\dagger_I]]|0\rangle + \langle 0|[[0_{I'}, H], 0^\dagger_I]|0\rangle\}, \tag{9.10}$$

where the symmetrization has been carried out to assure hermiticity. If this matrix is diagonal, it then gives the excitation energy. In any event, it can be computed from the information previously yielded by the solution of the variational problem and then diagonalized to yield the excitation energies. The eigenvectors provide us with a unitary transformation which carry us from the set $|I\rangle$ to a set $|K\rangle$, such that the $X(K)$ are RPA amplitudes in the sense noted above.

A final remark: It is possible to force the lowest order variational problem to yield the RPA equations of motion, but in fact we see little advantage in doing so. As we have seen, for weak coupling, numerical differences between the RPA eigenvalues and amplitudes and those of the variational problem are slight. When the coupling increases, deviations from the RPA can be serious, whereas a suitably formulated nonlinearized variational treatment can continue to be accurate, as we shall demonstrate by future example.

---

*Supported in part by the U.S. Atomic Energy Commission.

†Based in part on a thesis submitted by F. K. to the University of Pennsylvania, August 1972.

‡Now at the University of Washington, Seattle.

[1]W. Bohne, H. Homeyer, H. Morgenstern, and J. Scheer, Nucl. Phys. A. **113,** 97 (1968).
[2]K. H. Purser, W. P. Alford, D. Cline, H. W. Fulbright, H. E. Gove, and M. S. Krick, Nucl. Phys. A. **132,** 75 (1969).
[3]A. P. Zucker, B. Buck, and J. B. McGrory, Phys. Rev. Lett. **21,** 39 (1968).
[4]C. D. Latorre and J. M. Irvine, Nucl. Phys. A. **188,** 524 (1972).
[5]E. A. Sanderson, Phys. Lett. **19,** 141 (1965).
[6]D. J. Rowe, Phys. Rev. **175,** 1283 (1968).
[7]R. E. Johnson, R. M. Dreizler, and A. Klein, Phys. Rev. **186,** 1289 (1969).
[8]P. J. Ellis and L. Zamick, Ann. Phys. (N.Y.) **55,** 61 (1969).
[9]P. J. Ellis, Nucl. Phys. A. **155,** 635 (1970).
[10]R. M. Dreizler, A. Klein, F. R. Krejs, and G. J. Dreiss, Nucl. Phys. A. **166,** 624 (1971).
[11]S. Y. Li and A. Klein, Phys. Rev. C **3,** 1871 (1971).
[12]G. Schalow and M. Yamamura, Nucl. Phys. A. **161,** 93 (1971).
[13]M. Iwasaki and M. Yamamura, Prog. Theor. Phys. **46,** 1440 (1971).
[14]M. V. Mihailovic and M. Rosina, *Nuclear Many Body Problems; Herceg Novi*, edited by M. V. Mihailovic *et al*. (Yugoslav Nuclear Energy Commission, 1967); M. V. Mihailovic and M. Rosina, Nucl. Phys. A. **130,** 386 (1969).
[15]G. J. Dreiss, A. Klein, and S. C. Pange, Phys. Lett. **29,** 465 (1969); R. M. Dreizler and A. Klein, Phys. Lett. B **30,** 236 (1969).
[16]A. J. Coleman, Rev. Mod. Phys. **35,** 668 (1973); H. Kummer, J. Math. Phys. **8,** 2063 (1967); C. Garrod and J. K. Percus, J. Math. Phys. **5,** 1756 (1969); F. Weinhold and E. B. Wilson, J. Chem. Phys. **47,** 2298 (1969); E. R. Davidson, J. Math. Phys. **10,** 725 (1969).
[17]M. Rosina, *Direct Variational Calculation of the Two-Body Density Matrix*, preprint, University of Ljubljana.
[18]B. E. Chi, Nucl. Phys. A. **164,** 449 (1970).
[19]N. Ullah and K. K. Gupta, J. Math. Phys. **13,** 1163 (1972).
[20]P. O. Lowden, Phys. Rev. **97,** 1474 (1955).
[21]E. R. Davidson, Rev. Mod. Phys. **44,** 451 (1972).
[22]D. H. Kobe, J. Chem. Phys. **50,** 5183 (1969).
[23]A. E. L. Dieperink, P. J. Brussard, and R. Y. Cusson, Nucl. Phys. A. **180,** 110 (1972).
[24]D. J. Rowe, *Nuclear Collective Motion* (Methuen, London, 1970).
[25]D. J. Rowe, in *Dynamic Structure of Nuclear States*, edited by D. J. Rowe, L. E. H. Trainor, S. S. M. Wong, and T. W. Donnelly (University of Toronto Press, Toronto, 1972).
[26]R. Padjen and G. Ripka, Nucl. Phys. A. **149,** 273 (1970).
[27]A. Faessler and A. Plastino, Z. Phys. **220,** 88 (1969).
[28]A. Faessler, A. Plastino, and K. W. Schmid,
[29]Q. Ho-Kim, Phys. Rev. C **4,** 1077 (1971). Phys. Lett. B **34,** 31 (1971).
[30]N. W. Kirson, Ann. Phys. (N.Y.) **66,** 624 (1971) and references contained therein.

Covariance of the conserved current equations and transformations of Noether's current

F. Bayen and P. Hillion

Institut Henri Poincaré, Paris, France

(Received 29 November 1971; revised manuscript received 10 January 1972)

The covariance of the conserved current equations is first proved in a direct way without using the transformation laws of Noether currents. Then the transformation of these currents is investigated and we find, when the action integral is invariant up to a divergence, a result which does not seem to have been noted previously.

1. INTRODUCTION

The covariance problem for the conserved current equations under the invariance group of the action integral has been recently set and treated in particular cases.[1] In this work, we generalize this result which was concerned separately with external and internal groups, and a simplified proof of covariance is given without using the transformation laws of currents. It is then shown that currents co-transform under the invariance group in a non-trivial way.

Let $\psi(x) = (\psi_1(x), \ldots, \psi_n(x))$ be any fields defined on the vector space $\mathbb{R}^m$ ($\psi_i(x) \in \mathbb{R}$, ψ_i is assumed to be twice continuously differentiable for $i = 1, \ldots, n$) and $\mathcal{L}(x, \psi(x), \partial\psi(x)) \equiv \mathcal{L}(x^\mu, \psi_i(x), \partial_\mu\psi_i(x))$, $x = (x^1, \ldots, x^m) \in \mathbb{R}^m$, is a local Lagrangian on this set of fields. We assume that $\mathcal{L}$ has continuous first and second partial derivatives with respect to all arguments x^μ, ψ_i, and $\partial_\mu\psi_i$.

The action integral

$$\mathcal{I} = \int_\Omega \mathcal{L}(x, \psi(x), \partial\psi(x))dx, \qquad dx = dx^1 \cdots dx^m$$

on an arbitrary simply-connected domain $\Omega \subset \mathbb{R}^m$ is assumed to be invariant up to a divergence under a real r-dimensional connected Lie group G acting (on the left) on $\mathbb{R}^m$ and on fields in the following way ($g \in G$):

$$x \xrightarrow{g} x' = f(g,x) = g\cdot x, \tag{1.1}$$

$$\psi(x) \xrightarrow{g} \psi'(x') = S(g,x,\psi(x)) \tag{1.2}$$

with the group law (e is the identity of G, $g, g' \in G$)

$$x = f(e,x), \quad g'\cdot(g\cdot x) = (g'g)\cdot x, \tag{1.1'}$$

$$\psi(x) = S(e,x,\psi(x)), \quad S(g'_i g\cdot x, S(g,x,\psi(x))) = S(g'g,x,\psi(x)), \tag{1.2'}$$

for any $x \in \mathbb{R}^m$ and $\psi(x) \in \mathbb{R}^n$.

Throughout this paper all expressions are assumed to have derivatives to a sufficient order to make computations meaningful, and we use the following notation:

$$J(g,x) = \det(\partial_\nu x'^\mu),$$

$$J^\mu{}_\nu(g,x) = \partial_\nu x'^\mu, \quad A^\mu{}_\nu(g,x) = J(g,x)^{-1}\partial_\nu x,^\mu,$$

$$\partial'_\mu = \frac{\partial}{\partial x'^\mu} = J^{-1\nu}{}_\mu(g,x)\partial_\nu,$$

where ∂_μ is the total derivative $\partial/\partial x^\mu$ while $D_\mu, D_i, D_{i\mu}$ are partial derivatives with respect to $x^\mu, \psi_i, \partial_\mu\psi_i$ respectively; the summation convention on dummy indices will be used. The Lie algebra of G will be denoted $\mathfrak{g}$; we assume that (a_s), $1 \leq s \leq r$, defines a basis of $\mathfrak{g}$ and we write for any $a \in \mathfrak{g}$:

$$\varphi_a(x) = \left[\frac{d}{du} f(\exp au, x)\right]_{u=0}, \tag{1.3}$$

$$s_a(x,\psi(x)) = \left[\frac{d}{du} S(\exp au, x, \psi(x))\right]_{u=0}. \tag{1.3'}$$

The invariance of the action integral up to a divergence may be written ($\Omega' = \{g\cdot x : x \in \Omega\}$)

$$\int_{\Omega'} \mathcal{L}(x', \psi'(x'), \partial'\psi'(x'))dx' = \int_\Omega [\mathcal{L}(x, \psi(x), \partial\psi(x)) + \partial_\mu R^\mu(g,x,\psi(x))]dx.$$

We bear in mind that in this definition the function $\mathcal{L}$ is the same on both sides. The arbitrariness of Ω then implies

$$\mathcal{L}(x', \psi'(x'), \partial'\psi'(x')) = J(g,x)^{-1}[\mathcal{L}(x, \psi(x), \partial\psi(x)) + \partial_\mu R^\mu(g,x,\psi(x))]. \tag{1.4}$$

Let us note that there is no *a priori* reason why the term $R(g,x,\psi(x))$ should not depend explicitely on x even if $\mathcal{L}$ does not[2]; we assume for R^μ the same conditions of continuity and differentiability as those satisfied by S.

We now briefly review Noether's theorem in the framework of the previous assumptions, which are not the most general ones, but nevertheless the most frequently encountered in physics.[3] Equation (1.4) then gives, if we write, assuming that $a \in \mathfrak{g}$,

$$x^* = (\exp au)\cdot x, \qquad \psi^*(x^*) = S(\exp au, x, \psi(x)),$$

$$J(\exp au, x)\mathcal{L}\left(x^*, \psi^*(x^*), \frac{\partial}{\partial x^*}\psi^*(x^*)\right) - \mathcal{L}(x, \psi(x), \partial\psi(x)) = \partial_\mu R^\mu(\exp au, x, \psi(x)).$$

Differentiating this relation with respect to u for $u = 0$, after some simplification, if we write

$$j^\mu_a(x) = j^\mu_a(\mathcal{L}; x, \psi(x)\partial\psi(x)) = \varphi^\mu_a(x)\mathcal{L}(x, \psi(x), \partial\psi(x)) + (s_{ai}(x,\psi(x)) - \varphi^\rho_a(x)\partial_\rho\psi_i(x))D_{i\mu}\mathcal{L}(x, \psi(x), \partial\psi(x)), \tag{1.5}$$

$$r^\mu_a(x) = r^\mu_a(\mathcal{L}; x, \psi(x)) = \left[\frac{d}{du} R^\mu(\exp au, x, \psi(x))\right]_{u=0}, \tag{1.6}$$

$$I^\mu_a(x) = I^\mu_a(\mathcal{L}; x, \psi(x), \partial\psi(x)) = j^\mu_a(x) - r^\mu_a(x), \tag{1.7}$$

we get[4]

$$\partial_\mu I^\mu_a(\mathcal{L}; x, \psi(x), \partial\psi(x)) + (s_{ai}(x,\psi(x)) - \varphi^\rho_a(x)\partial_\rho\psi_i(x))(D_i - \partial_\mu D_{i\mu})\mathcal{L}(x, \psi(x), \partial\psi(x)) = 0. \tag{1.8}$$

Thus, for those fields satisfying the Euler–Lagrange equations, there are conservation equations,

$$\partial_\mu I^\mu_{as}(\mathcal{L}; x, \psi(x), \partial\psi(x)) = 0. \tag{1.9}$$

Copyright © 1973 by the American Institute of Physics

In the following, since it is sufficient to consider only one current, we will write a for a_s.

At this point it is worth noticing that if the Langrangian $\mathcal{L}_1(x,\psi(x),\partial\psi(x)) = \mathcal{L}(x,\psi(x),\partial\psi(x)) + \partial_\mu L^\mu(x,\psi(x))$, where the functions $L^\mu(x,\psi(x))$ are arbitrary, is substituted for $\mathcal{L}(x,\psi(x),\partial\psi(x))$ we get the same equations of motion.

It is known that this possibility is not the most general one, see Rund[4] and Edelen.[5] This implies that the current $I_a(x)$ is not uniquely defined. In fact, if we write

$$R_1(g,x,\psi(x)) = R(g,x,\psi(x)) + A(g^{-1},x')L(x',\psi'(x')) - L(x,\psi(x))$$

and if $\mathcal{L}$ satisfies (1.4), we get by the lemma of sec. 2

$$\mathcal{L}_1(x',\psi'(x'),\partial'\psi'(x')) = J(g,x)^{-1}[\mathcal{L}(x,\psi(x),\partial\psi(x)) + \partial_\mu R^\mu{}_1(g,x,\psi(x))].$$

Therefore, we have r currents $I_a(\mathcal{L}_1;x,\psi(x),\partial\psi(x))$ satisfying (1.8) with $\mathcal{L}_1$ instead of $\mathcal{L}$. An easy computations shows that if $L(x,\psi(x),\partial\psi(x)) = \partial_\mu L^\mu(x,\psi(x))$, the new current $I_a(\mathcal{L}+L;x,\psi(x),\partial\psi(x))$ may be written

$$I^\mu_a(\mathcal{L}+L;x,\psi(x),\partial\psi(x)) = I^\mu_a(\mathcal{L};x,\psi(x),\partial\psi(x)) + \partial_\rho(\varphi^\mu_a(x)L^\rho(x,\psi(x)) - \varphi^\rho_a(x)L^\mu(x,\psi(x))).$$

We will prove in a straightforward way the covariance of equation (1.9) under the group G, that is for any $g \in G$:

$$\partial_\mu I^\mu_a(\mathcal{L};x,\psi(x),\partial\psi(x)) = 0 \quad \text{for any } a \in \mathfrak{g}$$

implies

$$\partial'_\mu I^\mu_a(\mathcal{L};x',\psi'(x'),\partial'\psi'(x')) = 0 \quad \text{for any } a \in \mathfrak{g}.$$

After having proved in Sec. 2 the covariance of conserved current equations, we consider in sec. 3 the transformations of currents and first show that when the Lagrangian density transforms without any divergence,

$$j^\mu_a(\mathcal{L};x,\psi(x),\partial\psi(x)) \overset{g}{\to} j^\mu_a(\mathcal{L};x',\psi'(x'),\partial'\psi'(x')) = A^\mu_\nu(g,x)j^\nu_{\mathrm{ad}(g^{-1})a}(\mathcal{L};x,\psi(x)\partial\psi(x)),$$

where $g \to \mathrm{ad}(g)$ is the adjoint representation of G. This result does not seem to have been given in the literature

In particular, let $G = P$, the Poincaré group, be the invariance of $\mathcal{L}$ and $(p_\mu)_{0\le\mu\le3}$, $m_{\mu\nu} = -m_{\nu\mu}$, $0 \le \mu,\nu \le 3$ be the usual Lie algebra basis of P. If we write $j^\mu_{p_\rho}(x) = -T^{\rho\mu}(x)$ and $j^\mu_{m_{\rho\sigma}}(x) = J^{\rho\sigma\mu}(x)$, one can easily find (see appendix) that if $g = (b,\Lambda) \in P$, one obtains the well-known result

$$T'^{\mu_1\mu_2}(x') = \Lambda^{\mu_1}{}_{\nu_1}\Lambda^{\mu_2}{}_{\nu_2}T^{\nu_1\nu_2}(x), \tag{1.10}$$

$$J'^{\mu_1\mu_2\mu_3}(x') = \Lambda^{\mu_1}{}_{\nu_1}\Lambda^{\mu_2}{}_{\nu_2}\Lambda^{\mu_3}{}_{\nu_3}J^{\nu_1\nu_2\nu_3}(x) - \Lambda^{\mu_3}{}_{\nu_3}(\Lambda^{\mu_1}{}_{\nu_2}b^{\mu_2} - \Lambda^{\mu_2}{}_{\nu_2}b^{\mu_1})T^{\nu_2\nu_3}(x), \tag{1.11}$$

and this means that under a general Poincaré transformation the new components of the total angular momentum density are a linear combination of the previous ones and of the energy momentum density. So $J^{\sigma\mu}(x)$ is not strictly speaking a 3-index tensor with respect to P, though as can be easily checked, the formula (1.11) is consistent with the group structure of P.

When the Lagrangian density transforms according to (1.4), it is then shown that the current cotransforms according to the following law:

$$I^\mu_a(\mathcal{L}+R_{g^{-1}};x',\psi'(x'),\partial'\psi'(x') = A^\mu_\nu(g,x)I^\nu_{\mathrm{ad}(g^{-1})a}(\mathcal{L};x,\psi(x),\partial\psi(x)),$$

where

$$R_g(x,\psi(x),\partial\psi(x)) = \partial_\mu R^\mu(g,x,\psi(x)).$$

2. COVARIANCE OF CONSERVED CURRENT EQUATION

Equation (1.8) holds for any ψ, in particular for ψ',

$$\partial_\mu I^\mu_a(\mathcal{L};x,\psi'(x),\partial\psi'(x)) + (s_{ai}(x,\psi'(x)) - \varphi^\nu_a(x)\partial_\nu\psi'_i(x))(D_i - \partial_\mu D_{i\mu}) \times \mathcal{L}(x,\psi'(x),\partial\psi'(x)) = 0.$$

Substituting the variable x' for x, we obtain, with an obvious definition for $D'_i, D'_{i\mu}$,

$$\partial'_\mu I^\mu_a(\mathcal{L};x',\psi'(x'),\partial'\psi'(x')) + (s_{ai}(x',\psi'(x')) - \varphi^\nu_a(x')\partial'_\nu\psi'_i(x'))(D'_i - \partial'_\mu D'_{i\mu})\mathcal{L}(x',\psi'(x'),\partial'\psi'(x')) = 0. \tag{2.1}$$

We now compute $(s_{ai}(x',\psi'(x')) - \varphi^\nu_a(x')\partial'_\nu\psi'_i(x'))(D'_i - \partial'_\mu D'_{i\mu})\mathcal{L}(x',\psi'(x'),\partial'\psi'(x'))$ and $(D'_i - \partial'_\mu D'_{i\mu})\mathcal{L}(x',\psi'(x'),\partial'\psi'(x'))$.

Equation (1.4) may be written

$$\mathcal{L}(x',\psi'(x'),\partial'\psi'(x')) = J(g^{-1},x')(\mathcal{L}+R_g)(g^{-1}\cdot x', S(g^{-1},x'\psi'(x')), J^\mu_\nu(g,x)\partial'_\mu S(g^{-1},x',\psi'(x'))), \tag{2.2}$$

remembering that $R_g(x,\psi(x),\partial\psi(x)) = \partial_\mu R^\mu(g,x,\psi(x))$. The differentiation of (2-2) with respect to ψ'_i, $\partial'_\mu\psi'_i$ gives

$$D'_i\mathcal{L}(x',\psi'(x'),\partial'\psi'(x')) = J(g,x)^{-1}\{[D'_iS_j(g^{-1},x',\psi'(x'))] \times Dj(\mathcal{L}+R_g)(x,\psi(x),\partial\psi(x)) + J^\mu{}_\nu(g,x)[D'_i\partial'_\mu S_j(g^{-1},x',\psi'(x'))] \times D_{j\nu}(\mathcal{L}+R_g)(x,\psi(x),\partial\psi(x))\}, \tag{2.3}$$

$$D'_{i\mu}\mathcal{L}(x',\psi'(x'),\partial'\psi'(x')) = A^\mu{}_\nu(g,x)[D'_iS_j(g^{-1},x',\psi'(x'))] \times D_{j\nu}(\mathcal{L}+R_g)(x,\psi(x),\partial\psi(x)) = A^\mu{}_\nu(g,x)K^\nu(x). \tag{2.4}$$

If we write

$$K^\nu(x) = [D'_iS_j(g^{-1},x',\psi',(x'))\,D_{j\nu}(\mathcal{L}+R_g)(x,\psi(x),\partial\psi(x)),$$

the computation of $\partial'_\mu D'_{i\mu}\mathcal{L}(x',\psi'(x'),\partial'\psi'(x'))$ will be based on the following lemma which will also be used in the next section.

Lemma: If under the transformation (1.1) a vector $K(x)$ cotransforms according to

$$K^\mu(x) \overset{g}{\to} K'^\mu(x') = A^\mu{}_\nu(g,x)K^\nu(x),$$

then

$$\partial'_\mu K'^\mu(x') = J(g,x)^{-1}\partial_\mu K^\mu(x).$$

Proof: With the notations already used we may write

$$\partial'_\mu K'^\mu(x') = J(g,x)^{-1}\,[J(g,x)\,(\partial_\nu J(g,x)^{-1}) + J^{-1\rho}{}_\mu(g,x)\partial_\rho J^\mu{}_\nu(g,x)] \times K^\nu(x) + J(g,x)^{-1}\partial_\mu K^\mu(x).$$

Now Jacobi's lemma[6] states that if $C^\mu{}_\rho$ is the cofactor of $\partial_\rho x'^\mu$ in the Jacobian $J(g,x)$, then the following identities hold:

$$\sum_{\nu=1}^{m} \partial_\nu C^\mu{}_\nu = 0, \qquad \mu = 1,\ldots,m.$$

But $C^\mu{}_\rho = J(g,x)J^{-1\rho}{}_\mu(g,x)$ and, therefore,

$$J(g,x)^{-1}(\partial_\rho J(g,x))J^{-1\rho}{}_\mu(g,x) + \partial_\rho J^{-1\rho}{}_\mu(g,x) = 0.$$

Multiplying both sides by $J^\mu{}_\nu(g,x)$ and summing over μ, we get

$$-J(g,x)\partial_\nu J(g,x)^{-1} - J^{-1\rho}{}_\mu(g,x)\partial_\rho J^\mu{}_\nu(g,x) = 0,$$

which proves the lemma.

Applying the lemma to equation (2.4) gives

$$\partial'_\mu D'_{i\mu}\mathfrak{L}(x',\psi'(x'),\partial'\psi'(x')) = J(g,x)^{-1}\partial_\mu\{[D'_i S_j(g^{-1},x',\psi',(x'))] \times D_{j\mu}(\mathfrak{L}+R_g)(x,\psi(x),\partial\psi(x))\}.$$

From this equation and (2.3), it follows that

$$(D'_i - \partial'_\mu D'_{i\mu})\mathfrak{L}(x',\psi'(x'),\partial'\psi'(x')) = J(g,x)^{-1}[D'_i S_j(g^{-1},x',\psi'(x'))] \times (D_j - \partial_\mu D_{j\mu})(\mathfrak{L}+R_g)(x,\psi(x),\partial\psi(x)) + A^\mu{}_\nu(g,x)\,[D_{j\nu}(\mathfrak{L}+R_g)(x,\psi(x),\partial\psi(x))] \times (D'_i\partial'_\mu - \partial'_\mu D'_i)S_j(g^{-1},x',\psi'(x')).$$

But it is easily seen that for any function $\Phi(x,\psi(x))$ we have $(D_i\partial_\mu - \partial_\mu D_i)\,\Phi(x,\psi(x)) = 0$ and, therefore, we get

$$(D'_i - \partial'_\mu D'_{i\mu})\mathfrak{L}(x',\psi'(x'),\partial'\psi'(x')) = J(g,x)^{-1}[D'_i S_j(g^{-1},x',\psi'(x'))] \times (D_j - \partial_\mu D_{j\mu})\mathfrak{L}(x,\psi(x),\partial\psi(x)). \tag{2.5}$$

In order to compute $s_{ai}(x',\psi'(x')) - \varphi^\rho_a(x')\partial'_\rho\psi'_i(x')$

we first notice that if $t \in \mathbb{R}$,

$$t\,s_a(x',\psi'(x')) = S(\exp at, x', \psi'(x')) - \psi'(x') + o(t).$$

But if we write $a' = \mathrm{ad}(g^{-1})\,a$, gives equation (1.2′)

$$\begin{aligned} tsa'(x,\psi(x)) &= S(g^{-1}\exp(at)g, x, \psi(x)) - \psi(x) + o(t)\\ &= S(g^{-1}, (\exp at)g\cdot x, S(\exp(at)g, x, \psi(x))) - \psi(x) + o(t)\\ &= S(g^{-1}, x' + t\varphi_a(x') + o(t), \psi'(x')\\ &\quad + t\,s_a(x',\psi'(x')) + o(t)) - \psi(x) + o(t)\\ &= tD'_\mu S(g^{-1},x',\psi'(x'))\varphi^\mu_a(x')\\ &\quad + tD'_i S_j(g^{-1},x',\psi'(x'))s_{ai}(x',\psi'(x')) + o(t), \end{aligned}$$

from which we get

$$s_{ai}(x',\psi'(x'))D'_i S_j(g^{-1},x'\psi'(x')) = s_{a'j}(x,\psi(x)) - \varphi^\mu_a(x')D'_\mu S_j(g^{-1},x'\psi'(x')). \tag{2.6}$$

In the same way, if $t \in \mathbb{R}$, we get from eq. (1.1′)

$$\begin{aligned} t\varphi_a(x') &= (\exp at)\cdot x' - x' + o(t)\\ &= g\cdot(\exp t\,\mathrm{d}(g^{-1})a\cdot x) - g\cdot x + o(t)\\ &= g\cdot(x + t\varphi_a,(x) + o(t)) - g\cdot x + o(t), \end{aligned}$$

so that

$$\varphi^\mu_a(x') = J^\mu{}_\nu(g,x)\varphi^\nu_a,(x). \tag{2.7}$$

Then, using relation (2.7), we get

$$\varphi^\mu_a(x')\partial'_\mu\psi'_i(x') = \varphi^\nu_a,(x')[D_\nu S_i(g,x,\psi(x)) + (D_K S_i(g,x,\psi(x)))\partial_\nu\psi_K(x))]. \tag{2.8}$$

From the Eqs. (2.6) and (2.8), we have

$$\begin{aligned} &[s_{ai}(x',\psi'(x')) - \varphi^\rho_a(x')\partial'_\rho\psi'_i(x')]\,D'_i S_j(g^{-1},x',\psi'(x'))\\ &\quad = s_{a'j}(x,\psi(x)) - \varphi^\rho_a,(x)\partial_\rho\psi_j(x)\\ &\quad\quad - \varphi^\rho_a,(x)[J^\sigma_\rho(g,x)D'_\sigma S_j(g^{-1},x',\psi'(x'))\\ &\quad\quad + D_\rho S_i(g,x,\psi(x))D'_i S_j(g^{-1},x',\psi'(x'))]. \end{aligned} \tag{2.9}$$

To get this equality, we used the relation

$$[D'_i S_j(g^{-1},x',\psi'(x'))]\,D_K S_i(g,x,\psi(x)) = \delta_{jK},$$

obtained from the derivation with respect to ψ_K of the identity

$$S(g^{-1},g\cdot x, S(g,x,\psi(x))) = \psi(x).$$

But applying D_ρ to this identity gives

$$D'_\sigma S(g^{-1},x',\psi'(x'))J^\sigma{}_\rho(g,x) + [D'_i S(g^{-1},x',\psi'(x'))]\,D_\rho S_i(g,x,\psi(x)) = 0,$$

and this supplies an obvious simplification of (2.9) (the term between the square brackets is zero):

$$[s_{ai}(x',\psi'(x')) - \varphi^\rho_a(x')\partial'_\rho\psi'_i(x')]\,D'_i S_j(g^{-1},x',\psi'(x')) = s_{a'j}(x,\psi(x)) - \varphi^\rho_a,(x)\partial_\rho\psi_j(x). \tag{2.10}$$

Finally, from (2.5) and (2.10), we get

$$\begin{aligned} &[s_{ai}(x',\psi'(x')) - \varphi^\rho_a(x')\partial'_\rho\psi'_i(x')]\\ &\quad\times(D'_i - \partial'_\mu D'_{i\mu})\mathfrak{L}(x',\psi'(x'),\partial'\psi'(x'))\\ &= J(g,x)^{-1}[s_{a'i}(x,\psi(x)) - \varphi^\rho_a,(x)\partial_\rho\psi_i(x)]\\ &\quad\times(D_i - \partial_\mu D_{i\mu})\,\mathfrak{L}\,(x,\psi(x),\partial\psi(x)), \end{aligned} \tag{2.11}$$

so that by (2.1)

$$\begin{aligned} &J(g,x)\partial'_\mu I^\mu_a(\mathfrak{L};x',\psi'(x'),\partial'\psi'(x'))\\ &\quad + (s_{a',i}(x,\psi(x)) - \varphi^\rho_a,(x)\partial_\rho\psi_i(x))(D_i - \partial_\mu D_{i\mu})\\ &\quad\times\mathfrak{L}(x,\psi(x),\partial\psi(x)) = 0. \end{aligned}$$

If we observe that equation (1.8) holds for any $a \in \mathfrak{g}$ and in particular for a', we have

$$\partial'_\mu I^\mu_a(\mathfrak{L};x',\psi'(x'),\partial'\psi'(x')) = J(g,x)^{-1}\partial_\mu I^\mu_a,(\mathfrak{L};x,\psi(x),\partial\psi(x)). \tag{2.12}$$

This clearly proves the covariance of the r equations (1.9). Let us notice that this proof does not need any knowledge about transformation laws of the current

under the group G, and it is more general than the one previously given[1]. Of course, the same proof is valid, if we have $I_a(\mathfrak{L}+L; x, \psi(x)\partial\psi(x))$ instead of $I_a(\mathfrak{L}; x, \psi(x), \partial\psi(x))$.

3. TRANSFORMATION OF THE CURRENT

A. First case $R(g, x, \psi(x)) = 0$

Let us first investigate how the various terms of $I_a = j_a$ cotransform under $g \in G$. The transformation of the first term in (1.5) is written, according to (1.4) and (2.7)

$$\mathfrak{L}(x', \psi'(x'), \partial'\psi'(x'))\varphi_a(x') = A(g,x)\mathfrak{L}(x, \psi(x), \partial\psi(x))\varphi_{a'}(x). \tag{3.1}$$

For the second term in (1.5), according to (2.4) and (2.10), we have

$$[s_{ai}(x', \psi'(x')) - \varphi^\rho_a(x')\partial'_\rho\psi_i(x')]D'_{i\mu}\mathfrak{L}(x', \psi'(x'), \partial'\psi'(x'))$$
$$= A^\mu{}_\nu(g,x)[s_{a'i}(x, \psi(x)) - \varphi^\rho_{a'}(x)\partial_\rho\psi_i(x)]$$
$$\times D_{i\nu}\mathfrak{L}(x, \psi(x), \partial\psi(x)). \tag{3.2}$$

So that with obvious matrix notations

$$j_a(\mathfrak{L}; x', \psi'(x'), \partial'\psi'(x')) = A(g,x)j_{a'}(\mathfrak{L}; x, \psi(x), \partial\psi(x)). \tag{3.3}$$

Of course, this transformation makes it easy to prove in another way the covariance of (1.9), by the lemma of sec. 2:

$$\partial'_\mu j^\mu_a(\mathfrak{L}; x', \psi'(x'), \partial'\psi'(x')) = J(g,x)^{-1}\partial_\mu j^\mu_a(\mathfrak{L}; x, \psi(x), \partial\psi(x)).$$

Let us notice that (3.3) implies $j_{a_s}(\mathfrak{L}; x', \psi'(x')) = A(g,x)\mathrm{ad}\,(g^{-1})_{ts}\, j_{at}(\mathfrak{L}; x, \psi(x), \partial\psi(x))$, i.e., if we let ${}^tA$ be the transposed matrix to A,

$$j(\mathfrak{L}; x', \psi'(x'), \partial'\psi'(x')) = (A(g,x) \otimes {}^t\mathrm{d}(g^{-1}))j(\mathfrak{L}; x, \psi(x), \partial\psi(x)).$$

B. Second case $R(g, x, \psi(x)) \neq 0$

Let us first discuss the properties of the function $R_g(x, \psi(x), \partial\psi(x))$. Two consecutive transformations performed on x and on the fields $\psi(x)$ lead to

$$x \xrightarrow{g} x' = g\cdot x \xrightarrow{g'} x'' = g'\cdot x' = (g'g)\cdot x,$$
$$\psi(x) \xrightarrow{g} \psi'(x') = S(g,x,\psi(x)) \xrightarrow{g'} \psi''(x'') = S(g'g, x, \psi(x)).$$

Besides, with Eq. (1.4) the computation of $\mathfrak{L}(x'', \psi''(x''), \partial/\partial x''\psi''(x''))$ carried out in two different ways gives

$$R_{g'}(x', \psi'(x'), \partial'\psi'(x')) = J(g,x)^{-1}(Rg'g - Rg)(x, \psi(x), \partial\psi(x)). \tag{3.4}$$

This relation gives for the Lagrangian $\mathfrak{L} + R_h, h \in G$, the following transformation law

$$J(g,x)(\mathfrak{L} + R_h)(x', \psi'(x'), \partial'\psi'(x'))$$
$$- (\mathfrak{L} + R_h)(x, \psi(x), \partial\psi(x))$$
$$= (R_{hg} - R_h)(x, \psi(x), \partial\psi(x)), \tag{3.5}$$

and for the two consecutive transformations mentioned above

$$J(g', x')(\mathfrak{L} + R_h)\left(x'', \psi''(x''), \frac{\partial}{\partial x''}\psi''(x'')\right)$$
$$- (\mathfrak{L} + R_h)(x', \psi'(x'), \partial'\psi'(x'))$$
$$= J(g,x)^{-1}(R_{hg'g} - R_{hg})(x, \psi(x), \partial\psi(x)). \tag{3.6}$$

We now consider the various terms of I_a. An easy computation gives for $j_a(\mathfrak{L}; x', \psi'(x'), \partial'\psi'(x'))$, instead of equation (3.3)

$$j_a(\mathfrak{L}; x', \psi'(x'), \partial'\psi'(x')) = A(g,x)j_{a'}(\mathfrak{L} + R_g; x, \psi(x), \partial\psi(x))$$

[we have used Eq. (2.10)] and, therefore, according to (3.5)

$$j_a(\mathfrak{L} + R_{g^{-1}}; x', \psi'(x'), \partial'\psi'(x')) = A(g,x)j_{a'}(\mathfrak{L}; x, \psi(x), \partial\psi(x)). \tag{3.7}$$

Equation (3.6) with the help of the lemma then shows that

$$J(g', x')(\mathfrak{L} + R_{g^{-1}})\left(x'', \psi''(x''), \frac{\partial}{\partial x''}\psi''(x'')\right)$$
$$- (\mathfrak{L} + R_{g^{-1}})(x', \psi'(x'), \partial'\psi'(x'))$$
$$= \partial'_\mu(A^\mu{}_\nu(g,x)R^\nu(g^{-1}g'g, x, \psi(x)))$$

and this gives

$$r_a(\mathfrak{L} + R_{g^{-1}}; x', \psi'(x')) = A(g,x)r_{a'}(\mathfrak{L}; x, \psi(x)). \tag{3.8}$$

Using (3.7) and (3.8), we finally get

$$I_a(\mathfrak{L} + R_{g^{-1}}; x', \psi'(x'), \partial'\psi'(x')) = A(g,x)I_{a'}(\mathfrak{L}; x, \psi(x), \partial\psi(x)). \tag{3.9}$$

This transformation is easily seen to be consistent with the group structure of G and, as in subsection A, may be written

$$I(\mathfrak{L} + R_{g^{-1}}; x', \psi'(x'), \partial'\psi'(x')) = (A(g,x) \otimes {}^t\mathrm{d}(g^{-1}))I(\mathfrak{L}; x, \psi(x), \partial\psi(x)).$$

Remark: The explicit computation of $I_a(\mathfrak{L}; x', \psi'(x'), \partial'\psi'(x'))$ gives

$$I_a(\mathfrak{L}; x', \psi'(x'), \partial'\psi'(x')) = A(g,x)[I_{a'}(\mathfrak{L}; x, \psi(x), \partial\psi(x))$$
$$+ M(a,g,x,\psi(x),\partial\psi(x)) + C(a,g,x)],$$

where

$$M^\mu(a,g,x,\psi(x),\partial\psi(x)) = \partial_\rho(R^\rho(g,x,\psi(x))\varphi^\mu_{a'}(x) - R^\mu(g,x,\psi(x))\varphi^\rho_{a'}(x))$$

and $C^\mu(a,g,x)$ which does not depend on ψ is divergenceless.[7]

4. CONCLUSION

The first part of this paper can be summed up in the following theorem:

Let us agree to call the equations $\partial_\mu j^\mu_a(x) = 0$ and $\partial_\mu j^\mu_a(x) = s_a(x)$, respectively, conservation equations

(more correctly, continuity equations), partial conservation equations when $s_a(x) = \partial_\mu r^\mu_a(x)$, and nonconservation equations in other cases. Let G be an r-parameter Lie group and $\mathcal{L}(x, \psi(x), \partial\psi(x))$ be a Lagrangian density on fields $\psi(x)$. Then we have

Theorem: If G is an invariance group for the action integral, i.e. if $\mathcal{L}$ transforms like a scalar density: $\mathcal{L}(x, \psi(x), \partial\psi(x))dx = \mathcal{L}(x', \psi'(x'), \partial'\psi'(x'))dx'$ if $\mathcal{L}$ is form invariant (up to a divergence), there exist r conservation equations (r partial conservation equations) covariant for G transformations.

In the second part we obtained transformation laws of currents when $\mathcal{L}$ is form-invariant up to a divergence. As stated in the introduction, it does not seem that these last results were noticed before and the particular case considered in the Appendix is worth studying.

Now, it is well-known (see for instance Ref. 8) that one can obtain (partial) conservation equations under some transformations not necessarily belonging to a group, without having form invariance (up to a divergence) of $\mathcal{L}(x, \psi, \partial\psi)$. This result together with the previous theorem suggests the following conjecture:

Conjecture: If a (partial) conservation equation is covariant under a group G, then the Lagrangian density $\mathcal{L}(x, \psi(x)\partial\psi(x))$ is form invariant (up to a divergence) under G.

Notice that such a conjecture is not true for the Euler equations. For instance, let E be the set of solutions of the Euler–Lagrange equations (assumed to be linear and G be the infinite group defined by endowing E with a module structure. Then of course, the Euler–Lagrange equations are covariant for the transformations $\psi(x)$ $\psi'(x) = \psi(x) + \psi_i(x)$, $\psi_i(x) \in E$, but $\mathcal{L}$ is not an invariant.

More generally, if no assumption is made about the transformation of the Lagrangian density $\mathcal{L}$ under G, we obtain a nonconservation equation $\partial_\mu j^\mu_a(x) = s_a(x)$ so that the previous conjecture can be extended in the following way:

Conjecture: If the action of a Lie group G on the variables of a Lagrangian leads to nonconservation equations, these equations are not covariant under G.

Let us, for instance, consider Touschek's group G_T, acting on a Dirac spinor:

$$\psi(x) \to \psi'(x) = e^{\lambda\gamma^5}\psi(x), \qquad \tilde{\psi}'(x) = \tilde{\psi}(x)e^{\lambda\gamma^5}$$

$$\tilde{\psi}(x) = \psi^t(x)\gamma^0, \qquad \lambda \in \mathbb{R}.$$

The Lagrangian density for Dirac field is

$$\mathcal{L}(\psi(x), \partial\psi(x)) = i\tilde{\psi}(x)\gamma^\mu\partial_\mu\psi(x) - m\tilde{\psi}(x)\psi(x),$$

so that an infinitesimal transformation of G_T leads to a

$$\mathcal{L}(\psi'(x), \partial\psi'(x)) = \mathcal{L}(\psi(x), \partial\psi(x)) - 2\lambda m\tilde{\psi}(x)\gamma^5\psi(x),$$

which yields the nonconservation equation

$$\partial_\mu(\tilde{\psi}(x)\gamma^\mu\gamma^5\psi(x)) = 2im\tilde{\psi}(x)\gamma^5\psi(x),$$

only covariant under G_T for $m = 0$.

Some interesting attempts were made[9] to justify, in weak interactions, the nonconservation equations introduced in a phenomenological way:

$$\partial_\mu P^\mu(x) = k\varphi(x), \qquad k = Cte$$

or, in the presence of an electromagnetic field[10] $F^{\mu\nu}$

$$\partial_\mu P^\mu(x) = k\varphi(x) + (e^2/4\pi)\epsilon_{\mu\nu\alpha\beta}F^{\mu\nu}(x)F^{\alpha\beta}(x),$$

where $\varphi(x)$ is the pion field and $P^\mu(x)$ the axial vector $P^\mu(x) = \tilde{\psi}(x)\gamma^\mu\gamma^5\psi(x)$, $\psi(x)$ belonging to a multiplet of $SU(2)$ or to a supermultiplet of $SU(3)$. This was obtained with an appropriate Lagrangian density together with the generalized Touschek group:

$$\psi(x) \to \psi'(x) = e^{i\lambda j\tau^j\gamma^5}\psi(x), \quad \lambda j \in \mathbb{R},$$

where τ_j denotes a basis in the Lie algebra of $SU(2)$ or $SU(3)$, provided either with a special transformation on $\varphi(x)$, or with the introduction of a new field (Schwinger's σ model[11]) of course, these nonconservation equations are not covariant.

Thus, as a general result, since noncovariant equations have a limited interest in physics, if both previous conjectures are true, it might be difficult to justify the actual phenomenology used in weak interactions on theoretical grounds.

ACKNOWLEDGMENT

The authors are very grateful to Professor M. Flato for many suggestions and helpful advice.

APPENDIX: THE CASE OF THE POINCARÉ GROUP

The Lagrangian $\mathcal{L}$ is assumed to be Poincaré invariant, i.e., $x \in M$ (Minkowski space) transforms into $x' = \Lambda x + b = g\cdot x$ and $\psi(x)$ cotransforms according to a finite-dimensional representation S of $SL(2, C)$: $\psi(x) \to \psi'(x') = S(\tilde{\Lambda})\psi(x)$ [where we write $g = (b, \Lambda)$, $\Lambda \in SO_0(3,1)$, $b \in M$ and $\pm\Lambda \to \Lambda$ the homomorphism $(1/2, 1/2) : SL(2, \mathbb{C}) \to SO_0(3,1)$], then

$$\mathcal{L}(\psi'(x'), \partial'\psi'(x')) = \mathcal{L}(\psi(x), \partial\psi(x)).$$

We denote by $\mathcal{P}$ the Lie algebra of P. It is the semidirect product of the Lorentz Lie algebra $\mathcal{so}(3,1)$ by a four-dimensional commutative ideal $\mathcal{P}_0$; if $a \in \mathcal{P}$ we can write $a = \omega + \alpha$ with $\alpha \in \mathcal{P}_0$, $\omega \in \mathcal{so}(3.1)$) We identify ω with a matrix of order 4, α with a 4-vector and $(p_\mu)_{0\le\mu\le3}, m_{\mu\nu} = -m_{\nu\mu}$ will be the usual basis of $\mathcal{P}$ ($m_{ij}, i, j > 0$ corresponding to rotations and to boosts along x^i).

If $a = \alpha = \alpha^e p_\rho \in \mathcal{P}_0$ according to (1.5) we get

$$j^\mu_\alpha(x) = -\alpha_\rho T^{\rho\mu}(x),$$

where

$$T^{\rho\mu}(x) = -j^\mu_{p\rho}(x) = \partial^\rho\psi_i(x)D_{i\mu}\mathcal{L}(\psi(x), \partial\psi(x)) - g^{\mu\rho}\mathcal{L}(\psi(x), \partial\psi(x))$$

and, if $a = \omega = \frac{1}{2}\omega^{\rho\sigma}m_{\rho\sigma}$, we get in the same manner

$$j^\mu_\omega(x) = \tfrac{1}{2}\omega_{\rho\sigma}J^{\rho\sigma\mu}(x), \; J^{\rho\sigma\mu}(x) = j^\mu_{m\rho\sigma}(x)$$

$$= (M_{\rho\sigma}\psi(x))_i D_{i\mu}\mathcal{L}(\psi(x), \partial\psi(x)) + x^\rho T^{\sigma\mu}(x) - x^\sigma T^{\rho\mu}(x),$$

where we have denoted $M_{\rho\sigma} = [d/du S(\exp u\tilde{m}_{\rho\sigma})]_{u=0}$.

The formula (3.3) then gives, if $g = (b, \Lambda) \in P$,

$$j^{\mu}_{p}{}^{\nu}(x') = -T'^{\mu\nu}(x') = \Lambda^{\mu}{}_{\rho} j^{\rho}{}_{\mathrm{ad}(g-1)p}{}^{\nu}(x),$$

so that

$$T'^{\mu\nu}(x') = \Lambda^{\mu}{}_{\rho}\Lambda^{\mu}{}_{\sigma} T^{\rho\sigma}(x).$$

For the angular momentum density, if we write

$$j'^{\mu}_{m\rho\sigma}(x') = J'^{\rho\sigma\mu}(x'),$$

we obtain again from formula (3.3), remembering that

$$\mathrm{d}(g^{-1})\omega = \Lambda^{-1}\omega\Lambda + \Lambda^{-1}\omega b,$$

$$J'^{\mu\nu\rho}(x') = \Lambda^{\rho}{}_{\sigma}(j^{\sigma}_{\Lambda^{-1}m\mu\nu\Lambda}(x) + j^{\sigma}_{\Lambda^{-1}m\mu\nu b}(x)).$$

A straightforward computation, using the relation

$$\Lambda^{-1}m^{\mu\nu}\Lambda = \Lambda^{\mu}{}_{\rho}\Lambda^{\nu}{}_{\sigma} m^{\rho\sigma},$$

finally gives

$$J'^{\mu_1\mu_2\mu_3}(x') = \Lambda^{\mu_3}{}_{\nu_3}[\Lambda^{\mu_1}{}_{\nu_1}\Lambda^{\mu_2}{}_{\nu_2} J^{\nu_1\nu_2\nu_3}(x) - (\Lambda^{\mu_1}{}_{\nu_2} b^{\mu_2} - \Lambda^{\mu_2}{}_{\nu_2} b^{\mu_1}) T^{\nu_2\nu_3}(x)].$$

[1] P. Hillion, C.R. Acad. Sci. A **272**, 975 (1971).

[2] An example of this fact is the transformation, under the special conformal group of space–time of the massless Lagrangian for the Klein Gordon equation: $\mathcal{L} = 1/2\,\partial_{\mu}\phi(x)\partial^{\mu}\phi(x)$. When $x \to x' = (1+2cx+c^2x^2)^{-1}(x+cx^2)$ [x, cx^2 are 4-vectors, $cx = c_{\mu}x^{\mu}$, $x^2 = x_{\mu}x^{\mu}$, and $\phi'(x') = (1+2cx+c^2x^2)\phi(x)$] one has $R(c,x,\phi(x)) = (1+2cx+c^2x^2)^{-1}(c+c^2x)\phi(x)^2$.

[3] Extensions of Noether's theorem are possible in several directions; see for instance I. M. Gelfand and S. V. Fomin, *Calculus of Variations* (Prentice-Hall, New York, 1963); Ref. 4; B. F. Plybon, J. Math. Phys. **12**, 57 (1971), and Ref. 8.

[4] See, for instance, P. Funk, *Variationsrechnung und Ihre Anwendung in Physik und Technik* (Springer, Berlin, 1962); H. Rund, *The Hamilton Jacobi Theory in the Calculus of Variations* (Van Nostrand, New York, 1966), pp. 295 and 257.

[5] D. G. B. Edelen, Arch. Ration. Mech. Anal. **2**, 117 (1962).

[6] T. Muir, *A Treatise on the Theory of Determinants* (Dover, New York, 1960); or H. W. Turnbull, *The Theory of Determinants, Matrices and Invariants* (Dover, New York, 1960).

[7] The vector $C(a,g,x)$ is not always equal to zero. Take for example the Lagrangian of Footnote 2 and consider $G = \mathbf{R}^r (r > I)$ acting in the following way on x and $\phi(x)$: under $t = (t_1, \ldots, t_r) \in \mathbf{R}^r$, $x' = x$, $\phi'(x) = \phi(x) + t f_i(x)$, where $\Box f_i(x) = 0$ for $i = I, \ldots, r$.

[8] J. Rosen, Int. J. Theor. Phys. **4**, 287 (1971).

[9] M. Gell-Mann, M. Lévy, Nuovo Cimento **16**, 705 (1960).

[10] B. Zumino, "Topical Conference on Weak Interactions" (preprint), C.E.R.N., 1969, p. 366.

[11] J. Schwinger, Ann. Phys. (N.Y.) **2**, 407 (1957).

Statistical mechanics of parafermi systems

J. L. Richard* and E. C. G. Sudarshan

Center for Particle Theory, University of Texas, Austin, Texas 78712
(Received 3 February 1972)

The statistical mechanics of a collection of parafermi harmonic oscillators considered as a canonical ensemble is examined. The factorization property of the corresponding partition function selects the representations of the parafermi ring allowed to furnish the energy levels of the oscillators. While the representations corresponding to Green's ansatz are in an obvious way perfect candidates, it is shown however that, contrary to expectations, they do not exhaust all the possibilities. The general conditions on the representations are given and special families of representations are outlined.

I. INTRODUCTION

Following the classic work of Green, paraquantization has received special attention.[1] The representation of parafermi rings for an arbitrary finite number of degrees of freedom has been completely worked out by Ryan and Sudarshan.[2] They have shown that a parafermi ring for ν degrees of freedom is realized by the representations of the orthogonal group B_ν in $2\nu+1$ dimensions. In particular, the Green ansatz of order p is the pth Kronecker power of the fundamental spinor representation ${}^\nu\Delta$ of B_ν. For $p=1$, we get the standard fermi system represented by ${}^\nu\Delta$ of dimension 2^ν; this representation is, of course, irreducible.

For $p>1$, the Green ansatz does not furnish irreducible representations. In a sense the Green ansatz for a parafermi system of order p may be thought of as a fermi ring with $\nu\cdot p$ degrees of freedom with a "hidden" label which can take on distinct values. This may be seen in the original construction of Green.[1]

It is therefore not surprising that if we are to consider the statistical mechanics of a system of parafermi oscillators realized by the Green ansatz no Gibbs paradox[3] arises: The partition function for $\nu=\nu_1+\nu_2$ noninteracting oscillators at any fixed temperature β^{-1} equals the product of the partition functions for ν_1 oscillators at the same temperature β^{-1} and that for ν_2 oscillators. The entropy in this case is thus simply additive.

It is therefore relevant to ask what are the general requirements on the representations of the orthogonal group realizing the parafermi oscillator system for which statistical mechanics can be studied without the Gibbs paradox arising. We propose to investigate this question in this paper. Specifically, we consider a collection of ν parafermi harmonic oscillators, the energy levels of which being given by some representation of the parafermi ring involving ν degrees of freedom. As a canonical ensemble, this system is seen to be in equilibrium with a heat bath at a given temperature. If we consider this system as constituted by two separate subsystems, then the partition function of the whole system can be expressed as the product of the partition function of the two subsystems. This factorization property of the partition function will select the representations of the parafermi ring allowed to describe the ν parafermi harmonic oscillators.

As mentioned above, one can see easily that the representations corresponding to Green's ansatz are perfect candidates. But in this case the system under consideration is essentially of *Fermi type* since it looks like a collection of $\nu\cdot p$ Fermi harmonic oscillators, p being the order of the Green ansatz. If the only solutions were given by Green's ansatz, no system of "essentially parafermi" oscillators should exist. It is worthwhile to see whether or not Green's ansatz furnishes the only representations compatible with the factorization property of the partition function.

In the next section, we show that indeed Green's ansatz leads trivially to possible solutions. Then, in Sec. III we give a general treatment, and it is seen that in fact there exists representations distinct from Green's ansatz for which the factorization property of the partition function holds. As an illustration, some examples are investigated. In the last section, we summarize our results.

II. SYSTEM OF PARAFERMI HARMONIC OSCILLATORS

A collection of ν parafermi harmonic oscillators is defined through the hamiltonian

$$H=\frac{1}{2}\sum_{k=1}^{\nu}\omega_k(a_k^+a_k-a_ka_k^+), \tag{II.1}$$

where a_k^+ and a_k are creation and annihilation operators satisfying the parafermi commutation relations

$$[[a_k,a_l^+],a_j]=2\delta_{lj}a_k, \tag{II.2}$$

$$[[a_k,a_l],a_j]=0 \tag{II.3}$$

for all $j,k,l=1,2,\ldots,\nu$.

From the representation theory[2] of parafermi rings, we know that the representations of the operators are obtained by means of the isomorphism between the Lie algebra of the orthogonal group B_ν in $2\nu+1$ dimensions and the Lie algebra generated by the operators

$$H_k=\tfrac{1}{2}[a_k^+,a_k], \tag{II.4a}$$

$$Q_{+k}=a_k^+, \tag{II.4b}$$

$$Q_{-k}=a_k, \tag{II.4c}$$

together with

$$N_{\pm k}\equiv[Q_{\pm k},Q_{\pm k}],\quad N_{\pm k,\pm k'}\equiv[Q_{\pm k},Q_{\mp k'}],\quad k\neq k'.$$

We use here the standard basis of the Lie algebra of B_ν for which the H_k are diagonal in the adjoint representation. In other words, each unitary representation of B_ν leads through (II.4) to a representation of the parafermi ring (generated by the a_k and their adjoints a_k^+).

Using (II.4a), the Hamiltonian (II.1) reads

Copyright © 1973 by the American Institute of Physics

$$H=\sum_{k=1}^{\nu}\omega_k H_k, \tag{II. 5}$$

the operators H_k being a complete set of commuting operators which generate a maximal Abelian subalgebra of B_ν. Thus from the definition of the partition function of a canonical ensemble, we see that the partition function of the ν parafermi harmonic oscillators $Z(\beta,\nu)$ takes the form

$$Z(\beta,\nu)=\mathrm{Tr}\left(\prod_{k=1}^{\nu} e^{-\beta\omega_k H_k}\right) \quad \text{with } \beta=\frac{1}{KT} \tag{II. 6}$$

in some representation of B_ν.

Let us denote by (ν) the system of ν parafermi harmonic oscillators and consider two systems (ν_1) and (ν_2) both in equilibrium with a heat bath at a given temperature T. Then, from the additivity of the entropy,[3,4] one deduces that the partition function of the system $(\nu_1+\nu_2)$ is the product of the partition functions $Z(\beta,\nu_1)$ and $Z(\beta,\nu_2)$ of (ν_1) and (ν_2), repectively, that is

$$Z(\beta,\nu_1+\nu_2)=Z(\beta,\nu_1)\times Z(\beta,\nu_2). \tag{II. 7}$$

This factorization property will select the representations of the parafermi ring capable of describing the system of ν parafermi oscillators.

As a particular case, let us consider the ensemble of ν *Fermi* harmonic oscillators. The corresponding representation of B_ν is the fundamental spinor representation ${}^{\nu}\Delta$ in which the eigenvectors of the Hamiltonian (I. 1) reads $|\pm\frac{1}{2},\pm\frac{1}{2},\ldots,\pm\frac{1}{2}>$, where the plus and minus signs are *arbitrarily* distributed, the labels $\pm\frac{1}{2}$ referring to the eigenvalues of the H_k defined in (II. 4a). Hence using the definition (II. 6) of the partition function, its expression reads

$$Z(\beta,\nu)=2^{\nu}\prod_{k=1}^{\nu}\cosh\beta\omega_k,$$

which shows that, in fact, the partition function factorizes entirely, each parafermi harmonic oscillator being in the fundamental spinor representation ${}^{1}\Delta\equiv{}^{1}D_{1/2}$ of $O(3)$. Since the Green ansatz of order p is given by the Kronecker product $[{}^{\nu}\Delta]^p$, it immediately follows that Green's ansatz leads to the factorization of the corresponding partition function.

So, the representations corresponding to Green's ansatz are in an obvious way solutions for our problem. We shall show in the next section that in fact the complete factorization of the partition function into the product of the "particle" partition functions is a necessary condition.

III. GENERAL TREATMENT

Let us investigate the consequences of the factorization (II. 7) of the partition function. A representation ${}^{\nu_1+\nu_2}D$ of the system $(\nu_1+\nu_2)$ being given, we denote by $|L>$ the eigenstates of the corresponding Hamiltonian (II. 1) with $L=(l_1,l_2,\ldots,l_{\nu_1+\nu_2})$, l_k being an eigenvalue of H_k. The energy of the state $|L>$ then reads

$$E_L=\sum_{k=1}^{\nu_1+\nu_2}\omega_k l_k.$$

Likewise, let $|L_1>$ and $|L_2>$ be the eigenstates of the Hamiltonians of the systems (ν_1) and (ν_2), respectively. Since the Hamiltonian of the whole system is simply the sum of the Hamiltonians for each subsystem, to the state $|L>$ there corresponds two states $|L_1>$ and $|L_2>$ such that $L_1=(l_1,l_2,\ldots,l_{\nu_1})$ and $L_2=(l_{\nu_1+1},\ldots,l_{\nu_1+\nu_2})$. But the factorization property (II. 7) written in the form

$$Z(\beta,\nu_1+\nu_2)=\sum_{L_1}\sum_{L_2}\exp[-\beta(E_{L_1}+E_{L_2})], \tag{III. 1}$$

where the summations on L_1 and L_2 are *independent*, shows that, conversely, to the states $|L_1>$ and $|L_2>$ there corresponds a state $|L>$ such that $L=(L_1,L_2)$. In other words, the weight space of the representation ${}^{\nu_1+\nu_2}D$ must be the direct product of the weight spaces of the representations ${}^{\nu_1}D$ and ${}^{\nu_2}D$ of the system (ν_1) and (ν_2). This property can be recasted in terms of characters. Indeed, let ${}^{\nu_1+\nu_2}\chi$ be the character of the representation ${}^{\nu_1+\nu_2}D$. It is defined[5] by

$$\begin{aligned}&{}^{\nu_1+\nu_2}\chi(\phi_1,\phi_2,\ldots,\phi_{\nu_1+\nu_2})\\&=\sum_{(l_1,l_2,\ldots,l_{\nu_1+\nu_2})}\exp[i(l_1\phi_1+l_2\phi_2+\cdots+l_{\nu_1+\nu_2}\phi_{\nu_1+\nu_2})]\end{aligned} \tag{III. 2}$$

with $\phi_i\in[0,2\pi]$. It then follows that ${}^{\nu_1+\nu_2}\chi$ can be written as

$$\begin{aligned}&{}^{\nu_1+\nu_2}\chi(\phi_1,\ldots,\phi_{\nu_1+\nu_2})\\&=\sum_{(l_1,\ldots,l_{\nu_1})}\exp[i(l_1\phi_1+\cdots+l_{\nu_1}\phi_{\nu_1})\\&\quad\times\sum_{(l_{\nu_1+1},\ldots,l_{\nu_1+\nu_2})}\exp[i(l_{\nu_1+1}\phi_{\nu_1+1}+\cdots+l_{\nu_1+\nu_2}\phi_{\nu_1+\nu_2})].\end{aligned}$$

The right-hand side is seen to be the product of the characters ${}^{\nu_1}\chi$ and ${}^{\nu_2}\chi$ of the representations ${}^{\nu_1}D$ and ${}^{\nu_2}D$. We thus have

$$\begin{aligned}{}^{\nu_1+\nu_2}\chi(\phi_1,\ldots,\phi_{\nu_1+\nu_2})&={}^{\nu_1}\chi(\phi_1,\ldots,\phi_{\nu_1})\\&\quad\times{}^{\nu_2}\chi(\phi_{\nu_1+1},\ldots,\phi_{\nu_1+\nu_2}).\end{aligned} \tag{III. 3}$$

Therefore, the factorization property (II. 7) is equivalent to the factorization (III. 3) of the character of the representation ${}^{\nu_1+\nu_2}D$ describing the whole system. From (III. 3) we deduce easily that the character ${}^{\nu_1+\nu_2}\chi$ factorizes entirely into the product $\prod_{k=1}^{\nu=\nu_1+\nu_2}{}^{1}\chi(\phi_k)$ in which ${}^{1}\chi$ is the character of the representation ${}^{1}D$ describing each parafermi harmonic oscillator. But then the partition function $Z(\beta,\nu)$ factorizes entirely, namely,

$$Z(\beta,\nu)=\prod_{k=1}^{\nu}Z_k(\beta); \tag{III. 4}$$

$Z_k(\beta)$ being the partition function of the kth oscillator. This result is not surprising since it shows that the factorization property (II. 7) is actually independent of the choice of the two subsystems (ν_1) and (ν_2) as it was expected.

Finally, the system of ν parafermi harmonic oscillators will be conveniently described by the representation ${}^{\nu}D$ such that their corresponding character ${}^{\nu}\chi$ can be written as

$${}^{\nu}\chi(\phi_1,\ldots,\phi_\nu)=\prod_{k=1}^{\nu}{}^{1}\chi(\phi_k); \tag{III. 5}$$

${}^{1}\chi$ being the character of the representation ${}^{1}D$ of $O(3)$ which gives the energy levels of each oscillator. In fact,

Eq. (III. 5) gives a way to construct the representations ${}^{\nu}D$ if we know the representation ${}^{1}D$ for each oscillator. Therefore, a representation ${}^{1}D$ and its character ${}^{1}\chi$ being given, we have to see if the product $\Pi_{k=1}^{\nu}\,{}^{1}\chi(\phi_k)$ is either a character for some representation ${}^{\nu}D$ of B_ν or not.

First of all, let us note that if we consider a representation ${}^{1}D$ which contains *both* integer and half integer spins, then the product $\Pi_{k=1}^{\nu}\,{}^{1}\chi(\phi_k)$ will *not* be a character for B_ν. Indeed, we know that a character ${}^{\nu}\chi$ of B_ν reads

$$ {}^{\nu}\chi(\phi_1,\ldots,\phi_\nu)=\sum \exp[i(j_1\phi_1+\cdots+j_\nu\phi_\nu)], \tag{III. 6}$$

where the components $j_1, j_2, \ldots, j_\nu$ of the weights are either all integers or all half integers. But Eq. (III. 5) tells us that $j_1, j_2, \ldots, j_\nu$ are, in fact, the weights contained in ${}^{1}\chi$ so that ${}^{1}\chi$ must contain either integer or half integer spins.

So we define ${}^{1}D$ as

$$ {}^{1}D\equiv\sum_{j=\epsilon}^{l} a_j\,{}^{1}D_j, \tag{III. 7}$$

where ${}^{1}D$ denotes the irreducible representation of spin j of $O(3)$, a_j being its multiplicity and ϵ is 0 or $\frac{1}{2}$ according to whether we are considering integer or half-integer spins, respectively, l being the highest spin in ${}^{1}D$ assuming $a_l\neq 0$. The character ${}^{1}\chi$ of ${}^{1}D$ reads

$$ {}^{1}\chi(\phi)=\sum_{k=\epsilon}^{l} c_k\cos k\phi \tag{III. 8}$$

with

$$ c_k=2\sum_{j=k}^{l} a_j \quad \text{if } k\geq 1 \text{ or } k=\epsilon=\tfrac{1}{2},$$

$$ c_0=\sum_{j=0}^{l} a_j \quad \text{when } \epsilon=0.$$

In order to identify a character of B_ν with the product (III. 5), let us consider its general expression recalling that a finite-dimensional unitary representation of B_ν is fully reducible, that is, of the form

$$ {}^{\nu}D=\sum_{\{k\}} a_{\{k\}}\,{}^{\nu}D_{\{k\}}, \tag{III. 9}$$

where ${}^{\nu}D_{\{k\}}$ denotes the irreducible representation of B_ν with highest weight

$$ \{k\}=\{k_1,\ldots,k_n\}k_1\geq k_2\cdots\geq k_\nu\geq 0; \tag{III. 10}$$

$k_1, k_2, \ldots, k_\nu$ being either all integers or all half integers. From (III. 9), we see that the character ${}^{\nu}\chi$ of ${}^{\nu}D$ reads

$$ {}^{\nu}\chi(\phi_1,\ldots,\phi_\nu)=\sum_{\{k\}} a_{\{k\}}\frac{A_{\{k\}}(\phi_1,\ldots,\phi_\nu)}{A_0(\phi_1,\ldots,\phi_\nu)}\,, \tag{III. 11}$$

where

$$ {}^{\nu}\chi_{\{k\}}(\phi_1,\ldots,\phi_\nu)=\frac{A_{\{k\}}(\phi_1,\ldots,\phi_\nu)}{A_0(\phi_1,\ldots,\phi_\nu)} \tag{III. 12}$$

is the simple character of B_ν corresponding to the representation ${}^{\nu}D_{\{k\}}$, $A_{\{k\}}$ and A_0 being alternating elementary sums[6] (A. E. S.) corresponding to the highest weight $\{k\}$ and to the weight $\{0,0,\ldots,0\}$ of the identify representation, respectively. Let us recall that an A. E. S. of B_ν is given by

$$ A_{\{k\}}(\phi_1,\ldots,\phi_\nu)=\sum_{\sigma\in S_\nu} P(\sigma)\sin l_{\sigma 1}\phi_1\sin l_{\sigma 2}\phi_2\cdots\sin l_{\sigma\nu}\phi_\nu,$$

where S_ν is the permutation group of ν objects, $P(\sigma)$ the partity of the permutation σ, and $l_i=k_i+\nu+\frac{1}{2}-i$.

In order to satisfy Eq. (III. 5), we have to prove that the quantity

$$ A(\phi_1,\ldots,\phi_\nu)\equiv A_0(\phi_1,\ldots,\phi_\nu)\prod_{k=1}^{\nu}{}^{1}\chi(\phi_k), \tag{III. 13}$$

where ${}^{1}\chi$ is given in (III. 8), is a sum of A. E. S. *with positive coefficients*. A simple but lengthy computation shows that the quantity (III. 13) is indeed a sum of A. E. S. Let us indicate the proof: From the expression of $A_0(\phi_1,\ldots,\phi_\nu)$ and Eq. (III. 8), we get

$$\begin{aligned} A(\phi_1,\ldots,\phi_\nu)=\frac{1}{2^\nu}\sum_{j_1,j_2,\ldots,j_\nu=\epsilon}^{l} c_{j_1}c_{j_2}\cdots c_{j_\nu}\sum_{\sigma\in S_\nu} P(\sigma)\\ \times[\sin(r_{\sigma_1}+j_1)\phi_1+\sin(r_{\sigma_1}-j_1)\phi_1]\cdots\\ \times[\sin(r_{\sigma_\nu}+j_\nu)\phi_\nu+\sin(r_{\sigma_\nu}-j_\nu)\phi_\nu]\end{aligned}$$

with $r_i=\nu+\frac{1}{2}-i$.

After some algebraic manipulations, we get the following results:

$$ A(\phi_1,\ldots,\phi_\nu)=\sum_{\{k\}=\{\epsilon,\epsilon,\ldots,\epsilon\}}^{\{l,l,\ldots,l\}} a_{\{k\}}A_{\{k\}}(\phi_1,\ldots,\phi_\nu), \tag{III. 14}$$

where the coefficients $a_{\{k\}}$ expressed in terms of the a_j occurring in the representation (III. 7) of each oscillator are given by

$$ a_{\{k\}}=\sum_{\sigma\in S_\nu} P(\sigma)\prod_{i=1}^{\nu}\sum_{j_i=|k_i+\sigma i-i|}^{2\nu+k_i-i-\sigma i} a_{j_i}. \tag{III. 15}$$

Therefore from Eqs. (III. 14) and (III. 12), the representation of B_ν we are looking for is seen to be

$$ {}^{\nu}D=\sum_{\{k\}=\{\epsilon,\epsilon,\ldots,\epsilon\}}^{\{l,l,\ldots,l\}} a_{\{k\}}\,{}^{\nu}D_{\{k\}} \tag{III. 16a}$$

provided of course that the coefficients $a_{\{k\}}$ given in (III. 15) are all nonnegative. This is the *only* condition the a_j introduced in (III. 7) must fulfill.

In the following subsections, we shall investigate successively the case for which the representations of the parafermi ring involving ν degrees of freedom are irreducible, the case for which the representations are given by a "sum" of Green's ansatz, and finally the most general representations.

A. Irreducible representations

In this case, following our notation, the representations of B_ν are denoted by ${}^{\nu}D_{\{k\}}$ where $\{k\}$ is defined in (III. 10). From Eq. (III. 16a) we see that if this representation satisfies the factorization condition (III. 5), then it must be characterized by the weight $\{k\}=\{l,l,\ldots,l\}$. Indeed from (III. 15), we deduce that

$$ a_{\{l,l,\ldots,l\}}=a_l^{\nu}, \tag{III. 16b}$$

and since we assumed that $a_l\neq 0$, this coefficient does not vanish so that in (III. 16a) the representation ${}^{\nu}D_{\{l,l,\ldots,l\}}$ appears in any case. But since we assume ${}^{\nu}D$ irreducible, a_l must be equal to 1 and all the coeffi-

cients $a_{\{k\}}$ with $\{k\} \neq \{l, l, \ldots, l\}$ must vanish. Let us compute the coefficients $a_{\{l,l,\ldots,l,k\}}$ with $\epsilon \leq k < l$. From (III. 15), we get

$$a_{\{l,l,\ldots,l,k\}} = a_l^{\nu-1} a_k = a_k. \tag{III. 17}$$

These coefficients vanish if $a_k = 0$ for $\epsilon \leq k < l$, which means that the representation ${}^1D$(III. 7) is irreducible and contains the spin l. So, at this stage the representations ${}^\nu D$ and ${}^1D$ are seen to be

$${}^\nu D \equiv {}^\nu D_{\{l,l,\ldots,l\}}, \quad {}^1D \equiv {}^1D_l.$$

If we now look directly at the factorization condition (III. 5), we see that the restriction of ${}^\nu D$ to $O(3)$ must be a multiple of the representation ${}^1D$. Using the branching theorems which give the reduction of an irreducible representation of B_ν in terms of irreducible representations of $B_{\nu-1}$, one can be easily convinced that the restriction of ${}^\nu D$ to $O(3)$ will be a multiple of the representation ${}^1D$ *only* in the case $l = \frac{1}{2}$ for which we have

$${}^\nu D \equiv {}^\nu \Delta \to 2^\nu {}^1D_{1/2};$$

in other words, only for the fundamental spinor representation.

Thus the result is the following: If we are considering irreducible representations for a parafermi ring, then the only representation satisfying the factorization condition corresponds in fact to a Fermi ring.

B. Green's ansatz

In this section, we investigate the case for which the representation ${}^\nu D$ is given by a "sum" of Green ansatz, that is to say,

$${}^\nu D = \sum_{i=1}^{n} \alpha_i [{}^\nu \Delta]^{p_i}, \tag{III. 18}$$

where α_i is the multiplicity of the Green ansatz of order p_i, the order being either all even or all odd and such that

$$p_1 < p_2 < \cdots < p_n. \tag{III. 19}$$

The character of the representation (III. 18) reads

$${}^\nu \chi(\phi_1, \ldots, \phi_\nu) = \sum_{i=1}^{n} \alpha_i [{}^\nu \chi_{\{1/2, 1/2, \ldots, 1/2\}}(\phi_1, \ldots, \phi_\nu)]^{p_i}. \tag{III. 20}$$

Using the factorization property of the fundamental spinor representation we get

$${}^\nu \chi(\phi_1, \ldots, \phi_\nu) = \sum_{i=1}^{n} \alpha_i \prod_{j=1}^{\nu} [{}^1\chi_{1/2}(\phi_j)]^{p_i} \tag{III. 21}$$

from which we deduce

$${}^\nu \chi(\phi, 0, \ldots, 0) = \sum_{i=1}^{n} \alpha_i 2^{p_i - 1} [{}^1\chi_{1/2}(\phi)]^{p_i}. \tag{III. 22}$$

From the factorization condition (III. 5), it is easily derived that the character ${}^1\chi$ for each oscillator must be of the form

$${}^1\chi(\phi) = \sum_{i=1}^{n} \beta_i [{}^1\chi_{1/2}(\phi)]^{p_i}, \quad \beta_n \neq 0, \tag{III. 23}$$

so that the factorization condition reads

$$\sum_{i=1}^{n} \alpha_i \prod_{j=1}^{\nu} [{}^1\chi_{1/2}(\phi_j)]^{p_i} = \prod_{j=1}^{\nu} \sum_{i=1}^{n} \beta_i [{}^1\chi_{1/2}(\phi_j)]^{p_i}. \tag{III. 24}$$

Via (III. 19), we see that the simple character ${}^1\chi_{p_n/2}(\phi_1)$ will appear on the right-hand side of (III. 24) with the coefficient

$$\alpha_n \prod_{j=2}^{n} [{}^1\chi_{1/2}(\phi_j)]^{p_n} \tag{III. 25}$$

and on the left-hand side with the coefficient

$$\beta_\nu \prod_{j=2}^{\nu} \sum_{i=1}^{n} \beta_i [{}^1\chi_{1/2}(\phi_j)]^{p_i}. \tag{III. 26}$$

Identifying these coefficients, we get

$$\alpha_n = \beta_n^\nu, \tag{III. 27}$$
$$\beta_i = 0 \quad \text{for } i < n;$$

by using again the fact that in $[{}^1\chi_{1/2}(\phi_j)]^{p_n}$ the simple character ${}^1\chi_{p_n/2}(\phi_j)$ appears only once.

Hence, it follows from (III. 23) and (III. 27) that each oscillator is described by a representation which contains only the Green ansatz of order p_n and so is the representation for the whole system.

C. General case

Let us go back to the general treatment we gave at the beginning of this section. We shall show that there exists representations other than Green's ansatz which are compatible with the factorization condition.

Let us define the polynomials $P_{\{k\}}(x_\epsilon, \ldots, x_{l-1})$ of l variables by

$$P_{\{k\}}(x_\epsilon, \ldots, x_{l-1}) \equiv \sum_{\sigma \in S_\nu} P(\sigma) \prod_{i=1}^{\nu} \sum_{j_i = |k_i + \sigma i - i|}^{2\nu + k_i - i - \sigma i} x_{j_i} \tag{III. 28}$$

with $x_l = 1$.

Then the coefficients (III. 15) occurring in the representation ${}^\nu D$ (III. 16a) are given by

$$a_{\{k\}} = a_l^\nu P_{\{k\}}(x_\epsilon, \ldots, x_{l-1}) \tag{III. 29}$$

with

$$x_i = a_i / a_l, \quad i = \epsilon, \ldots, l-1.$$

Our problem is to find the sets of nonnegative rational numbers $(x_\epsilon, \ldots, x_{l-1})$ for which all the polynomials $P_{\{k\}}$ are nonnegative. Then if we choose a_l to be the least common multiple of the denominators of the rational numbers $(x_\epsilon, \ldots, x_{l-1})$, it follows that we obtain a set of nonnegative integers $(a_\epsilon, \ldots, a_l)$ for which $a_{\{k\}}$ are all nonnegative integers.

First of all, let us note that there exists a set of integers $y = (y_\epsilon, \ldots, y_{l-1})$ for which all the polynomials are strictly positive, namely, the set corresponding to Green ansatz[7] of order $2l$. Thus, for each polynomial there exists a neighborhood $\mathcal{O}_{\{k\}}(y)$ of y such that

$$P_{\{k\}}(x) > 0 \quad \text{for all } x \in \mathcal{O}_{\{k\}}(y).$$

Then defining

$$\mathcal{O}(y) = \bigcap_{\{k\} = \{\epsilon, \epsilon, \ldots, \epsilon\}}^{\{l, l, \ldots, l\}} \mathcal{O}_{\{k\}}(y),$$

which is obviously not empty, it is clear that all the polynomials $P_{\{k\}}$ are strictly positive in $\mathcal{O}(y)$. Hence, to every set of rational numbers in $\mathcal{O}(y)$ there corresponds a representation ${}^{\nu}D$ for which the partition function factorizes.

IV. EXAMPLES

Before considering some examples, it is useful to derive a recurrence formula expressing the coefficients $a_{\{k\}}$ occurring in the representation ${}^{\nu}D$ of ν oscillators in terms of the coefficients occurring in the representation ${}^{\nu-1}D$ of $\nu-1$ oscillators. From Eq. (III.15), using the decomposition of a permutation $\sigma \in S_\nu$ in terms of a product of transpositions, we get the following recurrence formula:

$$a_{\{k_1,\ldots,k_\nu\}} = \sum_{i=1}^{\nu} (-1)^{i+1} a_{\{k_1+1,\ldots,k_{i-1}+1,k_{i+1}+1,\ldots,k_\nu\}} \times \sum_{j_i=|k_i+1-i|}^{k_i+2\nu-1-i} a_{j_i}, \tag{IV.1}$$

where the coefficients $a_{\{k_1',k_2',\ldots,k_{\nu-1}'\}}$ appearing on the right-hand side belong to the representation ${}^{\nu-1}D$ of $\nu-1$ oscillators. This relation is useful to construct step by step the representation of ν oscillators.

For example, let us consider the case for which the representation of the parafermi ring corresponding to one oscillator is given by

$${}^{1}D = a_0\,{}^{1}D_0 + a_1\,{}^{1}D_1 + a_2\,{}^{1}D_2, \quad a_2 \neq 0. \tag{IV.2}$$

According to Eq. (III.16a), the representation of ν oscillators will read

$${}^{\nu}D = \sum_{\{k\}=\{0,0,\ldots,0\}}^{\{2,2,\ldots,2\}} a_{\{k\}}\,{}^{\nu}D_{\{k\}}. \tag{IV.3}$$

For $\nu=2$ oscillators, the coefficients appearing in (IV.3) are given by (III.15) as well as (IV.1). We get

$$\begin{aligned} &a_{22} = a_2^2, \quad a_{21} = a_2 a_1, \quad a_{20} = a_2 a_0, \\ &a_{11} = a_1^2 - a_0 a_2 - a_2^2, \\ &a_{10} = a_2 a_0 - a_2 a_1 + a_1 a_0, \\ &a_{00} = a_0^2 + a_0 a_1 + a_0 a_2 - a_1^2. \end{aligned} \tag{IV.4}$$

From the decomposition

$$({}^{1}\Delta)^4 = 2\,{}^{1}D_0 + 3\,{}^{1}D_1 + {}^{1}D_2, \tag{IV.5}$$

we see that Green's ansatz of order 4 corresponds to the values

$$a_0 = 2, \quad a_1 = 3, \quad a_2 = 1, \tag{IV.6}$$

from which the decomposition of $({}^{2}\Delta)^4$ can be obtained through (IV.3) and (IV.4). According to the discussion of Sec. III.C, the positiveness of the coefficients (IV.4) reads

$$x_1^2 - x_0 - 1 \geq 0, \tag{IV.7a}$$

$$x_0 - x_1 + x_1 x_0 \geq 0, \tag{IV.7b}$$

$$x_0^2 + x_0 x_1 + x_0 - x_1^2 \geq 0 \quad \text{with } x_0 = a_0/a_2, \quad x_1 = a_1/a_2. \tag{IV.7c}$$

From this set of conditions, we see that x_1 and x_0 cannot vanish which means that in the representation (IV.2) all the spins up to 2 must appear. In fact this is quite general as it can be seen directly in (IV.1) with $\nu=2$. Moreover from (IV.7a) we must have $x_1 > 1$, that is, a_1 is always greater than a_2. From these simple considerations, we thus see that many representations do not lead to the factorization property. However from the next example, we shall see that there exists an arbitrarily large number of solutions even for a large number of oscillators.

Let us consider the case for which the representation ${}^{1}D$ contains only the spins 0 and 1, namely, we put

$${}^{1}D \equiv a_0\,{}^{1}D_0 + a_1\,{}^{1}D_0. \tag{IV.8}$$

Then the representation ${}^{\nu}D$ for ν oscillators is given by

$${}^{\nu}D = \sum_{k=0}^{\nu} a_{\{1^k 0^{\nu-k}\}}\,{}^{\nu}D_{\{1^k 0^{\nu-k}\}}, \tag{IV.9}$$

where $\{1^k 0^{\nu-k}\}$ denotes the weight in which the first k components are equal to 1 and the last $\nu-k$ components are zero. Using the recurrence formula (IV.1), we get

$$a_{\{1^k 0^{\nu-k}\}} = a_1^k a_{\{0^{\nu-k}\}}, \quad k \geq 1, \tag{IV.10a}$$

$$a_{\{0^\nu\}} = (a_0 + a_1) a_{\{0^{\nu-1}\}} - a_1^2 a_{\{0^{\nu-2}\}} \tag{IV.10b}$$

in which the coefficients $a_{\{0^{\nu-k}\}}$ refer to the representation ${}^{\nu-k}D$ for $\nu-k$ oscillators. Therefore, the conditions for the existence of the representations (IV.9) are simply

$$P_{\{0^k\}}(x) \geq 0 \quad \text{for } 2 \leq k \leq \nu \quad \text{with } x = a_0/a_1, \tag{IV.11}$$

where $P_{\{0^k\}}(x)$ is defined according to (III.28) and (III.29) by

$$P_{\{0^k\}}(a_0/a_1) = a_1^{-k} a_{\{0^k\}} \tag{IV.12}$$

and obey the recurrence formula

$$P_{\{0^\nu\}}(x) = (x+1) P_{\{0^{\nu-1}\}}(x) - P_{\{0^{\nu-2}\}}(x). \tag{IV.13}$$

For $\nu=2$ oscillators, a direct computation or the use of Eq. (III.15) lead to the following representation

$${}^{2}D = a_1^2\,{}^{2}D_{\{1,1\}} + a_1 a_0\,{}^{2}D_{\{1,0\}} + (a_0^2 + a_1 a_0 - a_1^2)\,{}^{2}D_{\{0,0\}} \tag{IV.14}$$

provided that

$$P_{\{0,0\}}(x) = x^2 + x - 1 \geq 0.$$

We shall show that the condition

$$x \geq 1 \tag{IV.15}$$

is strong enough in order to satisfy (IV.11) *for any number of oscillators*. Indeed, let us assume that for a given k, $P_{\{0^{k-1}\}} > P_{\{0^{k-2}\}} > 0$. Then from (IV.14) with the condition (IV.15), we get

$$P_{\{0^k\}} > x P_{\{0^{k-1}\}} > P_{\{0^{k-1}\}} > 0. \tag{IV.16}$$

Since this is true for $k=3$, then this is true for $k=4$ and finally for any k.

So, provided that $a_0 \geq a_1$, the representation

$${}^{1}D = a_0\,{}^{1}D_0 + a_1\,{}^{1}D_1$$

leads to a representation ${}^{\nu}D$ for any ν which fulfills the factorization property. Let us note that Green's ansatz of order 2 is given by $a_0 = a_1 = 1$.

As a direct consequence, the Kronecker product

$$\prod_{i=1}^{p} (a_0^{(i)}\,{}^{1}D_0 + a_1^{(i)}\,{}^{1}D_1) \quad \text{with } a_0^{(i)} \geq a_1^{(i)} \tag{IV.17}$$

will lead to new representations of the system of parafermi harmonic oscillators in the same way the fundamental spinor representation leads to Green's ansatz. Indeed, since the character of a Kronecker product of representations is the product of the characters of the representations, it follows that the character of the representation (IV. 17) reads

$$^1\chi(\phi) = \prod_{i=1}^{p} [a_0^{(i)} + a_1^{(i)}\,{}^1\chi_1(\phi)], \tag{IV.18}$$

so that the product $\Pi_{k=1}^{\nu}\,{}^1\chi(\phi_k)$ will be

$$\prod_{k=1}^{\nu} {}^1\chi(\phi_k) = \prod_{i=1}^{p}\prod_{k=1}^{\nu} [a_0^{(i)} + a_1^{(i)}\,{}^1\chi_1(\phi_i)]. \tag{IV.19}$$

But we have seen that the product $\Pi_{k=1}^{\nu}[a_0^{(i)} + a_1^{(i)}\,{}^1\chi_1(\phi_k)]$ is a character for a representation, say ${}^{\nu}D^{(i)}$, of the form (IV. 9) so that the product $\Pi_{k=1}^{\nu}\,{}^1\chi(\phi_k)$ is the character of the Kronecker product $\Pi_{i=1}^{p}\,{}^{\nu}D^{(i)}$. In particular, for $p=2$ we obtain solutions of the first example we were considering at the beginning of this section, solutions which are valid for any number of oscillators.

V. CONCLUSION

Let us summarize our results following the type of representations of the parafermi ring involving ν degrees of freedom we considered.

(1) Under the assumption of an irreducible representation (which implies a unique vacuum), it turns out that only system leading to the factorization of the partition function is in fact the standard Fermi system.

(2) It has been shown that a sum of Green ansatz of different orders does *not* satisfy the factorization condition so that among all the representations constructed in terms of Green ansatz, the order plays the role of a *selection rule*.

(3) We have shown that there exists infinitely many reducible representations distinct from Green's ansatz for which the factorization condition is fulfilled. Starting with particular representations, we have seen that we can generate other representations by considering Kronecker products of the "basic" representations as one does starting with the fundamental spinor representation. One can be convinced then that the number of possible representations is in fact arbitrarily large.

For these general representations the statistical weights of the various energy levels are very large in contrast with the standard Fermi representation; and to that extent they seem to have no immediate physical application.

---

*On leave from the Centre de Physique Théorique, C. N. R. S., Marseille.

[1]H. S. Green, Phys. Rev. **90,** 270 (1953); parabose quantization has been pointed out previously by E. P. Wigner, Phys. Rev. **77,** 711 (1950); D. V. Volkov, Zh. Eksp. Teor. Fiz. **36,** 1560 (1959) [Sov. Phys.-JETP **9,** 1107 (1959)]; Zh. Eksp. Teor. Fiz. **38,** 518 (1960) [Sov. Phys.-JETP **11,** 375 (1960)]; S. Kametuchi and Y. Y. Takahashi, Nucl. Phys. **36,** 177 (1962); G. F. Dell'Antonio, O. W. Greenberg, and E. C. G. Sudarshan, *Proceedings of the Istambul Summer School (1962)*, edited by F. Gürsey (Gordon and Breach, 1963), T. F. Jordan, N. Mukunda, and S. V. Pepper, J. Math. Phys. **4,** 1089 (1963); O. W. Greenberg and L. Messiah, Phys. Rev. **138,** B1155 (1965).

[2]C. Ryan and E. C. G. Sudarshan, Nucl. Phys. **47,** 207 (1963).

[3]J. W. Gibbs, *Collected Works* (Yale U. P., New Haven, Conn., 1948), Vol. II; E. C. G. Sudarshan and J. Mehra, Int. J. Theor. Phys. **3,** 245 (1970).

[4]See for instance R. Kubo, *Statistical Mechanics* (North-Holland, Amsterdam, 1963), Chap. 1, Sec. 9.

[5]All the properties of the characters of B_ν which will be used are contained for instance in H. Boerner's *Representations of Groups* (North-Holland, Amsterdam, 1963), especially Chaps. VII and VIII.

[6]H. Boerner, Ref. 5, Chap. VII, Sec. 9, p. 240.

[7]This property can be shown by using the reduction of the Kronecker product ${}^{\nu}\Delta \times {}^{\nu}D_{\{k\}}$ in terms of irreducible components. See, for instance, F. D. Murnaghan, *The Theory of Group Representations* (John Hopkins, Baltimore, 1958), Chap. 10, Sec. 4, p. 313.

The symmetric group: Characters, products and plethysms

P. H. Butler* and R. C. King

Mathematics Department, The University, Southampton, England
(Received 29 November 1971; final revised manuscript received 5 November 1972)

It is demonstrated following the work of Murnaghan and Littlewood that the symmetric group may with advantage be considered as a subgroup, albeit finite, of the general linear group. This approach involves in a natural way the use of a reduced notation for the specification of irreducible representations of Σ_n and leads to the determination of closed formulas for the dimension, characters, and Kronecker products of such representations. The relationship between the characters and S functions is discussed in detail and some results on symmetrized Kronecker squares are obtained.

1. INTRODUCTION

In 1937[1] and 1938[2] Murnaghan gave tables of the characters and Kronecker products of irreducible representations of the symmetric groups, Σ_n, which were valid for all values of n. It was not until more recently that he suggested a "reduced" notation for labelling the representations. In a series of papers[3-5] he discussed in an n-independent manner the dimensions of irreducible representations of Σ_n, the corresponding characters, their ordinary Kronecker products, and their symmetrized Kronecker products. In 1958 Littlewood used the same "reduced" notation to discuss inner plethysms[6] and went on to derive closed formulas[7] for both the ordinary and symmetrized Kronecker products by treating the symmetric group Σ_n as a subgroup of the linear group L_n.

In this paper we review these results, stress the importance of the reduced notation, discuss in more detail the link between the finite and continuous groups, and derive a number of results which could only be obtained with difficulty using the standard notation.

It is well known that outer products of S functions correspond to Kronecker products of representations of the linear groups, because an S function on the appropriate variables is a simple character of such a group. It is also well known that inner products of S functions correspond to Kronecker products of representations of the symmetric groups, because of the combinatorial manner in which S functions are composed. However, this correspondence involves the standard Young tableau notation for irreducible representations and the evaluation of inner products is a complex process quite distinct from that used in the evaluation of outer products. The symmetric groups are thus treated quite differently from the linear and other continuous groups: the orthogonal, rotation, and symplectic groups. The characters of these groups are known[8] in terms of S functions and the usual method of calculating such things as Kronecker products of the representations of these groups is to use S-functional expressions for their characters and the powerful algebra of S functions associated with the n-independent outer product rule. The labels that arise from this approach are the same as those that arise from tensorial arguments.[7,9] The aim of this paper is to show that the symmetric groups, Σ_n, may be treated in an n-independent manner similar to that used for the restricted groups O_n and Sp_n, rather than in the usual n-dependent manner requiring a development of the somewhat complicated algebra of inner products of S functions.[10]

In the "reduced" notation the irreducible representation of Σ_n usually labelled by the symbol $[\nu] = [n-m, \mu_1, \mu_2, \cdots]$, where (μ) is a partition of m, will simply be labelled by the symbol $\langle\mu\rangle = \langle\mu_1, \mu_2, \cdots\rangle$. In this "reduced" notation the use of "triangular" brackets to specify irreducible representations of Σ_n should give rise to no confusion with the same notation which is conventionally used to specify irreducible representations of Sp_n. Rather it serves to underline the view of the symmetric group as a subgroup of the full linear group.

After some preliminary comments on special series of S functions and their inverses we give, in Sec. 3, the important branching rule of Littlewood appropriate to $L_n \to \Sigma_n$. It is then shown that the use of the reduced notation, which arises naturally from this branching rule, far from complicating the statement of well-established results in the theory of Σ_n, leads to very simple and concise formulas for the dimensions and characters of irreducible representations of Σ_n. These are obtained in Sec. 4.

In contrast to this, Sec. 5 is devoted to using the branching rule appropriate to $L_n \to \Sigma_n$ to find expressions for the characters of Σ_n. It is perhaps here that the similarity with the restricted continuous groups shows up most clearly, even though the S-functional approach entirely obviates the necessity of evaluating characters explicitly.

We then give Littlewood's theorems on Kronecker products and use them in the final section to discuss the question of the occurrence of the totally symmetric and totally antisymmetric representations in the ordinary and symmetrized Kronecker squares of an arbitrary representation.

We include some n-independent tables of branching rules for $O_{n-1} \to \Sigma_n$, of the characters of Σ_n as S functions on $n-1$ variables, and of the symmetrized Kronecker squares of representations of Σ_n.

2. S FUNCTIONS AND SERIES OF S FUNCTIONS

Various series of S functions (see Ref. 8, p. 238) are used to express the characters of the restricted groups in terms of S functions and to evaluate branching rules, which is the inverse operation. Littlewood[7] has introduced further series in order to give the branching rule from L_{n-1} to Σ_n. It is our purpose in this section to discuss these series and to find their inverses.

The notation used for S functions and operations upon them is standard, but it is worth remarking that Greek letters denote general partitions of the integers denoted by the corresponding Latin letters, e.g., λ is any partition of l. Latin letters are also used to denote partitions into one part only. Summations are to be taken over all possible partitions consistent with this notation.

Copyright © 1973 by the American Institute of Physics

Curly brackets $\{\ \}$ are used to distinguish an S function $\{\lambda\}$ from an associated partition (λ) and throughout the paper, except where otherwise indicated, algebraic operations are to be understood as operations on S functions. These consist of the operations of outer multiplication, division, inner multiplication, outer plethysm, and inner plethysm denoted by $\cdot, /, \circ, \otimes$ and $\odot$, respectively,[11] although the symbol for outer multiplication is often omitted.

It is always to be understood that $(\nu) = (\nu_1, \nu_2, \cdots)$ and $(\tilde{\nu}) = (\tilde{\nu}_1, \tilde{\nu}_2, \cdots)$ are mutually conjugate partitions of n, so that for example $\{\tilde{n}\} = \{1^n\}$.

We start by reminding readers of the context in which some of the simpler series of S functions appear, that is in the evaluation of branching rules and their inverses. These series may be defined as a sum of symmetrized outer products, or outer plethysms, of S functions.

For example, under the restriction from L_n to L_{n-1} the linear function $\{1\}$ associated with the defining representation of L_n contains an invariant of L_{n-1} whose symmetrized powers must be contracted out of every other S function $\{\lambda\}$ by the operation of division to give the branching rule:

$$L_n \to L_{n-1}\{\lambda\} \to \{\lambda\}/\Big(\sum_r \{1\} \otimes \{r\}\Big). \tag{2.1a}$$

In this case the set of S functions in the divisor is just the set of all $\{r\}$.

Likewise for the restriction from L_n to O_n and Sp_n the bilinear forms $\{2\}$ and $\{1^2\}$ respectively contain invariants and we obtain the branching rules

$$L_n \to O_n\{\lambda\} \to \Big[\{\lambda\}/\Big(\sum_r \{2\} \otimes \{r\}\Big)\Big], \tag{2.1b}$$

$$L_n \to Sp_n\{\lambda\} \to \Big\langle\{\lambda\}/\Big(\sum_r \{1^2\} \otimes \{r\}\Big)\Big\rangle, \tag{2.1c}$$

where the brackets $[\ ]$ and $\langle\ \rangle$ indicate that the enclosed expressions, when evaluated using the algebra of S functions, are to be considered as characters of O_n and Sp_n, respectively. The sets of S functions in the divisors of (2.1b) and (2.1c) are just the sets of all $\{\delta\}$ and of all $\{\beta\}$ defined by Littlewood (Ref. 8, p. 238). The inverse series are then associated with the sets of all $\{\gamma\}$ and of all $\{\alpha\}$, while the series inverse to that of the divisor of (2.1a) is the set of all $\{1^r\}$. Using the algebra of plethysm (Ref. 8, p. 290) which implies that

$$(-\{\lambda\}) \otimes \{\rho\} = (-)^r\{\lambda\} \otimes \{\tilde{\rho}\}, \tag{2.2}$$

it is easy to see[12] that the inverse series are given by

$$\sum_r (-)^r\{1^r\} = \sum_r (-\{1\}) \otimes \{r\}, \tag{2.3a}$$

$$\sum_r (-)^r\{2\} \otimes \{1^r\} = \sum_r (-\{2\}) \otimes \{r\}, \tag{2.3b}$$

$$\sum_r (-)^r\{1^2\} \otimes \{1^r\} = \sum_r (-\{1^2\}) \otimes \{r\}. \tag{2.3c}$$

These latter series allow one to express the characters of the respective subgroups in terms of the S functions on the variables of L_n. In the past they have generally been written out in full, but it is instructive to write them in this way in order to see their simple combinatorial structure.

More generally for a series of S functions A to have as its inverse B it is required that for all $\{\lambda\}$

$$\{\lambda\}/A/B = \{\lambda\},$$

but since the sequential division by two series is equal to the division by the product of the series, we have

$$\{\lambda\}/A/B = \{\lambda\}/AB = \{\lambda\},$$

i.e.,

$$AB = \{0\},$$

where the S function $\{0\}$ is none other than the integer 1.

It is well known that

$$\{0\} \otimes \{r\} = \{0\},$$

so that writing

$$\{0\} = \{0\} + \{\lambda\} - \{\lambda\}$$

and using the algebra of plethysm we have

$$\{0\} = (\{0\} + \{\lambda\} - \{\lambda\}) \otimes \{r\}$$
$$= \sum_{s,t} \{0\} \otimes \{r - s - t\} \cdot \{\lambda\} \otimes \{s\} \cdot (-\{\lambda\}) \otimes \{t\};$$

r can be taken as large as we wish, so we have in general

$$\{0\} = \sum_s \{\lambda\} \otimes \{s\} \cdot \sum_t (-\{\lambda\}) \otimes \{t\}. \tag{2.4}$$

This is the general result which verifies the inverse series given in (2.3). In the following section the mutually inverse series of (2.4) will be required for all $\{\lambda\} = \{l\}$, but regretably this will not be all. Later we produce the terms of a different type of series altogether.

3. THE SYMMETRIC GROUP AS A SUBGROUP OF THE LINEAR GROUP

The main results of this paper rest on the fact that the symmetric group on n variables, Σ_n, is a subgroup of all the linear transformations of those variables, L_n. The matrices of all representations of Σ_n may be chosen to be orthogonal including those of the defining representation $\langle 1\rangle$ and thus the symmetric group is also a subgroup of the orthogonal group O_n, i.e., there exists the subgroup chain

$$L_n \to O_n \to \Sigma_n. \tag{3.1}$$

However, some of the matrices of the representation $\langle 1\rangle$ of Σ_n have negative determinant so the special orthogonal, or rotation, group SO_n, or R_n, cannot be included in this chain.

The orthogonal group leaves invariant the quadratic form $\sum_{i=1}^n x_i^2$, whereas the symmetric group clearly leaves invariant all such algebraic forms $\sum_{i=1}^n x_i^p$. The existence of the invariant linear form, with $p = 1$, corresponds to the restriction from n variables to $n - 1$ new variables, so that we may consider two alternative embeddings of Σ_n in L_n:

$$\begin{array}{ccccc} & & O_n & & \\ & \nearrow & & \searrow & \\ L_n & & & & O_{n-1} \to \Sigma_n. \\ & \searrow & & \nearrow & \\ & & L_{n-1} & & \end{array} \tag{3.2}$$

Combining the branching rules (2.1a) and (2.1b), we obtain the branching rule

$$L_n \to O_{n-1}\{\lambda\} \to \left[\{\lambda\}/\left(\sum_r \{1\}\otimes\{r\} \cdot \sum_s \{2\}\otimes\{s\}\right)\right]. \quad (3.3)$$

In restricting to the symmetric group there are many more terms to consider, and Littlewood gives,[7] in a different notation, the branching rule:

$$L_{n-1} \to \Sigma_n\{\lambda\} \to \left\langle \sum (\{\lambda\}/(\{2\}\otimes\{s\} \cdot \{3\}\otimes\{t\} \cdot \{4\}\otimes\{u\}\cdots\{2\}\otimes\{\xi\} \cdot \{3\}\otimes\{\eta\} \cdot \{4\}\otimes\{\xi\}\cdots))\{\xi\}\{\eta\}\{\zeta\}\cdots\right\rangle, \quad (3.4)$$

where the sum is taken over all possible $s, t, u, \ldots, \xi, \eta, \zeta, \cdots$ and the brackets $\langle\cdots\rangle$ indicate that the enclosed set of S functions are to be interpreted as specifying irreducible representations of Σ_n in the "reduced" notation. Thus the S function $\{\mu\}$ appearing within the brackets $\langle\cdots\rangle$ is to be written as $\langle\mu\rangle$ and the corresponding irreducible representation of Σ_n would be denoted in the standard notation by $[\nu] = [n-m, \mu_1, \mu_2, \cdots]$. For small values of n, i.e., $n < m + \mu_1$, this symbol will still not be in the standard form. It may be brought to standard form by reordering the parts in the usual way[13] in accordance with the formula

$$[\nu_1, \ldots, \nu_i, \nu_{i+1}, \ldots \nu_r] = -[\nu_1, \ldots, \nu_{i+1} - 1, \nu_i + 1, \ldots \nu_r]. \quad (3.5)$$

This modification rule is very easy to apply in practice but it is to be noted that it may produce null terms, e.g., $[2,3] = 0$, or terms that cancel others, e.g., $[1,4] = -[3,2]$. Of course the modification rules are only needed when dealing with representations of a specific group Σ_n, that is when n takes on a specific value. On the other hand (3.4) is valid for all values of n and clearly lends itself to the adoption of the n-independent "reduced" notation involving the symbols $\langle\mu\rangle$.

Some specific examples of (3.4) are of interest, namely:

$$L_{n-1} \to \Sigma_n\{1\} \to \langle 1\rangle \quad (3.6a)$$

$$\{2\} \to \langle 2\rangle + \langle 1\rangle + \langle 0\rangle \quad (3.6b)$$

$$\{1^2\} \to \langle 1^2\rangle \quad (3.6c)$$

$$\{1^k\} \to \langle 1^k\rangle \quad (3.6d)$$

$$\{1^{n-1}\} \to \langle 1^{n-1}\rangle. \quad (3.6e)$$

These results all follow in a trivial way from (3.4); the first (3.6a), justifies our use of the term "defining representation" in referring to $\langle 1\rangle$. It also defines the embedding of Σ_n in L_{n-1} completely and as such confirms that Σ_n is a subgroup of L_{n-1}. The second result (3.6b) implies that there exists a symmetric bilinear invariant, so that Σ_n is then necessarily a subgroup of O_{n-1}. The result (3.6c) corresponds to the converse statement that Σ_n is not a subgroup of Sp_{n-1} since there exists no antisymmetric bilinear invariant. In the same way (3.6e) implies that Σ_n is not a subgroup of either SU_{n-1} or $SO_{n-1} \sim R_{n-1}$ since the representation $\langle 1^{n-1}\rangle$ is not the same as $\langle 0\rangle$; they are instead mutually conjugate since $\langle 1^{n-1}\rangle = [1^n] = [\bar{n}]$ and $\langle 0\rangle = [n]$ in the standard notation for representations of Σ_n.

Let us return however to the rather formidable expression on the right-hand side of (3.4). For purposes of computation we may separate this branching rule into two parts:

$$L_{n-1} \to Q_{n-1}\{\lambda\} \to \left\{\{\lambda\}/\prod_{i=2}\left(\sum_r \{i\}\otimes\{r\}\right)\right\}, \quad (3.7)$$

$$Q_{n-1} \to \Sigma_n\{\rho\} \to \left\langle \sum_{\xi,\eta,\cdots} (\{\rho\}/(\{2\}\otimes\{\xi\} \cdot \{3\}\otimes\{\eta\}\cdots)\{\xi\}\{\eta\}\cdots)\right\rangle, \quad (3.8)$$

where Q_{n-1} is just a convenient label. The functions $\{\rho\}$ defined here have well-defined properties although they will not, in general, be simple characters of any group. However, they are certainly compound characters of Σ_n as shown by (3.8). To obtain the branching rules appropriate to $L_n \to Q_{n-1}$ and $O_{n-1} \to Q_{n-1}$ it is only necessary in (3.7) to include the case $i = 1$ and exclude the case $i = 2$, respectively.

From the discussion of the previous section we know the inverse of (3.7), namely,

$$Q_{n-1} \to L_{n-1}\{\rho\} \to \{\rho\}/\prod_{i=2}\left(\sum_r (-1)^r\{i\}\otimes\{1^r\}\right). \quad (3.9)$$

Unfortunately (3.8) is not as simple as (3.7) either to use or to invert. The presence of the multiplicative factors $\{\xi\} \cdot \{\eta\}\cdots$ is such that it is not sufficient simply to invert series of the form $\sum_\tau \{j\}\otimes\{\tau\}$. However, by forming the products of the expansions of such series and writing the multiplicative factors as a sum of S functions, we can write

$$Q_{n-1} \to \Sigma_n\{\rho\} \to \left\langle \sum_\alpha (\{\rho\}/\{\alpha\}) \cdot A\right\rangle. \quad (3.10)$$

If we further introduce a semicolon into the notation and write

$$Q_{n-1} \to \Sigma_n\{\rho\} \to \left\langle \{\rho\}/\sum_\alpha (\{\alpha\};A)\right\rangle, \quad (3.11)$$

then the operations associated with the branching rule may be specified without reference to $\{\rho\}$. Littlewood[7] uses the operator notation of Foulkes[14] to achieve this separation. In the notation of (3.11), the rule (3.8) then gives

$$\begin{aligned}\sum_\alpha \{\alpha\};A = {} & \{0\};\{0\} + \{2\};\{1\} + \{3\};\{1\} + \{4\};(\{2\} + \{1\}) \\ & + \{31\};\{1^2\} + \{2^2\};\{2\} + \{5\};(\{2\} + \{1^2\} + \{1\}) + \{41\}; \\ & (\{2\} + \{1^2\}) + \{32\};(\{2\} + \{1^2\}) + \{6\}; \\ & (\{3\} + 2\{2\} + \{1^2\} + \{1\}) + \{51\};(\{21\} + \{2\} + 2\{1^2\}) \\ & + \{42\};(\{3\} + \{21\} + 2\{2\} + \{1^2\}) + \{41^2\};\{1^3\} + \{3^2\}; \\ & \times(\{1^3\} + \{1^2\}) + \{321\};\{21\} + \{2^3\};\{3\} + \cdots, \end{aligned} \quad (3.12)$$

where we include all terms $\{\alpha\}$ with $a \leq 6$.

These results have been used to calculate the branching rules of Table 1 appropriate to $O_{n-1} \to \Sigma_n$ for all representations specified by partitions of six and less. Using these branching rules and those for $L_{n-1} \to O_{n-1}$ we are able to invert these examples to obtain the expressions for the simple characters of Σ_n in terms of S functions on $n-1$ variables. These are given in Table II. It is worth pointing out that similar results for $L_{n-1} \to \Sigma_n$ and some of the corresponding inversions have been given by Murnaghan.[5]

From the results of Table II it is possible to find the first few terms of the form $\{\beta\};B$ defined such that

$$\Sigma_n \to L_{n-1}\langle\mu\rangle \to \{\mu\}/\sum_\beta \{\beta\};B, \quad (3.13)$$

i.e.

$$\sum_\beta \{\beta\};B = \left\{\prod_{i=2}\left(\sum_r \{i\}\otimes\{r\}\right) \cdot \sum_\alpha \{\alpha\};A\right\}^{-1}. \quad (3.14)$$

In the same way series have also been found for the

TABLE I. Branching rules $O_{n-1} \to \Sigma_n$.

$[0] \to \langle 0 \rangle$

$[1] \to \langle 1 \rangle$

$[2] \to \langle 2 \rangle + \langle 1 \rangle$
$[1^2] \to \langle 1^2 \rangle$

$[3] \to \langle 3 \rangle + \langle 2 \rangle + \langle 1^2 \rangle + \langle 1 \rangle + \langle 0 \rangle$
$[21] \to \langle 21 \rangle + \langle 2 \rangle + \langle 1^2 \rangle$
$[1^3] \to \langle 1^3 \rangle$

$[4] \to \langle 4 \rangle + \langle 3 \rangle + \langle 21 \rangle + 2\langle 2 \rangle + \langle 1^2 \rangle + 2\langle 1 \rangle + \langle 0 \rangle$
$[31] \to \langle 31 \rangle + \langle 3 \rangle + 2\langle 21 \rangle + \langle 1^3 \rangle + \langle 2 \rangle + 2\langle 1^2 \rangle + \langle 1 \rangle$
$[2^2] \to \langle 2^2 \rangle + \langle 3 \rangle + \langle 21 \rangle + \langle 2 \rangle$
$[21^2] \to \langle 21^2 \rangle + \langle 21 \rangle + \langle 1^3 \rangle$
$[1^4] \to \langle 1^4 \rangle$

$[5] \to \langle 5 \rangle + \langle 4 \rangle + \langle 31 \rangle + 2\langle 3 \rangle + 2\langle 21 \rangle + 3\langle 2 \rangle + 2\langle 1^2 \rangle + 3\langle 1 \rangle + \langle 0 \rangle$
$[41] \to \langle 41 \rangle + \langle 4 \rangle + 2\langle 31 \rangle + \langle 2^2 \rangle + \langle 21^2 \rangle + 2\langle 3 \rangle + 4\langle 21 \rangle + 2\langle 1^3 \rangle + 3\langle 2 \rangle + 3\langle 1^2 \rangle + 2\langle 1 \rangle$
$[32] \to \langle 32 \rangle + \langle 4 \rangle + 2\langle 31 \rangle + \langle 2^2 \rangle + \langle 21^2 \rangle + 2\langle 3 \rangle + 3\langle 21 \rangle + \langle 1^3 \rangle + 2\langle 2 \rangle + \langle 1^2 \rangle + \langle 1 \rangle$
$[31^2] \to \langle 31^2 \rangle + \langle 31 \rangle + \langle 2^2 \rangle + 2\langle 21^2 \rangle + \langle 1^4 \rangle + 2\langle 21 \rangle + 2\langle 1^3 \rangle + \langle 1^2 \rangle$
$[2^21] \to \langle 2^21 \rangle + \langle 31 \rangle + \langle 2^2 \rangle + \langle 21^2 \rangle + \langle 3 \rangle + \langle 21 \rangle$
$[21^3] \to \langle 21^3 \rangle + \langle 21^2 \rangle + \langle 1^4 \rangle$
$[1^4] \to \langle 1^4 \rangle$

$[6] \to \langle 6 \rangle + \langle 5 \rangle + \langle 41 \rangle + 2\langle 4 \rangle + 2\langle 31 \rangle + \langle 2^2 \rangle + 4\langle 3 \rangle + 3\langle 21 \rangle + \langle 1^3 \rangle + 5\langle 2 \rangle + 3\langle 1^2 \rangle + 4\langle 1 \rangle + 2\langle 0 \rangle$
$[51] \to \langle 51 \rangle + \langle 5 \rangle + 2\langle 41 \rangle + \langle 32 \rangle + \langle 31^2 \rangle + 2\langle 4 \rangle + 5\langle 31 \rangle + 2\langle 2^2 \rangle + 3\langle 21^2 \rangle + 4\langle 3 \rangle + 8\langle 21 \rangle + 3\langle 1^3 \rangle + 5\langle 2 \rangle + 6\langle 1^2 \rangle + 3\langle 1 \rangle$
$[42] \to \langle 42 \rangle + \langle 5 \rangle + 2\langle 41 \rangle + 2\langle 32 \rangle + \langle 31^2 \rangle + \langle 2^21 \rangle + 3\langle 4 \rangle + 5\langle 31 \rangle + 4\langle 2^2 \rangle + 3\langle 21^2 \rangle + \langle 1^4 \rangle + 5\langle 3 \rangle + 7\langle 21 \rangle + 2\langle 1^3 \rangle + 5\langle 2 \rangle + 3\langle 1^2 \rangle + 2\langle 1 \rangle + \langle 0 \rangle$
$[41^2] \to \langle 41^2 \rangle + \langle 41 \rangle + \langle 32 \rangle + 2\langle 31^2 \rangle + \langle 2^21 \rangle + \langle 21^3 \rangle + 3\langle 31 \rangle + 2\langle 2^2 \rangle + 5\langle 21^2 \rangle + 2\langle 1^4 \rangle + \langle 3 \rangle + 4\langle 21 \rangle + 4\langle 1^3 \rangle + \langle 2 \rangle + 2\langle 1^2 \rangle$
$[3^2] \to \langle 3^2 \rangle + \langle 41 \rangle + \langle 32 \rangle + \langle 31^2 \rangle + \langle 4 \rangle + 3\langle 31 \rangle + 2\langle 21^2 \rangle + 2\langle 3 \rangle + 2\langle 21 \rangle + 2\langle 1^3 \rangle + \langle 2 \rangle + 2\langle 1^2 \rangle + \langle 1 \rangle$
$[321] \to \langle 321 \rangle + \langle 41 \rangle + 2\langle 32 \rangle + 2\langle 31^2 \rangle + 2\langle 2^21 \rangle + \langle 21^3 \rangle + \langle 4 \rangle + 4\langle 31 \rangle + 3\langle 2^2 \rangle + 4\langle 21^2 \rangle + \langle 1^4 \rangle + \langle 3 \rangle + 4\langle 21 \rangle + \langle 1^3 \rangle + \langle 2 \rangle + \langle 1^2 \rangle$
$[31^3] \to \langle 31^3 \rangle + \langle 31^2 \rangle + \langle 2^21 \rangle + 2\langle 21^3 \rangle + \langle 1^5 \rangle + \langle 2^2 \rangle + 2\langle 21^2 \rangle + 2\langle 1^4 \rangle + \langle 1^3 \rangle$
$[2^3] \to \langle 2^3 \rangle + \langle 32 \rangle + \langle 2^21 \rangle + \langle 4 \rangle + \langle 31 \rangle + \langle 2^2 \rangle + \langle 3 \rangle$
$[2^21^2] \to \langle 2^21^2 \rangle + \langle 31^2 \rangle + \langle 2^21 \rangle + \langle 21^3 \rangle + \langle 31 \rangle + \langle 21^2 \rangle$
$[21^4] \to \langle 21^4 \rangle + \langle 21^3 \rangle + \langle 1^5 \rangle$
$[1^6] \to \langle 1^6 \rangle$

TABLE II. The characters of Σ_n as S functions on $n-1$ variables.

$\langle 0 \rangle = \{0\}$

$\langle 1 \rangle = \{1\}$

$\langle 2 \rangle = \{2\} - \{1\} - \{0\}$
$\langle 1^2 \rangle = \{1^2\}$

$\langle 3 \rangle = \{3\} - \{2\} - \{1^2\} - \{1\}$
$\langle 21 \rangle = \{21\} - \{2\} - \{1^2\} + \{0\}$
$\langle 1^3 \rangle = \{1^3\}$

$\langle 4 \rangle = \{4\} - \{3\} - \{21\} - \{2\} + \{1^2\} + \{1\}$
$\langle 31 \rangle = \{31\} - \{3\} - 2\{21\} - \{1^3\} + \{2\} + \{1\}$
$\langle 2^2 \rangle = \{2^2\} - \{3\} - \{21\} + 2\{1^2\} + 2\{1\}$
$\langle 21^2 \rangle = \{21^2\} - \{21\} - \{1^3\} + \{2\} - \{0\}$
$\langle 1^4 \rangle = \{1^4\}$

$\langle 5 \rangle = \{5\} - \{4\} - \{31\} - \{3\} + \{21\} + \{1^3\} + \{2\} + \{1^2\}$
$\langle 41 \rangle = \{41\} - \{4\} - 2\{31\} - \{2^2\} - \{21^2\} + \{3\} + 2\{21\} + \{1^3\} + \{2\} - \{1\}$
$\langle 32 \rangle = \{32\} - \{4\} - 2\{31\} - \{2^2\} - \{21^2\} + \{3\} + 3\{21\} + 2\{1^3\} + \{2\} + \{1^2\} - \{1\}$
$\langle 31^2 \rangle = \{31^2\} - \{31\} - \{2^2\} - 2\{21^2\} - \{1^4\} + 2\{3\} + 2\{21\} - \{2\} - \{1^2\} - 2\{1\}$
$\langle 2^21 \rangle = \{2^21\} - \{31\} - \{2^2\} - \{21^2\} + \{3\} + 2\{21\} + 2\{1^3\} - 2\{1\}$
$\langle 21^3 \rangle = \{21^3\} - \{21^2\} - \{1^4\} + \{21\} - \{2\} + \{0\}$
$\langle 1^5 \rangle = \{1^5\}$

other relevant cases: $\Sigma_n \to Q_{n-1}$, $\Sigma_n \to O_{n-1}$, and $\Sigma_n \to L_n$. The simplest series, i.e., the series with the fewest terms, is that of (3.13) and is given by

$$\sum_{\beta} \{\beta\};B = \{0\};\{0\} - \{2\};(\{1\} + \{0\}) + \{21\};(\{1\} + \{0\}) + (\{4\} + \{2^2\});(\{1^2\} + \{1\}) + \{31\};\{2\} - \{21^2\};(\{1\} + \{0\}) - \{41\};(\{1^2\} + \{1\}) + (-\{32\} + \{31^2\} + \{2^21\});$$
$$\times (\{2\} + \{1^2\} - \{1\}) + \{21^3\};(\{1\} + \{0\}) + \cdots. \quad (3.15)$$

4. DIMENSIONS AND CHARACTERS OF REPRESENTATIONS OF Σ_n

One of the most remarkable formulas associated with the symmetric group is that which gives the dimensions or degree of an irreducible representation $[\nu]$ of Σ_n, namely (Ref. 15, p. 44),

$$f^{[\nu]} = \frac{n!}{H(\nu)} \quad (4.1)$$

with the hook length factor $H(\nu)$ given by

$$H(\nu) = \prod_{i,j} (\nu_i - j + \tilde{\nu}_j - i + 1),$$

where $(\nu) = (\nu_1, \nu_2, \cdots)$ and $(\tilde{\nu}) = (\tilde{\nu}_1, \tilde{\nu}_2, \cdots)$ are mutually conjugate partitions of n.

It would be most unfortunate if the use of the reduced notation led to the abandonment of a result of this type. However, a straightforward calculation involving the explicit expressions for the hook lengths of the boxes in the first row of the Young tableau defined by (ν), and a cancellation of all of these factors with terms in $n!$ yields the formula

$$f^{[\nu]} = f_n^{<\mu>} = \frac{1}{H(\mu)} \frac{n!}{(n-m+r)!} \prod_{i=1}^{r} (n - m - \mu_i + i), \quad (4.2)$$

where $[\nu] = [n-m, \mu_1, \mu_2, \ldots, \mu_r]$.

This formula may also be derived by noting that the hook lengths of the boxes in the first row of the tableau defined by (ν) are all distinct so that the cancellation referred to must always take place to give for $f_n^{<\mu>}$ a polynomial in n of degree m, divided by the hook length factor associated with the reduced tableau defined by (μ). The factors of the polynomial are determined by noting further that if $n - m < \mu_1$, $[\nu]$ is defined by (3.5). Repeated application of (3.5) gives

$$[n-m, \mu_1, \mu_2, \ldots, \mu_i, \cdots] = -[\mu_i - i, \mu_1, \mu_2, \ldots, n-m+i, \cdots]. \quad (4.3)$$

Therefore the dimension of the corresponding representation of Σ_n must vanish identically whenever

$$n - m = \mu_i - i.$$

From this it follows that

$$f_n^{<\mu>} = \prod_{i=1}^{r} (n - m - \mu_i + i)/H(\mu), \quad (4.4)$$

where of course $\mu_i = 0$ if $i > r$. This formula is then identical to (4.2) and here takes a form which is only slightly more complicated than that of (4.1) while being considerably more general.

For example,

$$f^{[n-5,2^21]} = f_n^{<2^21>} = \frac{1}{24} n(n-1)(n-3)(n-5)(n-6). \quad (4.5)$$

The factors in this expression correspond to the vanishing of the S functions $\{-5, 2^21\}, \{-4, 2^21\}, \{-2, 2^21\}, \{0, 2^2, 1\}$, and $\{1, 2^2, 1\}$. Substituting $n = 2$ and $n = 4$ in (4.5) gives $f^{[1^2]} = f^{[1^4]} = 1$ by virtue of the identities $\{-3, 2^21\} = -\{1^2\}$ and $\{-1, 2, 1\} = \{1^4\}$. Setting $n = 7, 8, \cdots$ yields $f^{[2^31]} = 14, f^{[32^21]} = 70, \cdots$, etc.

The n-dependence of $f_n^{<\mu>}$ may be obtained in a quite distinct form by making use of the identity (see Ref. 15, p. 42)

$$[\nu] = |[\nu_i - i + j]|.$$

Expansion of this determinant with respect to the first row yields, for $[\nu] = [n-m, \mu_1, \mu_2, \cdots]$, the result

$$[\nu] = \sum_{r=0}^{m} (-1)^r [n-m+r] \cdot [\mu]/[1^r] \tag{4.6}$$

as can be seen by expanding $[\mu]/[1^r]$ in accordance with the general formula (Ref. 15, p. 48)

$$[\mu]/[\rho] = |[\mu_i - \rho_j - i + j]|^{\cdot}.$$

Since (Ref. 15, p. 48),

$$f^{[\alpha]\cdot[\beta]} = \frac{(a+b)!}{a!b!} f^{[\alpha]} f^{[\beta]},$$

it follows from (4.6) with the substitution $k = m - r$, that

$$f^{[\nu]} = f_n^{<\mu>} = \sum_{k=0}^{m} (-1)^{m-k} \binom{n}{k} f^{[\mu]/[1^{m-k}]}. \tag{4.7}$$

This formula is clearly not as easy to use as (4.4) as it involves a summation; however, it will be used in what follows to establish a formula for the characters of Σ_n.

The trace of the matrix representation $[\nu]$ of any element of Σ_n in the class (α) of cycle structure $(1^{\alpha_1}2^{\alpha_2}\cdots k^{\alpha_k}\cdots n^{\alpha_n})$ is the character

$$\chi_{(\alpha)}^{[\nu]} = \chi_{(1^{\alpha_1}2^{\alpha_2}\cdots)}^{[\nu_1,\nu_2,\cdots]},$$

where

$$\sum_k k\alpha_k = n.$$

In general[1]

$$\chi_{(\alpha)}^{[\nu]} = \sum_i \chi_{(1^{\alpha_1}2^{\alpha_2},\ldots,k^{\alpha_{k-1}},\ldots n^{\alpha_n})}^{[\nu_1,\nu_2,\ldots,\nu_i-k\cdots]}, \tag{4.8}$$

where the use of the rearrangement formula (3.5) may be needed to relate nonstandard S functions to those specified by partitions.

Using this formula (4.8) in association with the reduced notation and treating the first row of the corresponding Young tableau somewhat differently from the others, since it is of undetermined length, leads to the formula

$$\chi_{(\alpha)}^{[\nu]} = \chi_{(\alpha)}^{<\mu>} = \sum_{\lambda,\beta} \chi_{(\beta)}^{[\mu]/[\lambda]} f_{\alpha_1}^{<\lambda>} \prod_{k=2}^{n} \binom{\alpha_k}{\beta_k}, \tag{4.9}$$

where $(\beta) = (1^{0}2^{\beta_2}3^{\beta_3}\cdots)$, $\sum_k k\beta_k = m - l$, (λ) is a partition of l, and $\chi_{(\beta)}^{[\mu]/[\lambda]}$ can be considered to be defined either by the division procedure and the formula (4.8), or directly by a recurrence formula like (4.8) but with the subtraction procedure leaving a residue of l boxes associated with $\{\lambda\}$.

To establish (4.9) it is only necessary to use (4.8), and to note that

$$f_{\alpha_1}^{<\lambda>} = \chi_{(1^{\alpha_1})}^{[\alpha_1-l,\lambda_1,\lambda_2,\cdots]} \tag{4.10}$$

while the factor $\binom{\alpha_k}{\beta_k}$ is just the number of distinct ways of selecting β_k k cycles from a set of α_k k cycles.

The resulting formula (4.9) is the general formula appropriate to the specific results obtained by Murnaghan.[3]

As an application it is easily seen that

$$\chi_{\alpha}^{<2^21>} = f_{\alpha_1}^{<2^21>} - f_{\alpha_1}^{<1^3>}\binom{\alpha_2}{1} + f_{\alpha_1}^{<1>}\binom{\alpha_2}{2} - f_{\alpha_1}^{<2>}\binom{\alpha_3}{1} + f_{\alpha_1}^{<1>}\binom{\alpha_4}{1} - f_{\alpha_1}^{<0>}\binom{\alpha_3}{1}\binom{\alpha_2}{1}.$$

So using (4.5) and similar results, we have

$$\chi_{(\alpha)}^{<2^21>} = \frac{1}{24}\alpha_1(\alpha_1-1)(\alpha_1-3)(\alpha_1-5)(\alpha_1-6) - \frac{1}{6}(\alpha_1-1)(\alpha_1-2)(\alpha_1-3)\alpha_2 - \frac{1}{2}\alpha_1(\alpha_1-3)\alpha_2 + (\alpha_1-1) \times \frac{1}{2}\alpha_2(\alpha_2-1) + (\alpha_1-1)\alpha_4 - \alpha_2\alpha_3,$$

in agreement with the result given by Murnaghan.[3]

The formula (4.9) is similar to, but not identical with, a formula due to Gamba[16] which may be obtained from (4.9) by making use of (4.7) with $n = \alpha_1$ and $k = \beta_1$. This gives

$$\chi_{\alpha}^{[\nu]} = \chi_{(\alpha)}^{<\mu>} = \sum_{\beta} \chi_{(\beta)}^{[\mu]/[1^p]} (-1)^p \prod_{k=1}^{n} \binom{\alpha_k}{\beta_k}, \tag{4.11}$$

where now $(\beta) = (1^{\beta_1}2^{\beta_2}\cdots)$ with $\sum k\beta_k = m - p$. While Gamba's formula (4.11) has the advantage over (4.9) of possessing full symmetry in $\alpha_1, \alpha_2, \ldots, \alpha_n$, the use of (4.4) in conjunction with (4.9) yields expressions for characters containing fewer terms than those arising from the use of (4.11). It is to be noted that (4.7) is just a special case of (4.11) (see Ref. 17, p. 201).

5. CLASS PARAMETERS AND THE NUMERICAL VALUE OF CHARACTERS

In the preceding section is has been shown that formulas for the dimensions and characters of representations of Σ_n may be conveinently written down using the reduced notation. However, Sec. 3 makes it clear that these results must also follow from a study of the related S functions on n or $n-1$ variables. In the discussion of these S functions we have at no time required any information on these variables. However, the evaluation of the variables for the classes of a group is a trivial matter and can be carried out for the symmetric group in much the same way that Littlewood carried out the evaluation for the orthogonal and rotation groups (Ref. 7, Chap. XI.) In this instance, however, the variables are discrete since the classes of Σ_n are labelled by discrete, as opposed to continuous, parameters.

From a knowledge of the characters in terms of S functions and the explicit values of the variables associated with the S functions for each class it is a straightforward matter to obtain the explicit values of the characters for each class. We show that although this technique is quite different from that used in Sec. 4 identical results are obtained in the form of functions of the cycle structure parameters $\alpha_1, \alpha_2, \cdots$.

The class of an element S_0 of a group is defined as the set of all elements, S, which can be written in the form TS_0T^{-1} where T is any element of the group. For the unitary group there always exists an element T whose matrix representation diagonalizes the matrix representation of S_0 to give a diagonal matrix D. All other matrix representatives of elements S in the same class as S_0 may also be diagonalized to D. In the case

of the n-dimensional unitary, defining representation $\{1\}$ of U_n, D will be of the form

$$D = \text{diag}\,(e^{i\phi_1}, e^{i\phi_2}, \ldots, e^{i\phi_n}). \tag{5.1}$$

As the parameters vary independently and continuously from 0 to 2π all classes are covered once. Thus for each value of the set of parameters $(\phi_1 \cdots \phi_n)$ the S function $\{1\}$ on the n variables $e^{i\phi_j}$, that is, the function

$$\{1\} = \sum_{j=1} e^{i\phi_j} \tag{5.2}$$

will give the characteristic (trace) of the matrices of this class of this defining representation of U_n. Littlewood, (Ref. 8, p. 222), goes on to prove that the S functions $\{\lambda\}$ on the n variables $e^{i\phi_j}$ are the simple characters of all the unitary higher order representations of U_n.

Coming back to the symmetric group it is only necessary in determining the characters to study one matrix representative of each class $(\alpha) = (1^{\alpha_1} 2^{\alpha_2} \cdots)$ of permutations. For Σ_5 the following suffice:

$$(1^5) \sim \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & 1 \end{pmatrix}, \quad (1^3 2) \sim \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & 0 & 1 \\ & & & 1 & 0 \end{pmatrix}, \quad (1^2 3) \sim \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & 0 & 1 & 0 \\ & & 0 & 0 & 1 \\ & & 1 & 0 & 0 \end{pmatrix}, \quad (1 2^2) \sim \begin{pmatrix} 1 & & & & \\ & 0 & 1 & & \\ & 1 & 0 & & \\ & & & 0 & 1 \\ & & & 1 & 0 \end{pmatrix},$$

$$(14) \sim \begin{pmatrix} 1 & & & & \\ & 0 & 1 & 0 & 0 \\ & 0 & 0 & 1 & 0 \\ & 0 & 0 & 0 & 1 \\ & 1 & 0 & 0 & 0 \end{pmatrix}, \quad (23) \sim \begin{pmatrix} 0 & 1 & & & \\ 1 & 0 & & & \\ & & 0 & 1 & 0 \\ & & 0 & 0 & 1 \\ & & 1 & 0 & 0 \end{pmatrix}, \quad (5) \sim \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The sum of their characteristic roots may be read off immediately but to obtain the class parameters it is necessary to find the individual roots. We do this by diagonalizing with a unitary matrix. It is found that the various cycles contribute

$$(1) \to 1, \quad (2) \to \pm 1, \quad (3) \to 1, e^{\pm 2\pi i/3},$$

$$(4) \to \pm 1, e^{\pm \pi i/2}, \quad (5) \to 1, e^{\pm 2\pi i/5}, e^{\pm 4\pi i/5}$$

with the general result

$$(p) \to e^{(2j/p)\pi i} \quad \text{with } j = 1, 2, \ldots, p. \tag{5.3}$$

Note that on giving the representation matrices of Σ_5 we have used a reducible basis of dimension 5 to give the permutation representation of the group which reduces to $\langle 1 \rangle + \langle 0 \rangle$. To avoid confusion with the embedding of Σ_n in L_{n-1} we shall attach to the S functions a label giving the number of variables.

We have

$$\{1\}_n = \langle 1 \rangle + \langle 0 \rangle \tag{5.4a}$$

or, written in full,

$$\chi^{\{1\}_n}_{(\alpha)} = \chi^{\langle 1 \rangle}_{(\alpha)} + \chi^{\langle 0 \rangle}_{(\alpha)}. \tag{5.4b}$$

But $\chi^{\langle 0 \rangle}_{(\alpha)} = 1$ and it follows from the example that

$$\chi^{\langle 1 \rangle}_{(\alpha)} = \alpha_1 - 1 \tag{5.5a}$$

if use is made of (5.3). Here α_1 is just the number of 1 cycles in the class.

The S function $\{2\}_n$ on the roots $x_1, x_2, \ldots, x_n$ is the function

$$\{2\}_n = \sum_{i=1}^{n} x_i^2 + \sum_{i<j=1}^{n} x_i x_j.$$

Substituting in the values of the roots given by (5.3) we obtain

$$\{2\}_n = \tfrac{1}{2}\alpha_1(\alpha_1 + 1) + \alpha_2.$$

From the branching rules of Sec. 3

$$\{2\}_n = \{2\}_{n-1} + \{1\}_{n-1} + \{0\}_{n-1}$$
$$= \langle 2 \rangle + 2\langle 1 \rangle + 2\langle 0 \rangle,$$

so that

$$\chi^{\langle 2 \rangle}_{(\alpha)} = \tfrac{1}{2}\alpha_1(\alpha_1 - 3) + \alpha_2. \tag{5.5b}$$

Similarly,

$$\chi^{\langle 1^2 \rangle}_{(\alpha)} = \tfrac{1}{2}\alpha_1(\alpha_1 - 3) - \alpha_2 + 1. \tag{5.5c}$$

These formulas (5.5) are just special cases of those furnished by (4.9) or (4.11) and have been given many years ago by Murnaghan.[1] However, it is felt that the derivation given in this section demonstrates that the symmetric group can with advantage be viewed as a subgroup of the linear group.

6. ORDINARY AND SYMMETRIZED KRONECKER PRODUCTS

It is clearly possible to calculate the reduction of products and branching rules using the results of the previous sections and the algebra of S functions. The results so obtained would be given independently of n in a form

analogous to those obtained for the restricted continuous groups.

However, for the computation of Kronecker products Littlewood[7] has treated possible contractions with the set of all invariants associated with Σ_n from first principles to establish the following remarkably simple results.

Firstly, the Kronecker product of symmetric group representations $\langle\lambda\rangle$ and $\langle\mu\rangle$ is

$$\langle\lambda\rangle\langle\mu\rangle = \sum_{\alpha,\beta,\gamma} \langle(\{\lambda\}/\{\alpha\}\{\beta\}) \cdot (\{\mu\}/\{\alpha\}\{\lambda\}) \cdot (\{\beta\}\circ\{\gamma\})\rangle, \quad (6.1)$$

where, of course, (β) and (γ) are necessarily both partitions of the same number. It is to be noted that although $\circ$ signifies an inner product of S functions we can write $\{\beta\}\circ\{\gamma\} = \langle\sigma\rangle\langle\tau\rangle$ where $st \ll lm$, so that (6.1) is a useful formula giving an n-independent Kronecker product in terms of an n-dependent Kronecker product which may be evaluated from a knowledge of an n-independent product of much lower weight.

If $\langle\lambda\rangle = \langle\mu\rangle = [\nu] = [n-l, \lambda_1, \lambda_2 \cdots]$ then the product (6.1) may be reduced into its symmetric and antisymmetric parts $\langle\lambda\rangle \otimes \{2\} = [\nu] \odot \{2\}$ and $\langle\lambda\rangle \otimes \{1^2\} = [\nu] \odot \{1^2\}$, respectively. Littlewood[7] gives the result

$$\begin{aligned}[\nu] \odot \{\tau\} &= \langle\lambda\rangle \otimes \{\tau\} \\ &= \sum_{\alpha} \sum_{\beta<\gamma} \langle(\{\lambda\}/\{\alpha\}\{\beta\}) \cdot (\{\lambda\}/\{\alpha\}\{\gamma\}) \cdot (\{\beta\}\circ\{\gamma\})\rangle \\ &\quad + \sum_{\alpha,\beta,\sigma} \langle(\{\lambda\}/\{\alpha\}\{\beta\}) \otimes \{\sigma\} \cdot \{\beta\} \odot (\{\sigma\}\circ\{\tau\})\rangle, \quad (6.2)\end{aligned}$$

where $\{\tau\} = \{2\}$ or $\{1^2\}$ and the summation over σ only includes $\{\sigma\} = \{2\}$ and $\{\sigma\} = \{1^2\}$. The occurrence of all the S function operations $\cdot, /, \circ, \otimes$ and $\odot$ in this result is to be noticed.

Murnaghan[2,4] has tabulated Kronecker products of the form (6.1) to give results sufficient to determine all the products of Σ_n with $n \leq 8$. In Table III we give the resolution of the Kronecker squares into their symmetric and antisymmetric parts, $\langle\lambda\rangle \otimes \{2\}$ and $\langle\lambda\rangle \otimes \{1^2\}$, respectively, for $l \leq 3$.

It is of interest to determine in which part of the resolution of a Kronecker square the scalar 1-dimensional representations of $\sum_n$ occur, that is the totally symmetric and antisymmetric respresentations $[n] = \langle 0\rangle$ and $[1^n] = \langle 1^{n-1}\rangle$. They both occur but rarely in Kronecker products; the former in products of the form $[\nu]\circ[\nu]$, and the latter in $[\nu]\circ[\tilde{\nu}]$. This reflects the statement that every irreducible representation $[\nu]$ of Σ_n is self adjoint so that its square must contain the invariant $[n]$ once and once only. The remark regarding $[\nu]\circ[\tilde{\nu}]$ then follows by noting that $[\nu]\circ[\nu]\circ[1^n] = [\nu]\circ[\tilde{\nu}]$.

Using (6.2) it is clear that $\langle\lambda\rangle \otimes \{\tau\}$ with $\{\tau\} = \{2\}$ or $\{1^2\}$ will contain $\langle 0\rangle$ if and only if in the second summation we have $\{\beta\} = \{0\}$, $\{\alpha\} = \{\lambda\}$ and the corresponding term $\{0\} \otimes \{\sigma\} \cdot \{0\} \odot (\{\sigma\}\circ\{\tau\})$ contains $\{0\}$. This requires $\{\sigma\} = \{\tau\} = \{2\}$, so that

$$\langle\lambda\rangle \otimes \{2\} \supset \langle 0\rangle, \quad (6.3a)$$

while

$$\langle\lambda\rangle \otimes \{1^2\} \not\supset \langle 0\rangle. \quad (6.3b)$$

This trivial result merely reflects the fact, commented upon earlier, that every representation of Σ_n may be made orthogonal but not of course symplectic.

The above remarks imply that $\langle\lambda\rangle \otimes \{\tau\}$ with $\{\tau\} = \{2\}$ or $\{1^2\}$ will contain $\langle 1^{n-1}\rangle$ only if $[\nu] = [n-l, \lambda_1, \lambda_2, \cdots]$ is self conjugate, i.e., $[\nu] = [\tilde{\nu}]$. Furthermore, $\langle\lambda\rangle \otimes \{\tau\}$ contains $\langle 1^{n-1}\rangle$ only once at most, so that such a term cannot arise in the first summation of (6.2) which contributes equally to the cases $\{\tau\} = \{2\}$ and $\{\tau\} = \{1^2\}$. For $\langle 1^{n-1}\rangle$ to arise in the second summation we require that both

$$(\{\lambda\}/\{\alpha\}\{\beta\}) \otimes \{\sigma\} \supset \{1^{2(l-a-b)}\} \quad (6.4)$$

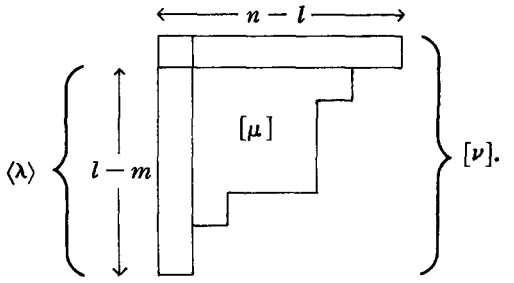
and

$$\{\beta\} \odot (\{\sigma\}\circ\{\tau\}) \supset \{1^b\} \quad (6.5)$$

with

$$n - 1 = 2l - 2a - b. \quad (6.6)$$

Consider the Young tableau of $[\nu]$ with its subtableaux corresponding to $\langle\lambda\rangle$ and $[\mu]$, where $[\nu] = [n-l, \lambda_1, \lambda_2, \cdots]$ and $\langle\tilde{\lambda}\rangle = \langle l-m, \tilde{\mu}_1, \tilde{\mu}_2, \cdots\rangle$ as illustrated below:



From the condition $[\nu] = [\tilde{\nu}]$ we have $[\mu] = [\tilde{\mu}]$ and $m = 2l - n + 1$. Comparison with (6.6) then gives

$$m = 2a + b. \quad (6.7)$$

Now $\{\pi\}\{\pi\}$ contains $\{1^{2p}\}$ if and only if $\{\pi\} = \{1^p\}$ and

$$\{1^p\} \otimes \{2\} \supset \{1^{2p}\} \quad \text{if } p \text{ even}, \quad (6.8a)$$

$$\{1^p\} \otimes \{1^2\} \supset \{1^{2p}\} \quad \text{if } p \text{ odd}; \quad (6.8b)$$

thus (6.4) will only be satisfied if

$$\{\lambda\}/\{\alpha\}\{\beta\} \supset \{1^{l-a-b}\} \quad (6.9)$$

TABLE III. "Symmetrized Kronecker squares for the symmetric groups".

$\langle 0\rangle \otimes \{2\} = \langle 0\rangle$

$\langle 0\rangle \otimes \{1^2\} = 0$

$\langle 1\rangle \otimes \{2\} = \langle 0\rangle + \langle 1\rangle + \langle 2\rangle$

$\langle 1\rangle \otimes \{1^2\} = \langle 1^2\rangle$

$\langle 2\rangle \otimes \{2\} = \langle 0\rangle + \langle 1\rangle + 2\langle 2\rangle + \langle 21\rangle + \langle 3\rangle + \langle 2^2\rangle + \langle 4\rangle$

$\langle 2\rangle \otimes \{1^2\} = \langle 1^2\rangle + \langle 1^3\rangle + \langle 21\rangle + \langle 31\rangle$

$\langle 1^2\rangle \otimes \{2\} = \langle 0\rangle + \langle 1\rangle + 2\langle 2\rangle + \langle 21\rangle + \langle 3\rangle + \langle 1^4\rangle + \langle 2^2\rangle$

$\langle 1^2\rangle \otimes \{1^2\} = \langle 1^2\rangle + \langle 1^3\rangle + \langle 21\rangle + \langle 21^2\rangle$

$\langle 3\rangle \otimes \{2\} = \langle 0\rangle + \langle 1\rangle + 2\langle 2\rangle + \langle 21\rangle + 2\langle 3\rangle + 2\langle 2^2\rangle + \langle 31\rangle + 2\langle 4\rangle + \langle 2^21\rangle + \langle 32\rangle + \langle 41\rangle + \langle 5\rangle + \langle 2^3\rangle + \langle 42\rangle + \langle 6\rangle$

$\langle 3\rangle \otimes \{1^2\} = \langle 1^2\rangle + \langle 1^3\rangle + \langle 21\rangle + \langle 21^2\rangle + 2\langle 31\rangle + \langle 31^2\rangle + \langle 32\rangle + \langle 41\rangle + \langle 3^2\rangle + \langle 51\rangle$

$\langle 21\rangle \otimes \{2\} = \langle 0\rangle + 2\langle 1\rangle + \langle 1^2\rangle + 4\langle 2\rangle + \langle 1^3\rangle + 5\langle 21\rangle + 4\langle 3\rangle + 2\langle 1^4\rangle + 3\langle 21^2\rangle + 5\langle 2^2\rangle + 4\langle 31\rangle + 3\langle 4\rangle + \langle 1^5\rangle + 2\langle 21^3\rangle + 3\langle 2^21\rangle + 2\langle 31^2\rangle + 3\langle 32\rangle + 2\langle 41\rangle + \langle 5\rangle + \langle 2^3\rangle + \langle 31^3\rangle + \langle 321\rangle + \langle 42\rangle$

$\langle 21\rangle \otimes \{1^2\} = 3\langle 1^2\rangle + 4\langle 1^3\rangle + 4\langle 21\rangle + \langle 3\rangle + \langle 1^4\rangle + 6\langle 21^2\rangle + \langle 2^2\rangle + 5\langle 31\rangle + 2\langle 21^3\rangle + 2\langle 2^21\rangle + 4\langle 31^2\rangle + 2\langle 32\rangle + 2\langle 41\rangle + \langle 2^21^2\rangle + \langle 321\rangle + \langle 3^2\rangle + \langle 41^2\rangle$

$\langle 1^3\rangle \otimes \{2\} = \langle 0\rangle + \langle 1\rangle + 2\langle 2\rangle + \langle 21\rangle + 2\langle 3\rangle + \langle 1^4\rangle + 2\langle 2^2\rangle + \langle 31\rangle + \langle 4\rangle + \langle 1^5\rangle + \langle 21^3\rangle + \langle 2^21\rangle + \langle 32\rangle + \langle 21^4\rangle + \langle 2^3\rangle$

$\langle 1^3\rangle \otimes \{1^2\} = \langle 1^2\rangle + \langle 1^3\rangle + \langle 21\rangle + 2\langle 21^2\rangle + \langle 31\rangle + \langle 21^3\rangle + \langle 2^21\rangle + \langle 31^2\rangle + \langle 1^6\rangle + \langle 2^21^2\rangle$

and comparison with the diagram shows that

$$a + b \geq m. \tag{6.10}$$

Hence $a = 0$ and $b = m$, so that $\{\alpha\} = \{0\}$. Further comparison of (6.9) with the figure shows that $\{\beta\} = \{\mu\}$. The relevant term of (6.2) takes the form

$$[\nu] \odot \{\tau\} \supset \sum_{\sigma} \langle \{1^{l-m}\} \otimes \{\sigma\} \cdot \{\mu\} \odot (\{\sigma\} \circ \{\tau\}) \rangle. \tag{6.11}$$

Clearly $l - m = \nu_1 - 1$ and we conclude that

(i) if $\nu_1 - 1$ is even $[\nu] \odot \{\tau\} \supset [1^n]$ if and only if $[\mu] \odot \{\tau\} \supset [1^m]$,

(ii) if $\nu_1 - 1$ is odd $[\nu] \odot \{\tau\} \supset [1^n]$ if and only if $[\mu] \odot \{\tilde{\tau}\} \supset [1^m]$.

Applying this argument to $[\mu] \odot \{\tau\}$, the relevant factor is $\mu_1 - 1 = \nu_2 - 2$. Continuing in this way we arrive at $[0] \odot \{2\} = [0]$ and $[0] \odot \{1^2\} = 0$.

If r is the rank of the partition (ν) in Frobenius notation, that is the number of boxes on the leading diagonal of the corresponding tableau, then clearly

$$\sum_{i=1}^{r} (\nu_i - i) = \tfrac{1}{2}(n - r);$$

so we have shown, in the standard notation, if $[\nu] = [\tilde{\nu}]$, then

$$[\nu] \odot \{2\} \supset [1^n] \text{ only if } \tfrac{1}{2}(n - r) \text{ is even,} \tag{6.12a}$$

$$[\nu] \odot \{1^2\} \supset [1^n] \text{ only if } \tfrac{1}{2}(n - r) \text{ is odd.} \tag{6.12b}$$

The smallest value of n for which both "symmetries" occur for $[1^n]$ is $n = 9$. In this case we have

$$[51^4] \odot \{2\} \supset [1^9] \quad \text{and} \quad [3^3] \odot \{1^2\} \supset [1^9].$$

It is perhaps worth pointing out that the factor $\frac{1}{2}(n - r)$ occurred before in the consideration of the S function series arising in the discussion of the spin characters of the rotation and orthogonal groups.[18] The relevant identities were

$$\left(\sum_{\lambda} (\pm)^l \{\lambda\}\right) \cdot \left(\sum_{\nu = \tilde{\nu}} (-)^{(n \mp r)/2} \{\nu\}\right) = \{0\}. \tag{6.13}$$

7. CONCLUSIONS

It is to be hoped that the outline presented here, and the not insignificant number of general results, will serve to convince all who use representations of the symmetric group of the advantages to be found in the use of the n-independent reduced notation.

It is worth pointing out that the branching rule (3.4) corresponding to the evaluation of the inner plethysm $\langle 1 \rangle \otimes \{\lambda\}$ has been used already[19] by nuclear physicists while further tables of such inner plethysms would be of use in extending some nuclear shell model calculations.

Quite apart from physical applications, we have tried to stress the role of the symmetric group as just another subgroup, albeit finite, of the general linear group.

In a later paper it is hoped to make much use of the reduced notation and the results obtained here, to obtain the symmetries of and general formulas for the Clebsch–Gordan coefficients of the symmetric group.

ACKNOWLEDGMENT

One of the authors (P. H. B.) wishes to acknowledge the financial support of the New Zealand University Grants Committee.

---

*Permanent address: Physics Department, University of Canterbury, Christchurch, New Zealand.

[1] F. D. Murnaghan, Am. J. Math. **59**, 739 (1937).
[2] F. D. Murnaghan, Am. J. Math. **60**, 761 (1938).
[3] F. D. Murnaghan, Proc. Natl. Acad. Sci. USA **37**, 55 (1951); An. Acad. Bras. Cienc. **23**, 141 (1951); Proc. Natl. Acad. Sci. USA **41**, 396 (1955).
[4] F. D. Murnaghan, Proc. Natl. Acad. Sci. USA **41**, 515 (1955).
[5] F. D. Murnaghan, Proc. Natl. Acad. Sci. USA **41**, 514 (1955); Proc. Natl. Acad. Sci. USA **41**, 1096 (1955).
[6] D. E. Littlewood, Can. J. Phys. **10**, 1 (1958).
[7] D. E. Littlewood, Can. J. Phys. **10**, 17 (1958).
[8] D. E. Littlewood, *The Theory of Group Characters* (Oxford U.P., Oxford, 1950), 2nd ed.
[9] R. C. King, J. Math. Phys. **12**, 1588 (1971).
[10] G. de B. Robinson and O. E. Taulbee, Proc. Natl. Acad. Sci. USA **40**, 723 (1954).
[11] P. H. Butler and B. G. Wybourne, J. Phys. (Paris) **30**, 795 (1969).
[12] R. C. Read, Can. J. Phys. **20**, 808 (1968).
[13] F. D. Murnaghan, *The Theory of Group Representations* (John Hopkins, Baltimore, 1938).
[14] H. O. Foulkes, Philos. Trans. R. Soc. Lond. **246**, 555 (1954).
[15] G. de B. Robinson, *Representation Theory of the Symmetric Group* (University Press, Edinburgh, 1961).
[16] A. Gamba, Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Nat. Rend. **12**, 167 (1952).
[17] H. Boerner, *Representations of Groups* (North Holland, Amsterdam, 1963).
[18] P. H. Butler and B. G. Wybourne, J. Phys. (Paris) **30**, 655 (1969).
[19] M. Kretzchmar, Z. Phys. **157**, 433 (1960); Z. Phys. **158**, 284 (1960). L. Sabaliauskas, Liet Fiz. Rink. **9**, 477 (1969). V. Vanagas, *Algebraic Methods in Nuclear Theory* (Mintis, Vilnius, 1971) (in Russian).

Stieltjes summability and convergence of the Padé approximants for the vacuum polarization by an external field

Sandro Graffi[*†]

Department of Physics, New York University, New York, New York
(Received 17 November 1971; revised manuscript received 13 December 1971)

It is proven that the divergent perturbation expansion of the exactly solvable vacuum polarization by an external constant electromagnetic field is Stieltjes summable. This implies the convergence of the Padé approximants to the solution. The convergence of the approximation method based on Borel summability is also proven.

1. INTRODUCTION

It is well known that perturbation theory often gives rise to divergent power series expansion, in quantum mechanical problems[1] as well as in quantum field theory.[2] There exist, however, some regular summation techniques for divergent power series, such as Stieltjes or the Borel ones,[3] and in recent times there has been some interest in proving their applicability to singular perturbation theory, in order to recover the solution starting from its divergent perturbation expansion.

The Stieltjes method, being equivalent to the convergence of the Padé approximants, also allows direct computations. It is then the most powerful and elegant but also the most restrictive technique: till now its applicability has been rigorously proven only in quantum mechanical problems, such as the anharmonic oscillators[4] $p^2 + x^2 + \beta x^4$ and $p^2 + x^2 + \beta x^6$, and the exactly soluble Peres model.[5]

On the other hand the Borel method, whose applicability requires less information about the analyticity properties of the solution as a function of the coupling constant, has been proven valid for a larger class of problems: several years ago for the vacuum polarization by an external constant electromagnetic field,[6] and recently for any anharmonic oscillator[7] (βx^{2m} and in any finite number of dimensions) and for the ground state energy of a spatially cut-off field theory of the type $\beta : \phi^4(x)$: in two-dimensional space–time.[8] The Borel method, however, does not define in itself a convergent approximation procedure, since in this framework one is faced with the problem of performing the analytic continuation of the appropriate Borel transform outside its circle of convergence, continuation that may be accomplished in many different ways.[9] One way consists in applying the Padé approximants to the Borel transform: this method, although not proven to be convergent, gives better results than the Padé approximants in cases in which both the Stieltjes and the Borel method are applicable.[7]

The purpose of this paper is to show that the divergent perturbation expansion of the vacuum polarization by an external constant electromagnetic field is also Stieljtes summable to the exact solution,[10] thus providing a first example of convergence of the Padé approximants in a field theoretical problem. In addition to that, it will be shown that in this case it is possible to prove also the convergence of the approximation method consisting in applying the Padé approximants to the Borel transform.

2. STIELTJES SUMMABILITY AND CONVERGENCE OF THE PADÉ APPROXIMANTS

The vacuum polarization by a prescribed external constant electromagnetic field has been computed exactly by Schwinger.[10] After charge and field strength renormalization, the complete Lagrangian is given by the following formula:

$$L = -F - \frac{1}{8\pi^2}\int_0^\infty ds\, s^{-3}e^{-m^2 s} \times\left((es)^2 G\,\frac{\mathrm{Re}\cosh(esX)}{\mathrm{Im}\cosh(esX)} - 1 - \frac{2}{3}(es)^2 F\right), \quad (2.1)$$

where, as usual, m is the electron mass, e is the electron charge; $F = \frac{1}{4}F_{\mu\nu}^2 = \frac{1}{2}(\mathbf{H}^2 - \mathbf{E}^2)$ is the free Maxwell field Lagrangian; $G = \frac{1}{4}\epsilon_{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma} = \mathbf{E}\cdot\mathbf{H}$ is the pseudoscalar electromagnetic field invariant; $X = \sqrt{2(F + iG)}$.

For a pure magnetic field, invariantly characterized by $G = 0$, $2F = \mathbf{H}^2 > 0$, the Lagrangian (2.1) simplifies to

$$L = -\frac{1}{2}H^2 - \frac{1}{8\pi^2}\int_0^\infty ds\, s^{-3}e^{-m^2 s} \times\left((eHs)\coth(eHs) - 1 - \frac{1}{3}(eHs)^2\right) \quad (2.2)$$

and for a pure electric field ($G = 0$, $-2F = \mathbf{E}^2 > 0$)

$$L = \frac{1}{2}E^2 - \frac{1}{8\pi^2}\int_0^\infty ds\, s^{-3}e^{-m^2 s} \times\left((eEs)\cot(eEs) - 1 + \frac{1}{3}(eEs)^2\right). \quad (2.3)$$

Writing $L = L_0 + L_1$, where L_0 is the free field Lagrangian, and assuming for simplicity, without loss of generality, $H = 1$, $E = 1$, $m = 1$, we have from (2.2) and (2.3)

$$L_1^H = -\frac{1}{8\pi^2}\int_0^\infty \frac{e^{-s}}{s^3}\left((es)\coth(es) - 1 - \frac{1}{3}(es)^2\right)ds, \quad (2.4)$$

$$L_1^E = -\frac{1}{8\pi^2}\int_0^\infty \frac{e^{-s}}{s^3}\left((es)\cot(es) - 1 + \frac{1}{3}(es)^2\right)ds. \quad (2.5)$$

Expanding (2.4) and (2.5) in power series we obtain the perturbation series for the vacuum polarization

$$L_1^H(\alpha) = -\frac{1}{8\pi^2}\sum_{n=2}^{\infty}(8\pi)^{2n}B_{2n}\frac{(2n-3)!}{(2n)!}\alpha^n, \quad (2.6)$$

$$L_1^E(\alpha) = -\frac{1}{8\pi^2}\sum_{n=2}^{\infty}(8\pi)^{2n}B_{2n}\frac{(2n-3)!}{(2n)!}(-\alpha)^n, \quad (2.7)$$

where $\alpha = e^2/4\pi$ and B_{2n} are the Bernoulli numbers. Since $B_{2n} \sim (-1)^{n-1}(2n)!/2^{2n-1}\pi^{2n}$ for $n \gg 1$ (see Ref. 11), the power series (2.6) and (2.7) have radius

Copyright © 1973 by the American Institute of Physics

of convergence zero. Since, however, they are Borel summable to the solution, {the Borel sum is intended to be the generalized Borel sum of order 2 [(B. 2) sum], since the coefficients of (2.6) and (2.7) diverge as $(2n)!$} they uniquely determine the solution exactly as if they had a nonzero radius of convergence. It will be proven now that (2.6) and (2.7) are Stieltjes summable, which implies the convergence to the solution of their Padé approximants. To this end, use will be made of the following theorem.

Theorem:[12]

Let $f(z)$ a function with the following properties:

(a) $f(z)$ is analytic in the complex z plane cut along the negative real axis from $-\infty$ to 0; and is real on the positive real axis.

(b) $|f(z)| < |z|^{\alpha}$ for all z such that $|z| > R_0$. Here $\alpha < k$, $k \geq 0$ is integer, and R_0 is some positive number.

(c) $f(z)$ has an asymptotic expansion $f(z) \sim \sum a_n z^n$ valid uniformly in $\arg(z)$.

(d) Imf has limits above the cut along the negative real axis and Im$f \geq 0$ there.

(e) $\sum_{n=k}^{\infty} a_n^{-1/2n} = \infty$.

Then the series $\sum_{n=k}^{\infty} a_n z^n$ is Stieltjes summable to the function

$$\bar{f}(z) = \frac{f(z) - \sum_{n=0}^{k-1} a_n z^n}{(-z)^k}.$$

This implies that for any $j \geq k-1$ the diagonal Padé approximants sequences $f^{[N,N+j]}(z)$ converge to $f(z)$ as $N \to \infty$, uniformly in any compact of the cut z plane.

We shall consider only L_1^H, because for L_1^E the proof goes exactly in the same way.

(a) From (2.4) we have

$$L_1^H(\alpha) = -\frac{1}{8\pi^2}\int_0^{\infty} \frac{e^{-s}}{s^3} \times\left(2\sqrt{\pi}\sqrt{\alpha}\, s \coth(2\sqrt{\pi}\sqrt{\alpha}s) - 1 - \frac{4\pi}{3}\alpha s^2\right) ds. \quad (2.8)$$

This formula clearly displays the analyticity of L_1^H as a function of α, as long as $\sqrt{\alpha}$ is not pure imaginary, i.e., for α not on the negative real axis.

(b) To find the behavior of $L_1^H(\alpha)$ as $\alpha \to \infty$ in any direction of the complex plane, let us write, performing in (2.8) the substitution $s\sqrt{\alpha} \to x$,

$$L_1^H(\alpha) = -\frac{1}{8\pi^2}\alpha[I_1(\alpha) + I_2(\alpha)], \quad (2.9)$$

where

$$I_1(\alpha) = \int_0^1 \frac{e^{-x/\sqrt{\alpha}}}{x^3}\left(2\sqrt{\pi}\, x \coth(2\sqrt{\pi}x) - 1 - \frac{4\pi}{3}x^2\right) dx, \quad (2.10a)$$

$$I_2(\alpha) = \int_1^{\infty} \frac{e^{-x/\sqrt{\alpha}}}{x^3}\left(2\sqrt{\pi}\, x \coth(2\sqrt{\pi}x) - 1 - \frac{4\pi}{3}x^2\right) dx. \quad (2.10b)$$

Now it is evident from (2.10a) that $I_1(\alpha) \to$ const as $\alpha \to \infty$ in any direction within the cut plane, and from (2.10b) that $I_2(\alpha) \to -4\pi/3\, Ei(1/\sqrt{\alpha}) \sim (2\pi/3)\log\alpha$. To find the behavior as $\alpha \to -\infty$, we perform the analytic continuation of $L_1^H(\alpha)$ by means of $\alpha - \pi/2$ rotation of the integration path in (2.8), i.e., we put $s = it$:

$$L_1^H(\alpha) = \frac{1}{8\pi^2 i}\int_0^{\infty} \frac{e^{-it}}{t^3} \times\left(2\sqrt{\pi}\sqrt{\alpha}\, it \coth(2\sqrt{\pi}\sqrt{\alpha}it) - 1 + \frac{4\pi}{3}\alpha t^2\right) dt \quad (2.11)$$

thus obtaining the analytic continuation of $L_1^H(\alpha)$ defined by (2.8) as an analytic function of α in the first sheet $-\pi < \alpha < \pi$, up to $\arg\alpha = -3\pi/2$.

The behavior of (2.11) for $\alpha \to -\infty$ can be discussed in exactly the same way as that of (2.8). Hence we can conclude that $|L_1^H(\alpha)| < A|\alpha \log\alpha|$, as $\alpha \to \infty$ in any direction of the complex plane, A being some positive constant, requiring $k = 2$ for this case.

(c) That the series (2.6) is the uniform asymptotic expansion of $L_1^H(\alpha)$ for $|\arg\alpha| < \pi$ follows trivially from its Borel summability, and it is also evident from (2.8).

(d) We may compute explicitly from (2.8) the discontinuity across the cut, so that we have

$$\mathrm{Im}L_1^H(\alpha) = \frac{1}{2i}\,\mathrm{disc}L_1^H(\alpha) = \frac{|\alpha|}{\pi}\sum_{n=1}^{\infty}\frac{1}{n^2}e^{-n\pi/2\sqrt{\pi}|\sqrt{\alpha}|} > 0. \quad (2.12)$$

(e) Since $|B_n| \sim (2n)!$ as $n \to \infty$, the condition is fulfilled.

We can thus conclude that all the $[N, N+j]$, $j \geq 1$, sequences of Padé approximants to the divergent series (2.6) converge to $L_1^H(\alpha)$, uniformly in any compact of the cut α plane, since the series

$$-\frac{1}{8\pi^2}\sum_{n=0}^{\infty}\frac{B_{2(n+2)}(8\pi)^{2(n+2)}}{[2(n+2)]!}(2n+1)!\alpha^n$$

is Stieltjes summable to $L_1^H(\alpha)/\alpha^2$.

For $L_1^E(\alpha)$ the proof goes exactly in the same way, with some obvious modifications to account for the fact that $L_1^E(\alpha)$ is analytic in the α plane cut along the positive real axis, as it is evident from (2.5). In this case then the uniform convergence of the $[N, N+j]$, $j \geq 1$ sequences of Padé approximants will take place in any compact of the α plane cut along the positive real axis.

3. PADÉ APPROXIMANTS ON THE BOREL TRANSFORM

Let us first express Ogievetsky's result[6] on Borel summability in the notations of Ref. 7. From (2.6), dividing by α^2, we get the divergent expansion

$$\frac{L_1^H(\alpha)}{\alpha^2} \sim -\frac{1}{8\pi^2}\sum_{n=0}^{\infty}\frac{(8\pi)^{2(n+2)}}{[2(n+2)]!}B_{2(n+2)}(2n+1)!\alpha^n. \quad (3.1)$$

The Borel transform of order 2 of the above series is defined to be

$$F(\alpha) = -\frac{1}{8\pi^2}\sum_{n=0}^{\infty}\frac{(8\pi)^{2(n+2)}}{[2(n+2)]!}B_{2(n+2)}\alpha^n \quad (3.2)$$

and has radius of convergence π. Since it is easy to obtain

$$F(\alpha) = -\frac{1}{8\pi^2\alpha}\left(2\sum_{n=1}^{\infty}\frac{1}{\alpha + n^2\pi^2} - \frac{1}{3}\right), \quad (3.3)$$

we see that $F(\alpha)$ is a meromorphic function, regular in the whole α plane except for simple poles along the negative real axis at $\alpha_n = -n^2\pi^2$, $n = 1, 2, \cdots$ with positive residues. Then, the integral $\int_0^\infty e^{-\sqrt{x}} F(\alpha x)dx$, which defines the Borel sum of (3.1), represents an analytic function in the whole α plane cut along the negative real axis. Taking into account the expansion

$$\coth(z) = \frac{1}{z} + 2z \sum_{n=1}^{\infty} \frac{1}{z^2 + n^2\pi^2},$$

it is trivial to show that the Borel sum actually is $L_1^H(\alpha)/\alpha^2$, i.e.,

$$\frac{L_1^H(\alpha)}{\alpha^2} = \int_0^\infty e^{-\sqrt{x}} F(\alpha x)dx. \tag{3.4}$$

Taking, as in Ref. (7), the Padé approximants $F^{[N,N+j]}(z)$ on the Borel transform, we define the approximants to $L_2^H(\alpha)/\alpha^2$ as

$$f_B^{[N,N+j]}(\alpha) = \int_0^\infty e^{-\sqrt{x}} F^{[N,N+j]}(\alpha x)dx. \tag{3.5}$$

We have to prove now that

$$\lim_{N\to\infty} f_B^{[N,N+j]}(\alpha) = L_1^H(\alpha)/\alpha^2, \quad j \geq -1, \tag{3.6}$$

uniformly in any compact subset of the cut α plane.

Proof: As usual, we have to prove (3.6) only for $j = -1$, because for $j > -1$ one can apply the same procedure to the function

$$M_1^H(\alpha) = \left[\frac{L_1^H(\alpha)}{\alpha^2} - \sum_{n=0}^{j+1} a_n(-\alpha)^n\right] / (-\alpha)^{j+1}$$

whose $[N, N-1]$ approximants are the $[N, N+j]$ ones of $L_1^H(\alpha)/\alpha^2$. As far as the convergence of $F^{[N,N+j]}(\alpha)$ to $F(\alpha)$ is concerned, one has

$$\lim_{N\to\infty} F^{[N,N+j]}(\alpha) = F(\alpha), \quad j \geq 1, \tag{3.7}$$

uniformly in any compact containing none of its poles, because $F(\alpha)$ is meromorphic in the whole complex α plane having only simple poles with positive residues along the negative real axis.[13]

Now for α in the cut plane the boundedness of $|F(\alpha)|$ as $|\alpha| \to \infty$, as well as that of $F^{[N,N-1]}(\alpha)$ for any N are clearly sufficient to justify in (3.5) the interchange of the limit $N \to \infty$ with the integral, in spite of the non-uniformity of the convergence of $F^{[N,N-1]}(\alpha)$ to $F(\alpha)$ at infinity. We can thus conclude

$$\lim_{N\to\infty} f_B^{[N,N+j]}(\alpha) = L_1^H(\alpha)/\alpha^2, \quad j \geq 1, \tag{3.8}$$

uniformly in any compact of the cut α plane, so that we have for $L_1^H(\alpha)$

$$\lim_{N\to\infty} f_B^{[N,N+j]}(\alpha) = L_1^H(\alpha), \quad j \geq -1, \tag{3.9}$$

uniformly in any compact of the cut α plane. In this case also, the proof for $L_1^H(\alpha)$ goes exactly in the same way.

In addition, it may also be noted that the first Borel transform of $L_1^H(\alpha)/\alpha^2$, defined by the divergent expansion

$$-\frac{1}{8\pi^2} \sum_{n=0}^{\infty} \frac{(8\pi)^{2(n+2)}}{[2(n+2)]!} B_{2(n+2)} \frac{(2n+1)!}{n!} \alpha^n, \tag{3.10}$$

is Stieltjes summable. It is indeed easy to see that (3.10) is the divergent Taylor expansion near $\alpha = 0$ of the function

$$F_1(\alpha) = \frac{2}{\sqrt{\pi}} \int_0^\infty \sqrt{x}\, e^{-x} F(4\alpha x)dx \tag{3.11}$$

[$F(\alpha)$ being defined by (3.3) and its Taylor expansion near $\alpha = 0$ by (3.2)], so that to prove its Stieltjes summability we have only to proceed in exactly the same way as in Sec. 2. This result implies the convergence to the solution also of the generalized Padé approximants $f_1^{[N,N+j]}(\alpha)$, as defined in Refs. 14 and 15, for $j \geq 1$ and uniformly in any compact of the cut plane.

As far as the rate of convergence of the approximants $f_B^{[N,N+j]}(\alpha)$ and $f_1^{[N,N+j]}(\alpha)$, is concerned, as well as their monotonicity properties, the reader is referred to the discussion given for the case of the anharmonic oscillators.[14]

ACKNOWLEDGMENT

It is a pleasure to thank Professor W. Zimmermann for his interest in this work and for reading the manuscript.

---

*Fulbright Fellow, on leave from Istituto Nazionale di Fisica Nucleare, Sezione di Bologna, Italy.

†This work supported in part by funds from the National Science Foundation.

[1] C. M. Bender and T. T. Wu, Phys. Rev. **184,** 1231 (1969).
[2] A. M. Jaffe, Commun. Math. Phys. **1,** 127 (1965). This paper also contains references to earlier work by C. A. Hurst, W. Thirring, and A. Petermann.
[3] E. Borel, *Leçons sur les series divergentes* (Gauthier-Villars, Paris, 1928); H. Wall, *Continued Fractions* (Van Nostrand, New York, 1947); G. H. Hardy, *Divergent Series* (Oxford U.P., Oxford, 1948).
[4] J. J. Loeffel, A. Martin, B. Simon, and A. S. Wightman, Phys. Lett. B **30,** 656 (1969).
[5] G. A. Baker and J. S. R. Chisholm, J. Math. Phys. **7,** 1900 (1966).
[6] V. I. Ogievesky, Dokl. Akad. Nauk SSSR **109,** 919 (1956).
[7] S. Graffi, V. Grecchi, and B. Simon, Phys. Lett. B **32,** 631 (1970).
[8] B. Simon, Phys. Rev. Lett. **25,** 1583 (1970).
[9] J. Gunson and P. H. Ng, University of Birmingham preprint, July 1971.
[10] J. Schwinger, Phys. Rev. **82,** 664 (1951).
[11] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1965).
[12] B. Simon, Ann. Phys. (N.Y.) **58,** 76 (1970), Lemma IV. 2.2.
[13] H. S. Wall, *Continued Fractions*, Chap. XVIII.
[14] S. Graffi, V. Grecchi, and G. Turchetti, Nuovo Cimento B **4,** 313 (1971).
[15] G. A. Baker, Phys. Rev. **161,** 434 (1967).

Stochastic symmetry breaking of time reversal invariance

Ronald Forrest Fox

School of Physics, Georgia Institute of Technology, Atlanta, Georgia
(Received 7 February 1973)

An equation for the time evolution of the density matrix for a nonrelativistic quantum mechanical system is presented. It involves a Hamiltonian which contains a stochastic contribution. Before stochastic averaging is performed, it is shown that the density matrix equation is time reversal invariant, whereas after stochastic averaging is performed, it is shown that the averaged density matrix equation is not time reversal invariant.

In a series of three papers[1,2,3] a new approach to the theory of non-equilibrium statistical mechanics has been presented. The approach is based upon the mathematical theory of multiplicative stochastic processes.[1] Within this context it has been possible to derive an averaged density matrix equation which shows time irreversible behavior,[1] to prove an H-theorem,[2] and to characterize the approach to equilibrium for microcanonical and canonical ensembles.[3] In this paper, the relationship between time reversible and time irreversible behavior will be considered from the viewpoint of the time reversal operation of quantum mechanics.

The Schrödinger–Heisenberg picture of quantum mechanics may be expressed by the equation

$$i\frac{d}{dt}C_\alpha(t) = \sum_\beta M_{\alpha\beta}C_\beta(t), \tag{1}$$

where $M^*_{\beta\alpha} = M_{\alpha\beta}$. In the following the repeated index summation convention will be used. When a stochastic contribution to the Hamiltonian is considered, (1) becomes

$$i\frac{d}{dt}C_\alpha(t) = M_{\alpha\beta}C_\beta(t) + \tilde{M}_{\alpha\beta}(t)C_\beta(t), \tag{2}$$

where $\tilde{M}^*_{\beta\alpha}(t) = \tilde{M}_{\alpha\beta}(t)$. The stochastic properties of $\tilde{M}_{\alpha\beta}(t)$ are those appropriate for a purely random, Gaussian stochastic matrix.[1] The first two averaged moments are given by

$$\langle\tilde{M}_{\alpha\beta}(t)\rangle = 0, \tag{3}$$

$$\langle\tilde{M}_{\alpha\beta}(t)\tilde{M}_{\mu\nu}(S)\rangle = 2Q_{\alpha\beta\mu\nu}\delta(t-s). \tag{4}$$

Equation (2) leads to a density matrix equation. The density matrix $\rho_{\alpha\beta}(t)$ is defined by

$$\rho_{\alpha\beta}(t) \equiv C^*_\alpha(t)C_\beta(t), \tag{5}$$

and (2) with (5) gives

$$i\frac{d}{dt}\rho_{\alpha\beta}(t) = L_{\alpha\beta\mu\nu}\rho_{\mu\nu}(t) + \tilde{L}_{\alpha\beta\mu\nu}(t)\rho_{\mu\nu}(t), \tag{6}$$

where $L_{\alpha\beta\mu\nu}$ and $\tilde{L}_{\alpha\beta\mu\nu}(t)$ are defined by

$$L_{\alpha\beta\mu\nu} \equiv \delta_{\alpha\mu}M_{\beta\nu} - \delta_{\beta\nu}M^*_{\alpha\mu}, \tag{7}$$

$$\tilde{L}_{\alpha\beta\mu\nu}(t) \equiv \delta_{\alpha\mu}\tilde{M}_{\beta\nu}(t) - \delta_{\beta\nu}\tilde{M}^*_{\alpha\mu}(t). \tag{8}$$

It has been proved[1] that stochastic averaging of (6) leads to

$$\frac{d}{dt}\langle\rho_{\alpha\beta}(t)\rangle = -iL_{\alpha\beta\mu\nu}\langle\rho_{\mu\nu}(t)\rangle - R_{\alpha\beta\mu\nu}\langle\rho_{\mu\nu}(t)\rangle, \tag{9}$$

where $R_{\alpha\beta\mu\nu}$ is defined by

$$R_{\alpha\beta\mu\nu} \equiv \delta_{\alpha\mu}Q_{\beta\theta\theta\nu} + \delta_{\beta\nu}Q_{\theta\alpha\mu\theta} - Q_{\beta\nu\mu\alpha} - Q_{\mu\alpha\beta\nu} \tag{10}$$

and $\sum_\alpha\langle\rho_{\alpha\alpha}(t)\rangle = 1$ for all t. It has also been proved,[2] on the basis of the properties of $R_{\alpha\beta\mu\nu}$, that if $H(t)$ is defined by

$$H(t) \equiv \mathrm{Tr}\,[\langle\rho(t)\rangle\log_e\langle\rho(t)\rangle], \tag{11}$$

then

$$\frac{d}{dt}H(t) \leq 0. \tag{12}$$

While (12) clearly demonstrates the time irreversibility of (9), the time reversibility of (6) remains to be demonstrated.

TIME REVERSAL INVARIANCE

If the Schrödinger equation for a single particle in a potential field is given by

$$ih\frac{\partial}{\partial t}\Psi(\mathbf{r}t) = \left(-\frac{h^2}{2m}\nabla^2 + V(\mathbf{r})\right)\Psi(\mathbf{r},t), \tag{13}$$

then the time reversal operation is accomplished by simultaneously complex conjugating every term in (13) and replacing all instances of t by $-t$.[4] Doing this to (13) gives

$$-i\hbar\frac{\partial}{-\partial t}\Psi^*(\mathbf{r},-t) = \left(-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r})\right)\Psi^*(\mathbf{r},-t). \tag{14}$$

It is seen that $\Psi^*(\mathbf{r},-t)$ also satisfies the original Schrödinger equation. This is the condition of time reversal invariance.

Suppose that a complete, orthonormal set of complex basis functions, $\varphi_k(\mathbf{r})$, are introduced satisfying the conditions

$$\int\varphi^*_k(\mathbf{r})\varphi_l(\mathbf{r})d\mathbf{r} = \delta_{kl}. \tag{15}$$

By introducing the definitions

$$\Psi(\mathbf{r},t) \equiv \sum_l C_l(t)\varphi_l(\mathbf{r}) \tag{16}$$

and

$$H_{kl} \equiv \int\varphi^*_k(\mathbf{r})\left[-\left(\frac{\hbar^2}{2m}\right)\nabla^2 + V(\mathbf{r})\right]\varphi_l(\mathbf{r})d\mathbf{r}, \tag{17}$$

Eq. (13) becomes, upon multiplication by $\varphi^*_k(\mathbf{r})$ and integration over all $\mathbf{r}$,

$$i\hbar\frac{d}{dt}C_k(t) = H_{kl}C_l(t). \tag{18}$$

Generally, (17) also implies $H^*_{lk} = H_{kl}$.

When applied to (18), the time reversal operation leads to

$$-i\hbar\frac{d}{-dt}C^*_k(-t) = H^*_{kl}C^*_l(-t). \tag{19}$$

Copyright © 1973 by the American Institute of Physics

This is the same as the result achieved from (14) by multiplication by $\varphi_k(\mathbf{r})$ followed by integration over all $\mathbf{r}$. The complex conjugate of (17) provides for the H^*_{kl} in (19).

Therefore, in matrix notation, the condition for time reversal invariance of (18) is that $C^*_k(-t)$ satisfies

$$i\frac{d}{dt}C^*_k(-t) = H^*_{kl}C^*_l(-t), \tag{20}$$

which is equivalent to (19). The presence of H^*_{kl} in (20) instead of H_{kl} should be noted, and reflects the antilinear character of the time reversal operation.

KUBO OSCILLATOR

Before proceeding to the consideration of the time reversal properties of (6), it is instructive to consider a simple, special case, the Kubo oscillator[1]:

$$\frac{d}{dt}a(t) = i\,[\omega_0 + \tilde{\varphi}(t)]a(t), \tag{21}$$

where $a(t)$ is complex and $\tilde{\varphi}(t)$ is a purely random, Gaussian stochastic frequency fluctuation with average value zero and second moment given by

$$\langle\tilde{\varphi}(t)\tilde{\varphi}(s)\rangle = 2\lambda\delta(t-s). \tag{22}$$

The oscillator equation (21) is a one-component special case of (18). Because the H_{kl} in (18) is Hermitian, in the one-component case $H_{kl} \to H$ which must be real. In (21) the analog of H is $\omega_0 + \tilde{\varphi}(t)$ which is also real.

The time reversal operation applied to (21) gives

$$-\frac{d}{dt}a^*(-t) = -i\,[\omega_0 + \tilde{\varphi}(-t)]a^*(-t). \tag{23}$$

If the $\tilde{\varphi}(t)$ in (21) and the $\tilde{\varphi}(-t)$ in (23) were not present, then time reversibility of (21) would be proved by (23). However, because of the presence of $\tilde{\varphi}(t)$ in (21) and $\tilde{\varphi}(-t)$ in (23), it is necessary that $\tilde{\varphi}(t) = \tilde{\varphi}(-t)$ in order for time reversal invariance to obtain.

If $\tilde{\varphi}(t)$ were an ordinary, nonstochastic function, then $\tilde{\varphi}(t) = \tilde{\varphi}(-t)$ would require that $\tilde{\varphi}(t)$ be an even function of t. It will be shown that $\tilde{\varphi}(t)$ is not an even function of t, but that $\tilde{\varphi}(t) = \tilde{\varphi}(-t)$ anyway, because of the stochastic properties of $\tilde{\varphi}(t)$.

Consider the time interval from $-T$ to $+T$. $\tilde{\varphi}(t)$ may be given a Fourier representation in the interval $(-T, +T)$ by[5]

$$\tilde{\varphi}(t) = \sum_k [\tilde{a}_k \cos(\omega_k t) + \tilde{b}_k \sin(\omega_k t)], \tag{24}$$

where $\omega_k = k\pi/T$, and the $\tilde{a}_k$'s and $\tilde{b}_k$'s are all independent, Gaussianly distributed stochastic coefficients. Now, from (24) it is seen that

$$\tilde{\varphi}(-t) = \sum_k [\tilde{a}_k \cos(\omega_k t) - \tilde{b}_k \sin(\omega_k t)]. \tag{25}$$

Therefore, the equality $\tilde{\varphi}(t) = \tilde{\varphi}(-t)$ requires that $\tilde{b}_k = -\tilde{b}_k$.

For ordinary, nonstochastic variables, $X = -X$ implies $X = 0$. For stochastic variables other possibilities exist. The $\tilde{b}_k$'s are completely characterized by their Gaussian distribution functions:

$$W(b_k) = (1/\sqrt{4\pi\lambda})\exp(b_k^2/4\lambda). \tag{26}$$

Two stochastic variables are equal if and only if their distribution functions are equal. Clearly, for Gaussian distributions such as (26) it follows that $\tilde{b}_k = -\tilde{b}_k$, without $\tilde{b}_k = 0$ having to hold. Therefore

$$\tilde{\varphi}(t) = \tilde{\varphi}(-t) \tag{27}$$

and using (27) in (23) leads to

$$\frac{d}{dt}a^*(-t) = i[\omega_0 + \tilde{\varphi}(t)]a^*(-t), \tag{28}$$

which is a special case of (20), and confirms time reversal invariance of (21).

It should be observed that if (21) is averaged,[1] the result is

$$\frac{d}{dt}\langle a(t)\rangle = i\omega_0\langle a(t)\rangle - \lambda\langle a(t)\rangle, \tag{29}$$

which is definitely not time reversible since the time reversal operation leads to

$$\frac{d}{dt}\langle a(-t)\rangle^* = i\omega_0\langle a(-t)\rangle^* + \lambda\langle a(-t)\rangle^* \tag{30}$$

and $\langle a(-t)\rangle^*$ does not satisfy the appropriate time reversed equation because of the dissipative λ-dependent term.

DENSITY MATRIX

The results for the Kubo oscillator may be straightforwardly generalized for the consideration of the density matrix as follows. The time reversal operation, when applied to (6), leads to

$$-i\frac{d}{-dt}\rho^*_{\alpha\beta}(-t) = L^*_{\alpha\beta\mu\nu}\rho^*_{\mu\nu}(-t) + \tilde{L}^*_{\alpha\beta\mu\nu}(-t)\rho^*_{\mu\nu}(-t). \tag{31}$$

If the stochastic terms in (6) and (31) are omitted, consideration of (5), (7), and (20) verifies time reversal invariance. However, the presence of the stochastic terms requires that

$$\tilde{L}^*_{\alpha\beta\mu\nu}(-t) = \tilde{L}^*_{\alpha\beta\mu\nu}(t) \tag{32}$$

in order for time reversal invariance to be satisfied. This is the density matrix analog of (27).

In order to check the validity of (32), (8) may be used to see that (32) is equivalent to

$$\tilde{M}_{\alpha\beta}(-t) = \tilde{M}_{\alpha\beta}(t). \tag{33}$$

Therefore, it remains to prove the validity of (33). Again consider the time interval $(-T, +T)$. A Fourier representation for $\tilde{M}_{\alpha\beta}(t)$ is possible:

$$\tilde{M}_{\alpha\beta}(t) \equiv \sum_k [\tilde{a}^k_{\alpha\beta}\cos(\omega_k t) + \tilde{b}^k_{\alpha\beta}\sin(\omega_k t)], \tag{34}$$

where $\omega_k = k\pi/T$, and the $\tilde{a}_{\alpha\beta}$'s and $\tilde{b}_{\alpha\beta}$'s are Gaussianly distributed stochastic matrix coefficients satisfying the conditions

$$\langle\tilde{a}^k_{\alpha\beta}\tilde{b}^l_{\mu\nu}\rangle = 0 \quad \text{for all } k, l, \alpha, \beta, \mu, \text{and } \nu, \tag{35}$$

$$\langle\tilde{a}^k_{\alpha\beta}\tilde{a}^l_{\mu\nu}\rangle = 2\delta_{kl}Q_{\alpha\beta\mu\nu}, \qquad \langle\tilde{b}^k_{\alpha\beta}\tilde{b}^l_{\mu\nu}\rangle = 2\delta_{kl}Q_{\alpha\beta\mu\nu},$$

These conditions lead to (4). Returning to (34) we see that (33) will be satisfied if and only if

$$\tilde{b}^k_{\alpha\beta} = -\tilde{b}^k_{\alpha\beta}, \tag{36}$$

which is the analog of $\tilde{b}_k = -\tilde{b}_k$ in the Kubo oscillator case. Equation (36) is true because the $\tilde{b}^k_{\alpha\beta}$'s are Guassianly distributed. Therefore, it has been demonstrated that (31) leads to

$$i\frac{d}{dt}\rho^*_{\alpha\beta}(-t) = L^*_{\alpha\beta\mu\nu}\rho^*_{\mu\nu}(-t) + \tilde{L}^*_{\alpha\beta\mu\nu}(t)\rho^*_{\mu\nu}(-t), \qquad (37)$$

which is the time reversal invariant equation corresponding with (6). Application of the time reversal operation to (9), however, leads to time reversal noninvariance because of the dissipative $R_{\alpha\beta\mu\nu}$ term in (9). Of course, the time irreversible nature of (9) is already evident in (12).

SUMMARY

It has been proved that the addition of a purely random, Guassian stochastic contribution to the Hamiltonian of a quantum mechanical system leads to a density matrix equation which is time reversal invariant if stochastic averaging is not performed, but which is time reversal noninvariant if stochastic averaging is performed. The key reason behind this result is the special property of Gaussian stochastic variables that they are equal to their own negatives without having to be zero. This follows from consideration of the distribution function for a Gaussian stochastic variable.

The implications of this theory for nonequilibrium statistical mechanics have already been discussed in earlier publications.[1,2,3] Whether or not stochastic symmetry breaking of time reversal invariance has significance in relativistic quantum mechanics and particle physics invites consideration.

---

[1]R. F. Fox, J. Math. Phys. **13**, 1196 (1972).
[2]R. F. Fox, J. Math. Phys. **13**, 1726 (1972).
[3]R. F. Fox, J. Math. Phys. **14**, 20 (1973).
[4]E. P. Wigner, *Group Theory* (Academic, New York, 1959), Chap. 26.
[5]M. C. Wang and G. E. Uhlenbeck, Rev. Mod. Phys. **17**, 323 (1945), Sec. 6.

Eigenvalues of λx^{2m} anharmonic oscillators

S. N. Biswas, K. Datta, R. P. Saxena, P. K. Srivastava, and V. S. Varma

Department of Physics and Astrophysics, University of Delhi, Delhi-7, India
(Received 1 November 1971; revised manuscript received 4 January 1972)

The ground state as well as excited energy levels of the generalized anharmonic oscillator defined by the Hamiltonian $H_m = -d^2/dx^2 + x^2 + \lambda x^{2m}$, $m = 2,3,\cdots$, have been calculated nonperturbatively using the Hill determinants. For the λx^4 oscillator, the ground state eigenvalues, for various values of λ, have been compared with the Borel–Padé sum of the asymptotic perturbation series for the problem. The agreement is excellent. In addition, we present results for some excited states for $m = 2$ as well as the ground and the first even excited states for $m = 3$ and 4. The behaviour of all the energy levels with respect to the coupling parameter shows a qualitative similarity to the ground state of the λx^4 oscillator. Thus the results are model independent, as is to be expected from the WKB approximation. Our results also satisfy the scaling property that $\epsilon_n^{(m)}(\lambda)/\lambda^{1/(m+1)}$ tend to a finite limit for large λ, and always lie within the variational bounds, where available.

1. INTRODUCTION

Recently there has been a great deal of interest in the analytical as well as numerical study of the one-dimensional anharmonic oscillator of the type λx^{2m} $^{1-7}$ (m being a positive integer). In particular, the analyticity with respect to the coupling constant λ of the energy levels of the Hamiltonian $p^2 + x^2 + \lambda x^4$ has been studied quite exhaustively.[2,4] Interest in this type of investigation stems from the belief that the nature of solutions of such a Hamiltonian may lead to a fuller understanding of an equivalent one-dimensional model Hamiltonian in field theory.[2] In addition, it is well known that the knowledge of the exact eigenvalues of a λx^4-anharmonic potential is of particular interest in molecular physics. We will be concerned in this paper with the determination of the energy levels of arbitrary λx^{2m} anharmonic oscillators.

To obtain the energy levels for Hamiltonians which are not exactly solvable, one has to use some approximate scheme such as a variational method or the techniques of perturbation theory. As is well known, rigorous upper bounds on the energy levels can be determined using a Rayleigh–Ritz variational method with a proper orthonormal set of trial wavefunctions. An interesting method for the determination of lower bounds of these levels has been given by Bazley and Fox.[8,9] Their procedure is to construct a set of exactly solvable intermediate Hamiltonians $H^k (k = 0,1,2,\ldots)$ such that $H^0 < H^1 < H^2 \cdots < H$ (H being the exact Hamiltonian). The determination of the eigenvalues of the successive H^k lead sequentially to the rigorous lower bounds to the eigenvalues of H. In particular Bazley and Fox have calculated the lower and upper bounds for the first five energy levels of a λx^4-anharmonic oscillator. Their variational calculation is based upon a five-parameter trial wavefunction of the type

$$\sum_{n=0}^{4} C_n H_n(x) \exp(-x^2/2),$$

where $H_n(x)$ are the Hermite polynomials.

A perturbative calculation of the λx^4-energy levels, on the other hand, gives rise to a singular perturbation series.[10] From an exhaustive numerical analysis of the perturbation series for the ground-state energy level of the one dimensional anharmonic oscillator, Bender and Wu[2] have shown that the power series in λ is divergent for all λ though each term of the series is finite. Further, they conclude that the energy level for the system originally defined for real positive λ can be analytically continued into the complex λ plane and that the continuation has an infinite number of branch points with a limit point at $\lambda = 0$. Simon[4] has, however, pointed out that the conclusions arrived at by Bender and Wu are based upon "arguments of doubtful validity". He has studied the analytic properties of the singular perturbation theory for the Hamiltonian $p^2 + x^2 + \lambda x^4$ and proved rigorously several of the properties of the energy levels previously found empirically by Bender and Wu. He has shown in particular that the nth energy level $\epsilon_n^{(2)}(\lambda)$ has a third order branch point at $\lambda = 0$; further that $\lambda = 0$ is not the only singularity of $\epsilon_n^{(2)}(\lambda)$ on the three sheeted surface—rather there are infinitely many singularities. Such perturbation series are quite common in relativistic quantum mechanics and the usual belief is that they are asymptotic in nature. It is well known in mathematical literature that such series can often be summed uniquely through the use of summability techniques such as the Stieltjes–Padé or the Borel methods. Simon has investigated the anharmonic oscillator with the general anharmonic term $\lambda x^{2m}, m$ integer and > 0 and has shown that the nth energy level is analytic in a certain region of the λ plane and that the perturbation series is asymptotic to the value $\epsilon_n^{(m)}(\lambda)$. In certain cases, one may obtain the analytic continuation by appropriate manipulations on the power series. The main trouble is that one does not always know the location of all the singular points of the function being studied. This trouble could be avoided if we could introduce a sequence of approximants to the function which are invariant under the group of homographic transformations. One would expect such a sequence to converge at least as well as the best power series obtainable for the function. The sequence of $[N,M]$ Padé approximants has the property that it is invariant under the above mentioned group of homographic transformations. In general, a Padé approximation consists in replacing a power series by a sequence of rational functions of the form of a polynomial (of degree M) divided by another polynomial (of degree N). Loeffel *et al.*[3] have proved that the perturbation series for the energy level of one-dimensional λx^4 and λx^6-oscillators sums under Padé approximations to the actual level. Their proof, however, is not known for oscillators of the type $\lambda x^{2m}, m > 3$. Simon[4] has calculated $\epsilon_0^{(2)}(\lambda)$ by converting the perturbation series into a series of Padé approximants for various values of λ. In a recent communication, Graffi *et al.*[5] have shown how improved values of this ground state energy level for arbitrary λ can be obtained by using Padé approximants to the Borel transform of the asymptotic perturbation series. In essence, their method consists in replacing the series

Copyright © 1973 by the American Institute of Physics

$\sum_{n=0}^{\infty} a_n z^n$ by the Borel sum $\int_0^\infty e^{-t}\phi(tz)dt$,

where $\phi(z) = \sum_{n=0}^{\infty}(a_n z^n/n!)$. Within their region of convergence both the series are identical, but for values of z for which the series $\sum a_n z^n$ diverges, the integral representation gives the value of the series provided the integral exists for that z. To facilitate numerical computation, Graffi *et al.* used Padé approximants for $\phi(tz)$.

We wish to point out that in addition to those described above, there are other methods which may be used to obtain approximate wavefunctions and energy levels; in particular, the method of difference equations may be mentioned. It is well known that only under special circumstances does the substitution of an infinite series for the wavefunction lead to a two term difference equation for the coefficients, but a technique can be developed which permits the numerical computation of the exact energy levels even when a three term difference equation is obtained. We have already applied this method[11] to the calculation of the ground state eigenvalues and eigenfunction of the λx^4 anharmonic oscillator. In the present work, we consider in addition the application of this method to the calculation of the ground state and excited state eigenvalues of the general λx^{2m} anharmonic oscillator. We will discuss this method in Sec. 2 and apply the techniques to the λx^4 anharmonic oscillator. In the various accompanying tables we present our numerical results. In Table I we give the ground state energy levels from $\lambda = 0.1$ to $\lambda = 100$, and compare them with the results of the Borel–Padé and of the variational calculations. Table II presents the corresponding results for the first seven excited levels of the λx^4 oscillator. These results are displayed in Fig. 1, where these eigenvalues have been plotted as a function of λ. An important feature of our work is that for large λ our eigenvalues satisfy the condition that $\epsilon_n^{(2)}(\lambda)/\lambda^{1/3}$ tends to a finite limit as is expected from the scaling property of the λx^4 anharmonic oscillator Hamiltonian.[4] It is worth mentioning that both the Padé as well as the modified Borel–Padé approximants have the defect that $\epsilon_0^{(2)}(\lambda)$ becomes constant for large λ.[5] We have also verified that our results always lie within the bounds obtained by Bazley and Fox,[8] and that our results for $\lambda = 100$ are within 1% of the infinite λ limit results for the λx^4 oscillator obtained by Schiff[12] and Chasman[13] (see Table III).

In Sec. 3 the generalization of our technique to arbitrary λx^{2m} oscillators have been discussed, and in Table IV we present the ground-state and the first even excited state energy levels of the λx^{2m} oscillator for $m = 3$ and 4. These results show that all the excited energy levels behave as

$$\epsilon_n^{(m)}(\lambda) \sim (n + \tfrac{1}{2})^{2m/(m+1)}\lambda^{1/(m+1)}$$

as is to be expected from a straight forward application of the WKB approximation and Bohr quantization rule. In Fig. 2 we have displayed the behavior of the ground state energy levels for $m = 2, 3$ and 4.

In Sec. 4 the normalization and the overlap of the eigenfunctions of the λx^4 oscillator have been discussed.

2. EIGENVALUES FROM THE THREE TERM DIFFERENCE EQUATION FOR THE λx^4 OSCILLATOR

In this section we discuss how exact values of the energy levels of the anharmonic oscillator can be obtained without any recourse to the standard perturbation series and associated summability techniques. Our approach is essentially nonperturbative and is based upon solving the Hill determinant for finding eigenvalues.

For simplicity we examine initially the level shifts of the even partity states and consider the differential equation $H\psi = \epsilon\psi$, where

$$H = -\frac{d^2}{dx^2} + x^2 + \lambda x^4. \tag{1}$$

We now make the ansatz

$$\psi = \exp(-x^2/2)\sum_{n=0}^{\infty} c_n x^{2n}. \tag{2}$$

Substituting this ansatz in the differential equation, we find that the c_n satisfy the following three term difference equation:

$$2(n+1)(2n+1)c_{n+1} + (\epsilon - 1 - 4n)c_n - \lambda c_{n-2} = 0. \tag{3}$$

The condition that a nontrivial solution for the c_n exist is given by the vanishing of the following infinite determinant

$$D = \begin{vmatrix} \epsilon-1 & 2 & 0 & 0 & \cdot & \cdot & \cdot & \cdot & \cdots \\ 0 & \epsilon-5 & 12 & 0 & \cdot & \cdot & \cdot & \cdot & \cdots \\ -\lambda & 0 & \epsilon-9 & 30 & \cdot & \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdot & -\lambda & 0 & (\epsilon-1-4n) & 2(n+1)(2n+1) & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \end{vmatrix}. \tag{4}$$

This is the well-known Hill determinant of the problem. Our eigenvalue problem is in principle solved; now we only need to find the roots of $D = 0$, to obtain the allowed energy levels of the anharmonic oscillator, and substitute these values in the set of Eqs. (3) to evaluate the coefficients c_n and obtain the wave functions. The lowest root will correspond to the ground state energy level and the excited levels will be given by the sequence of higher roots.

It is interesting to note that if D_n stands for the $(n+1) \times (n+1)$ approximant to D, then the D_n satisfy the following difference equation:

$$D_n = (\epsilon - 1 - 4n)D_{n-1} - 16\lambda n(n-\tfrac{1}{2})(n-1)(n-\tfrac{3}{2})D_{n-3}. \tag{5}$$

To facilitate numerical calculations we extract out the full asymptotic behaviour of D_n, writing

$$D_n = (-1)^n 6^{4n/3}\lambda^{n/3} \Gamma\left(\frac{n}{3}+\frac{1}{2}\right)\Gamma\left(\frac{n}{3}+\frac{2}{3}\right)\Gamma\left(\frac{n}{3}+\frac{5}{6}\right)\Gamma\left(\frac{n}{3}+1\right)P_n. \quad (6)$$

The recurrence relation satisfied by P_n is then

$$P_n = (4n + 1 - \epsilon)\alpha_n P_{n-1} + P_{n-3}, \quad (7)$$

where

$$\alpha_n = \frac{1}{6(6\lambda)^{1/3}} \frac{\Gamma\left(\frac{n}{3}+\frac{1}{6}\right)\Gamma\left(\frac{n}{3}+\frac{1}{3}\right)}{\Gamma\left(\frac{n}{3}+\frac{5}{6}\right)\Gamma\left(\frac{n}{3}+1\right)}. \quad (8)$$

Equation (7) is the basis of our numerical analysis for the determination of the eigenvalues. The eigenvalues are the zeroes of D_n, i.e., of P_n in the limit $n \to \infty$. The lowest root of D will correspond to the ground state energy level and the various excited energy levels will be given by the sequence of higher roots. To determine the energy levels we therefore need to obtain the roots of characteristic polynomials associated with the determinants P_n for large n. Equation (7) affords a very simple procedure for generating characteristic polynomials of all higher degrees in a recursive way. In this connection we would like to point out some interesting properties of the characteristic polynomials P_n:

(i) In a characteristic polynomial of any given order the coefficients of successive powers of ϵ alternate in sign, showing that there are no real negative eigenvalues.

(ii) Near the lowest root the derivatives of the characteristic polynomials P_{n-1} and P_{n-3} for large n and $\lambda > 0$ are of the same sign. Hence from Eq. (7) we can conclude that for sufficiently large n, the nth order characteristic polynomial P_n will have a zero between the zeroes of P_{n-1} and P_{n-3}, showing that the lowest root of P_n will stabilize as $n \to \infty$. Similar arguments can be used to establish the stability of all higher roots.

We now search for this stable root numerically by computing successively the lowest zeroes of the sequence of polynomials $\cdots P_n, P_{n+1}, P_{n+2}, \cdots$. For small values of λ, the anharmonicity parameter, the stability sets in at comparatively low order polynomials and the energy level is given by the first few terms of the asymptotic perturbation series. In Table I we compare our stabilized values of the ground state energy level for λ between 0.1 and 1.0 with the calculations of Graffi *et al.*,[5] who used the Padé approximants to the Borel sum of the perturbation series, and with the variational bounds calculated by Bazley and Fox.[8] The agreement of our results with the Borel–Padé results is remarkable, as is evident from the tables. In addition, we list the ground state eigenvalues for λ between 1 and 100.

In Table II we present results which are entirely new—the energy eigenvalues for the first seven excited states of the oscillator for values of λ lying between 0.1 and 100, comparing them with the variational bounds whenever these are available. We note that the equations for the odd parity eigenstates can be obtained from those for the even parity ones by the simple replacement $n \to n + \frac{1}{2}$. The behavior of the energy of the excited states as a function of λ is qualitatively similar to that of the ground state as is evident from the plot in Fig. 1. For all the excited states we have studied, we find no evidence of any level crossing. Further, they all obey the scaling law of the λx^4 oscillator, i.e., $\epsilon_n^{(2)}(\lambda)/\lambda^{1/3} \to$ const. for large λ.

We list these ratios $(\epsilon_n^{(2)}(\lambda)/\lambda^{1/3})$ from our results for $\lambda = 100$ for n, the excitation level, between 0 and 7 in Table III. In the same table we also list the excited to ground state energy ratios $(\epsilon_n^{(2)}/\epsilon_0^{(2)})$ for $\lambda = 100$ and compare them with these ratios obtained by Chasman[13] by extending the techniques of Heisenberg matrix mechanics and applying it to study the general equation $H = p^2/2m + \alpha q^n/n$. This would correspond to the large λ limit of the anharmonic oscillator when the anharmonic term would swamp out the harmonic term altogether. The agreement to within 1% of his results with ours demonstrates the rapid convergence of his approximation scheme. A similar result was also obtained earlier by Schiff[12] in his discussion of the lattice space quantization of a $\lambda\phi^4$ theory.

TABLE I: Ground state energy levels of the λx^4 anharmonic oscillator for values of λ between 0.1 and 100. $\epsilon_0^{(2)}$ are the results of the present calculation, ϵ_{BP} are the results of the Borel–Padé method (see Ref. 5) and ϵ_{oviar} are the variational bounds obtained by Bazley and Fox (see Ref. 8).

λ	$\epsilon_0^{(2)}$	ϵ_{BP}	ϵ_{oviar}
0.1	1.065 285 509 543 71	1.065 285 509 543 70	1.065 286 1.065 278
0.2	1.118 292 654 367 03	1.118 292 654 35(85)	1.118 293 1.118 255
0.3	1.164 047 157 353 84	1.164 047 157 0(754)	1.164 055 1.163 987
0.4	1.204 810 327 372 49	1.204 810 324 (7674)	1.204 848 1.204 738
0.5	1.241 854 059 651 49	1.241 854 04(6 6782)	1.241 957 1.241 746
0.6	1.275 983 566 342 55	1.275 983 5(21 8545)	1.276 195 1.275 773
0.7	1.307 748 651 120 03	1.307 748 5(31 5493)	1.308 110 1.307 324
0.8	1.337 545 208 148 17	1.337 544 9(37 0465)	1.338 096 1.336 760
0.9	1.365 669 825 784 43	1.365 669 2(83 1623)	1.366 442 1.364 349
1.0	1.392 351 641 530 29	1.392 350 (653 6791)	1.393 371 1.390 301

λ	$\epsilon_0^{(2)}$	λ	$\epsilon_0^{(2)}$
2	1.607 541 302 468 54	20	3.009 944 815 557 78
3	1.769 588 844 280 39	30	3.410 168 532 636 82
4	1.903 136 945 459 00	40	3.731 391 602 053 10
5	2.018 340 649 365 31	50	4.003 992 768 277 62
6	2.120 532 929 394 27	60	4.243 081 446 423 64
7	2.212 914 211 174 15	70	4.457 408 192 303 19
8	2.297 577 828 252 07	80	4.652 551 847 306 33
9	2.375 978 549 783 10	90	4.832 314 406 233 05
10	2.449 174 072 118 38	100	4.999 417 545 137 58

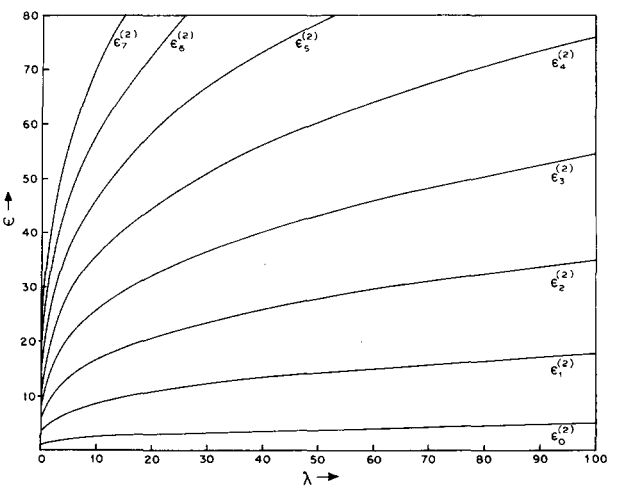


FIG. 1. The ground and excited energy levels $\epsilon_n^{(2)}$ of the λx^4 oscillator as a function of the anharmonicity parameter λ.

TABLE II(a): The even excited energy levels for the λx^4 oscillator along with the variational bounds of Bazley and Fox (see Ref. 8) for λ between 0.1 and 1.0.

λ	$\epsilon_{2_{var}}$	$\epsilon_2^{(2)}$	$\epsilon_{4_{var}}$	$\epsilon_4^{(2)}$	$\epsilon_{6_{var}}$	$\epsilon_6^{(2)}$
0.1	5.748 178		11.100 38		17.515 24	
		5.747 9592		11.098 5956		16.954 7946
	5.746 596		10.953 33		16.172 79	
	6.278 820		12.480 16		21.873 39	
0.2		6.277 2486		12.440 6018		19.315 6799
	6.260 404		12.225 85		16.908 45	
	6.708 557		13.678 53		26.410 21	
0.3		6.705 7193		13.488 8813		21.123 9549
	6.655 885		13.259 90		17.643 13	
	7.075 869		14.828 28		31.030 13	
0.4		7.072 5987		14.368 9125		22.626 4770
	6.979 830		14.030 37		18.631 19	
	7.400 376		15.968 21		35.692 20	
0.5		7.396 9006		15.136 8457		23.929 0872
	7.258 083		14.554 30		19.880 68	
	7.694 107		17.110 54		40.378 15	
0.6		7.689 5652		15.823 5054		25.088 5173
	7.505 763		14.906 30		21.000 00	
	7.965 074		18.258 89		45.078 85	
0.7		7.957 5684		16.447 9293		26.139 2565
	7.732 038		15.155 26		21.000 00	
	8.218 847		19.413 90		49.789 25	
0.8		8.205 6773		17.022 8270		27.104 0690
	7.942 661		15.344 32		21.000 00	
	8.459 408		20.575 19		54.506 37	
0.9		8.437 3184		17.557 1690		27.998 8825
	8.141 353		15.497 81		21.000 00	
	8.689 663		21.742 03		59.228 33	
1.0		8.655 0499		18.057 5574		28.835 3384
	8.330 586		15.629 53		21.000 00	

(b): The odd excited energy levels of the λx^4 oscillator for λ between 0.1 and 1.0.

λ	$\epsilon_1^{(2)}$	$\epsilon_3^{(2)}$	$\epsilon_5^{(2)}$	$\epsilon_7^{(2)}$
0.1	3.306 87201	8.352 67782	13.969 9261	20.043 8636
0.2	3.539 00528	9.257 76561	15.799 5344	22.974 6311
0.3	3.732 48427	9.975 31279	17.212 9823	25.203 9616
0.4	3.901 08705	10.582 5370	18.392 6790	27.049 8140
0.5	4.051 93232	11.115 1542	19.418 2957	28.646 5303
0.6	4.189 28397	11.593 1472	20.332 9777	30.065 5424
0.7	4.315 93924	12.029 0158	21.163 1324	31.350 0478
0.8	4.433 85153	12.431 1826	21.926 2749	32.528 4506
0.9	4.544 44891	12.805 6348	22.634 7055	33.620 5672
1.0	4.648 81270	13.156 8038	23.297 4414	34.640 8483

(c): Even and odd excited levels for the λx^4 oscillator for λ between 1 and 100.

λ	$\epsilon_1^{(2)}$	$\epsilon_2^{(2)}$	$\epsilon_3^{(2)}$	$\epsilon_4^{(2)}$	$\epsilon_5^{(2)}$	$\epsilon_6^{(2)}$	$\epsilon_7^{(2)}$
2	5.475 7845	10.358 583	15.884 807	21.927 166	28.406 278	35.268 098	42.472 870
3	6.086 8964	11.600 658	17.859 316	24.715 035	32.075 093	39.876 587	48.073 337
4	8.585 7356	12.607 761	19.454 646	26.962 551	35.028 264	43.581 912	52.572 250
5	7.013 4791	13.467 730	20.813 966	28.874 996	37.538 815	46.729 704	56.392 169
6	7.391 3260	14.225 181	22.009 467	30.555 406	39.743 353	49.492 502	59.743 658
7	7.731 8318	14.906 304	23.083 323	32.063 806	41.721 298	51.970 463	62.748 807
8	8.043 1313	15.527 960	24.062 594	33.438 623	43.523 422	54.227 549	65.485 519
9	8.330 8363	16.101 721	24.965 808	34.706 127	45.184 396	56.307 403	68.006 918
10	8.599 0034	16.635 921	25.806 276	35.885 171	46.729 080	58.241 298	70.351 051
20	10.643 216	20.694 111	32.180 293	44.817 502	58.422 969	72.873 817	88.080 202
30	12.094 733	23.565 623	36.682 747	51.120 398	66.668 459	83.185 793	100.569 24
40	13.256 904	25.860 921	40.278 829	56.151 981	73.248 553	91.412 909	110.531 31
50	14.241 707	27.803 962	43.321 550	60.408 032	78.813 286	98.369 454	118.953 88
60	15.104 567	29.505 240	45.984 800	64.132 529	83.682 335	104.455 68	126.322 12
70	15.877 488	31.028 414	48.368 661	67.465 802	88.039 491	109.901 63	132.914 84
80	16.580 824	32.413 919	50.536 658	70.496 898	92.001 319	114.853 18	138.908 77
90	17.228 425	33.689 233	52.531 933	73.286 248	95.646 946	119.409 32	144.423 85
100	17.830 192	34.873 984	54.385 291	75.877 004	99.032 837	123.640 69	149.545 65

3. EIGENVALUES OF THE λx^{2m} ANHARMONIC OSCILLATOR

In this section we would like to discuss how our Hill-determinant method could be utilized for the general anharmonic term λx^{2m}. Detailed numerical evaluation of the eigenvalues and eigenfunctions for this general anharmonic case has not been previously performed. Simon has pointed out that in the general case the perturbation theory is again singular and converges to the exact eigenvalues only asymptotically. It is also not known whether the various $[N, M]$ Padé approximants of the divergent perturbation series in the general λx^{2m} case converge at all to the actual levels. The construction of the various Padé approximants in this case becomes extremely involved and the various successive coefficients in the perturbation series grow enormously fast. Nonetheless, the Borel summability method can still be utilized, again the evaluation of the energy level requiring an analytic continuation of the Borel series through the construction of various Padé approximants. Again, it is not known at all whether the Padé approximants of the Borel sum of the perturbation series converge.

TABLE III: The ratios $(\epsilon_n^{(2)}/\epsilon_0^{(2)})_\infty$ are the results of Chasman.[13] We compare them with our values of $(\epsilon_n^{(2)}/\epsilon_0^{(2)})$ and the converged values of $(\epsilon_n^{(2)}/\lambda^{1/3})$ for $\lambda = 100$.

n	$(\epsilon_n^{(2)}/\epsilon_0^{(2)})_\infty$	$\epsilon_n^{(2)}/\epsilon_0^{(2)}$	$\epsilon_n^{(2)}/\lambda^{1/3}$
0	1	1	1.07
1	3.58	3.57	3.84
2	7.04	6.98	7.51
3	10.98	10.88	11.7
4	15.31	15.18	16.3
5	20.01	19.81	21.3
6	24.99	24.73	26.6
7	30.24	29.91	32.8

We can very easily avoid all these analytic as well as numerical difficulties in the calculation of the energy levels for the λx^{2m} oscillator by a straightforward generalization of the difference equation and the associated Hill-determinant technique used for the λx^4 oscillator. It is of particular interest to note that if we assume again the same ansatz [Eq. (2)] for ψ in the case of the Hamiltonian

$$H = -\frac{d^2}{dx^2} + x^2 + \lambda x^{2m},$$

the associated $(n+1) \times (n+1)$ determinant $D_n^{(m)}$, for the even parity solutions, satisfies a three term difference equation, viz.,

$$\begin{aligned} D_n^{(m)} &= (\epsilon - 1 - 4n)D_{n-1}^{(m)} \\ &\quad + (-1)^{m+1}\lambda 2^{2m} n(n-\tfrac{1}{2})(n-1)(n-\tfrac{3}{2}) \\ &\quad \ldots (n-m+1)(n-m+\tfrac{1}{2})D_{n-m-1}^{(m)} . \end{aligned} \tag{9}$$

Thus the problem of eigenvalue determination in this general case reduces to finding the roots of $D_n^{(m)} = 0$ when $n \to \infty$ as in the case of the λx^4 oscillator. As an application of our method, we have obtained the ground state and the first even excited state eigenvalues of the λx^{2m} oscillator for $m = 3$ and 4. Our results are displayed in Table IV. In Fig. 1 we have plotted for comparison the ground state eigenvalues for $m = 2, 3$ and 4. The behavior of the various characteristic polynomials and the convergence of roots follow the same pattern as discussed in the case of the λx^4 anharmonic oscillator. Our solutions are consistent with the scaling property that $\epsilon_n^{(m)}(\lambda)/\lambda^{1/(m+1)} \to$ const. for large λ. It may also be noticed that convergence of the eigenvalues for any fixed λ, to any specified degree, occurs at higher and higher order of the truncated determinant both for increasing m as well as increasing n—an aspect which is reflected in the entries of Table IV.

It is interesting to mention here that numerical computations by Graffi, Greechi and Turchetti (see Ref. 5) suggest that Padé does not converge to the eigenvalues for an x^8 oscillator, however the mixed Borel–Padé method converges.

4. NORMALIZATION AND OVERLAP INTEGRAL

Before concluding this paper we would like to make a few observations on the nature and normalization of the wavefunctions. In our earlier paper[11] we have shown by solving the recurrence relation for the c_n for large n that the wave functions, for the ground as well as for the excited states of the λx^4 oscillator, are entire functions for all λ.[14] In fact we obtained bounds on our wave functions which clearly established the asymptotic normalizations of our wavefunctions. We have now carried out a numerical investigation of the amount of normalization and the extent of orthogonality of the wave functions by evaluating the appropriate overlap integrals for the λx^4 oscillator. To carry out the overlap and normalization integrals we use the expression

$$\psi_k(x) = \exp(-x^2/2)\sum_{n=0}^{\infty} c_n^{(k)}(\epsilon_k, \lambda)x^{2n}, \tag{10}$$

so that

$$I_{kl} = \int_{-\infty}^{\infty} \psi_k(x)\psi_l(x)dx = \sum_{n,m=0}^{\infty} c_n^{(k)}(\epsilon_k,\lambda)c_m^{(l)}(\epsilon_l,\lambda) \times \Gamma(n+m+\tfrac{1}{2}). \tag{11}$$

Equation (11) is evaluated numerically by truncating the double sum at $n = m = N$. The coefficients $c_n^{(k)}(\epsilon_k, \lambda)$ are evaluated for a given λ by using the recurrence relation of Eq. (3) with ϵ_k set equal to the values corresponding to the roots of the characteristic polynomial P_n, defined in Eq. (7), for $n = N$. We find that for values up to $N = 40$ a stable value of the ratio of the overlap and normalization integrals occur, i.e.,

$$\frac{I_{kl}}{(I_{kk}I_{ll})^{1/2}} \sim 10^{-14},$$

TABLE IV: The ground state and first even excited energy levels for $m = 3$ and 4 for λ between 0.1 and 100.

λ	$\epsilon_0^{(3)}$	$\epsilon_2^{(3)}$	$\epsilon_0^{(4)}$	$\epsilon_2^{(4)}$
0.1	1.109 0870	6.644 391	1.168	7.639-40
0.2	1.173 8893	7.381 647	1.240-1	8.452-3
0.3	1.223 6871	7.909 026	1.291-2	9.001-2
0.4	1.265 0993	8.330 571	1.332-3	9.426-7
0.5	1.300 9869	8.686 393	1.367-8	9.776-8
0.6	1.332 8959	8.996 752	1.397-8	10.077-9
0.7	1.361 7725	9.273 480	1.423-4	10.342-4
0.8	1.388 2449	9.524 158	1.447-9	10.580-1
0.9	1.412 7543	9.753 966	1.470-1	10.795-7
1	1.435 6246	9.966 622	1.490-1	10.993-4
2	1.609 9319	11.543 93	1.640-3	12.41-3
3	1.732 8571	12.623 40	1.742-5	13.36-8
4	1.830 4373	13.467 06	1.822-6	14.08-9
5	1.912 4538	14.169 09	1.888-92	14.67-8
6	1.983 7805	14.775 27	1.945-9	15.181-9
7	2.047 2390	15.311 65	1.994-9	15.625-31
8	2.104 6259	15.794 62	2.038-44	16.020-8
9	2.157 1630	16.235 20	2.078-85	16.380-6
10	2.205 7232	16.641 21	2.115-22	16.707-15
20	2.564 6446		2.383-91	
30	2.809 3811		2.560-9	
40	3.000 3148-57		2.69-70	
50	3.159 0208-15		2.80-1	
60	3.295 9516-22		2.90-1	
70	3.417 0453-61		2.98-9	
80	3.526 0301-10		3.06-7	
90	3.625 4144-53		3.12-3	
100	3.716 9743-50		3.18-9	

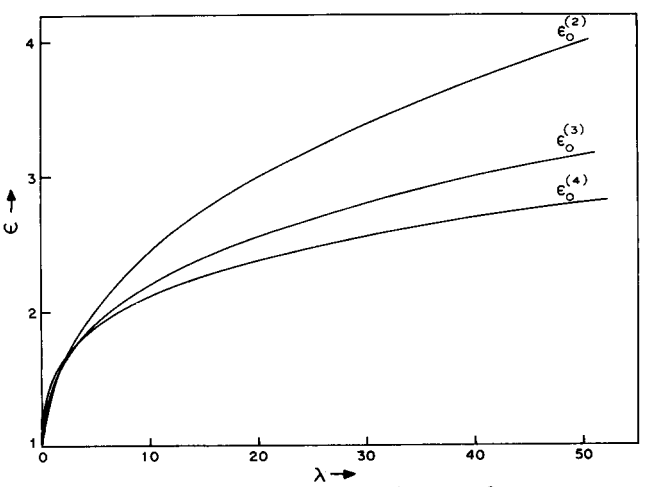


FIG. 2. The ground state energy levels $\epsilon_0^{(m)}$ of the λx^{2m} oscillator for $m = 2, 3$ and 4 as a function of the anharmonicity parameter λ.

showing the approximate correctness of the behavior of the eigenfunctions of the energy levels in our calculations. This numerical evaluation of the overlap integrals can of course be done for large N, however the numerical estimates become unreliable beyond typically $N = 40$ for low λ. This is because some of the terms involved in Eq. (11) become exceedingly large due to the presence of the gamma function and cancellation between such terms leads to large round off error even with double precision arithmetic on the computer.

5. CONCLUDING REMARKS

We have developed a procedure for a straight forward evaluation of the eigenvalues of the generalized λx^{2m} anharmonic oscillator. The essence of our method consists in discovering a suitable basis within which the Schrödinger equation for such an anharmonic oscillator reduces to a difference equation involving only three terms. The condition that the set possesses a nontrivial solution is given by the vanishing of an infinite tridiagonal determinant. Our calculational scheme consists of requiring the truncated $(n \times n)$ determinants of successive orders to vanish. Their tridiagonal form permits a simple recursive generation of the corresponding characteristic polynomials whose roots we determine numerically. The sequence of lowest roots of the successive polynomials for increasing n, oscillates about the true lowest eigenvalue with decreasing amplitude. The eigenvalues corresponding to the excited states are limits of similar sequences of the higher roots of these polynomials. It bears pointing out that the sequence of approximants to the true eigenvalue do not form a monotonic sequence as is the case in a Rayleigh–Ritz calculation, or the Bazley–Fox calculation with intermediate Hamiltonians. In fact if a variational calculaton were attempted with a nonorthogonal basis such as we have used, the resulting matrix would be entirely different—none of its elements would be zero, in contrast to the simple tridiagonal form we obtain. The evaluation of the eigenvalues of such a matrix would be entirely untractable and no statement could be made about the convergence of the sequence of eigenvalues.

Although our method is applicable to the general anharmonic oscillator we present results for the ground state and seven excited states for $m = 2$, and only ground states and the first even excited state for $m = 3$ and 4. Calculations for larger m were not attempted in view of the large amount of computer time required in order to get even three figure stability in the computed eigenvalues. The nature of the dependence of the eigenvalues of all the states that we have investigated are qualitatively similar, in agreement with the conjecture made by Bender[7] on the basis of his WKB solutions.

Further, we wish to point out that in our method unlike the Padé and the Borel–Padé summability techniques, no preliminary perturbation series needs to be constructed. In fact, the accuracy of these techniques is crucially restricted by the accuracy to which the coefficients of the perturbation series are known. Furthermore, we can evaluate all the excited energy levels and the corresponding eigenfunctions simultaneously, whereas a separate perturbation series for each excited state has to be developed before the Padé or Borel–Padé methods can be applied. Our results can be taken as further empirical evidence of the rapid convergence of the Borel–Padé technique for the λx^4 oscillator.

Finally, we note that the method we have used can also be applied to a model one dimensional nonpolynomial interaction Lagrangian of the kind $\lambda x^2/(1 + gx^2)$. The perturbation theory for such a Lagrangian presents difficult problems both in principle and in actual computation, however, our method applies in a straight forward way for the numerical computations of eigenvalues and eigenfunctions. The interest in such a system derives from quantum field theory: the Schrödinger equation with such an interaction Lagrangian is the analogue of a zero-dimensional field theory with a nonlinear Lagrangian, of a kind which currently finds extensive use in elementary particle physics. Our method, applied in conjunction with perturbation theory, may be expected to answer questions related to finiteness or otherwise of mass renormalization in such a model nonlinear field theory and the nature of the perturbation series. We hope to answer such questions in a future investigation.

---

[1] C. Bender and T. T. Wu, Phys. Rev. Lett. **21**, 406 (1968).
[2] C. Bender and T. T. Wu, Phys. Rev. **184**, 1231 (1969).
[3] J. Loeffel, et al., Phys. Lett. B **30**, 656 (1969). The first computation of the energy levels of the x^4 oscillator using Padé approximants appeared in C. Reid., Int. J. Quantum Chem. **1**, 521 (1967).
[4] B. Simon, Ann. Phys. (N.Y.) **58**, 76 (1970).
[5] S. Graffi *et al.*, Phys. Lett. B 32, 631 (1970). See also S. Graffi, V. Greechi, and G. Turchetti, Nuovo Cimento **43**, 313 (1971).
[6] J. J. Loeffel and A. Martin, Rep. TH 1167 - CERN, 1970.
[7] C. Bender, J. Math. Phys. **11**, 796 (1970).
[8] N. Bazley and D. Fox, Phys. Rev. **124**, 483 (1961).
[9] A. Weinstein, Mem. Sci. Math. ,(1937), No. 88.
[10] A. Jaffe, Commun. Math. Phys. **1**, 127 (1965).
[11] S. N. Biswas, K. Datta, R. P. Saxena, P. K. Srivastava, and V. S. Varma, Phys. Rev. D 4, 3617 (1971). Another nonperturbative method has recently been pointed out by W. McLory (University of Toronto, preprint). An application of Euler summability methods to anharmonic oscillators has been made by J. Gunson and P. Ng (University of Birmingham, preprint).
[12] L. I. Schiff, Phys. Rev. **92**, 766 (1953).
[13] R. Chasman, J. Math. Phys. **2**, 733 (1961).
[14] See the appendix to Ref. 4 by A. Dicke, and Hsich, and Sibuya, J. Math. Anal. Appl. **16**, 84 (1966).

Duality of a Young diagram describing a representation and dimensionality formulas

G. Murtaza

Institute of Physics, University of Islamabad, Islamabad, Pakistan

M. A. Rashid

Institute of Mathematics, University of Islamabad, Islamabad, Pakistan

(Received 22 February 1972)

We discuss the standard dimensionality formulas for irreducible representations of continuous Lie groups $SU(n)$, $O(n)$, and $Sp(n)$. Every Young diagram describes not only an irreducible representation of any of the groups $SU(n)$, $O(n)$, and $Sp(n)$ of dimension n, but it also describes an irreducible representation of the symmetric group S_N, where N is the number of boxes in the diagram. This property of a Young diagram leads to the factorization of the dimensionality formulas for irreducible representations of any of these continuous groups into two factors. One factor which depends upon the dimension n of the group appears as a polynomial (with integral roots) of degree N. If we choose the leading coefficient of this polynomial as unity, the other factor turns out to be just $1/N!$ times the dimensionality of the associated irreducible representation of the symmetric group on N symbols. The polynomial depends upon the type of the group under study, i.e., whether it is an $SU(n)$ or $O(n)$ or $Sp(n)$. We also give simple recipes to read the dimensionality of the representations of these groups from the Young diagram following our formulas.

1. INTRODUCTION

We start by considering the general group of linear transformations $GL(n)$ in an n-dimensional space R_n. A vector x in space R_n undergoes transformation

$$x' = ax \quad \text{or} \quad x_i' = a_{ij}x_j \quad (i,j = 1, 2, \ldots, n) \tag{1}$$

and a tensor of rank α transforms like the product of α factors $a \times a \times \cdots \times a$. These tensors form a vector space of n^α dimension and constitute the basis for a representation of the group $GL(n)$. This representation can be reduced into irreducible representations by using the permutation operators (Young symmetrizers). Each of these irreducible representations of $GL(n)$ is described by a Young diagram with pattern $[\lambda] = [\lambda_1, \lambda_2, \ldots, \lambda_n]$ where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ and $\sum_{i=1}^{n} \lambda_i$ is the Number N of blocks arranged in r rows. The pattern reflects the particular symmetry type of the associated irreducible tensor. The dimensionality $d[\lambda]$ of IR of $GL(n)$,which is the number of independent components of irreducible tensors of definite symmetry type, is given by the Weyl formula (1):

$$d[\lambda] = \frac{D(l_1, l_2, \ldots, l_n)}{D(n-1, n-2, \ldots, 2, 1, 0)} = \prod_{i=1}^{n} \frac{1}{(n-i)!} \prod_{j=i+1}^{n} (l_i - l_j), \tag{2}$$

where $l_i = \lambda_i + n - i$ and $D(l_1, l_2 \ldots l_n) = \prod_{i<j} (l_i - l_j)$ and (as always in this paper) an empty product like $\prod_{n+1}^{n}$ is taken as 1.[1,2]

For the general $GL(n)$, there is no restriction on the matrices a. If, however, we do put restrictions like reality, unitarity, and unimodularity, we can realize subgroups $RL(n)$, $SL(n)$, $U(n)$, and $SU(n)$ of the group $GL(n)$. For all these subgroups the dimensionality formula (2) does not change, because the irreducible representations of $GL(n)$ remain irreducible as we go to these subgroups.

We may note in passing that for $GL(n)$ and its subgroups mentioned above, the following identities are satisfied:

$$d[S^n] = 1,$$

$$d[\lambda_1, \lambda_2, \ldots, \lambda_n] = d[\lambda_1 + S, \ldots, \lambda_n + S] \quad (S \text{ an integer})$$
$$= d[\lambda_1 - \lambda_n, \lambda_2 - \lambda_n, \ldots, \lambda_{n-1} - \lambda_n, 0]$$
$$= d[\lambda_1 - \lambda_n, \lambda_1 - \lambda_{n-1}, \ldots, \lambda_1 - \lambda_2].$$

When we go to the orthogonal subgroup $O(n)$, the irreducible representations of $GL(n)$ in general do not remain irreducible. The reason is that, in addition to the operation of symmetrization which one uses to construct irreducible representations of $GL(n)$, a new operation of contraction appears which commutes with the orthogonal transformations and further reduces the representation. The resulting irreducible representation (after symmetrization and contraction have been performed) can *still* be associated with a Young diagram with pattern described as before by $[\lambda] = [\lambda_1, \lambda_2, \ldots, \lambda_n]$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ and $\lambda_1 + \lambda_2 + \cdots + \lambda_n = N$, N being the number of boxes in the diagram. However, for $O(n)$ only those diagrams are permissible for which the sum of lengths of the first two columns is $\leq n$. Also the permissible associated patterns (with first column having μ and $n-\mu$ boxes) have the same dimensionality. The associated patterns when different correspond to equivalent representations of the subgroup $SO(n)$.

Representations of $O(n)$ corresponding to self-associated diagrams, which occur only when n is even, split into two nonequivalent irreducible representations of equal dimension of the subgroup $SO(n)$[3]. Thus to compute the dimensionality of an irreducible representation of $O(n)$, we can restrict ourselves to the patterns $[\lambda_1, \lambda_2, \ldots, \lambda_k]$ where $k = n/2$ if n is even and $k = (n-1)/2$ if n is odd.

The dimensionalities of irreducible representations of $SO(2k)$ and $SO(2k+1)$ are given respectively by[4]

$$d[\lambda] = \frac{1}{2} \prod_{i=1}^{k} \frac{2}{(2k-2i)!} \prod_{j=i+1}^{k} (l_i - l_j)(l_i + l_j) \tag{3}$$

and

$$d[\lambda] = \prod_{i=1}^{k} \frac{(2l_i + 1)}{(2k-2i+1)!} \prod_{j=i+1}^{k} (l_i - l_j)(l_i + l_j + 1), \tag{4}$$

where $l_i = \lambda_i + k - i$

Copyright © 1973 by the American Institute of Physics

Next we pass on to the symplectic group $Sp(n)$[5]. The procedure for obtaining its irreducible representations is very similar to the method used for the orthogonal groups. However, the diagrams described by patterns having more than $n/2$ rows are not permissible, and consequently we are restricted to the patterns $[\lambda_1, \lambda_2, \ldots \lambda_k]$ where $k = n/2$. The dimensionality formula for $Sp(2k)$ is

$$d[\lambda] = \prod_{i=1}^{k} \frac{(l_i + 1)}{(2k - 2i + 1)!} \prod_{j=i+1}^{k} (l_i - l_j)(l_i + l_j + 2) \tag{5}$$

with

$$l_i = \lambda_i + k - i.$$

Lastly a word about the symmetric group S_N which is the group of all permutations on N symbols. This group plays quite an important role in physics. Whenever we deal with a system of N identical particles, the total symmetry of the Hamiltonian will contain the group S_N. Also the classification of atomic and nuclear states depends essentially on the properties of the symmetric group. Besides, the knowledge of finding the irreducible representations of S_N is used as a tool for constructing irreducible representations of other groups.

Each irreducible representation of S_N is associated with a Young pattern $[\lambda_1, \lambda_2, \ldots, \lambda_r]$ where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r \geq 0$ and $\lambda_1 + \lambda_2 + \cdots + \lambda_r = N$, the dimensionality being given by the formula[6]

$$d[\lambda] = N! D(l_1, l_2, \ldots, l_r)/l_1!\, l_2! \cdots l_r!, \tag{6}$$

where

$$l_i = \lambda_i + r - i.$$

It is interesting to observe that a Young diagram simultaneously describes an irreducible representation of various types of groups $SU(n), O(n)$, and $Sp(n)$. However what is more interesting is the fact that each pattern describing an irreducible representation of any of these groups is necessarily associated with an irreducible representation of some symmetric group S_N, where N is the number of the boxes in the diagram. This everpresence of the symmetric group in the background of any of the groups $SU(n), O(n)$, and $Sp(n)$ is suggestive. One is tempted to speculate that all irreducible representations of the said groups exhibit two facets of behavior, one corresponding to the type of the group and the other corresponding to the shape or pattern of the Young diagram the latter being related with some symmetric group.

Here in this paper we shall be interested in investigating this dichotomous character of a Young tableau insofar as the dimensionality of an irreducible representation of the groups $SU(n), O(n)$, and $Sp(n)$ is concerned. The dimensionality formulas depend both upon the dimension of the group and the symmetry type of the pattern that describes the irreducible representation. A question arises: Can one factorize the dimensionality expression into parts such that the part which is independent of the group dimension stands separate from the other? The answer is found in the affirmative. Pietenpol and Speiser[7] exhibited this feature for the case of $SU(n)$ group. We have extended the argument to other groups as well. The part which is dependent on the group-dimension n turns out to be a polynomial in n. *Once this polynomial is chosen to have the leading coefficient unity, the group-dimension-independent part is found to be the same for all the groups SU(n), O(n) and Sp(n). This part is equal to 1/N! times the dimensionality of the associated irreducible representation of the symmetric group S_N, where N is the number of boxes in the diagram.*

2. RECASTING THE DIMENSIONALITY FORMULAS

The dimensionality formulas of irreducible representations of the groups $SU(n), SO(2k), SO(2k+1)$and $Sp(2k)$ are given by Eqs. (2), (3), (4), and (5). We try to recast them in a form such that the part which is independent of the dimension of the group stands separate from the rest. The procedure to do this is straightforward. We first describe the argument which is *generally* valid. We try to split the i product into two parts $i > r$ and $i \leq r$ and compute the $i > r$ part explicitly. Clearly this is only necessary when $r <$ upper limit of any summation. In this case, for each $i > r$, the terms in the product reduce to unity. This part of the argument fails only for the group $SO(2k)$ where (note that $r < k$), for $i = k$, the product reduces to 2 (and not 1) and enables us to cancel the factor $\frac{1}{2}$ outside (note also that, in the above argument, empty products are always taken as 1 in accordance with the convention described before). For $r = k$, the factor $\frac{1}{2}$ can again be cancelled in $SO(2k)$ formula by considering $O(2k)$ in place of $SO(2k)$ where such patterns are self-associated.

Next we break the j product in the same way and note that the product of the terms with $j > r$ can be computed to yield a simple i-dependent expression consistent with our convention. As a result we obtain dimensionality formulas as summarized below:

$$SU(n), \quad d[\lambda] = C \prod_{i=1}^{r} \frac{(\lambda_i + n - i)!}{(n - i)!}, \tag{7}$$

$$O(n), \quad d[\lambda] = C \prod_{i=1}^{r} \frac{(\lambda_i + n - r - i - 1)!}{(n - 2i)!} \times \prod_{j=1}^{r} (\lambda_i + \lambda_j + n - i - j), \tag{8}$$

$$Sp(n), \quad d[\lambda] = C \prod_{i=1}^{r} \frac{(\lambda_i + n - r - i + 1)!}{(n - 2i + 1)} \times \prod_{j=i+1}^{r} (\lambda_i + \lambda_j + n - i - j + 2), \tag{9}$$

where

$$C = \prod_{i=1}^{r} \frac{1}{(\lambda_i + r - i)!} \prod_{j=i+1}^{r} (\lambda_i - \lambda_j - i + j). \tag{10}$$

At this stage we wish to prove that in formulas (7)–(9) the n-dependent parts multiplying C are all polynomials in n with leading coefficient unity. This is manifest in Eq. (7). To prove it for the remaining two equations, we note that C is a fixed rational number for each pattern. In fact it is just $1/N!$ times the dimensionality of the associated representation of the symmetric group S_N. Since on the left-hand sides of these equations we have functions of n which take integral values for $n = 0, 1, 2, \cdots$, the n-dependent parts must be polynomials in n. Thus all the factors in the denominators must cancel some of the factors in the numerators. This cancellation, which, unfortunately, does not seem to be easily expressible, will be presented elsewhere. However, we note that the split in these equations is such that the n-dependent polynomials are *chosen with the leading coefficient unity*. With this choice the other factor turns out to be the same C in all these cases.

We wish to emphasize that not only we have been able to obtain an affirmative answer to the question raised in the introduction, but we have also been able to express the two apparently different type of formulas for $SO(2k+1)$ and $SO(2k)$ into a single formula for $O(n)$. Also from computational point of view, our formulas are far superior.

Incidentally the degree of these polynomials is equal to N, the number of boxes in the diagram.

The determination of the C part which is common to all the groups poses no problem because, as mentioned before, it is $1/N!$ times the dimensionality of an irreducible representation of the associated symmetric group S_N for which a thumb rule is known.[8] For the other parts, however, we deal with the various groups separately and give the following recipes for them.

$SU(n)$

We allot the boxes of the Young diagram numbers starting with n in the top left-hand corner and then increasing or decreasing by unity as we move along a row or down a column, respectively. The polynomial in n is simply given by the product of numbers allotted to the boxes of the diagram.

For the groups $O(2k)$, $O(2k+1)$, and $Sp(2k)$, the group-dimension-dependent parts are not so simple. They themselves consist of two parts. The second part which involves sums of the type $(l_i + l_j)$ is calculated by first writing all l's and then taking a product of terms involving sums of any two l's *consistent* with the variation of i and j. For the first part we can have the following thumb rules.

$O(2k)$ and $O(2k+1)$

Like $SU(n)$ group, we do the numbering of the boxes in the diagram starting this time with $R = n - r - 1$ in the top left-hand corner. Next we define a principal pattern $[r, r-1, \ldots, 1]$ of r rows.[9] Now we look at the given diagram. If it is a principal pattern, the expression immediately reduces to unity. If otherwise, we proceed to mark out in the given diagram a principal part of r rows (complete the pattern to the principal form if necessary) and observe that this part is given by a product of R's corresponding to the boxes which are in excess of the principal pattern divided by those R's corresponding to the boxes that are short of it.

$Sp(2k)$

For this group, the procedure is exactly the same as for $O(2k)$ and $O(2k+1)$ except that now $R = n - r + 1$ and the principal pattern of reference is of $r - 1$ rows.

Note added in proof: We are thankful to Dr. R. C. King for pointing out to us that he had independently obtained the results contained in our paper earlier and published them in Can. J. Math. **33**, 176 (1972).

ACKNOWLEDGMENT

The authors wish to thank Dr. J. L. Pietenpol for having taken great pains in reading in detail the first version of this paper. His many valuable suggestions have resulted in a considerable improvement in our work.

---

[1]Herman Weyl, *The Classical Groups* (Princeton U.P., Princeton, N.J., 1946), p. 201, (5.14).

[2]Obviously $l_1 > l_2 > \cdots > l_n$.

[3]See, e.g., M. Hamermesh, *Group Theory* (Addison-Wesley, Reading, Mass., 1962), pp. 396–97.

[4]For the dimensionality formulas,Eqs. (3), (4), and (5), see, e.g., F. D. Murnaghan, Commun. Dublin Inst. Adv. Stud. A, No. 13 (1958).

[5]The group $Sp(n)$ exists only for even n.

[6]See, e.g., Ref. 1, Theorem 7.7 to p. 213.

[7]J. L. Pietenpol and D. Speiser, J. Math. Phys. **12**, 1 (1971).

[8]The dimensionality of an irreducible representation of the symmetric group S_N is given by $N!$ divided by the factorials of the numbers allotted to the boxes in the Young pattern describing the representation; the allotment of numbers being done in the following way: We start with the top left-hand box and give it a number which is equal to the total number of boxes in the row and column which meet there including the reference box only once. Then we fill the second box, the third and so on in the first row the same way. Next we repeat the process with the second row starting from the extreme left corner. We keep doing this till we have filled the whole diagram. See, e.g., G. Murtaza, Ph.D. thesis (Imperial College, London, 1966) (unpublished).

[9]The principal patterns satisfies the relation $\lambda_i = r - i + 1$.

Local products as operators

Mary Beth Ruskai* and John R. Klauder

Bell Laboratories, Murray Hill, New Jersey 07974

(Received 6 September 1972; revised manuscript received 7 March 1973)

It is shown for a large class of domains that the bilinear forms associated with formal expressions of the type $\int h(x)a_{s_0}^{\dagger}(x)\ldots a_{s_1}^{\dagger}(x)a_{t_1}(x)\ldots a_{t_q}(x)dx$ define closable operators if and only if p, $q\in\{0,1\}$.

I. INTRODUCTION

Formal products of conventional annihilation and creation operators taken at the same point arise in numerous quantum field problems. In particular the special features of ultralocal quantum field theory[1] lead one to consider expressions of the form

$$\int h(x)a_{s_p}^{\dagger}(x)\cdots a_{s_1}^{\dagger}(x)a_{t_1}(x)\cdots a_{t_q}(x)dx, \tag{1}$$

where $a_s(x)$ and $a_s^{\dagger}(x)$ are conventional (Fock representation) annihilation and creation operators with a distinguished vector Ω, such that

$$a_s(x)\Omega = 0$$

for all $s = 1,\ldots,M \le \infty$ and all $x \in R$. In this paper we consider the problem of determining when such products lead to closable operators. Since one expects the Hamiltonian to be represented by a closable operator, restrictions on the expressions (1) can be interpreted physically as the restrictions on the nature of the allowable interactions. Our results confirm the heuristic expectation[2] that Eq. (1) can be interpreted as a closable operator if and only if $p \le 1$ and $q \le 1$, i.e., if (1) is at most bilinear in the fields $a_s^{\dagger}(x)$ and $a_t(x)$.

Special techniques have been used[3] to study expressions of the form (1) in the case of Bose commutation relations for a and $a^{\dagger}$. Here we exploit more general techniques appropriate to either Bose or Fermi commutation relations, or even to a combination of such fields. Bosons and fermions can be treated uniformly if we write the commutation relations as

$$a_s(x)a_s(x') - \epsilon_s a_s(x')a_s(x) = 0,$$

$$a_s(x)a_s^{\dagger}(x') - \epsilon_s a_s^{\dagger}(x')a_s(x) = \delta(x-x'),$$

where $\epsilon_s = 1$ for a boson and $\epsilon_s = -1$ for a fermion.

In addition, a boson field commutes with any other (i.e., different spinor indices) field, while any two different fermion fields anticommute.

We first prove the theorem on a specific dense domain $\mathfrak{D}$, defined below, and then show how our results arise under less restrictive assumptions in Sec. 4.

II. THEOREM

In order to study expressions of the form (1), we look at the corresponding bilinear forms on Fock space $\mathfrak{F}$, which are always well defined, even if not well behaved, i.e., arise from closable operators. We write the Fock space vectors as

$$\Phi = (\varphi_0, \varphi_1, \ldots, \varphi_n \ldots) \in \mathfrak{F},$$

and let

$$\mathfrak{D} = \{\Phi \text{ in } \mathfrak{F}\colon \varphi_n \text{ is a continuous function with compact support and } \exists \text{ an } N_\Phi \text{ such that } \varphi_n = 0 \text{ if } n > N_\Phi\}.$$

For any real continuous function $h(x)$ wtih compact support we define the bilinear form f_{pq}^{ST} as

$$f_{pq}^{ST}(\Phi,\Psi) = \sum_{n=p}^{\infty} A_n \int h(x) \times \langle \varphi_n(xs_1\cdots xs_p, x_1r_1\cdots x_{n-p}r_{n-p}), \psi_{n-p+q}(xt_1\ldots xt_q, x_1r_1\ldots x_{n-p}r_{n-p})\rangle dx, \tag{2}$$

where $S \equiv \{s_1,\ldots,s_p\}$, $T \equiv \{t_1,\ldots,t_q\}$, and

$$A_n = \sqrt{n(n-1)\cdots(n-p+1)}\,\sqrt{(n-p+q)\cdots(n-p+1)}.$$

The forms can be defined for a rather large domain $\mathfrak{D}(f_{pq}^{ST})$ which we do not specify. In our study we merely require $\mathfrak{D}(f_{pq}^{ST}) \supset \mathfrak{D}$ and note that this can always be satisfied; e.g., f_{pp}^{SS} can be extended to

$$\Big\{\Phi\colon \sum_{n=p}^{\infty} A_n \int |h(x)|\langle \varphi_n(xs_1\cdots xs_p, x_1r_1\cdots x_{n-p}r_{n-p}), \varphi_n(xs_1\cdots xs_p, x_1r_1\cdots x_{n-p}r_{n-p})\rangle dx < \infty\Big\}.$$

We want to know when f_{pq}^{ST} defines a closable operator Θ_{pq}^{ST} with $\mathfrak{D}(f_{pq}^{ST}) \supset \mathfrak{D}(\Theta_{pq}^{ST}) = \mathfrak{D}$ such that

$$f_{pq}^{ST}(\Phi,\Psi) = \langle \Phi, \Theta_{pq}^{ST}\Psi\rangle \quad \forall \Phi,\Psi \text{ in } \mathfrak{D}. \tag{3}$$

The result we obtain is contained in the following.

Theorem: The form f_{pq}^{ST} defined by (2) determines a closable operator Θ_{pq}^{ST} satisfying (3) if and only if $p, q \in \{0,1\}$.

The most general field operator constructed in the manner described is thus given formally by:

$$C + \sum_s \int g_s(x)a_s^{\dagger}(x)dx + \sum_t \int h_t(x)a_t(x)dx + \sum_{st} \int k_{st}(x)a_s^{\dagger}(x)a_t(x)dx. \tag{4}$$

III. PROOF

In our treatment, most of the symmetry of the problem is contained in the choice of Fock space $\mathfrak{F}$, i.e., in the (mixed) symmetry of φ_n, ψ_n. Therefore, we will, for simplicity, drop explicit reference to the index variables. At appropriate points, we will indicate what changes, if any, occur because of the indices.

We present the proof in several parts:

(a) We first note that it follows from the Riesz representation theorem that $f(u,v)$ defines an operator A with domain $\mathfrak{D}(A)$ such that $f(u,v) = \langle u, Av\rangle$ if and only if $u \to f(u,v)$ is continuous on $\mathfrak{D}(A)$ for each fixed v in $\mathfrak{D}(A) \subset \mathfrak{D}(f)$.[4] In particular, if $|f(u,v)| \le C_v\|u\|$,

Copyright © 1973 by the American Institute of Physics

there exists a unique x such that $f(u,v) = \langle u, x\rangle$ and $Av = x$.

(b) We show that p cannot be > 1 by constructing a sequence Φ^K such that $\Phi^K \to 0$, but $f_{pq}(\Phi^K, \Psi) \to R \neq 0$ if $p \neq 0, 1$. Let the sequence Φ^K be defined as follows:

$$\hat{\Phi} = (0, 0, \hat{\varphi}_2, \ldots, \hat{\varphi}_N, 0, 0, \ldots) \in \mathfrak{D},$$

$$\varphi_n^K = \prod_{i<j} e^{-K(x_i - x_j)^2} \hat{\varphi}_n(x_1, \ldots, x_n),$$

$$\Phi^K = (0, 0, \varphi_2^K, \ldots, \varphi_N^K, 0, 0, \ldots).$$

Then Φ^K is in $\mathfrak{D}$ for all $K \geq 0$ and

$$\lim_{K\to\infty} \Phi^K = 0.$$

Therefore, if f_{pq} defines an operator, then $f_{pq}(\Phi^K, \Psi) \to_{K\to\infty} 0$ for each fixed Ψ in $\mathfrak{D}$. But

$$\begin{aligned} f_{pq}(\Phi^K, \Psi) &= A_p \int h(x) \hat{\varphi}_p^*(x, \ldots, x) \psi_q(x, \ldots, x) dx \\ &+ \sum_{n=p+1}^{N} \int h(x) \Big\langle \prod_i e^{-Kp(x-x_i)^2} \prod_{i<j} e^{-K(x_i - x_j)^2} \\ &\times \hat{\varphi}_n(x \ldots x, x_1 \ldots x_{n-p}) \psi_{n-p+q}(x \ldots x, x_1 \ldots x_{n-p}) \Big\rangle dx \\ &\xrightarrow[K\to\infty]{} A_p \int h(x) \hat{\varphi}_p^*(x, \ldots, x) \psi_q(x, \ldots, x) dx = R. \end{aligned}$$

If $p = 0, 1$, $\hat{\varphi}_p = 0$ and $R = 0$ so that no conclusion follows in this case. But if $p \geq 2$, one can choose $\hat{\varphi}_p$ and ψ_q so that $R \neq 0$. In particular, let Λ be a region in $(-\infty, \infty)$ such that $h(x) > 0$ [or equivalently $h(x) < 0$] on Λ. Then choose $\hat{\varphi}_p \geq 0$ such that $\varphi_p \not\equiv 0$ and supp $\hat{\varphi}_p \subset \Lambda$ and ψ_q such that $\psi_q > 0$ on supp $\hat{\varphi}_p$. Then $R > 0$ (equivalently < 0). This can always be done by translating and contracting arbitrary, nonnegative $\hat{\varphi}_p, \psi_q$.

(c) We next show that f_{1q} and f_{0q} are, in fact, continuous in Φ on $\mathfrak{D}$. Let $\Psi = (\psi_0, \psi_1, \ldots, \psi_{N-1+q}, 0, 0, \ldots)$ be an arbitrary element of $\mathfrak{D}$, and suppose $q \neq 0$. Then

$$\begin{aligned} |f_{1q}(\Phi, \Psi)| &= \Big| \sum_{n=1}^{N} A_n \int h(x) \\ &\times \langle \varphi_n(x, x_1 \ldots x_{n-1}), \psi_{n-1+q}(x \ldots x, x_1 \ldots x_{n-1}) \rangle dx \Big| \\ &\leq \sum_{n=1}^{N} A_n \Big[\int |\varphi_n(x_1 \ldots x_n)|^2 d^n x \Big]^{1/2} \\ &\times \Big[\int |h(x_1) \psi_{n-1+q}(x_1 \ldots x_1, x_2 \ldots x_n)|^2 d^n x \Big]^{1/2} \\ &\leq A_N \sup_x |h(x)| \\ &\times \sum_{n=1}^{N} \|\varphi_n\| \Big[\int |\psi_{n-1+q}(x_1 \ldots x_1, x_2 \ldots x_n)|^2 d^n x \Big]^{1/2} \\ &\leq A_N \sup_x |h(x)| \\ &\times \Big(\sum_{n=1}^{N} \int |\psi_{n-1+q}(x_1 \ldots x_1, x_2 \ldots x_n)|^2 d^n x \Big)^{1/2} \\ &\times \Big(\sum_{n=1}^{N} \|\varphi_n\|^2 \Big)^{1/2} \\ &\leq C_\psi \|\Phi\|. \end{aligned}$$

Continuity of f_{0q} and f_{10} is proved similarly with $\sup|h(x)|$ replaced by $\|h\| = [\int |h(x)|^2 dx]^{1/2}$.

(d) Now define the operators Θ_{1q} and Θ_{0q} on $\mathfrak{D}$ by

$$\begin{aligned} &\Theta_{1q} \psi_n(x_1 \ldots x_n) \\ &\quad = B_n \sum_{j=1}^{n-q+1} h(x_j) \psi_n(x_1 \ldots x_j \ldots x_j \ldots x_{n-q+1}), \end{aligned} \tag{5}$$

where $B_n = \sqrt{n(n-1)\cdots(n-q+2)}$ and

$$\Theta_{0q} \psi_n(x_1 \ldots x_n) = C_n \int h(x) \psi_n(x \ldots x, x_1 \ldots x_{n-q}) dx \tag{6}$$

where $C_n = \sqrt{n(n-1)\cdots(n-q+1)}$. (Indices are treated below.) One easily checks that Θ_{0q} and Θ_{1q} are the operators associated with f_{0q} and f_{1q}, e.g.,

$$\begin{aligned} \langle \Phi, \Theta_{1q} \Psi \rangle &= \sum_{n=q}^{\infty} B_n \int \varphi_{n-q+1}^*(x_1 \ldots x_{n-q+1}) \\ &\times \sum_{j=1}^{n-q+1} h(x_j) \psi_n(x_1 \ldots x_j \ldots x_j \ldots x_{n-q+1}) d^{n-q+1} x. \end{aligned}$$

If we let $m = n - q + 1$, $x_j \to x$, and $x_{j+k} \to x_{j+k-1}$, then

$$\begin{aligned} \langle \Phi, \Theta_{1q} \Psi \rangle &= \sum_{m=1}^{\infty} B_{m+q} m \int h(x) \\ &\times \langle \varphi_m(x, x_1 \ldots x_{m-1}), \psi_{m+q-1}(x \ldots x, x_1 \ldots x_{m-1}) \rangle dx \\ &= \sum_{m=1}^{\infty} A_m \int h(x) \\ &\times \langle \varphi_m(x, x_1 \ldots x_{m-1}), \psi_{m+q-1}(x \ldots x, x_1 \ldots x_{m-1}) \rangle dx \\ &= f_{1q}(\Phi, \Psi). \end{aligned}$$

When indices are explicitly included Θ_{1q} becomes

$$\begin{aligned} &\Theta_{1q}^{S,T} \psi_n(x_1 r_1 \ldots x_n r_n) \\ &\quad = B_n \sum_{j=1}^{n-q+1} \gamma_j \sum_{t_1 \cdots t_q} \delta_{s r_j} h(x_j) \\ &\quad \times \psi_n(x_1 r_1 \ldots x_j t_1 \ldots x_j t_q \ldots x_{n-q+1} r_{n-q+1}), \end{aligned} \tag{7}$$

where

$$\gamma_j = \begin{cases} 1, & \text{if } \epsilon_s = 1 \text{ or } j = 1, \\ \epsilon_{r_1} \epsilon_{r_2} \cdots \epsilon_{r_{j-1}}, & \text{if } \epsilon_s = -1, \end{cases}$$

arises from the anticommutativity of fermion fields and the commutativity of all other fields.

(e) To show that Θ_{pq} is not closable if $q > 1$, we consider a sequence Ψ^K as in part b, i.e.,

$$\Psi^K = (0, 0, \psi_1^K \ldots \psi_N^K, 0 \ldots),$$

$$\psi_n^K = \prod_{i<j} e^{-K(x_i - x_j)^2} \hat{\psi}_n.$$

Then Ψ^K is in $\mathfrak{D}(\Theta_{pq}) = \mathfrak{D}$, $\Psi^K \to 0$, and $\|\Theta_{pq}(\Psi^K - \Psi^{K+L})\| \to 0$. To prove the last statement, note that, e.g.,

$$\begin{aligned} &\|\Theta_{1q}(\Psi^K - \Psi^{K+L})\| \\ &\quad = \sum_{n=q}^{N} B_n \Big\| \sum_{k=1}^{n-q+1} h(x_k) \Big(1 - \prod_{i\neq k} e^{-qL(x_i - x_k)^2} \prod_{\substack{i<j \\ \neq k}} e^{-L(x_i - x_j)^2} \Big) \\ &\quad \times \prod_{i\neq k} e^{-Kq(x_i - x_k)^2} \prod_{\substack{i<j \\ \neq k}} e^{-K(x_i - x_j)^2} \\ &\quad \times \hat{\psi}_n(x_1 \ldots x_k \ldots x_k \ldots x_{n-q+1}) \Big\| \\ &\quad \leq \sum_{n=q+1}^{N} B_n \Big\| h(x) \prod_i e^{-Kq(x - x_i)^2} \\ &\quad \times \prod_{i<j} e^{-K(x_i - x_j)^2} \hat{\psi}_n(x \ldots x, x_1 \ldots x_{n-q}) \Big\| \\ &\quad \xrightarrow[K\to\infty]{} 0. \end{aligned}$$

Therefore, if Θ_{pq} is closable, $\|\Theta_{pq} \Psi^K\| \xrightarrow[K\to\infty]{} 0$, and it follows that

$$|f_{pq}(\Phi, \Psi^K)| = |\langle \Phi, \Theta_{pq}\Psi^K\rangle|$$
$$\leq \|\Phi\| \, \|\Theta_{pq}\Psi^K\|$$
$$\xrightarrow{K\to\infty} 0.$$

But, as we saw in (b), one can choose $\Phi, \hat{\Psi}$ so that this is false unless $q = 0$ or 1.

(f) We have thus far proven that the only forms which can define closable operators are f_{00}, f_{01}, f_{10}, and f_{11}. We now show that the operators $\Theta_{00}, \Theta_{01}, \Theta_{10}$, and Θ_{11} are indeed closable. To show that an operator is closable, it suffices[5] to show that its adjoint is densely defined. Thus, it suffices to show that $\Theta^\dagger_{pq}$ is defined on $\mathfrak{D}$, i.e., $\Psi \to \langle \Phi, \Theta_{pq}\Psi\rangle = f_{pq}(\Phi, \Psi)$ is continuous on $\mathfrak{D}$ for each fixed Φ in $\mathfrak{D}$. To show this, let $\Phi = (\varphi_0, \ldots, \varphi_N, 0, \ldots)$ be an arbitrary element in $\mathfrak{D}$. Then, e.g.,

$$|f_{11}(\Phi, \Psi)|$$
$$= \sum_{n=1}^{N} n \int h(x_1)\varphi_n(x_1, x_2 \ldots x_n)\psi_n(x_1, x_2 \ldots x_n)d^n x$$
$$\leq \sum_{n=1}^{N} n[\int |h(x_1)\varphi_n(x_1 \ldots x_n)|^2 d^n x$$
$$\times \int |\psi_n(x_1 \ldots x_n)|^2 d^n x]^{1/2}$$
$$\leq N \sup_x |h(x)| \sum_{n=1}^{N} \|\varphi_n\| \|\psi_n\|$$
$$\leq \left(N \sup_x |h(x)| \, \|\Phi\|\right) \|\Psi\|.$$

For f_{01}, $\sup_x |h(x)|$ is replaced by $\|h\|$ since

$$\int h(x)\varphi_n^*(x_1 \ldots x_n)\psi_{n+1}(x, x_1 \ldots x_n)dx d^n x|$$
$$\leq [\int |h(x)\varphi_n(x_1 \ldots x_n)|^2 dx d^n x$$
$$\times \int |\psi_{n+1}(x, x_1 \ldots x_n)|^2 dx d^n x]^{1/2}$$
$$= \|h\| \, \|\varphi_n\| \, \|\psi_{n+1}\|,$$

and for f_{10} either $\sup_x |h(x)|$ or $\|h\|$ can be used. Thus we conclude that the forms f_{00}, f_{01}, f_{10}, and f_{11} do, in fact, define closable operators.

IV. GENERALIZATIONS

We now show that our results do not depend on the specific choice of $\mathfrak{D}$ but only on certain properties which we discuss below. Thus the theorem remains true for much larger classes of $h(x)$ and $\mathfrak{D}$. We discuss generalizations with respect to each part of the theorem separately; they may or may not be simultaneously applicable.

The exclusion of $p, q > 1$ uses only the following two facts which are certainly true for many other domains:

(i) $\Phi^K = \left\{\prod_{i<j} e^{-K(x_i - x_j)^2}\hat{\varphi}_n\right\} \in \mathfrak{D}$ for some suitable $\hat{\varphi}_n$.

(ii) $\exists\ \hat{\varphi}_p, \psi_q \in \mathfrak{D}$ such that $\int h(x)\hat{\varphi}_p^*(x \ldots x)\psi_q(x \ldots x)dx \neq 0$.

For continuity the restriction of $\mathfrak{D}$ to Φ, Ψ with only finitely many nonzero components is important although it could be weakened. Some of the other conditions could be also considerably weakened. For example, f_{11} will be separately continuous if $\sup_{x\in\Gamma}|h(x)|$ exists for all compact Γ; then one would replace $\sup_x |h(x)|$ by $\sup_{x\in\Gamma_\Phi}$ where $\Gamma_\Phi = \cup_{n=1}^N \text{supp}\varphi_n$. Continuity of f_{01} in Ψ and f_{10} in Φ will remain valid if $h(x)$ is merely square-integrable.

The restriction to ψ_n with compact support is used only for proving continuity of f_{1q} with $q > 1$. Since these forms are eventually excluded as being nonclosable, this restriction is ultimately not essential. Elsewhere, it suffices to consider φ_n, ψ_n square-integrable as befits a Fock space element.

Extensions to test functions and form domains forbidden to a specific monomial may be allowed when such a term appears in an appropriate sum. For example, $\int h(x)a_s^\dagger(x)dx$ fails to define an operator when $h(x) \notin L^2$, whereas the specific combination

$$\int [k(x)a_s^\dagger(x)a_s(x) + h^*(x)a_s(x) + a_s^\dagger(x)h(x)]dx$$

can be satisfactorily interpreted as an operator if $h(x)/k(x) \in L^2$. The domain of such an operator—or more to the point, of its defining form—is quite different from our original example in that each vector in the form domain has infinitely many nonzero components. Nevertheless, it is instructive to observe that even these operators can be obtained as formal limits of those we discussed earlier [i.e., choose a sequence of forms f^N defined by functions $h^N(x)$ and $k^N(x)$ with domains $\mathfrak{D}^N$ having the appropriate properties] and that the limiting operator may still be regarded as a bilinear combination of the creation and annihilation operators.

---

*Present address: Department of Mathematics, University of Oregon, Eugene, Oregon.

[1] J. R. Klauder, Commun. Math. Phys. **18**, 307 (1970).
[2] J. R. Klauder, Phys. Rev. Lett, **28**, 769 (1972).
[3] J. R. Klauder, Acta Phys. Austriaca, Suppl. 8, 227 (1971).
[4] By $u \to f(u,v)$ continuous on some domain $\mathfrak{D}$, we mean that if $\|u_n\| \to 0$ and $u_n \in \mathfrak{D}$, $\forall n$ then $f(u, v) \to 0$. An equivalent statement is $|f(u, v)| \leq C_v\|u\|$.
[5] T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1966), Theorem 5.28.

On a certain class of transformations in statistical mechanics*

C. Y. Shen

Department of Mathematics, Simon Fraser University, Burnaby 2, B.C., Canada
(Received 24 October 1972)

In statistical mechanics, the probability densities and the correlation functions are related by a pair of linear operators T, U, which are defined on a suitable subspace of the space of integrable functions and are inverse to each other. Our main result gives the conditions under which they commute with the members of another class $\mathfrak{F}$ of transformations which includes the usual exp and log mappings. Furthermore, the compatibility property possessed by the probability densities of an infinite system and the extension property possessed by the corresponding correlation functions are preserved, in some instances, by the members of $\mathfrak{F}$.

1. INTRODUCTION

In this paper we study the relationship between the linear operators T, U which connect the probability densities and correlation functions in statistical mechanics and a class $\mathfrak{F}$ of transformations which includes the usual exponential and logarithmic mappings.[1] We give, in particular, the conditions for which the operators T, U commute with an element F in $\mathfrak{F}$. Essentially, this is an improvement of Theorem 4.3 of Ref. 1, and involves the sharpening of certain useful results in that paper.[2]

As pointed out in many sources[3,4], the statistical state of an infinite system can be characterized by a family $\{w_\lambda\}$ of compatible densities. Furthermore, the corresponding correlation functions $\rho_\lambda = Tw_\lambda$ possess the extension property. It turns out that the compatibility property and the extension property are preserved, in some instances, by the members of the class $\mathfrak{F}$ of transformations.

The notations and terminologies in this paper follow closely those in Ref. 1. We shall summarize the ones that are pertinent to our present discussion here. Let $(X, \mathfrak{X}, \xi)$ be a σ-finite measure space and let $(X_e, \mathfrak{X}_e, \xi_e)$ be the corresponding exponential measure space. An element x in X_e is an unordered sequence of finite length in X. We shall use $l(x)$ to denote its length and sometime write it as a formal product $x = x_1 \cdots x_n$ of points in X. If we collect the equal factors, we can also express x in the form of $x = t_1^{r_1} t_2^{r_2} \cdots t_m^{r_m}$ where $t_1, \ldots, t_m$ are the distinct members of X. In this representation of x, we can define the index of x to be $I(x) = r_1! \cdots r_m!$. If $x = x_1 \cdots x_m, y = y_1 \cdots y_n$ are two members of X_e, we define their concatenation xy to be $x_1 \cdots x_m y_1 \cdots y_n$. An n-tuple $(x_1, \ldots, x_n)$ of elements of X_e is called a proper n-partition of $x \in X_e$ if $x = x_1 \cdots x_n$ and $x_i \neq 0$ for all $1 \leq i \leq n$ (0 is the empty sequence). The set of all proper partitions of an x that belongs to X_e is denoted by $Q(x)$. Let $\mathfrak{A}$ be the family of complex-valued functions defined on X_e. By introducing a star product on $\mathfrak{A}$ as in Ref. 1, it becomes a commutative algebra with identity 1^*. For each locally analytic function $F(z)$, let $\mathfrak{D}_F = \{\varphi \in \mathfrak{A}: \varphi(0) \in \mathrm{dom}(F)\}$, and define a mapping $F: \mathfrak{D}_F \to \mathfrak{A}$ by

$$F(\varphi) = \sum_{k=0}^{\infty} \frac{F^{(k)}(\varphi(0))}{k!} [\varphi - \varphi(0)1^*]^k. \tag{1.1}$$

This mapping is well defined in view of Theorem 2.2 of Ref. 1. Moreover, Eq. (1.1) is equivalent to

$$F(\varphi)(x) = \begin{cases} F(\varphi(0)), & \text{if } x = 0, \\ \sum{}^* \left\{ \dfrac{F^{(k)}(\varphi(0))}{k!} \varphi(x_1) \cdots \varphi(x_k) : (x_1, \ldots, x_k) \in Q(x) \right\}, & \text{if } x \neq 0. \end{cases} \tag{1.2}$$

The star at the upper right hand corner of the summation notation signifies that each term in the sum is multiplied by the index.

$$I(x_1, \ldots, x_k) = I(x)/I(x_1) \cdots I(x_k)$$

of the corresponding partition. For convenience, we use $\mathfrak{M}$ to denote the family of all measurable functions on X_e and $\mathfrak{L}_1$ the family of integrable functions. A useful fact to keep in mind for the proofs in Sec. 3 is that if $\varphi_1, \varphi_2 \in \mathfrak{L}_1$ then so is $\varphi_1 * \varphi_2$. Furthermore, $\int \varphi_1 * \varphi_2 d\xi_e = \int \varphi_1 d\xi \int \varphi_2 d\xi_e$ and $\int |\varphi_1 * \varphi_2| d\xi_e \leq \int |\varphi_1| d\xi_e \int |\varphi_2| d\xi_e$. Finally, for each $x \in X_e$, one can define an operator D_x: $\mathfrak{A} \to \mathfrak{A}$ by $D_x \varphi(y) = \varphi(xy)$.

2. THE CLASS $\mathfrak{F}$ OF TRANSFORMATIONS

Each locally analytic function $F(z)$ induces a mapping $F: \mathfrak{D}_F \to \mathfrak{A}$ as described in Sec. 1. The family of all such mappings, indexed by the set of all locally analytic functions, is denoted by $\mathfrak{F}$. When $F(z) = \log z$ and $\psi(x) = e^{-\beta H(x) - \tau l(x)}$ is the Gibbs grand canonical zero-density function, $F(\psi)$ is the usual Ursell–Mayer function.[5] When $F(z) = \exp z$, the induced mapping exp: $\mathfrak{A} \to \mathfrak{A}$ corresponds to the Γ-mapping as studied in Refs. 6, 7. The Ursell–Mayer function is useful in statistical mechanics, for it formally simplifies the calculation of the logarithm of the partition function through the expression $\mathrm{Log}\left(\int \psi d\xi_e\right) = \int \log \psi d\xi_e$.[8] It is our hope that by studying a larger class of mappings, one might find some other transformations which are physically significant. Our next theorem provides a useful criterion on the integrability of $F(\varphi)$ and also shows that integration commutes with the mapping operation.

Definition 2.1: Let $F(z)$ be a locally analytic function and let $\varphi \in \mathfrak{D}_F$. We define

$$D(F; \varphi(0)) = \left\{ z : \sum_{k=0}^{\infty} \frac{F^{(k)}(\varphi(0))}{k!} [z - \varphi(0)]^k \text{ converges} \right\}.$$

Theorem 2.1: Let $F(z)$ be a locally analytic function and let $\varphi \in \mathfrak{D}_F$. If $\varphi \in \mathfrak{L}_1$ and $\int \varphi d\xi_e \in D(F; \varphi(0))$, then $F(\varphi) \in \mathfrak{L}_1$ and $\int F(\varphi) d\xi_e = F\left(\int \varphi d\xi_e\right)$.

Proof: First we remark that the sufficiency part of the proof for the Vitali convergence theorem[9] can be easily modified to cover the case when the conditions of convergence in measure is replaced by a.e. convergence.

Next for each positive integer n, let $\psi_n = F_n(\varphi) - F(\varphi(0))1^*, \psi = F(\varphi) - F(\varphi(0))1^*$, and $\bar{\varphi} = \varphi - \varphi(0)1^*$, where $F_n(\varphi) = \sum_{k=0}^{n} F^{(k)}(\varphi(0))/k! [\varphi - \varphi(0)1^*]^k$ and $F(\varphi) = \sum_{k=0}^{\infty} [F^{(k)}(\varphi(0))/k!][\varphi - \varphi(0)1^*]^k$.

Copyright © 1973 by the American Institute of Physics

It is easy to see that ψ_n converges pointwise to ψ. Since $\tilde{\varphi} \in \mathfrak{L}_1$, for each $\epsilon > 0$, there exists a set $E \in \mathfrak{X}_e$ with finite measure such that $\int_{\sim E} |\tilde{\varphi}| d\xi_e < \epsilon$. For such a set E, we have

$$\int_{\sim E} |\psi_n| d\xi_n = \int_{\sim E} |F_n(\varphi) - F(\varphi(0))\mathbf{1}^*| d\xi_e \leq \sum_{k=1}^{\infty} \frac{|F^{(k)}(\varphi(0))|}{k!} \left(\int_{\sim E} |\tilde{\varphi}| d\xi_e\right)^k \leq \epsilon M,$$

where $M > 0$ is independent from ϵ provided ϵ is sufficiently small. By a similar argument, one can show that for every $\epsilon > 0$ there exists a $\delta > 0$ such that $\int_E |\psi_n| d\xi_e \leq \epsilon M$ if $\xi_e(E) < \delta$. Hence by the sufficiency part of the Vitali convergence theorem, we have ψ_n converges in $\mathfrak{L}_1$-norm to ψ which, of course, implies that $F_n(\varphi)$ converges to $F(\varphi)$ in the same sense. The second part of the theorem now follows from the inequality

$$\left| \int F(\varphi) d\xi_e - F\left(\int \varphi d\xi_e\right) \right| \leq \int |F(\varphi) - F_n(\varphi)| d\xi_e + \left| \int F_n(\varphi) d\xi_e - F\left(\int \varphi d\xi_e\right) \right|.$$

The theorem below is a generalization of Theorem 4.1 in Ref. 1 and Eq. 4.38 in Ref. 10. It is used to prove results in the next section. Since the proof is by induction on $l(x)$, and is essentially the same as that of Theorem 4.1 in Ref. 1, we shall omit the proof here.

Theorem 2.2: Let x be a member of X_e with $l(x) \geq 1$, and let $F(z)$ be a locally analytic function with derivatives $F^{(k)}(z)$, $k = 1, 2, \cdots$. Then for each $\varphi \in \mathfrak{D}_F$,

$$D_x F(\varphi) = \sum^* \{(1/k!) F^{(k)}(\varphi) * D_{x_1}\varphi * \cdots * D_{x_k}\varphi : (x_1, \ldots, x_k) \in Q(x)\}.$$

3. THE MAIN RESULTS

The probability density w and the correlation function ρ in statistical mechanics are related formally as

$$\rho(x) = Tw(x) = \int D_x w(y) d\xi_e(y), \qquad w(x) = U\rho(x) = \int (-1)^{l(y)} D_x \rho(y) d\xi_e(y). \tag{3.1}$$

For $x \in X_e$, we write $x = x_1 \cdots x_n$, where $x_1, \ldots, x_n$ are members of X, and interpret x as a distribution in X of n identical particles of a mechanical system. If Δx_i, $1 \leq i \leq n$, are small volumes containing each x_i, then $w(x_1 \cdots x_n)\xi_e(\Delta x_1) \cdots \xi_e(\Delta x_n)$ is approximately equal to the probability of finding exactly n particles, one in each Δx_i, and $\rho(x_1 \cdots x_n)\xi_e(\Delta x_1) \cdots \xi_e(\Delta x_n)$ is the probability of finding n particles, one in each Δx_i, without regard to what happens elsewhere. Thus, as suggested by its physical interpretation, the correlation function does not depend so much on the size of the system and is therefore often used to establish the infinite state.[11] The probability densities and correlation functions of infinite systems have many properties that they do not have in the case of finite systems. For example, suppose X is a finite region in E^3 and w is the Gibbs grand canonical density function defined on X_e. Let $\rho = Tw$ be the corresponding correlation function. Now suppose Y is a subregion of X.

Then the 'induced' probability density function w_Y of w to Y_e will not in general agree with the Gibbs grand canonical density function for Y_e. Moreover, $\rho_Y = Tw_Y$ may not agree with the restriction of ρ to Y_e. The next three theorems in this section furnish various alternative representations of w and ρ, and show that, in some cases, the relationships which exist between w and ρ for infinite systems are preserved among those representations.

In Ref. 1, Theorem 3.1 shows that Eq. (3.1) defines two linear operators, inverse to each other, on the subspace $\mathfrak{N}$ of $\mathfrak{L}_1$:

$\mathfrak{N} = \{\varphi \in \mathfrak{M}$: for each complex numer z and all $x \in X_e$, $z^{l} D_x \varphi \in \mathfrak{L}_1\}$. The following two examples show that $\mathfrak{N}$ is generally not closed in $\mathfrak{L}_1$ and T, U are not bounded. For simplicity, let X be the unit interval with Lebesgue measure.

(1) We define a sequence of functions $\{\varphi_k\}$ by

$$\varphi_k(x_1 \cdots x_n) = \begin{cases} (n-1)!/n + 1 & \text{if } 1 \leq n \leq k \\ 0 & \text{otherwise,} \end{cases}$$

and a function φ by

$$\varphi(x_1 \cdots x_n) = \begin{cases} 0 & \text{if } n = 0 \\ (n-1)!/n + 1 & \text{if } n \geq 1. \end{cases}$$

Clearly, $\varphi_k \in \mathfrak{N}$ for all k and $\int |\varphi_k - \varphi| d\xi_e \to 0$ as $k \to \infty$ but $\varphi \notin \mathfrak{N}$.

(2) We define $\{\varphi_k\}$ by

$$\varphi_k(x_1 \ldots x_n) = \begin{cases} n! & \text{if } n = k \\ 0 & \text{otherwise.} \end{cases}$$

One finds that $\int |T\varphi_k| d\xi_e = 2^k$ and $\int |\varphi_k| d\xi_e = 1$ for all integers $k \geq 1$.

Theorem 3.1: Let $F(z)$ be a locally analytic function and let $\varphi \in \mathfrak{D}_F$. If $\varphi \in \mathfrak{N}$ and $\int \varphi d\xi_e$, $\int (-1)^{l} \varphi d\xi_e \in D(F; \varphi(0))$, then $F(\varphi) \in \mathfrak{N}$ and $T(F(\varphi)) = F(T(\varphi))$, $U(F(\varphi)) = F(U(\varphi))$.

Proof: Since $\varphi \in \mathfrak{L}_1$ and $\int \varphi d\xi_e \in D(F; \varphi(0))$, in view of Theorem 2.1, $F^{(k)}(\varphi) \in \mathfrak{L}_1$ and $\int F^{(k)}(\varphi) d\xi_e = F^{(k)}\left(\int \varphi d\xi_e\right)$ for all $k = 0, 1, 2, \cdots$.

By using Theorem 2.2, we obtain

$$z^{l} D_x F(\varphi) = \sum^* \left\{ \frac{1}{k!} F^{(k)}(z^{l}\varphi) * z^{l} D_{x_1}\varphi * \cdots * z^{l} D_{x_n}\varphi : (x_1, \ldots, x_n) \in Q(x) \right\}.$$

From this equation, one sees that $F(\varphi) \in \mathfrak{N}$.

Clearly, $T\varphi$, $U\varphi \in \mathfrak{D}_F$ and $T(F(\varphi))(0) = F(T\varphi)(0)$, $U(F(\varphi))(0) = F(U\varphi)(0)$.

For $x \in X_e$, $l(x) \geq 1$, we have

$$\begin{aligned} T(F(\varphi))(x) &= \int D_x F(\varphi) d\xi_e \\ &= \sum^* \{(1/k!) \int F^{(k)}(\varphi) d\xi_e \int D_{x_1}\varphi d\xi_e \cdots \int D_{x_n}\varphi d\xi_e : (x_1, \ldots, x_n) \in Q(x)\} \\ &= \sum^* \{(1/k!) F^{(k)}\left(\int \varphi d\xi_e\right) T\varphi(x_1) \cdots T\varphi(x_n) : (x_1, \ldots, x_n) \in Q(x)\} \\ &= F(T\varphi)(x). \end{aligned}$$

A similar calculation works for U.

Definition 3.1: Let $\{X_\lambda\}$ be a strictly increasing family of measurable sets in X indexed by $[0, \infty)$, and let

$\{w_\lambda\}$ be a family of measurable functions such that for each λ, $\mathrm{dom}(w_\lambda) = (X_\lambda)_e$. This family of functions is said to be compatible (to possess the extension property) if for any pair of nonnegative reals $\lambda_1 < \lambda_2$, we have

$$w_{\lambda_1}(x) = \left(\int_{X_{\lambda_2} \sim X_{\lambda_1}}\right)_e w_{\lambda_2}(xy)d\xi_e(y)$$

for all $x \in (X_{\lambda_1})_e$ [the restriction of w_{λ_2} to $(X_{\lambda_1})_e$ is equal to w_{λ_1}].

Theorem 3.2: Let $\{w_\lambda\}$ be a family of compatible integrable functions indexed by $[0, \infty)$, and let $F(z)$ be a locally analytic function such that $w_\lambda \in \mathfrak{D}_F$ and $\int w_\lambda d\xi_e \in D(F; w_\lambda(0))$ for all λ. Then $\{Fw_\lambda\}$ are compatible and $\int F(w_\lambda)d\xi_e = F(\int w_\lambda d\xi_e)$.

Proof: Suppose $X_{\lambda_1} \subset X_{\lambda_2}$. Let $Z = X_{\lambda_2} \sim X_{\lambda_1}$. Then for all $x \in (X_{\lambda_1})_e$, $l(x) \geq 1$,

$$\begin{aligned}
&\int_{Z_e} Fw_{\lambda_2}(xy)d\xi_e(y) = \int_{Z_e} \dot{D}_x Fw_{\lambda_2}(y)d\xi_e(y) \\
&\quad = \int_{Z_e} \sum{}^*\{(1/k!)F^{(k)}(w_{\lambda_2})^* D_{x_1} w_{\lambda_2}{}^* \cdots {}^* D_{x_n} w_{\lambda_2} : \\
&\qquad (x_1, \ldots, x_n) \in Q(x)\}\, d\xi_e \\
&\quad = \sum{}^*\{(1/k!)F^{(k)}(w_{\lambda_1}(0))\, w_{\lambda_1}(x_1) \cdots w_{\lambda_1}(x_n) : \\
&\qquad (x_1, \ldots, x_n) \in Q(x)\} \\
&\quad = F(w_{\lambda_1})(x)
\end{aligned}$$

when $x = 0$, the calculation is trivial.

It follows easily from Eq. (1.2) that:

Theorem 3.3: Let $\{\rho_\lambda\}$ be a family of functions indexed by $[0, \infty)$ and possess the extension property. Let $F(z)$ be a locally analytic function such that $\rho_\lambda \in \mathfrak{D}_F$ for all λ. Then $\{F\rho_\lambda\}$ have the extension property.

---

*This work is partially supported by NRC Grant A8511.

[1]C. Y. Shen, J. Math. Phys. **13**, 754 (1972).
[2]See Ref. 1, Theorem 3.4, Theorem 4.1.
[3]C. Y. Shen and D. S. Carter, J. Math. Phys. **12**, 1263 (1971).
[4]D. S. Carter and P. M. Prenter, Z. Wahrschein Lichkeitstheorie **21**, 1 (1972).
[5]See Ref. 1.
[6]D. Ruelle, Rev. Mod. Phys. **36**, 580 (1964).
[7]G. Gallavotti and S. Miracle-Sole, Commun. Math. Phys. **7**, 274 (1968).
[8]J. Schwartz, "Statistical mechanics," lecture notes (Courant Institute of Mathematical Sciences, New York University, 1957). See p. 59.
[9]R. Bartle, *The Elements of Integration* (Wiley, New York, 1966), see Theorem 7.13.
[10]D. Ruelle, *Statistical Mechanics* (Benjamin, New York, 1968).
[11]D. Ruelle, Ann. Phys. (N.Y.) **25**, 109 (1963).

On nonrelativistic field theory: Interpolating fields and long-range forces

James R. Klein

Department of Physics, Iona College, New Rochelle, New York 10801

Irving I. Zinnes

Department of Physics, Fordham University, Bronx, New York 10458
(Received 31 January 1972; final revised manuscript received 15 January 1973)

A field-theoretic version of nonrelativistic, multichannel scattering theory is presented in which the particles interact via two-body potentials, which can include both long-range (Coulomb, etc.) and short-range potentials. By utilizing known first-quantized results, in and out fields and interpolating fields are constructed both for simple and composite particles, and an asymptotic condition using strong convergence is proved. Complexities present when long-range potentials are involved are discussed.

1. INTRODUCTION

In this paper we consider a rigorous formulation of a nonrelativistic field theory for a system of identical spinless bosons[1] which can interact via both short-range and long-range forces and which can form composite particles (bound states) that participate in scattering processes. Although we feel that this system is of sufficient intrinsic interest to merit study, our primary motivation is to investigate the effect of long-range interactions in a field-theoretic formalism that closely parallels the relativistic situation.

Recall that, in relativistic field theory, the particle interpretation (scattering theory) is customarily given in terms of "formally free" asymptotic fields which are obtained from an asymptotic condition as limits of an interpolating field as the time $t \to \pm\infty$.[2-6] It is usually assumed that there are no mass zero particles present; to our knowledge, there are no rigorous results for the zero-mass case. Since long-range interactions (e.g., Coulomb) would seem to be the nonrelativistic analog of zero-mass particles, difficulties due to the presence of long-range forces will be emphasized in the following. We shall be especially interested in defining interpolating fields appropriate to the scattering problem involving long-range as well as short-range forces and in studying the asymptotic forms of these fields.

An outline of the paper is as follows. In Sec. 2, a brief discussion of recent work in time-dependent scattering theory involving long-range forces is given; then the model studied is defined and the basic features of known results are abstracted and listed as postulates. In Sec. 3, the second-quantized form of the theory is given; asymptotic fields are constructed and their properties obtained. In Sec. 4, interpolating fields are constructed and are shown to satisfy a strong asymptotic condition. In Sec. 5, it is shown that the proposed interpolating fields yield known results[7-11] when restricted to the short-range case. Finally, a discussion of complexities due to long-range forces is given.

2. FIRST-QUANTIZED TIME-DEPENDENT SCATTERING THEORY

Consider first the case of potential scattering described by a self-adjoint Hamiltonian $H = H_0 + V$. In Jauch's formulation,[12] the asymptotic condition is expressed through the requirement of the existence of wave operators $\Omega_\pm$ where

$$\Omega_\pm = \underset{t\to\pm\infty}{\text{s-lim}}\ V(t)^* U(t), \tag{2.1}$$

$$V(t) = e^{-iHt}, \quad U(t) = e^{-iH_0 t}. \tag{2.2}$$

The wave operators $\Omega_\pm$ are isometries and their ranges $\mathfrak{R}_\pm$ satisfy $\mathfrak{R}_\pm \subseteq \mathfrak{B}^\perp$, where $\mathfrak{B}^\perp$ is the orthogonal complement of the bound states of H. The intertwining property

$$V(t)\Omega_\pm = \Omega_\pm U(t) \tag{2.3}$$

is satisfied. Furthermore, it is often assumed (and sometimes proved for a restricted class of potentials) that asymptotic completeness holds, i.e.,

$$\mathfrak{R}_+ = \mathfrak{R}_- = \mathfrak{B}^\perp. \tag{2.4}$$

Dollard[13] has shown that (2.1) does *not* hold for the Coulomb potential; however, by introducing a modified free propagator $\hat{U}(t)$ he was able to establish the existence of the limits in (2.1) if $U(t)$ is replaced by $\hat{U}(t)$. Furthermore, Dollard gave a physical justification for this modification and showed that (2.3) remains valid. For the case of pure Coulomb scattering, (2.4) was shown to hold also.

Recently several general approaches to time-dependent scattering theory which are also applicable to long-range potentials have been proposed.[13-19] These approaches differ in their basic physical assumptions, but in the case of potential scattering for long-range forces they all lead to a modification of (2.1) of the form

$$\Omega_\pm = \underset{t\to\pm\infty}{\text{s-lim}}\ V(t)^* \hat{U}(t), \tag{2.5}$$

where

$$\hat{U}(t) = e^{-i[H_0 t + H_0'(t)]} = U(t)U'(t) \tag{2.6}$$

and $H_0'(t)$ is a function of H_0 and t. In all these cases the intertwining property (2.3) is satisfied. A necessary and sufficient condition[18,20] for the validity of (2.3) is

$$\underset{|t|\to\infty}{\text{s-lim}}\ U'(t+s) - U'(t) = 0. \tag{2.7}$$

Specific forms for $\hat{U}(t)$ and their physical motivations are given for the Coulomb potential by Dollard[13]; for $V(\mathbf{r}) = 1/r^\beta$, $3/4 \le \beta < 1$, by Amrein, Martin, and Misra[15]; and for $V(\mathbf{r}) = O(1/r^\beta)$, $\beta > 0$, by Buslaev and Matveev.[17]

A generalization of the single-channel scattering theory to the multi-channel case was also given by Jauch.[21] This formulation is suitable for the nonrelativistic n-body problem with short-range forces in which composite particles undergoing rearrangement collisions

Copyright © 1973 by the American Institute of Physics

may be included in the scattering processes. With an appropriate modification of the channel propagators, long-range forces can be included in this formulation also.

In this paper we consider a system of identical spinless bosons of mass μ interacting through two-body potentials. The n-body Hamiltonian $H^{(n)}$ is a self-adjoint operator on $L^2(R^{3n})$ which has the form

$$H^{(n)} = \sum_{i=1}^{n} \frac{-\Delta_i}{2\mu} + \sum_{i<j} V(\mathbf{x}_i - \mathbf{x}_j) \tag{2.8}$$

on a suitable dense domain.[22] [We use the same symbol for the operator in (2.8) and its self-adjoint extension; this will cause no confusion here.] Furthermore, to keep the notation simple, we assume that the system is capable of forming only one composite particle: a one-level two-body bound state. Consider the two-particle Hamiltonian $H^{(2)}$; in terms of the c.m. coordinate $\mathbf{X} = 1/2(\mathbf{x}_1 + \mathbf{x}_2)$ and relative coordinate $\mathbf{y} = \mathbf{x}_1 - \mathbf{x}_2$, we can write

$$H^{(2)} = \frac{-\Delta(\mathbf{X})}{2M} + h(\mathbf{y}), \qquad h(\mathbf{y}) = \frac{-\Delta(\mathbf{y})}{M} + V(\mathbf{y}). \tag{2.9}$$

We assume $h(\mathbf{y})$ has one and only one bound state $\Phi_B(\mathbf{y})$, where

$$\begin{aligned} &h(\mathbf{y})\Phi_B(\mathbf{y}) = \epsilon\Phi_B(\mathbf{y}), \qquad \epsilon < 0, \\ &\int d\mathbf{y}\,|\Phi_B(\mathbf{y})|^2 = 1. \end{aligned} \tag{2.10}$$

Hence there exists in the model a composite particle with mass $M = 2\mu$, internal energy ϵ, and internal wavefunction $\Phi_B(\mathbf{y})$.

Next we state standard results for the multichannel scattering theory for n-body Hamiltonians which include those in (2.8). A channel α can be characterized by $\alpha = (H_\alpha, \hat{U}_\alpha(t), \mathfrak{D}_\alpha)$. H_α is the free channel Hamiltonian which governs the free motion of the fragments (moving without mutual interaction) in channel α and is just the sum of the kinetic energies and internal energies of the fragments. For example, for our Hamiltonian $H^{(n)}$ in (2.8),

$$\begin{aligned} &H_{lm} = \sum_{i=1}^{l} \frac{-\Delta(\mathbf{x}_i)}{2\mu} + \sum_{j=1}^{m} \frac{-\Delta(\mathbf{X}_j)}{2M} + m\epsilon, \\ &\mathbf{X}_j = \tfrac{1}{2}(\mathbf{x}_{l+2j-1} + \mathbf{x}_{l+2j}), \end{aligned} \tag{2.11}$$

describes the free motion in the channel $\alpha = \alpha_{lm}$, which corresponds to the free motion in which the *first* l particles are simple particles and the rest are paired *consecutively* to form m composite particles, so that $n = l + 2m$.

Let $U_\alpha(t)$ denote the free channel propagator

$$U_\alpha(t) = e^{-iH_\alpha t}. \tag{2.12}$$

$\hat{U}_\alpha(t)$ will denote the modified, renormalized or simply the channel propagator which we assume is unitary.[23] For short-range potentials, $\hat{U}_\alpha = U_\alpha$; for long-range potentials $\hat{U}_\alpha(t) = U_\alpha(t)U'_\alpha(t)$, where $U'_\alpha(t)$ is a suitable function of H_α and t. (See Dollard[13] for a specific example for Coulomb forces.)

Finally, $\mathfrak{D}_\alpha$ is a subspace consisting of all possible free states for the channel α. For example, for the channel α_{lm} [see Eq. (2.11)],

$$\mathfrak{D}_{lm} = \Big\{ g(\mathbf{x}_1, \ldots, \mathbf{x}_l, \mathbf{X}_1, \ldots, \mathbf{X}_m) \prod_{j=1}^{m} \Phi_B(\mathbf{y}_j) \Big\}, \tag{2.13}$$

where $g \in L^2(R^{3l+3m})$, and y_j is the relative coordinate $\mathbf{y}_j = \mathbf{x}_{l+2j-1} - \mathbf{x}_{l+2j}$.

We postulate that our theory possesses the following properties. {In the following $\mathfrak{H}$ is the Hilbert space of n spinless particles $[= L^2(R^{3n})]$.}

A. (asymptotic condition): Let $V(t) = e^{-iHt}$ and let E_α be the projector onto $\mathfrak{D}_\alpha$: $E_\alpha\mathfrak{H} = \mathfrak{D}_\alpha$. Then appropriate channel propagators $\hat{U}_\alpha(t)$ and wave operators $\Omega^\pm_\alpha$ exist such that

$$\Omega^\pm_\alpha = \operatorname*{s-lim}_{t\to\pm\infty} V(t)^* \hat{U}_\alpha(t) E_\alpha. \tag{2.14}$$

B. (orthogonality of ranges): Let $\mathfrak{R}^\pm_\alpha$ denote the ranges of $\Omega^\pm_\alpha$.

Then

$$\mathfrak{R}^+_\alpha \perp \mathfrak{R}^+_\beta, \qquad \mathfrak{R}^-_\alpha \perp \mathfrak{R}^-_\beta, \qquad \alpha \neq \beta. \tag{2.15}$$

C. (partial isometries): Let $F^\pm_\alpha$ be the projectors onto $\mathfrak{R}^\pm_\alpha$: $F^\pm_\alpha\mathfrak{H} = \mathfrak{R}^\pm_\alpha$. The wave operators $\Omega^\pm_\alpha$ are partial isometries and

$$\Omega^{\pm *}_\beta \Omega^\pm_\alpha = \delta_{\alpha\beta} E_\alpha, \qquad \Omega^\pm_\alpha \Omega^{\pm *}_\beta = F^\pm_\alpha. \tag{2.16}$$

D. (intertwining property):

$$V(t)\Omega^\pm_\alpha = \Omega^\pm_\alpha U_\alpha(t). \tag{2.17}$$

E. (asymptotic completeness): Let $\mathfrak{R}_\pm = \overline{\sum_\oplus \mathfrak{R}^\pm_\alpha}$, where the bar means closure. Then

$$\mathfrak{R}_+ = \mathfrak{R}_- = \mathfrak{B}^\perp, \tag{2.18}$$

where $\mathfrak{B}^\perp$ is the orthogonal complement of the bound state subspace of H.

Remarks:

1. We regard A–E as reasonable requirements for nonrelativistic scattering theory and will use them to construct the second-quantized theory for the Hamiltonian $H^{(n)}$ in (2.8). It is not implied that these requirements are independent. For example, for short-range potentials, B–D are implied by A.[21] It should be emphasized that A–D have been shown to hold for short-range potentials and for Coulomb plus short-range potentials[13] and are very likely satisfied for potentials $V = O(1/r^\beta)$, $\beta > 0$.[15-18] Only weak results are known concerning asymptotic completeness E for arbitrary n.[24] However, E is not crucial for our results (see Theorem 1 below).

2. Properties A–E have been stated in a form appropriate to the description of scattering of distinguishable particles. Although we are interested in identical particles in this paper, it is convenient to start with the unsymmetrized states. [Note that the channel Hamiltonians H_{lm} and subspaces $\mathfrak{D}_{lm}$ in (2.11) and (2.13) are *not* invariant under particle interchange.] The symmetrization will be taken care of in the second-quantized construction.

From now on in this paper we consider the system of identical particles whose n-particle Hamiltonian is $H^{(n)}$ in (2.8). We assume that the interparticle potential V is such that properties A–E are satisfied in the first-quantized n-particle theory in which the particles are treated as distinguishable. Furthermore, we make the additional technical assumption that only one type of composite particle can be formed and that it has mass $M = 2\mu$, even parity internal wavefunction Φ_B, and in-

ternal energy ϵ as in (2.10). It follows that any channel in the n-particle first-quantized theory can be obtained by a permutation of coordinates defining $\alpha_{lm} = (H_{lm}, \hat{U}_{lm}(t), \mathfrak{D}_{lm})$ for some l,m such that $n = l + 2m$. [See (2.11) and (2.13) for definitions of H_{lm} and $\mathfrak{D}_{lm}$.] To avoid superfluous notation, it will always be assumed that $l + 2m = n$ unless noted explicitly otherwise.

It should be noted that the technical assumption of the presence of only one type of composite particle is purely for simplicity of notation; it will be clear that the same construction can be carried out for any number of different composite particles.

3. SECOND QUANTIZATION AND ASYMPTOTIC FIELDS

We introduce the conventional second-quantized formalism.[25] States are vectors in a Fock space $\mathfrak{F}$. In the configuration space representation, the vectors are sequences $\Phi = (\Phi^{(0)}, \Phi^{(1)}(\mathbf{x}), \ldots, \Phi^{(j)}(\mathbf{x}_1, \ldots, \mathbf{x}_j), \ldots)$, where $\Phi^{(0)}$ is a complex number and $\Phi^{(j)}$ is a symmetric function of the variables $\mathbf{x}_1, \ldots, \mathbf{x}_j$ and is an element of $L^2(R^{3j})$. The scalar product is given by

$$(\Psi, \Phi) = \Psi^{(0)*}\Phi^{(0)} + \sum_{m=1}^{\infty} (\Psi^{(n)}, \Phi^{(n)})_n, \tag{3.1}$$

where $(\ ,\ )_n$ is the scalar product in $L^2(R^{3n})$ and the norm is $\|\Phi\| = (\Phi, \Phi)^{1/2}$. An important subset of $\mathfrak{F}$ is $\tilde{\mathfrak{F}}$, called the set of finite vectors, defined by $\tilde{\mathfrak{F}} = \{\Phi \in \mathfrak{F} \mid \Phi^{(j)} = 0 \text{ for } j > N \text{ for some } N\}$. $\tilde{\mathfrak{F}}$ is dense in $\mathfrak{F}$.

The Schrödinger field operator $\psi(\mathbf{x}, 0)$ satisfies the formal commutation relations

$$\begin{aligned} &[\psi(\mathbf{x},0), \psi(\mathbf{y},0)] = [\psi(\mathbf{x},0)^*, \psi(\mathbf{y},0)^*] = 0, \\ &[\psi(\mathbf{x},0), \psi(\mathbf{y},0)^*] = \delta(\mathbf{x} - \mathbf{y}). \end{aligned} \tag{3.2}$$

The second-quantized Hamiltonian $\mathcal{H}$ can be formally represented as

$$\begin{aligned} \mathcal{H} = &\int d\mathbf{x}\, \psi(\mathbf{x},0)^*(-\Delta/2\mu)\psi(\mathbf{x},0) \\ &+ \tfrac{1}{2} \int d\mathbf{x}d\mathbf{x}' \psi(\mathbf{x},0)^*\psi(\mathbf{x}',0)^* V(\mathbf{x} - \mathbf{x}')\psi(\mathbf{x}',0)\psi(\mathbf{x},0), \end{aligned} \tag{3.3}$$

and has the property that

$$\{e^{-i\mathcal{H}t} \mid \Phi\rangle\}^{(n)}(\mathbf{x}_1, \ldots, \mathbf{x}_n) = e^{-iH^{(n)}t}\{\mid \Phi\rangle\}^{(n)}(\mathbf{x}_1, \ldots, \mathbf{x}_n). \tag{3.4}$$

For a detailed mathematical discussion of properties (domain, self-adjointness, etc.) of $\mathcal{H}$ and related operators, the reader is referred to the papers of Cook[26] and Schroeck.[27]

The "smeared" field operators $\psi(f), \psi(f)^*$ can be represented by

$$\begin{aligned} &\psi(f) = \int d\mathbf{x} f(\mathbf{x})^* \psi(\mathbf{x},0), \quad f \in L^2(R^3), \\ &\psi(f)^* = \int d\mathbf{x} f(\mathbf{x}) \psi(\mathbf{x},0)^* \end{aligned} \tag{3.5}$$

and are unbounded operators on $\mathfrak{F}$. However, $\tilde{\mathfrak{F}}$ is an invariant domain for them. Hence we can apply polynomials and finite sums of polynomials in the smeared fields to the vectors in $\tilde{\mathfrak{F}}$ with impunity. The smeared fields have the following useful representations:

$$\begin{aligned} &\{\psi(f)^* \mid \Phi\rangle\}^{(n)}(\mathbf{x}_1, \ldots, \mathbf{x}_n) \\ &\quad = \begin{cases} 0, & n = 0, \\ n^{1/2} S_n [f(\mathbf{x}_1)\Phi^{(n-1)}(\mathbf{x}_2, \ldots, \mathbf{x}_n)], & n \neq 0, \end{cases} \end{aligned} \tag{3.6}$$

$$\begin{aligned} &\{\psi(f) \mid \Phi\rangle\}^{(n)}(\mathbf{x}_1, \ldots, \mathbf{x}_n) \\ &\quad = (n+1)^{1/2} \int d\mathbf{y} f(\mathbf{y})^* \Phi^{(n+1)}(\mathbf{y}_1 \mathbf{x}_1, \ldots, \mathbf{x}_n). \end{aligned} \tag{3.7}$$

In (3.6), $S_n = (n!)^{-1} \sum_p U_p$ is the symmetrizer, where $\sum_p$ means a sum over all the permutations of the n position variables; and U_p is the unitary operator which effects the permutations. S_n is a projection operator on $L^2(R^{3n})$.

The Heisenberg field operator $\psi(\mathbf{x}, t)$ is defined by

$$\psi(\mathbf{x}, t) = e^{i\mathcal{H}t}\psi(\mathbf{x}, 0)e^{-i\mathcal{H}t}. \tag{3.8}$$

In the following, we will use the first-quantized results to construct in and out fields that can be used to define the scattering states and the S matrix in the field-theoretic version. These fields will be shown to have the properties that one usually postulates in relativistic quantum field theory.

Let $|\Phi\rangle \in \mathfrak{F}$ and $f(\mathbf{x}) \in L^2(R^3)$; we *define* the operators $A_{ex}(f)^*$ and $B_{ex}(f)^*$ by (ex stands for either "in" or "out"):

$$\begin{aligned} &\{A^{\text{out}}_{\text{in}}(f)^* \mid \Phi\rangle\}^{(0)} = 0, \\ &\{A^{\text{out}}_{\text{in}}(f)^* \mid \Phi\rangle\}^{(n+1)}(\mathbf{x}_1, \ldots, \mathbf{x}_{n+1}) \\ &\quad = (n+1)^{1/2} \sum_{\substack{l,m \\ (n=l+2m)}} \frac{n!}{l!m!2^m} S_{n+1}\Omega_{l+1,m} \\ &\qquad \times [f(\mathbf{x}_1)\Omega^{\pm *}_{lm}\{\mid \Phi\rangle\}^{(n)}(\mathbf{x}_2, \ldots, \mathbf{x}_{n+1})], \end{aligned} \tag{3.9}$$

$$\begin{aligned} &\{B^{\text{out}}_{\text{in}}(f)^* \mid \Phi\rangle\}^{(i)} = 0, \quad i = 0, 1, \\ &\{B^{\text{out}}_{\text{in}}(f)^* \mid \Phi\rangle\}^{(n+2)}(\mathbf{x}_1, \ldots, \mathbf{x}_{n+2}) \\ &\quad = \left(\frac{(n+1)(n+2)}{2}\right)^{1/2} \sum_{\substack{l,m \\ (n=l+2m)}} \frac{n!}{l!m!2^m} S_{n+2}\Omega^{\pm}_{l,m+1} \\ &\qquad \times [(\Omega^{\pm *}_{lm}\{\mid \Phi\rangle\}^{(n)}(\mathbf{x}_1, \ldots, \mathbf{x}_n))\, f(\mathbf{X})\Phi_B(\mathbf{y})], \end{aligned} \tag{3.10}$$

where $\mathbf{X} = \frac{1}{2}(\mathbf{x}_{n+1} + \mathbf{x}_{n+2})$, $\mathbf{y} = \mathbf{x}_{n+1} - \mathbf{x}_{n+2}$.

It follows from the definition of the scalar product (3.1) that the operators adjoint to $A_{ex}(f)^*$ and $B_{ex}(f)^*$, called $A_{ex}(f)$ and $B_{ex}(f)$ respectively, have the representations:

$$\begin{aligned} &\{A^{\text{out}}_{\text{in}}(f) \mid \Phi\rangle\}^{(n)}(\mathbf{x}_1, \ldots, \mathbf{x}_n) \\ &\quad = (n+1)^{1/2} \sum_{\substack{l,m \\ (n=l+2m)}} \frac{n!}{l!m!2^m} S_n \Omega^{\pm}_{lm} \\ &\qquad \times \left[\int d\mathbf{x} f(\mathbf{x})^* (\Omega^{\pm *}_{l+1,m}\{\mid \Phi\rangle\}^{(n+1)})(\mathbf{x}, \mathbf{x}_1, \ldots, \mathbf{x}_n)\right], \end{aligned} \tag{3.11}$$

$$\begin{aligned} &\{B^{\text{out}}_{\text{in}}(f) \mid \Phi\rangle\}^{(n)}(\mathbf{x}_1, \ldots, \mathbf{x}_n) \\ &\quad = \left(\frac{(n+1)(n+2)}{2}\right)^{1/2} \sum_{\substack{l,m \\ (n=l+2m)}} \frac{n!}{l!m!2^m} S_n \Omega^{\pm}_{lm} \\ &\qquad \times \left[\int d\mathbf{X} d\mathbf{y} f(\mathbf{X})^* \Phi_B(\mathbf{y}) (\Omega^{\pm *}_{l,m+1}\{\mid \Phi\rangle\}^{(n+2)}(\mathbf{x}_1, \ldots, \mathbf{x}_{n+2}))\right], \end{aligned}$$

where $\mathbf{X} = \frac{1}{2}(\mathbf{x}_{n+1} + \mathbf{x}_{n+2})$, $\quad \mathbf{y} = \mathbf{x}_{n+1} - \mathbf{x}_{n+2}$. (3.12)

These operators are unbounded, but are well defined at least on $\tilde{\mathfrak{F}}$ (which is dense in $\mathfrak{F}$). $A_{ex}(f)^*$ $[A_{ex}(f)]$ and

$B_{ex}(f)^*$ [$B_{ex}(f)$] are one-particle and two-particle bound-state creation [annihilation] operators, respectively. The physical interpretation of these operators is transparently displayed in the representations (3.9)–(3.12). For example, $B_{in}(f)^*|\Phi\rangle$ is a scattering state with the same initial configuration ($t \to -\infty$) as $|\Phi\rangle$ except that it contains an extra composite particle whose "free" c.m. wavepacket is described by $f(\mathbf{X})$.

Consider the state

$$|\Psi_{lm}(\mathrm{ex})\rangle = A_{ex}(f_1)^* \ldots A_{ex}(f_l)^* B_{ex}(g_1)^* \ldots B_{ex}(g_m)^* |0\rangle. \quad (3.13)$$

Using the representations (3.9) and (3.10), we find that

$$\left\{\left|\Psi_{lm}\left(\begin{matrix}\mathrm{out}\\ \mathrm{in}\end{matrix}\right)\right\rangle\right\}^{(r)}(\mathbf{x}_1, \ldots, \mathbf{x}_r) = \delta_{rn} 2^{-m}(n!)^{1/2} S_n \Omega^{\pm}_{lm}\Big(f_1(\mathbf{x}_1)\ldots f_l(\mathbf{x}_l) g_1(\mathbf{X}_1)\ldots \times g_m(\mathbf{X}_m) \prod_{j=1}^{m} \Phi_B(\mathbf{y}_j)\Big). \quad (3.14)$$

Hence $\left\{\left|\Psi_{lm}\left(\begin{matrix}\mathrm{out}\\ \mathrm{in}\end{matrix}\right)\right\rangle\right\}^{(n)} \in S_n \mathcal{R}^{\pm}_{lm}$, since $\Omega^{\pm}_{lm}[\cdots] \in \mathcal{R}^{\pm}_{lm}$. It is clear that one can approximate any symmetrized scattering state by applying polynomials in the $A_{ex}(f)^*$'s and $B_{ex}(f)^*$'s to the vacuum. It is easy to demonstrate that the set of operators $\{A_{ex}(f), A_{ex}(f)^*, B_{ex}(f), B_{ex}(f)^*\}$ for all $f \in L^2(R^3)$ is irreducible in the sense of Ruelle.[4] (This depends critically on the asymptotic completeness property E. If E does not hold, the set is not irreducible; however, the ex fields can still be defined as above on the subspace of scattering states.) The S matrix can be defined as usual by the requirement that

$$S|\Psi_{lm}(\mathrm{out})\rangle = |\Psi_{lm}(\mathrm{in})\rangle \quad (3.15)$$

on states of the form (3.13). This implies that

$$A_{in}(f) = S A_{out}(f) S^{-1}, \quad B_{in}(f) = S B_{out}(f) S^{-1}. \quad (3.16)$$

S is unitary under the asymptotic completeness assumption E. (For a detailed justification of these points, the reader is referred to the thesis[28] of one of the authors.)

The representation (3.9) shows that $A_{ex}(\alpha f + \beta g)^* = \alpha A_{ex}(f)^* + \beta A_{ex}(g)^*$; hence we have a linear map from $L^2(R^3)$ into a set of operators on $\mathfrak{F}$, which is sufficient to determine a field $A_{ex}(\mathbf{x}, 0)^*$. Hence we can write

$$A_{ex}(f)^* = \int d\mathbf{x} f(\mathbf{x}) A_{ex}(\mathbf{x}, 0)^*, \qquad A_{ex}(f) = \int d\mathbf{x} f(\mathbf{x})^* A_{ex}(\mathbf{x}, 0). \quad (3.17)$$

Likewise, similar equations can be written for $B_{ex}(\mathbf{x}, 0)$ and its "adjoint" $B_{ex}(\mathbf{x}, 0)^*$. Formally, we can obtain the fields $A_{ex}(\mathbf{x}, 0)$ and $B_{ex}(\mathbf{x}, 0)$ as follows: let $\{f_i\}$ be any orthonormal basis for $L^2(R^3)$ and define

$$A_{ex}(\mathbf{x}, 0) = \sum_i f_i(\mathbf{x}) A_{ex}(f_i), \qquad A_{ex}(\mathbf{x}, 0)^* = \sum_i f_i(\mathbf{x})^* A_{ex}(f_i)^*. \quad (3.18)$$

Then we can write formally

$$A_{ex}(f) = \int d\mathbf{x} f(\mathbf{x})^* A_{ex}(\mathbf{x}, 0) = \sum_i c_i A_{ex}(f_i), \qquad A_{ex}(f)^* = \int d\mathbf{x} f(\mathbf{x}) A_{ex}(\mathbf{x}, 0)^* = \sum_i c_i^* A_{ex}(f_i)^*, \quad (3.19)$$

where $c_i = (f, f_i)$ is the Fourier coefficient of f with respect to $\{f_i\}$. The same considerations can be made for $B_{ex}(\mathbf{x}, 0)$ by simply replacing A_{ex} by B_{ex} in Eqs. (3.18)–(3.19).

It follows directly[28] from the representations (3.9)–(3.12) that the ex fields satisfy the commutation relations as stated in

Lemma 1:

$$[A_{ex}(d), A_{ex}(f)^*] = (d, f), \qquad [B_{ex}(g), B_{ex}(h)^*] = (g, h), \quad d, f, g, h \in L^2(R^3). \quad (3.20)$$

All other commutators between any two of the fields $A_{ex}(d), A_{ex}(f)^*, B_{ex}(g), B_{ex}(h)^*$ vanish. These commutation relations are valid at least on $\tilde{\mathfrak{F}}$.

In terms of the field distributions, Eqs. (3.20) become

$$[A_{ex}(\mathbf{x}, 0), A_{ex}(\mathbf{y}, 0)^*] = \delta(\mathbf{x} - \mathbf{y}), \qquad [B_{ex}(\mathbf{x}, 0), B_{ex}(\mathbf{y}, 0)^*] = \delta(\mathbf{x} - \mathbf{y}). \quad (3.21)$$

To make the analogy with the relativistic theory complete, it remains only to define time-dependent Heisenberg fields and to show that these are free fields in the sense that they satisfy free field equations of motion.

We define the time-dependent fields $A_{ex}(\mathbf{x}, t)$ and $B_{ex}(\mathbf{x}, t)$ by

$$A_{ex}(\mathbf{x}, t) = e^{i\mathcal{H}t} A_{ex}(\mathbf{x}, 0) e^{-i\mathcal{H}t}, \qquad B_{ex}(\mathbf{x}, t) = e^{i\mathcal{H}t} B_{ex}(\mathbf{x}, 0) e^{-i\mathcal{H}t}, \quad (3.22)$$

where $\mathcal{H}$ is the Fock space Hamiltonian defined in (3.3) or (3.4). The meaning of these equations is, for example,

$$A_{ex}(f, t) = \int d\mathbf{x} f(\mathbf{x})^* A_{ex}(\mathbf{x}, t) = e^{i\mathcal{H}t} A_{ex}(f) e^{-i\mathcal{H}t}. \quad (3.23)$$

The following lemma, which follows directly from (3.23), the representations of $A_{ex}(f)$ and $B_{ex}(f)$, and the intertwining property (2.17), enables us to demonstrate that $A_{ex}(\mathbf{x}, t)$ and $B_{ex}(\mathbf{x}, t)$ are free fields.

Lemma 2: Let $H_0 = -\Delta/2\mu$, $H_{B0} = -\Delta/2M + \epsilon$, $M = 2\mu$. Then

$$A_{ex}(f, t) = A_{ex}(e^{iH_0 t} f), \qquad B_{ex}(f, t) = B_{ex}(e^{tH_{B0} t} f). \quad (3.24)$$

(For the proof see the Appendix.)

Lemma 2 tells us that if we smear the fields with appropriate time-dependent wavefunctions, the resulting operators are time-independent; that is, if $f \to f_t = f(\mathbf{x}, t) = e^{-iH_0 t} f(\mathbf{x}, 0)$ or $f \to f_{Bt} = f_B(\mathbf{x}, t) = e^{-iH_{B0} t} f_B(\mathbf{x}, 0)$, then

$$A_{ex}(f_0) = \int d\mathbf{x} f(\mathbf{x}, t)^* A_{ex}(\mathbf{x}, t), \qquad B_{ex}(f_{B0}) = \int d\mathbf{x} f_B(\mathbf{x}, t)^* B_{ex}(\mathbf{x}, t) \quad (3.25)$$

are independent of time. In terms of the field distributions, Eqs. (3.24) mean

$$A_{ex}(\mathbf{x}, t) = e^{-iH_0 t} A_{ex}(\mathbf{x}, 0), \qquad B_{ex}(\mathbf{x}, t) = e^{-iH_{B0} t} B_{ex}(\mathbf{x}, 0). \quad (3.26)$$

Thus the fields $A_{ex}(\mathbf{x}, t)$ and $B_{ex}(\mathbf{x}, t)$ satisfy the free

equations of motion. Lemmas 1 and 2 together imply that

$$[A_{ex}(f,t),A_{ex}(g,t')^*] = (f,e^{-iH_0(t-t')}g),$$
$$[B_{ex}(f,t),B_{ex}(g,t')^*] = (f,e^{-iH_{B0}(t-t')}g), \tag{3.27}$$

and all other commutators between pairs of these Heisenberg fields are zero. In terms of the distributions themselves, Eqs. (3.27) can be summarized as

$$[A_{ex}(\mathbf{x},t),A_{ex}(\mathbf{x}',t')^*] = \mathfrak{D}_\mu(\mathbf{x}-\mathbf{x}',t-t'),$$
$$[B_{ex}(\mathbf{x},t),B_{ex}(\mathbf{x}',t')^*] = e^{-i\epsilon(t-t')}\mathfrak{D}_{2\mu}(\mathbf{x}-\mathbf{x}',t-t'), \tag{3.28}$$

where

$$\mathfrak{D}_\mu(\mathbf{x},t) = (2\pi)^{-3}\int d\mathbf{k}\, e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)}, \quad \omega = k^2/2\mu$$
$$= \begin{cases}(\mu/2\pi it)^{3/2}\, e^{i\mu x^2/2t}, & t\neq 0,\\ \delta(\mathbf{x}), & t=0,\end{cases} \tag{3.29}$$

is the well-known free nonrelativistic propagator.

The results of this section are summarized in

Theorem 1: In the nonrelativistic field theory characterized by the Hamiltonian $\mathfrak{K}$ [Eq. (3.3)], under the assumption that the interaction potential $V(\mathbf{x})$ allows properties A–D to be satisfied in the first-quantized theory and under the additional technical assumption that there is only one bound state (2.10), there exist "formally free" Heisenberg fields $A_{ex}(\mathbf{x},t)$ and $B_{ex}(\mathbf{x},t)$, associated with a simple particle and a composite particle, respectively. When smeared with solutions of the free one-particle equations [$f(\mathbf{x},t)$ and $f_B(\mathbf{x},t)$]—see (3.25)—the resulting operators can be used to construct all the scattering states of the theory with specified initial (ex = in) or final (ex = out) asymptotic configurations. An explicit representation of these operators is given in Eqs. (3.9)–(3.12).

If the asymptotic completeness property E holds, the operators form an irreducible set and the in fields are related to the out fields by a unitary S matrix:

$$A_{out}(\mathbf{x},t) = S^{-1}A_{in}(\mathbf{x},t)S,$$
$$B_{out}(\mathbf{x},t) = S^{-1}B_{in}(\mathbf{x},t)S. \tag{3.30}$$

It is worthwhile to emphasize that these results (Theorem 1) *do* hold for systems with Coulomb interactions. In fact, the results will hold for even longer-range interactions whenever generalized wave operators satisfying A–D exist. It will be seen in the next section that the "long-range" nature of the Coulomb interaction, for example, is reflected in the fact that the interpolating fields are not so simply related to the Heisenberg field operators $\psi(\mathbf{x},t)$ and $\psi(\mathbf{x},t)^*$ of (3.2) and (3.3) as they are in the short-range case.

As noted above, properties A–D have been shown to hold for a large class of interactions. Property E is reasonable but if E does not hold, the ex fields can still be constructed in the same way, but only on a subspace of $\mathfrak{F}$.

4. THE INTERPOLATING FIELDS

In this section we employ the results of the first-quantized theory and construct new fields which interpolate between the in and out fields of the previous section. What this means precisely is stated below in Theorem 2. The asymptotic condition proved in Theorem 2 can be regarded as the nonrelativistic analogue of the Haag–Ruelle[3,4] or LSZ[2] asymptotic conditions in relativistic field theory.

It is clear from trivial counterexamples that the set of interpolating fields cannot be unique. However, the fields we construct have the virtue that

(a) their structure is clearly reasonable and transparent,

(b) they reduce to known (and relatively simple) expressions in terms of the Heisenberg fields $\psi(\mathbf{x},t), \psi(\mathbf{x},t)^*$ when long-range forces are absent. Furthermore, it will be clearly evident why long-range potentials give rise to nontrivial difficulties when one passes from the first quantized formalism to the field-theoretic version. As in the previous section we confine our discussion to a system of identical particles which can form only one bound state, that of a pair of particles.

To facilitate the construction of the interpolating fields, it is convenient to define

$$F_{lm}(t) = \begin{cases}F^+_{lm}, & t>0,\\ F^-_{lm}, & t<0,\end{cases} \tag{4.1}$$

where $F^\pm_{lm}$ is defined in (2.16). Following closely the forms of the ex fields given in (3.9) and (3.10) we define the "smeared" interpolating fields $A(f,t)^*$ and $B(f,t)^*$ for $f\in L^2(R^3)$ by the representations

$$\{A(f,t)^*|\Phi\rangle^{(0)} = 0,$$
$$\{A(f,t)^*|\Phi\rangle\}^{(n+1)}(\mathbf{x}_1,\ldots,\mathbf{x}_{n+1})$$
$$= \sum_{\substack{l,m\\(n=l+2m)}} \frac{n!}{l!m!2^m}(n+1)^{1/2}S_{n+1}e^{iH^{(n+1)}t}\hat{U}_{l+1,m}(t)$$
$$\times [f(\mathbf{x}_1)\hat{U}_{lm}(t)^*F_{lm}(t)e^{-iH^{(n)}t}\{|\Phi\rangle\}^{(n)}(\mathbf{x}_2,\ldots,\mathbf{x}_{n+1})], \tag{4.2}$$

$$\{B(f,t)^*|\Phi\rangle\}^{(i)} = 0, \quad i=0,1,$$
$$\{B(f,t)^*|\Phi\rangle\}^{(n+2)}(\mathbf{x}_1,\ldots,\mathbf{x}_{n+2})$$
$$= \sum_{\substack{l,m\\(n=l+2m)}} \frac{n!}{l!m!2^m}\left(\frac{(n+1)(n+2)}{2}\right)^{1/2}S_{n+2}\,e^{iH^{(n+2)}t}\hat{U}_{l,m+1}(t)$$
$$\times [(\hat{U}_{lm}(t)^*F_{lm}(t)e^{-iH^{(n)}t}\{|\Phi\rangle\}^{(n)}(\mathbf{x}_1,\ldots,\mathbf{x}_n))f(\mathbf{X})\Phi_B(\mathbf{y})],$$
$$\mathbf{X} = \tfrac{1}{2}(\mathbf{x}_{n+1}+\mathbf{x}_{n+2}), \quad \mathbf{y} = \mathbf{x}_{n+1}-\mathbf{x}_{n+2}. \tag{4.3}$$

Note that $A(f,t)^*$ and $B(f,t)^*$ are time-dependent one-particle and two-particle bound state creation operators, respectively. It follows that the operators adjoint to $A(f,t)^*$ and $B(f,t)^*$, called $A(f,t)$ and $B(f,t)$, have the representations

$$\{A(f,t)|\Phi\rangle\}^{(n)}(\mathbf{x}_1,\ldots,\mathbf{x}_n)$$
$$= (n+1)^{1/2}\sum_{\substack{l,m\\(n=l+2m)}} \frac{n!}{l!m!2^m}S_n e^{iH^{(n)}t}F_{lm}(t)\hat{U}_{lm}(t)$$
$$\times \left(\int d\mathbf{y}\, f(\mathbf{y})^*[\hat{U}_{l+1,m}(t)^*e^{-iH^{(n+1)}t}\{|\Phi\rangle\}^{(n+1)}(\mathbf{y},\mathbf{x}_1,\ldots,\mathbf{x}_n]\right) \tag{4.4}$$

$$\{B(f,t)|\Phi\rangle\}^{(n)}(\mathbf{x}_1,\ldots,\mathbf{x}_n)$$
$$= \left(\frac{(n+1)(n+2)}{2}\right)^{1/2}\sum_{\substack{l,m\\(n=l+2m)}} \frac{n!}{l!m!2^m}S_n e^{iH^{(n)}t}F_{lm}(t)\hat{U}_{lm}(t)$$

$$\times \Big(\int d\mathbf{X}\, d\mathbf{y} f(\mathbf{X})^* \Phi_B(\mathbf{y}) \times [\hat{U}_{l,m+1}(t)^* e^{-iH^{(n+2)}t} \{|\Phi\rangle\}^{(n+2)}(\mathbf{x}_1, \ldots, \mathbf{x}_{n+2})] \Big), \quad (4.5)$$

$$\mathbf{X} = \tfrac{1}{2}(\mathbf{x}_{n+1} + \mathbf{x}_{n+2}), \quad \mathbf{y} = (\mathbf{x}_{n+1} - \mathbf{x}_{n+2}).$$

These operators are unbounded; however, $\tilde{\mathfrak{F}}$ is an invariant domain for them.

Next we prove that the fields defined in (4.2)–(4.5) are indeed interpolating fields, interpolating between the ex fields defined in (3.9) and (3.10).

Theorem 2: Let $|\Phi\rangle$ be any finite vector, $|\Phi\rangle \in \tilde{\mathfrak{F}}$. Then

$$\underset{t \to t_{\text{in}}^{\text{out}}}{\text{s-lim}} [A(f,t)^* - A_{\text{ex}}(f)^*]|\Phi\rangle = 0, \quad (4.6)$$

$$\underset{t \to t_{\text{in}}^{\text{out}}}{\text{s-lim}} [B(f,t)^* - B_{\text{ex}}(f)^*]|\Phi\rangle = 0, \quad (4.7)$$

where $t_{\text{in}}^{\text{out}} = \pm\infty$, and the limit is in the *strong* topology for $\mathfrak{F}$.

Proof: We report the proof for (4.7) only. We need only prove the strong convergence of $\{[B(f,t)^* - B_{\text{ex}}(f)^*]|\Phi\rangle\}^{(n)}$ to zero for arbitrary n.

According to (3.10) and (4.3), we have

$$\{[B(f,t)^* - B_{\text{ex}}(f)^*]|\Phi\rangle\}^{(i)} = 0, \quad i = 0,1,$$

$$\{[B(f,t)^* - B_{\text{ex}}(f)^*]|\Phi\rangle\}^{(n+2)}(\mathbf{x}_1, \ldots, \mathbf{x}_{n+2}) = \left(\frac{(n+1)(n+2)}{2}\right)^{1/2} \sum_{\substack{l,m \\ (n=l+2m)}} \frac{n!}{l!m!2^m} S_{n+2} \times [g_{lm}(\mathbf{x}_1, \ldots, \mathbf{x}_{n+2}, t) - g_{lm}^{\pm}(\mathbf{x}_1, \ldots, \mathbf{x}_{n+2})], \quad (4.8)$$

where

$$g_{lm}(\mathbf{x}_1, \ldots, \mathbf{x}_{n+2}, t) = e^{iH^{(n+2)}t} \hat{U}_{l,m+1}(t) \times [(\hat{U}_{lm}(t)^* F_{lm}(t) e^{-iH^{(n)}t} \{|\Phi\rangle\}^{(n)}(\mathbf{x}_1, \ldots, \mathbf{x}_n)) f(\mathbf{X}) \Phi_B(\mathbf{y})], \quad (4.9)$$

$$g_{lm}^{\pm}(\mathbf{x}_1, \ldots, \mathbf{x}_{n+2}) = \Omega_{l,m+1}^{\pm} [(\Omega_{lm}^{\pm *} \{|\Phi\rangle\}^{(n)}(\mathbf{x}_1, \ldots, \mathbf{x}_n)) f(\mathbf{X}) \Phi_B(\mathbf{y})],$$

$$\mathbf{X} = \tfrac{1}{2}(\mathbf{x}_{n+1} + \mathbf{x}_{n+2}), \quad \mathbf{y} = \mathbf{x}_{n+1} - \mathbf{x}_{n+2}.$$

Therefore,

$$\| \{(B(f,t)^* - B_{\text{ex}}(f)^*)|\Phi\rangle\}^{(n+2)} \|_{n+2} \leq \sum_{\substack{l,m \\ (n=l+2m)}} K_{lm}^n \| g_{lm}(t) - g_{lm}^{\pm} \|_{n+2}, \quad (4.10)$$

where K_{lm}^n is a constant independent of t, and $\|\cdot\|_j$ refers to the norm in $L^2(R^{3j})$. Hence it is sufficient to show that $\| g_{lm}(t) - g_{lm}^{\pm} \| \to 0$ as $t \to \pm\infty$. The intertwining relation (2.17) implies that

$$[F_{lm}(t), e^{-iH^{(n)}t}] = 0. \quad (4.11)$$

Thus we can write for $t \gtrless 0$

$$\| g_{lm}(t) - g_{lm}^{\pm} \|_{n+2} = \| e^{iH^{(n+2)}t} \hat{U}_{l,m+1}(t) \times [(\hat{U}_{lm}(t)^* e^{-iH^{(n)}t} \Omega_{lm}^{\pm} h_{lm}^{\pm}(\mathbf{x}_1, \ldots, \mathbf{x}_n)) f(\mathbf{X}) \Phi_B(\mathbf{y})] - \Omega_{l,m+1}^{\pm} [h_{lm}^{\pm}(\mathbf{x}_1, \ldots, \mathbf{x}_n) f(\mathbf{X}) \Phi_B(\mathbf{y})] \|_{n+2}, \quad (4.12)$$

where $h_{lm}^{\pm} = \Omega_{lm}^{\pm *} \{|\Phi\rangle\}^{(n)}(\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathfrak{D}_{lm}$.

Using the triangle inequality and the unitarity of $e^{iH^{(n+2)}t} \hat{U}_{l,m+1}(t)$, we get for $t \gtrless 0$

$$\| g_{lm}(t) - g_{lm}^{\pm} \|_{n+2} \leq \| [(\hat{U}_{lm}(t)^* e^{-iH^{(n)}t} \Omega_{lm}^{\pm} - I) h_{lm}^{\pm}(\mathbf{x}_1, \ldots, \mathbf{x}_n)] f(\mathbf{X}) \Phi_B(\mathbf{y}) \|_{n+2} + \| (e^{iH^{(n+2)}t} \hat{U}_{l,m+1}(t) - \Omega_{l,m+1}^{\pm}) h_{lm}^{\pm}(\mathbf{x}_1, \ldots, \mathbf{x}_n) f(\mathbf{X}) \Phi_B(\mathbf{y}) \|_{n+2}. \quad (4.13)$$

The asymptotic condition (2.14) and the fact that $h_{lm}^{\pm} \in \mathfrak{D}_{lm}$ implies the desired convergence as $t \to \pm\infty$

It follows from Theorem 2 that the restrictions of the operators $A(f,t)$ and $B(f,t)$ to $\tilde{\mathfrak{F}}$ converge weakly to $A_{\text{ex}}(f)$ and $B_{\text{ex}}(f)$ on $\tilde{\mathfrak{F}}$. In the case of short-range and/or Coulomb potentials, it can be shown[28] that this convergence is also strong convergence.

5. SHORT-RANGE CASE

In order to elucidate the role of the long-range forces and also to make contact with the work of previous authors, we will show that the interpolating fields simplify considerably when the interaction potential $V(\mathbf{x})$ is short-range. Recall that the only change in the usual formalism that we make when long-range potentials are present is to replace the channel propagator $e^{-iH_{lm}t}$ by $\hat{U}_{lm}(t)$, a modified channel propagator. We prove the following theorem:

Theorem 3: In the special case of short-range forces where $\hat{U}_{lm}(t) = e^{-iH_{lm}t}$, the interpolating fields can be expressed as

$$A(f_0,t)^* = \int d\mathbf{x} f(\mathbf{x},t) \psi(\mathbf{x},t)^*, \quad (5.1)$$

$$B(f_{B0},t)^* = 2^{-1/2} \int d\mathbf{x}_1 d\mathbf{x}_2 f_B(\mathbf{X}_1 t) \Phi_B(\mathbf{y}) \psi(\mathbf{x}_1,t)^* \psi(\mathbf{x}_2,t)^*, \quad (5.2)$$

$$\mathbf{X} = \tfrac{1}{2}(\mathbf{x}_1 + \mathbf{x}_2), \quad \mathbf{y} = \mathbf{x}_1 - \mathbf{x}_2,$$

where $\psi(\mathbf{x},t)$ is the conventional Heisenberg field operator (3.8) and

$$f_t = f(\mathbf{x},t) = e^{-iH_0 t} f(\mathbf{x},0), \quad H_0 = -\Delta/2\mu,$$
$$f_{Bt} = f_B(\mathbf{x},t) = e^{-iH_{B0}t} f(\mathbf{x},0), \quad (5.3)$$
$$H_{B0} = -\Delta/2M + \epsilon, \quad M = 2\mu.$$

Comparison of (5.1) with (3.5) and (3.8) shows that $A(f,t)^*$ is just the conventional Heisenberg field operator smeared with a freely propagating wavepacket. Also, $B(f,t)^*$ creates a freely moving bound state employing the same $\psi(\mathbf{x},t)^*$ field as $A(f,t)^*$.

Proof: [See (4.2).] If $\hat{U}_{lm}(t) = U_{lm}(t) = e^{-iH_{lm}t}$, we have the identities

$$U_{l+1,m}(t)[f(\mathbf{x}_1) U_{lm}(t)^* g(\mathbf{x}_2, \ldots, \mathbf{x}_{n+1})] = (e^{-iH_0 t} f(\mathbf{x}_1)) g(\mathbf{x}_2, \ldots, \mathbf{x}_{n+1}), \quad (5.4)$$

$$U_{l,m+1}(t)[(U_{lm}(t)^* h(\mathbf{x}_1, \ldots, \mathbf{x}_n)) f(\mathbf{X}) \Phi_B(\mathbf{y})] = h(\mathbf{x}_1, \ldots, \mathbf{x}_n)(e^{-iH_{B0}t} f(\mathbf{X})) \Phi_B(\mathbf{y}), \quad (5.5)$$

$$\mathbf{X} = \tfrac{1}{2}(\mathbf{x}_{n+1} + \mathbf{x}_{n+2}), \quad \mathbf{y} = \mathbf{x}_{n+1} - \mathbf{x}_{n+2}.$$

The symmetrizer satisfies the relations

$$S_{n+1}[f(\mathbf{x}_1)g(\mathbf{x}_2,\ldots,\mathbf{x}_{n+1})] = S_{n+1}[f(\mathbf{x}_1)S_n g(\mathbf{x}_2,\ldots,\mathbf{x}_{n+1})],$$

$$S_{n+2}[g(\mathbf{x}_1,\ldots,\mathbf{x}_n)h(\mathbf{x}_{n+1},\mathbf{x}_{n+2})] = S_{n+2}[(S_n g(\mathbf{x}_1,\ldots,\mathbf{x}_n))h(\mathbf{x}_{n+1},\mathbf{x}_{n+2})]. \quad (5.6)$$

Finally, on account of the asymptotic completeness property E, it follows that

$$\sum_{\substack{l,m \\ (n=l+2m)}} \frac{n!}{l!m!2^m} S_n F_{lm}(t)\chi = \chi, \qquad \chi \in S_n L^2(R^{3n}). \quad (5.7)$$

Using these relations, one can immediately express the representations (4.2) and (4.3) in the following forms:

$$\{A(f,t)^*|\Phi\rangle\}^{(n+1)}(\mathbf{x}_1,\ldots,\mathbf{x}_{n+1}) = (n+1)^{1/2}S_{n+1}e^{iH^{(n+1)}t} \times [(e^{-iH_0 t}f(\mathbf{x}_1))e^{-iH^{(n)}t}\{|\Phi\rangle\}^{(n)}(\mathbf{x}_2,\ldots,\mathbf{x}_{n+1})], \quad (5.8)$$

$$\{B(f,t)^*|\Phi\rangle\}^{(n+2)}(\mathbf{x}_1,\ldots,\mathbf{x}_{n+2}) = \left(\frac{(n+1)(n+2)}{2}\right)^{1/2} S_{n+2}e^{iH^{(n+2)}t} \times [(e^{-iH^{(n)}t}\{|\Phi\rangle\}^{(n)}(\mathbf{x}_1,\ldots,\mathbf{x}_n))e^{-iH_B 0^t}f(\mathbf{X})\Phi_B(\mathbf{y})], \quad (5.9)$$

$$\mathbf{X} = \tfrac{1}{2}(\mathbf{x}_{n+1}+\mathbf{x}_{n+2}), \qquad \mathbf{y} = \mathbf{x}_{n+1}-\mathbf{x}_{n+2},$$

which are equivalent to the representations of the operators in Theorem 3. ∎

Combining the result in (5.1) with the general convergence theorem (4.6), we may assert that, for short-range forces and for the operator-valued distribution $A(\mathbf{x},t)$ corresponding to our interpolating field operators $A(f,t)$,

$$A(\mathbf{x},t) = \psi(\mathbf{x},t) \overset{s}{\to} \psi_{\underset{\text{in}}{\text{out}}}(\mathbf{x},t) = A_{\underset{\text{in}}{\text{out}}}(\mathbf{x},t), \qquad t \to \pm\infty,$$

where $\psi(\mathbf{x},t)$ is the conventional Heisenberg field.

6. DISCUSSION

The interpolating fields of Theorem 3 are the ones found previously by others.[7-11] No such simple representation seems to exist in the long-range case, characterized in the first-quantized theory by the need to use generalized channel propagators to define the wave operators. Comparison of the representations of $A(f,t)^*$ for the general case in (4.2) and for the special short-range case (5.8) pinpoints the source of the complexity of the long-range interpolating field. In the short-range case $A(f,t)^*$ creates a particle in the state $f(\mathbf{x},t)$ essentially independent of the state to which $A(f,t)^*$ is applied, whereas in the long-range case $A(f,t)^*$ must add a particle in a state which takes into account the presence of the particles already there. Scrutiny of the proof of Theorem 3 shows that the simplification in the short-range case is due to the validity of (5.4), where a free one-particle propagator arises describing the time behavior of the particle created. On the other hand, when long-range potentials are involved, the crucial term $\hat{U}_{l+1,m}(t)f(\mathbf{x}_1)\hat{U}_{lm}(t)$ gives rise to a very complex one-particle propagator whose structure depends on l and m, i.e., on the particles already present. Thus, when long-range interactions are involved, and when a particle is added to those particles already present, even when they are both far removed from one another, it must begin to move under the influence of the distant structures. To add even a simple particle to the scattering process is not a trivial matter.

As our results show this phenomenon does not preclude the existence of a particle interpretation for long-range forces as outlined in the introduction. However, it does indicate that it is extremely unlikely that one can obtain interpolating fields for long-range forces by solving any reasonably simple field equations.

In conclusion, it should be mentioned that one can use the asymptotic condition (Theorem 2) to obtain the nonrelativistic analogue of the LSZ formalism.[28] Reduction formulas can be derived, nonrelativistic Yang–Feldman equations can be established, the S matrix can be expressed in terms of normal-ordered polynomials in the ex fields, and S-matrix elements for any scattering process involving both simple and composite particles can be related to vacuum expectation values of time-ordered products of the interpolating fields.

APPENDIX: PROOF OF LEMMA 2

We report the proof for $A_{\text{in}}(f,t)^*$. The other cases are no different. Consider the state $A_{\text{in}}(f,t)^*|\Phi\rangle$. Using the definition (3.23), the property (3.4) of the Fock space Hamiltonian $\mathcal{H}$, and the representation (3.9), we obtain

$$\{A_{\text{in}}(f,t)^*|\Phi\rangle\}^{(n+1)}(\mathbf{x}_1,\ldots,\mathbf{x}_{n+1}) = (n+1)^{1/2}\sum_{\substack{l,m \\ (n=l+2m)}} \frac{n!}{l!m!2^m} S_{n+1}\{e^{iH^{(n+1)}t}\Omega^-_{l+1,m} \times [f(\mathbf{x}_1)\Omega^{-*}_{lm}e^{-iH^{(n)}t}\{|\Phi\rangle\}^{(n)}(\mathbf{x}_2,\ldots,\mathbf{x}_{n+1})]\}, \quad (A1)$$

where we have used the fact that S_{n+1} commutes with $H^{(n+1)}$ in obtaining (A1). Using the intertwining property (2.17), one can express the quantity in the large curly brackets on the rhs in the form

$$\{\cdots\} = \Omega^-_{l+1,m}U_{l+1,m}(-t)[f(\mathbf{x}_1)U_{lm}(t) \times \Omega^{-*}_{lm}\{|\Phi\rangle\}^{(n)}(\mathbf{x}_2,\ldots,\mathbf{x}_{n+1}). \quad (A2)$$

Since $U_{l+1,m}(-t)f(\mathbf{x}_1)U_{lm}(t) = e^{iH_0 t}f(\mathbf{x}_1)$,

$$\{\cdots\} = \Omega^-_{l+1,m}[(e^{iH_0 t}f)(\mathbf{x}_1)\Omega^{-*}_{lm}\{|\Phi\rangle\}^{(n)}(\mathbf{x}_2,\ldots,\mathbf{x}_{n+1})]. \quad (A3)$$

Upon substitution of (A3) into (A1) and comparison with the representation (3.9), one sees that

$$\{A_{\text{in}}(f,t)^*|\Phi\rangle\}^{(n)} = \{A_{\text{in}}(e^{iH_0 t}f)^*|\Phi\rangle\}^{(n)} \quad (A4)$$

holds for all n. Thus $A_{\text{in}}(f,t) = A_{\text{in}}(e^{iH_0 t}f)$. ∎

[1]The restriction to spinless bosons is for convenience only. It should be clear from our construction that the extension to nonzero spin and different particle species is straightforward.
[2]H. Lehmann, K. Symanzik, and W. Zimmerman, Nuovo Cimento **1,** 205 (1955); Nuovo Cimento **6,** 319 (1957).
[3]R. Haag, Phys. Rev. **112,** 669 (1958).
[4]D. Ruelle, Helv. Phys. Acta **35,** 147 (1962).
[5]K. Hepp, Commun. Math. Phys. **1,** 95 (1965).
[6]H. Araki and R. Haag, Commun. Math. Phys. **4,** 77 (1967).
[7]W. Sandhas, Commun. Math. Phys. **3,** 358 (1966).
[8]O. W. Greenberg, thesis (Princeton University, 1956).
[9]P. Redmond and J. Uretsky, Ann. Phys. (N.Y.) **9,** 106 (1960).
[10]J. Dollard, thesis (Princeton University, 1963).
[11]K. Hepp, in *1965 Brandeis University Summer Institute in Theoretical Physics* (Gordon & Breach, New York, 1966), Vol. 1.
[12]J. Jauch, Helv. Phys. Acta **31,** 127 (1958).
[13]J. Dollard, J. Math. Phys. **5,** 729 (1964). See also Ref. 10.
[14]D. Mulherin and I. I. Zinnes, J. Math. Phys. **11,** 1403 (1970).
[15] W. Amrein, P. Martin, and B. Misra, Helv. Phys. Acta **43,** 313 (1970).
[16]R. Lavine, J. Funct. Anal. **5,** 368 (1970).
[17]V. S. Buslaev and V. B. Matveev, Teor. Mat. Fiz. 2, 266 (1970).
[18]E. Prugovečki, Nuovo Cimento B **4,** 105 (1971).
[19]J. V. Corbett, Phys. Rev. D **1,** 3331 (1970).
[20]W. Hunziker, Helv. Phys. Acta **40,** 1052 (1967).
[21]J. Jauch, Helv. Phys. Acta **31,** 661 (1958).
[22]T. Kato, Trans. Am. Math. Soc. **70,** 195 (1951).
[23]It is actually sufficient to assume that $\hat{U}(t)$ is only asymptotically unitary. See Ref. 15.
[24]K. Hepp, Helv. Phys. Acta **42,** 425 (1969).
[25]S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Harper & Row, New York, 1961), Chap. 6.
[26]J. M. Cook, Trans. Am. Math. Soc. **74,** 222 (1953).
[27]F. E. Schroeck, J. Math. Phys. **12,** 1849 (1971).
[28]J. Klein, thesis (Fordham University, 1969).

Related integral theorem. II. A method for obtaining quadratic constants of the motion for conservative dynamical systems admitting symmetries

Gerald H. Katzin

Department of Physics, North Carolina State University, Raleigh, North Carolina 27607
(Received 20 November 1972; revised manuscript received 19 March 1973)

By use of Lie derivatives, symmetry mappings for conservative dynamical systems are formulated in terms of continuous groups of infinitesimal transformations within the configuration space. Such symmetry transformations, called trajectory collineations, may be interpreted as point mappings which drag along coordinates and geometric objects as they map trajectories into trajectories. It is shown that if a conservative dynamical system admits a trajectory collineation, then in general a new quadratic (in the velocity) constant of the motion will result from the deformation of a given quadratic constant of the motion under such a symmetry mapping. The theory is applied to obtain the group of symmetry transformations and concomitant constants of the motion associated with the deformations of the energy integral for the Kepler problem and for the isotropic simple harmonic oscillator. The Runge-Lenz vector of the Kepler problem and the symmetric tensor constant of the motion for the three-dimensional oscillator are readily obtained by this method. The trajectory collineation group for the Kepler problem is a seven-parameter projective collineation group which is isomorphic to the similitude group in three-dimensional Euclidean space. For the oscillator the trajectory collineation group is the nine-parameter affine group. This dynamical symmetry group contains an eight-parameter subgroup which is shown to have the structure of SU_3.

1. INTRODUCTION

The related first integral theorem, published in an earlier paper,[1] is a unified method for obtaining "derived" mth order first integrals of the geodesic equations in a Riemannian space V_n. By this method additional constants of the motion can be generated from a given constant of the motion provided the space admits symmetries in the form of continuous groups of projective collineations[2] (which include affine collineations and motions as special cases). The derived constants of the motion were shown to result from the deformation of a given constant of the motion under the projective collineation symmetry transformations which mapped geodesics into geodesics.[3] In particular the quadratic first integrals of the geodesics previously obtained by direct construction from the conditions for projective collineations[4] were explained by the related integral theorem in terms of deformations in the (given) metrical first integral under such symmetry mappings.

In this present paper the domain of the theorem is extended. We show that if a conservative dynamical system admits a "trajectory collineation", i.e., a mapping which takes trajectories into trajectories, then in general a new quadratic constant of the motion will result from the deformation of a given one under this symmetry mapping. In every conservative dynamical system the energy integral is a known quadratic constant of the motion. It therefore follows that for such systems this extended theorem can be applied provided the appropriate symmetries exist.

Repeated application of this extended theorem results in "second derived" constants of the motion. Dependency relations between first and second derived constants of the motion are shown to depend upon the structure of the group of associated trajectory collineations.

The conditions for trajectory collineations are based upon transformations within the configuration space; hence, symmetries obtained by this method are amenable to immediate geometrical and/or physical interpretations. This fact along with the direct manner in which the symmetry mappings generate constants of the motion suggests the related integral theorem can provide useful insights in the analysis of systems which possess so called "accidental degeneracies."[5]

Two applications of the theory are given. It is shown that the three-dimensional Kepler problem admits a seven-parameter group of trajectory collineations. These parameters include three describing rotations, a homothetic motion (or scale change), and three proper projective collineations. This group is isomorphic to the similitude group of three-dimensional Euclidean space.[6] The deformation of the energy integral under the three proper projective collineations generates three derived quadratic constants of the motion which are the components of the Runge-Lenz vector.[7]

Application of the theory to the three-dimensional isotropic harmonic oscillator shows that it admits a nine-parameter trajectory collineation group which contains six proper affine collineations in addition to the expected three-parameter rotation group. This TC_9 contains an eight-parameter subgroup TC_8. By a change in basis it is shown how this TC_8 may be brought into a form which has commutation relations characteristic of SU_3. The deformations of the energy integral which result from the proper affine collineation symmetry mappings lead to six quadratic constants of the motion which are the components of the well-known symmetric tensor associated with the oscillator problem.[7]

2. PRELIMINARIES

We consider a conservative system (n degrees of freedom) with potential energy $V(x)$ whose motion in configuration space with metric $g_{ij}(x)$ is governed by the dynamical equation[8]

$$\frac{Dv^i}{dt} + g^{ij}V_{,j} = 0, \qquad v^i \equiv \frac{dx^i}{dt}. \tag{2.1}$$

Contraction of (2.1) by $g_{ik}v^k$ yields the energy integral

$$\tfrac{1}{2}g_{ij}v^iv^j + V = E, \tag{2.2}$$

where E is (an arbitrary)[9] constant along each trajectory.

Next we obtain the conditions for (2.1) to admit a quadratic first integral of the form

Copyright © 1973 by the American Institute of Physics

$$I \equiv B_{ij}(x)v^iv^j + A_i(x)v^i + \Omega(x) = k, \tag{2.3}$$

where $B_{ij} = B_{ji}$ and where k is (an arbitrary)[9] constant along trajectories. By use of (2.1) the requirement $DI/dt = 0$ leads to

$$B_{ij;k}v^iv^jv^k + A_{i;j}v^iv^j - (2B_{ij}g^{ik}V_{,k} - \Omega_{,j})v^j - A_ig^{ij}V_{,j} = 0. \tag{2.4}$$

If (2.4) holds without imposing additional restrictions on v^i, we obtain the necessary conditions[10]

$$B_{(ij;k)} = 0, \tag{2.5}$$

$$2B_{ij}g^{ik}V_{,k} - \Omega_{,j} = 0, \tag{2.6}$$

$$A_{(i;j)} = 0, \tag{2.7}$$

$$g^{ij}A_iV_{,j} = 0. \tag{2.8}$$

Inspection of (2.5)–(2.8) shows that a necessary condition for the existence of an inhomogeneous first integral of the form (2.3) is the independent existence of the quadratic first integral $B_{ij}v^iv^j + \Omega$ [from (2.5) and (2.6)] and the linear first integral A_iv^i [from (2.7) and (2.8)]. It immediately follows that the respective conditions are also sufficient for the existence of these integrals. Hence in our further considerations we recognize that use of (2.3) is merely a convenient formalism for handling both types of first integrals simultaneously.

3. TRAJECTORY COLLINEATIONS

A trajectory collineation is defined by a vector field $\xi^i(x)$ and scalar $\phi(x)$ such that the infinitesimal transformation ($\delta a \equiv$ infinitesimal)

$$\bar{x}^i = x^i + \xi^i(x)\delta a \tag{3.1}$$

along with an associated change in differential path parameter

$$d\bar{t} = \{1 + 2\phi[x(t)]\delta a\}dt, \tag{3.2}$$

defined by $\phi(x)$ evaluated along trajectories, maps trajectories into trajectories in that[11]

$$\pounds_\xi\left(\frac{Dv^i}{dt} + g^{ij}V_{,j}\right) = 0 \tag{3.3}$$

for all $x^i(t)$ which satisfy (2.1). Even though the accompanying change in path parameter is defined only along trajectories, we have taken the structure of the function ϕ which describes this change to be coordinate dependent because the change in path parameter is assumed to be dependent upon the mapping vector $\xi^i(x)$.

We define and then evaluate by use of (3.2)

$$\frac{\pounds_\xi dt}{dt} \equiv \lim_{\delta a \to 0}\frac{1}{\delta a}\left(\frac{d\bar{t} - dt}{dt}\right) = 2\phi[x(t)]. \tag{3.4}$$

From Yano[12] we obtain the well-known formulas (recall $v^i \equiv dx^i/dt$)

$$\pounds_\xi\frac{Dv^i}{dt} = \pounds_\xi\left\{{i \atop jk}\right\}v^jv^k - 2\frac{Dv^i}{dt}\frac{\pounds_\xi dt}{dt} - v^i\frac{D}{dt}\left(\frac{\pounds_\xi dt}{dt}\right), \tag{3.5}$$

$$\pounds_\xi v^i = -v^i\frac{\pounds_\xi dt}{dt}, \tag{3.6}$$

$$\pounds_\xi(g^{ij}V_j) = g^{ij}V_{;jk}\xi^k - g^{jk}V_j\xi^i_{;k}. \tag{3.7}$$

By use of (3.5) and (3.7) the expansion of (3.3) gives

$$\pounds_\xi\left\{{i \atop jk}\right\}v^jv^k - 2\frac{Dv^i}{dt}\frac{\pounds_\xi dt}{dt} - v^i\frac{D}{dt}\left(\frac{\pounds_\xi dt}{dt}\right) + g^{ij}V_{;jk}\xi^k - g^{jk}V_{,j}\xi^i_{;k} = 0. \tag{3.8}$$

Since (3.8) must hold for those v^i which satisfy (2.1), we obtain along each trajectory

$$(\pounds_\xi\left\{{i \atop jk}\right\} - \delta^i_j\phi_{,k} - \delta^i_k\phi_{,j})v^jv^k + 4\phi g^{ij}V_j + g^{ij}V_{;jk}\xi^k - g^{jk}V_{,j}\xi^i_{;k} = 0. \tag{3.9}$$

Equation (3.9) is to hold for all solutions of (2.1) and not act as a constraint; therefore, we obtain as conditions for a trajectory collineation the equations

$$\pounds_\xi\left\{{i \atop jk}\right\} = \delta^i_j\phi_{,k} + \delta^i_k\phi_{,j}, \tag{3.10}$$

$$4\phi g^{im}V_{,m} + g^{im}V_{;mk}\xi^k - g^{mk}V_{,m}\xi^i_{;k} = 0. \tag{3.11}$$

Equation (3.10) is recognized as the necessary and sufficient condition for the configuration space to admit a projective collineation (see Ref. 2).

If there exist functions ϕ_α and vectors $\xi^i_\alpha, \alpha = 1, \ldots, r$, which satisfy (3.10), then these vectors define an r-parameter group PC_r.[13] The additional conditions (3.11) select from these projective collineations those ξ^i_α and functions ϕ_α, $\alpha = 1, \ldots, s$, $s \leq r$, which define trajectory collineations (if they exist).

If we integrate the equation which results from the contraction of i and j in (3.10), we find that

$$\phi_\alpha = (n+1)^{-1}\xi^i_{\alpha;i} + h_\alpha, \tag{3.12}$$

where h_α are constants. Equation (3.12) together with (3.2) shows the explicit dependence of the associated change in differential parameter upon the mapping vector ξ^i_α.

The sufficiency of conditions (3.10) and (3.11) for trajectory collineation follows by inspection. Hence we may state

Theorem 3.1: Equations (3.10) and (3.11) are necessary and sufficient conditions for the infinitesimal transformation (3.1), with associated change in differential parameter (3.2), to define a trajectory collineation of a conservative dynamical system (2.1).

4. RELATED FIRST INTEGRAL THEOREM BASED ON TRAJECTORY COLLINEATIONS

We now assume that the dynamical equation (2.1) has a known first integral I of the form (2.3). In addition we assume that there exists a vector ξ^i_α which defines a trajectory collineation as described in Theorem 3.1. The process of Lie differentiation may be interpreted as a point mapping which drags along both coordinates and geometric objects.[14] We may therefore calculate the deformation $\pounds_\alpha$ (where $\pounds_\alpha \equiv \pounds_{\xi_\alpha}$) in the known first integral I under this trajectory collineation which drags trajectory into trajectory. This deformation is constant along the new trajectory. We thereby obtain an additional first integral $I_\alpha \equiv \pounds_\alpha I$ as a consequence of the deformation of the original one under the symmetry mapping. The form of such a "derived" quadratic first integral obtained from (2.3) is

$$I_\alpha \equiv \pounds_\alpha(B_{ij}v^iv^j + A_iv^i + \Omega) = k_\alpha, \tag{4.1}$$

where k_α is constant along trajectories. Expansion of (4.1) with use of (3.4) and (3.6) gives

$$I_\alpha = b_{\alpha ij}v^i v^j + a_{\alpha i}v^i + f_\alpha , \tag{4.2}$$

where

$$b_{\alpha ij} \equiv \pounds_\alpha B_{ij} - 4\phi_\alpha B_{ij}, \quad a_{\alpha i} \equiv \pounds_\alpha A_i - 2\phi_\alpha A_i,$$
$$f_\alpha \equiv \Omega_{,i}\xi^i_\alpha . \tag{4.3}$$

As an alternative proof of the validity of (4.2), we now give a direct verification that I_α is a constant of the motion. We will show that if a quadratic first integral of the form (2.3) is known and if there exists a vector ξ^i_α, which satisfies (3.10) and (3.11), then [refer to Eqs. (2.5)–(2.8)]

$$b_{\alpha(ij;k)} = 0, \tag{4.4}$$

$$2b_{\alpha ij}g^{ik}V_{,k} - f_{\alpha,j} = 0, \tag{4.5}$$

$$a_{\alpha(i;j)} = 0, \tag{4.6}$$

$$g^{ij}a_{\alpha i}V_{,j} = 0. \tag{4.7}$$

First we verify condition (4.4). By use of (3.10) and the identity (see Ref. 12)

$$(\pounds_\alpha B_{ij})_{;k} \equiv \pounds_\alpha(B_{ij;k}) + \pounds_\alpha\{{}^{m}_{ik}\}B_{mj} + \pounds_\alpha\{{}^{m}_{jk}\}B_{im}, \tag{4.8}$$

we obtain

$$b_{\alpha ij;k} = \pounds_\alpha(B_{ij;k}) - 2\phi_{\alpha,k}B_{ij} + \phi_{\alpha,i}B_{jk} + \phi_{\alpha,j}B_{ik} - 4\phi_\alpha B_{ij;k}. \tag{4.9}$$

If the indices (ijk) are symmetrized in (4.9) and if (2.5) is used in the resulting equation, we obtain a verification of (4.4).

We next consider the left side of (4.5) which expands to

$$2B_{ij;m}\xi^m_\alpha g^{ik}V_{,k} + 2B_{mj}\xi^m_{\alpha;i}g^{ik}V_{,k} + (2B_{im}g^{ki}V_{,k} - \Omega_{;m})\xi^m_{\alpha;j} - 8\phi_\alpha B_{ij}g^{ki}V_{,k} - \Omega_{;kj}\xi^k_\alpha . \tag{4.10}$$

From this expression subtract the Lie derivative of (2.6) with respect to ξ^m_α. The result is the same as condition (3.11) contracted with $-2B_{ij}$. Consequently (4.5) does hold.

We now consider condition (4.6). By an identity similar to (4.8) it follows with use of (3.10) that

$$(\pounds_\alpha A_i)_{;j} = \pounds_\alpha(A_{i;j}) + \phi_{\alpha,j}A_i + \phi_{\alpha,i}A_j . \tag{4.11}$$

It therefore follows that

$$a_{\alpha i;j} = \pounds_\alpha(A_{i;j}) - \phi_{\alpha,j}A_i + \phi_{\alpha,i}A_j - 2\phi_\alpha A_{i;j}, \tag{4.12}$$

which vanishes identically after symmetrization on indices (ij), provided (2.7) is used.

Finally we recognize that requirement (4.7) follows as a consequence of (2.8), the Lie derivative of (2.8) with respect to ξ^k_α, and the equation obtained by contraction of (3.11) with A_i.

This remark completes the alternative proof. Hence we may state

Theorem 4.1: If a conservative dynamical system (2.1) admits a quadratic constant of the motion (2.3) and if there exists a trajectory collineation based upon vector ξ^i_α (as described in Theorem 3.1), then in general there will exist an additional constant of the motion of the form (4.2), which is based upon the deformation of the original one under the trajectory collineation.

For the convenience of handling two examples treated later in the paper we specialize Theorem 4.1 to obtain additional constants of the motion which are based upon the deformation of the energy integral (2.2) under trajectory collineations in a flat configuration space with Cartesian coordinate system. Hence in (4.1) we take $B_{ij} \equiv \frac{1}{2}g_{ij} = \frac{1}{2}\delta_{ij}$, $A_i = 0$, $\Omega \equiv V(x)$, and obtain from (4.2) and (4.3) derived first integrals of the form

$$\tfrac{1}{2}(\xi^i_{\alpha,j} + \xi^j_{\alpha,i} - 4\phi_\alpha\delta_{ij})v^i v^j + V_{,i}\xi^i_\alpha = k_\alpha ,$$
$$\phi_\alpha = (n+1)^{-1}\xi^i_{\alpha,i} + h_\alpha . \tag{4.13}$$

5. DEPENDENCY RELATIONS BETWEEN DERIVED FIRST INTEGRALS

Let us now assume the existence of an s-parameter group of trajectory collineations TC_s defined by vectors ξ^i_α, $\alpha = 1, \ldots, s$. With a given constant of the motion I of the form (2.3), Theorem 4.1 generates derived constants of the motion I_α which are formulated in terms of the vectors ξ^i_α. Repeated application of Theorem 4.1 based upon the deformation of I_α with respect to ξ^i_β will in general lead to a new constant of the motion (second derived integral) $I_{\beta\alpha}$ given by

$$I_{\beta\alpha} = [\pounds_\beta(\pounds_\alpha B_{ij} - 4\phi_\alpha B_{ij}) - 4\phi_\beta(\pounds_\alpha B_{ij} - 4\phi_\alpha B_{ij})]v^i v^j + [\pounds_\beta(\pounds_\alpha A_i - 2\phi_\alpha A_i) - 2\phi_\beta(\pounds_\alpha A_i - 2\phi_\alpha A_i)]v^i + \pounds_\beta\pounds_\alpha\Omega . \tag{5.1}$$

We now consider dependency relations between such second derived quadratic first integrals for the case in which $\alpha \neq \beta$. From (5.1) we immediately obtain

$$I_{\beta\alpha} - I_{\alpha\beta} = [(\pounds_\beta\pounds_\alpha - \pounds_\alpha\pounds_\beta)B_{ij}]v^i v^j + [(\pounds_\beta\pounds_\alpha - \pounds_\alpha\pounds_\beta)A_i]v^i + (\pounds_\beta\pounds_\alpha - \pounds_\alpha\pounds_\beta)\Omega + 4(\pounds_\alpha\phi_\beta - \pounds_\beta\phi_\alpha)B_{ij}v^i v^j + 2(\pounds_\alpha\phi_\beta - \pounds_\beta\phi_\alpha)A_i v^i . \tag{5.2}$$

Since we have assumed the existence of a group of symmetry mappings, it follows that (see Ref. 12)

$$\pounds_\beta\pounds_\alpha - \pounds_\alpha\pounds_\beta = C^\gamma_{\beta\alpha}\pounds_\gamma , \tag{5.3}$$

$$\pounds_\beta\xi^i_\alpha = C^\gamma_{\beta\alpha}\xi^i_\gamma , \tag{5.4}$$

where $C^\gamma_{\alpha\beta}$ are the structure constants of the group of trajectory collineations. From (3.10)-(3.12) and an identity similar to (4.8) it follows by use of (5.4) that

$$\pounds_\alpha\phi_\beta - \pounds_\beta\phi_\alpha = -C^\gamma_{\beta\alpha}\phi_\gamma , \quad C^\gamma_{\alpha\beta}h_\gamma = 0. \tag{5.5}$$

By use of (5.3) and (5.5) in (5.2) we obtain

$$I_{\beta\alpha} - I_{\alpha\beta} = C^\gamma_{\beta\alpha}I_\gamma . \tag{5.6}$$

Hence we see that if a conservative dynamical system admits a trajectory collineation group, then the structure of this group is essential to the linear dependency relations between derived first integrals which are based upon the deformations of a given first integral under the action of these symmetry mappings.[15]

6. APPLICATION I: ISOTROPIC SIMPLE HARMONIC OSCILLATOR

We now consider the three-dimensional isotropic simple harmonic oscillator with both angular frequency and

mass taken to be unity. For simplicity we use Cartesian coordinates ($g_{ij} \equiv \delta_{ij}$).

We first find the trajectory collineation vectors for this dynamical system. Equation (3.10) now becomes the condition for a projective collineation in a Euclidean space E_3. In Cartesian coordinates it has the well-known solution (see Ref. 13)

$$\xi^k = a_m x^m x^k + s^k_m x^m + \omega^k_m x^m + c^k, \quad s^k_m = s^m_k, \quad \omega^k_m = -\omega^m_k, \tag{6.1}$$

where a_m, s^k_m, ω^k_m, and c^k are constants. From (6.1) we obtain for $n = 3$, the vectors ξ^i_α, $\alpha = 1, \ldots, 15$ [based upon the 15 arbitrary constants which appear in (6.1)] which define the 15-parameter group PC_{15}.

With the potential energy of the oscillator defined by $V \equiv \frac{1}{2}\delta_{ij}x^i x^j$ and with ξ^k given by (6.1), we find that (3.11) reduces to

$$4\phi x^i + c^i - a_m x^m x^i = 0. \tag{6.2}$$

From the divergence of (6.2) we obtain

$$\phi = \tfrac{1}{3}(a_m x^m - \phi_{,m} x^m). \tag{6.3}$$

Use of (6.1) and (3.12) gives $\phi_{,m} = a_m$. This result with (6.3) requires that $\phi = 0$. Consequently [from (6.2)] $c^i = 0$ and $a^i = 0$, $i = 1, 2, 3$; therefore, we obtain

$$\xi^i = s^i_j x^j + \omega^i_j x^j \quad (s^i_j = s^j_i, \ \omega^i_j = -\omega^j_i). \tag{6.4}$$

Substitution of each of the independent vectors ξ^i_α, $\alpha = 1, \ldots, 9$ (based upon the nine independent constants s^i_j, ω^i_j) into (3.12) dictates the values of the constants h_α such that $\phi_\alpha = 0$. It follows from (3.2) that $d\bar{t} = dt$ for this TC_9 group.

With $\phi_\alpha = 0$ we note from (3.10) that $\pounds_\alpha \{^i_{jk}\} = 0$. Hence the TC_9 admitted by the three-dimensional isotropic harmonic oscillator is the group of affine collineations[13] AC_9 (or linear homogeneous group[6]) in the configuration space E_3. This group contains three rotations (parameters ω^i_j) and six proper affine collineations (parameters s^i_j). The generators $\xi^i_\alpha \partial_i$ ($\partial_i \equiv \partial/\partial x^i$) are $R_\alpha \equiv (x^\beta \partial_\gamma - x^\gamma \partial_\beta)$, $A_\alpha \equiv (x^\beta \partial_\gamma + x^\gamma \partial_\beta)$, (α, β, γ) cyclic permutations of $(1, 2, 3)$; $\bar{S}_\alpha \equiv x^\alpha \partial_\alpha$, (no sum), $\alpha = 1, 2, 3$. The group structure for this TC_9 is given in Table I.

From (4.13) with ξ^j defined by (6.4), $\phi = 0$, and $V = \frac{1}{2}\delta_{ij}x^i x^j$ we obtain for the isotropic harmonic oscillator the quadratic constant of the motion

$$s^i_j(x^i x^j + v^i v^j). \tag{6.5}$$

Thus for each of the six independent parameters s^i_j of the proper affine collineations we find a concomitant quadratic constant of the motion based upon the deformation of the energy integral. These six quadratic first integrals are immediately recognized to be the components of the well-known[7] symmetric tensor associated with the oscillator problem.

We note that the three rotational symmetries of the TC_9 do not lead to (first) derived quadratic constants of the motion because they do not deform the energy integral.

From the TC_9 [(6.4)] we construct an eight-parameter subgroup by requiring $\mathrm{Tr}(s^i_j) = 0$. We take s^3_3 as dependent and obtain the eight generators $\xi^i_\alpha \partial_i$:

$$R_\alpha \equiv (x^\beta \partial_\gamma - x^\gamma \partial_\beta), \quad A_\alpha \equiv (x^\beta \partial_\gamma + x^\gamma \partial_\beta), \quad (\alpha, \beta, \gamma) \text{ a cyclic permutation of } (1, 2, 3);$$
$$S_\alpha \equiv (x^\alpha \partial_\alpha - x^3 \partial_3), \quad \text{(no sum)}, \quad \alpha = 1, 2. \tag{6.6}$$

The structure of this TC_8 is given in Table II.

By a change in basis this TC_8 may be brought into a form [Ref. 7 (1965), Eq. (26)] which has commutation relations characteristic of SU_3. The generators which exhibit the SU_3 group structure are obtained from those of (6.6) by $\bar{A}_\epsilon \equiv -\frac{1}{2}\epsilon(A_2 + \epsilon i A_1)$, $\bar{A}_{2\epsilon} \equiv (S_1 - S_2 + \epsilon i A_3)$, $\bar{L}_\epsilon \equiv -i(R_1 + \epsilon i R_2)$, $\epsilon = \pm 1$, $\bar{L}_3 \equiv -iR_3$, $\bar{A}_0 \equiv -(S_1 + S_2)$.

7. APPLICATION II: KEPLER PROBLEM

We consider now the three-dimensional Kepler problem for a unit mass. With the use of Cartesian coordinates ($g_{ij} = \delta_{ij}$) the potential energy is

$$V = -k_0/r, \quad r \equiv [\textstyle\sum_{i=1}^3 (x^i)^2]^{1/2}, \quad k_0 \equiv \text{const.}$$

A trajectory collineation vector ξ^k for this dynamical system must satisfy (3.10). Hence in Cartesian coordinates it must also be of the form (6.1). For the Kepler potential we find (3.11) takes the form

$$4\phi x^i - 4a_m x^m x^i - 3s^k_m x^m x^k x^i r^{-2} - 3c^k x^k x^i r^{-2} = 0. \tag{7.1}$$

The divergence of (7.1) and use of (3.12) imply

$$\phi = a_m x^m + \tfrac{3}{4} s^k_m x^m x^k r^{-2} + \tfrac{1}{2} c^k x^k r^{-2}. \tag{7.2}$$

By (7.2) and (7.1) we obtain $c^k = 0$, $k = 1, 2, 3$. We equate grad ϕ based on (7.2) (with $c^k = 0$) with grad ϕ based upon (3.12) and obtain $s^j_k = s\delta^j_k$. It then follows that all $h_\alpha = 0$ in (3.12). Thus the Kepler problem in three dimensions admits the TC_7 group defined by the vectors ξ^i_α, $\alpha = 1, \ldots, 7$, based on the remaining seven independent constants and written collectively in the form[16]

$$\xi^i = a_j x^j x^i + \omega^i_j x^j + s x^i, \quad \omega^i_j = -\omega^j_i. \tag{7.3}$$

TABLE I.[a] Isotropic harmonic oscillator TC_9 commutators $[X_\alpha, X_\beta]$.

X_α \ X_β	$\bar{S}_1$	$\bar{S}_2$	$\bar{S}_3$	R_1	R_2	R_3	A_1	A_2	A_3
$\bar{S}_1$	0	0	0	0	$-A_2$	A_3	0	$-R_2$	R_3
$\bar{S}_2$	0	0	0	A_1	0	$-A_3$	R_1	0	$-R_3$
$\bar{S}_3$	0	0	0	$-A_1$	A_2	0	$-R_1$	R_2	0
R_1	0	$-A_1$	A_1	0	$-R_3$	R_2	$\bar{W}_{23}$	A_3	$-A_2$
R_2	A_2	0	$-A_2$	R_3	0	$-R_1$	$-A_3$	$\bar{W}_{31}$	A_1
R_3	$-A_3$	A_3	0	$-R_2$	R_1	0	A_2	$-A_1$	$\bar{W}_{12}$
A_1	0	$-R_1$	R_1	$\bar{W}_{32}$	A_3	$-A_2$	0	$-R_3$	R_2
A_2	R_2	0	$-R_2$	$-A_3$	$\bar{W}_{13}$	A_1	R_3	0	$-R_1$
A_3	$-R_3$	R_3	0	A_2	$-A_1$	$\bar{W}_{21}$	$-R_2$	R_1	0

[a] The symbol $\bar{W}_{\alpha\beta} \equiv 2(\bar{S}_\alpha - \bar{S}_\beta)$, $\alpha, \beta = 1, 2, 3$.

TABLE II.[a] Isotropic harmonic oscillator TC_8 commutators $[X_\alpha, X_\beta]$.

X_α \ X_β	S_1	S_2	R_1	R_2	R_3	A_1	A_2	A_3
S_1	0	0	A_1	$-2A_2$	A_3	R_1	$-2R_2$	R_3
S_2	0	0	$2A_1$	$-A_2$	$-A_3$	$2R_1$	$-R_2$	$-R_3$
R_1	$-A_1$	$-2A_1$	0	$-R_3$	R_2	$2S_2$	A_3	$-A_2$
R_2	$2A_2$	A_2	R_3	0	$-R_1$	$-A_3$	$-2S_1$	A_1
R_3	$-A_3$	A_3	$-R_2$	R_1	0	A_2	$-A_1$	W_{12}
A_1	$-R_1$	$-2R_1$	$-2S_2$	A_3	$-A_2$	0	$-R_3$	R_2
A_2	$2R_2$	R_2	$-A_3$	$2S_1$	A_1	R_3	0	$-R_1$
A_3	$-R_3$	R_3	A_2	$-A_1$	W_{21}	$-R_2$	R_1	0

[a] The symbol $W_{\alpha\beta} \equiv 2(S_\alpha - S_\beta)$, $\alpha, \beta = 1, 2$.

From (7.2) and (7.3) we find

$$\phi = a_j x^j + 3s/4 \tag{7.4}$$

and hence from (3.2) it follows that $d\bar{t} = dt[1 + (2a_j x^j(t) + 3s/2)\delta a]$.

The transformations based on these ξ^i_α include three proper projective collineations (parameters a_j), a homothetic motion or scale change (parameter s), and a three-parameter rotation subgroup (parameters ω^i_j). The generators $\xi^i_\alpha \partial_i$ are $P_\alpha \equiv x^\alpha x^i \partial_i$, $\alpha = 1, 2, 3$, $S \equiv x^i \partial_i$, $R_\alpha \equiv (x^\beta \partial_\gamma - x^\gamma \partial_\beta)$, (α, β, γ) cyclic permutations of (1, 2, 3). The group structure for this TC_7 is given in Table III. We note this TC_7 is isomorphic to the similitude group in three-dimensional Euclidean space.[6]

TABLE III. Kepler TC_7 commutators $[X_\alpha, X_\beta]$.

X_α \ X_β	S	R_1	R_2	R_3	P_1	P_2	P_3
S	0	0	0	0	P_1	P_2	P_3
R_1	0	0	$-R_3$	R_2	0	$-P_3$	P_2
R_2	0	R_3	0	$-R_1$	P_3	0	$-P_1$
R_3	0	$-R_2$	R_1	0	$-P_2$	P_1	0
P_1	$-P_1$	0	$-P_3$	P_2	0	0	0
P_2	$-P_2$	P_3	0	$-P_1$	0	0	0
P_3	$-P_3$	$-P_2$	P_1	0	0	0	0

Based upon the potential $V = -k_0/r$, we obtain from (4.13) by use of (7.3) and (7.4) the (nontrivial) derived quadratic first integral

$$a_m[(\delta^m_j x^i - x^m \delta_{ij})v^i v^j + k_0 x^m/r] = k. \tag{7.5}$$

The derived constant of the motion based on parameter s was essentially the undeformed energy and hence was not included in (7.5). Again we observe that the rotation symmetries leave invariant the energy integral and differential path parameter and hence do not generate new quadratic constants of the motion. The three quadratic first integrals obtained from (7.5) are the well-known components of the Runge–Lenz vector.[7] Thus we have a second illustration which shows how quadratic constants of the motion are generated by the deformation of the energy integral as a result of a symmetry mapping within the configuration space.

The TC_7 of the Kepler problem admits a six-parameter subgroup TC_6 (refer to Table III) with generators R_α and P_α, $\alpha = 1, 2, 3$. (This projective TC_6 has the same group structure as the Euclidean group of rotations and translation.) The existence of such a dynamical symmetry subgroup is not surprising. By means of a canonical transformation it has been shown[17] that any dynamical system of three degrees of freedom with rotationally invariant Hamiltonian will admit six constants of the motion whose Poisson bracket algebra will exhibit this characteristic TC_6 structure.

Note added in proof: A phase space formulation of a related integral theorem for dynamical systems governed by Hamilton's equations has been obtained [G. H. Katzin, Lett. Nuovo Cimento (to appear 1973)]. The allowed dynamical symmetry mappings are a generalization of canonical transformations and the related integral theorem which results may be interpreted as a generalization of Poisson's theorem on constants of the motion.

ACKNOWLEDGMENTS

The author wishes to thank Dr. W. R. Davis, Dr. J. Levine, Dr. M. K. Moss, and Mr. A. P. Grishin for their helpful comments.

---

[1] G. H. Katzin and J. Levine, J. Math. Phys. **9**, 8 (1968).

[2] For a derivation of the conditions for projective collineations by use of Lie derivatives, see G. H. Katzin and J. Levine, Tensor, New Ser. **19**, 162 (1968).

[3] In an similar manner it can be shown that if V_n admits a conformal motion (which maps null geodesics into null geodesics) and a known constant of the motion along the null geodesics, then in general an additional constant of the motion of the null geodesics results from the deformation of the given one as a result of this conformal symmetry mapping. For an analysis of this technique and a general summary of various types of Riemannian space symmetries and concomitant constants of the motion for geodesics paths, see G. H. Katzin and J. Levine, Colloq. Math. (Wroclaw) **26**, 21 (1972).

[4] W. R. Davis and M. K. Moss, Nuovo Cimento **38**, 1558 (1965).

[5] For a discussion of accidental degeneracy refer to: J. M. Jauch and E. L. Hill, Phys. Rev. **57**, 641 (1940); D. F. Greenberg, Am. J. Phys. **34**, 1101 (1966); H. V. McIntosh, Am. J. Phys. **27**, 620 (1959).

[6] M. Hamermesh, *Group Theory* (Addison-Wesley, Reading, Mass., 1964).

[7] D. M. Fradkin, Am. J. Phys. **33**, 207 (1965); Prog. Theor. Phys. **37**, 789 (1967).

[8] Latin indices run from 1 to n. The Einstein summation notation is used unless stated otherwise. We indicate absolute total time differentiation by D/dt, total time differentiation by d/dt, partial differentiation with x^k by comma (,), and convariant differentiation (based upon Christoffel symbols $\{^i_{jk}\}$) by semicolon (;).

[9] We stress that the constant value of the integral is not specified to be the same for all trajectories. If it were so fixed, the first integral would act as a constraint for the dynamical symmetry mappings to be described in the next section.

[10] The symbol ( ) denotes symmetrization, so that $B_{(ij;k)} \equiv (1/3)(B_{ij;k} + B_{jk;i} + B_{ki;j})$.

[11] The symbol $\pounds_\xi$ indicates Lie differentiation with respect to the vector ξ^i.

[12] K. Yano, *The Theory of Lie Derivatives and Its Applications* (North-Holland, Amsterdam, 1957).

[13] L. P. Eisenhart, *Non-Riemannian Geometry*, Amer. Math. Soc., Providence, R.I. AMS Colloquium Publications, Vol. 8 (1972).

[14] J. A. Schouten, *Ricci-Calculus* (Springer-Verlag, Berlin, 1954), p. 102.

[15] We note that Eq. (5.9) of Ref. 1 is erronous; therefore, Theorem 5.2 of that paper should be deleted.

[16] The auther wishes to acknowledge the assistance of Mr. John A. Pasour in checking the calculations for the Kepler problem.

[17] N. Mukunda, Phys. Rev. **155**, 1383 (1967).

Charges as generators of unitary symmetry groups

H. Snellman

Department of Theoretical Physics, Royal Institute of Technology, S-10044 Stockholm, Sweden
(Received 29 November 1971)

We show, that in a local field theory of the Wightman type, with an irreducible set of finitely many, relatively local field operators $\phi_k, 1 \leq k \leq n$, and a set of local, locally conserved symmetric current operators $J_i^\mu(x)$, the charge operators G_i that are associated with these currents are essentially self-adjoint on the domain of strictly localized states, provided the G_i:s satisfy $[G_i, \phi_k(g)] = \sum_{l=1}^n T_{kl}^i \phi_l(g)$, $g \in D(\mathbb{R}^4)$, where the T^i:s are finite dimensional matrices. Furthermore when the T^i:s close to a faithful representation of a Lie algebra g, then g is finite-dimensional and compact and its representation can be integrated to a unitary representation of corresponding Lie group.

1. INTRODUCTION

The quantum-mechanical approach to symmetries follows the lines originated by Wigner.[1] In his scheme one considers invariances of transition probabilities under symmetry transformations of the state vectors. Wigner showed that such transformations can be realized as unitary (or antiunitary) representations of the symmetry group in question. In the case of a continuous symmetry group, the representation of the generators of the group consists of operators that are essentially self-adjoint on their common invariant Gårding domain. Since essentially self-adjoint operators have unique self-adjoint closures, the operators that represent the generators of the symmetry group are candidates for observables connected with the symmetry. The importance of having self-adjoint operators as observables lies of course in their spectral properties and in the completeness of their eigenfunction spaces.

In local field theory, we expect that the symmetry should manifest itself already at the local level. Thus, in local Lagrangian quantum field theory, Noether's theorem[2] states that, to any given group of point transformations of the fields that leaves the Lagrangian invariant, there corresponds a set of locally conserved current densities. The time-independent space integrals of the time components of these currents, which are called "formal charges," correspond to the generators of the symmetry group in the sense that their action on the fields under formal commutation is that expected of the generators.

Intuitively one would like to consider these charges as observables. From the previous discussion it is then necessary to show that they are in some sense at least essentially self-adjoint on some dense domain in Hilbert space. Following Ref. 3, this problem, how to come from the local current conservation law to the generator, will be called here "the converse of Noether's theorem."

At a more rigorous level of Wightman field theory, it turns out that a formal charge never exists as an operator in the Hilbert space $\mathcal{H}$ of states, but that it determines, by its matrix elements between vectors in a dense subdomain, the sesquilinear form of an unbounded operator G defined on this domain.[4] If the current operator is symmetric, then G is symmetric, and the problem is thus to show that G is at least essentially self-adjoint on some dense domain in $\mathcal{H}$. What is lacking in the Wightman framework is a dynamical principle, like Noether's theorem, which can give us the currents corresponding to a certain symmetry group. The best we can do at present is therefore to postulate the action of the charge, or rather the operator G, on the field algebra. This action will determine the equivalence class of those current operators that correspond to a certain operator G.

We will restrict ourselves to a study of this problem for the case of charges corresponding to internal (space–time decoupled) symmetries, assuming that there is in the theory a basic irreducible set $\{\phi_i\}_1^n$ of n local fields transforming linearly among themselves under the action of ad$G$. This approach still leaves open the question of whether such a set can always be chosen.

In Sec. 2 we recall some theorems of importance for a consistent formulation of the action of G in the field algebra. In Sec. 3 we show that under the assumption that ad$G$ acts as the particular derivation in the set $\{\phi_i\}_1^n$ mentioned above, then G is essentially self-adjoint on the domain of strictly localized states, and its closure can be integrated to a unitary representation of a one-parameter symmetry group.

In Sec. 4 we give conditions under which a set of G:s close under commutation to a Lie algebra. It turns out that the Lie algebra is compact and integrable to a unitary representation of a corresponding Lie group.

Section 5 is devoted to criticism of the assumptions and a discussion of the significance of the results.

2. CHARGES IN QUANTUM FIELD THEORY

Let us consider a local quantum field theory of the Wightman type with a set $\mathfrak{A} = \{\phi_k(g); g \in \mathcal{S}(\mathbb{R}^4); k \in K\}$ of local, relatively local field operators $\phi_k(g)$ ordered by some finite index set K. These operators act in a Hilbert space $\mathcal{H}$ of state vectors. We assume that the set $\mathfrak{A}$ is irreducible and generates the field algebra $\mathcal{F}(\mathfrak{A})$. This means that $\overline{\mathcal{F}(\mathfrak{A})\Omega} = \mathcal{H}$, where Ω is the unique vacuum state in $\mathcal{H}$. We will further assume that the energy-momentum spectrum has a mass-gap.

Two subsets $\mathcal{D}_{qL}$ and $\mathcal{D}_L$ of $\mathcal{H}$ are of importance. $\mathcal{D}_L$, the set of strictly localized states is defined by

$$\mathcal{D}_L = \{\Phi \in \mathcal{H}; \Phi = \sum_{m=0}^{M} \int d^4x_1 \cdots d^4x_m g_m(x_1, \ldots, x_m)\, \phi_{l_1}(x_1) \cdots \phi_{l_m}(x_m)\Omega \quad \forall\, M \text{ and } g_m \in \mathcal{D}(\mathbb{R}^{4m})\}. \tag{2.1}$$

($\mathcal{D}$ and $\mathcal{S}$ are the ordinary test function spaces defined by L. Schwartz.[5])

The set $\mathcal{D}_{qL}$ of quasilocal states is defined as $\mathcal{D}_L$ but with $g_m \in \mathcal{S}(\mathbb{R}^{4m})$.

Assume now that we are given a set

$\mathfrak{B} = \{J_i^\mu(g); i \in I; J_i^\mu(\partial_\mu g) = 0; g \in \mathcal{S}(\mathbb{R}^4)\}$ of local,

Copyright © 1973 by the American Institute of Physics

locally conserved four-vector current operators $J_i^\mu(g)$, symmetric on $\mathfrak{D}_L$ and relatively local to $\mathfrak{A}$.

To each one of these currents we associate a formal charge $Q_i = \lim_{R\to\infty} Q_{i,R}$, defined by

$$Q_{i,R} = J_i^0(f_T, f_R). \tag{2.2}$$

The test functions f_T and f_R are defined as follows[6]:

(a) $f_T(x^0) \in \mathfrak{D}(\mathbb{R}^1)$ and $f_R(\bar{x}) \in \mathfrak{D}(\mathbb{R}^3)$;

(b) $f_T(x^0) = 0$ for $|x^0| > T$,

$$f_T(x^0) = f_T(-x^0), \qquad \int f(x^0)dx^0 = 1,$$

$$f_R(\bar{x}) = \begin{cases} 1 & \text{for } |\bar{x}| < R \\ f_R(|\bar{x}|) & \text{for } R < |\bar{x}| < R + L,\ L > 0 \\ 0 & \text{for } |\bar{x}| > R + L. \end{cases}$$

The matrix elements of $Q_{i,R}$ between states in $\mathfrak{D}_{qL}$ define a sesquilinear form over $(\mathfrak{D}_{qL}, \mathfrak{D}_{qL})$. Thus

$$Q_{i,R}(\Psi, \chi) := (\Psi, J_i^0(f_T, f_R)\chi), \qquad \Psi, \chi \in \mathfrak{D}_{qL}.$$

We now have the following theorems:

Theorem 2.1[4,6,7]: The sesquilinear form $Q_{i,R}(\Psi, \chi)$ over $(\mathfrak{D}_{qL}, \mathfrak{D}_{qL})$ determines in the limit $R \to \infty$ the form of a charge operator G_i with domain $\mathfrak{D}_{qL}$.

For a proof see, e.g., Ref. 3.

Since the currents are assumed to be symmetric, the operators G_i are symmetric.

Under the assumptions of Theorem 2.1 it follows[6] that $Q_i(\Psi, \Omega) = 0$ since $\Omega \in \mathfrak{D}_{qL}$. Thus $(\Psi, G_i\Omega) = 0$ for $\Psi \in \mathfrak{D}_{qL}$. But since $\overline{\mathfrak{D}_{qL}} = \mathfrak{K}$ we get by continuity $G_i\Omega = 0$. For our considerations in next section we also need the following result.

Theorem 2.2[3,6]: $[G_i, \phi_k(g)]\mathfrak{D}_L \subset \mathfrak{D}_L, \forall\, i \in I$, whenever $g \in \mathfrak{D}(\mathbb{R}^4)$, i.e., the commutator $[G_i, \phi_k(g)]$ is a strictly localized operator.

3. ESSENTIAL SELF-ADJOINTNESS OF CHARGE OPERATORS

We are now prepared to postulate the action of the charge operators G_i on $\mathfrak{A}$. Our basic assumption, modeled after the situation in Lagrangian quantum field theory is that

$$[G_i, \phi_k(g)] = \sum_{l=1}^{n} T_{kl}^i \phi_l(g), \qquad g \in \mathfrak{D}(\mathbb{R}^4), \tag{3.1}$$

where T^i is a matrix of rank n. Relation (3.1) is understood to hold on $\mathfrak{D}_L$. That this kind of transformation law is not incompatible with the previous assumptions follows from Theorem 2.2 and the fact that $\phi_k(g)\mathfrak{D}_L \subset \mathfrak{D}_L$ and from a study of free-field models.

We now have the following.

Theorem 3.1: Given a set of symmetric charge operators $G_i, i \in I$, acting in $\mathfrak{A}$ according to (3.1) as a relation valid on $\mathfrak{D}_L$, the G_i are essentially self-adjoint on D_L and their closures can be integrated to unitary one-parameter groups.

Proof: G_i is by assumption a symmetric operator with domain $D_{G_i} = \mathfrak{D}_L$ which is invariant under G_i. Application of (3.1) to $\Omega \in \mathfrak{D}_L$ gives

$$G_i(\phi_k(g)\Omega) = T_{kl}^i(\phi_l(g)\Omega). \tag{3.2}$$

Thus to each $g \in \mathfrak{D}(\mathbb{R}^4)$ we can associate a vector space $V_g \subset \mathfrak{D}_L$, with $\dim V_g \le n$, which is invariant under G_i. The restriction $G_i|_{V_g}$ of G_i to V_g is given by the same T^i for every $g \in \mathfrak{D}(\mathbb{R}^4)$. The T^i are Hermitian operators since every G_i is symmetric on $\mathfrak{D}_L$ and thus $G_i|_{V_g}$ is a Hermitian operator since V_g is invariant under G_i.

Thus we can integrate $G_i|_{V_g}$ to a unitary representation of a one-parameter group in V_g:

$$\exp(itG_i|_{V_g}) = \exp(itT^i). \tag{3.3}$$

Now $\cup_{g\in\mathfrak{D}(\mathbb{R}^4)} V_g = \mathfrak{D}_L^1$, where $\mathfrak{D}_L^1$ is the linear span of vectors

$$\{\phi_k(g)\Omega; \forall k \in K; \forall g \in \mathfrak{D}(\mathbb{R}^4)\}.$$

Since V_g is finite-dimensional,

$$\bigcup_{g\in\mathfrak{D}(\mathbb{R}^4)} \exp(itG_i|_{V_g}) = \exp(itG_i|_{\mathfrak{D}_L^1}) = \bigcup_{g\in\mathfrak{D}(\mathbb{R}^4)} \exp(itT^i) \tag{3.4}$$

is an isometric representation of a one-parameter group in $\mathfrak{D}_L^1$.

We now consider the filtered subspaces $\mathfrak{D}_L^m \subset \mathfrak{D}_L$ obtained by applying polynomials of degree $\le m$ in the field operators in $\mathfrak{A}$ on Ω. Explicity we have

$$\mathfrak{D}_L^m = \Big\{\Phi \in \mathfrak{D}_L; \Phi = A_\Phi\Omega = \sum_{j=1}^{k} \phi_{l_1}(g_1)\cdots\phi_{l_j}(g_j)\Omega;\ \forall k \le m,\ \forall g_i \in \mathfrak{D}(\mathbb{R}^4)\Big\}.$$

For each fixed product of g_i:s we get a linear finite-dimensional subspace $V_{g_1\cdots g_k}$ of $\mathfrak{D}_L^m$, which is invariant by G_i.

Consider the operator $G_i|_{V_{g_1\cdots g_k}}$, which is Hermitian and can be integrated to a unitarity one-parameter group defined by

$$\exp(itG_i|_{V_{g_1\cdots g_k}})\Phi_{g_1\cdots g_k} = [\exp\{it(\mathrm{ad}G_i|_{V_{g_1\cdots g_k}})\}(A_{\Phi_{g_1\cdots g_k}})]\Omega, \tag{3.5}$$

where $\Phi_{g_1\cdots g_k} \in V_{g_1\cdots g_k}$. By gluing all spaces $V_{g_1\cdots g_k}$ together with respect to the test-functions and taking the union over all $k \le m$ we get $\mathfrak{D}_L^m$. Thus by similarly gluing all unitary operators on $V_{g_1\cdots g_k}$ together and taking the union over all $k \le m$ we get

$$\exp(itG_i|\mathfrak{D}_L^m) = \bigcup_{k=1}^{m} \bigcup_{g_1\cdots g_k \in \otimes_k^{\mathfrak{D}}(\mathbb{R}^4)} \exp(itG_i|_{V_{g_1\cdots g_k}}) \tag{3.6}$$

which is easily seen to be the isometric one-parameter groups on $\mathfrak{D}_L^m$.

Now

$$\overline{\lim_{m\to\infty} \operatorname{ind} \mathfrak{D}_L^m} = \mathfrak{K}$$

where bar denotes the closure in the Hilbert space topology. By linearity therefore the G_i:s integrate to unitary representations of one-parameter groups on $\mathfrak{K}$.

We have thus shown that each G_i, which is defined as a symmetric operator on $\mathfrak{D}_L$ can be integrated to a uni-

tary representation of a one-parameter group on $\mathcal{H}$ which leaves $\mathfrak{D}_L$ invariant. As a consequence of Stone's Theorem and a lemma by Nelson[8] it follows that G_i is essentially self-adjoint on $\mathfrak{D}_L$. (This follows after passing from $\otimes_m \mathfrak{D}(\mathbb{R}^4)$ to $\mathfrak{D}(\mathbb{R}^{4m})$ for every m.)

Remarks. 1. An alternative way to prove the essential self-adjointness of the G_i:s is to use Nelson's Theorem[9]. Thus we have for Φ^m the estimate

$$\sum_{k=0}^{\infty} \frac{s^k}{k!} \|G_i^k \Phi^m\| < C_m \exp(snm \|T^i\|),$$

where $\Phi^m = \phi_{l_1}(g_1) \ldots \phi_{l_m}(g_m)\Omega$.

This shows that there is a dense set of entire vectors for the G_i:s.

2. It is clear from the method of proof above that we can permit the T^i:s to depend upon g as an index. Thus we may take

$$[G_i, \phi_k(g)] = T^i_{kl}(g)\,\phi_l(g) \quad g \in \mathfrak{D}(\mathbb{R}^4). \tag{3.1'}$$

The proof proceeds as before. We have

$$\exp(itG_i|_{V_g}) = \exp(itT^i_{(g)}).$$

By compatibility with the weak topology induced from $\mathcal{H}$ on $\mathfrak{D}_L^1$, the $T^i(g)$:s must coincide when the g:s overlap since $G_i|_{\mathfrak{D}_L^1}$ is symmetric. Hence

$$\exp(itG_i|\mathfrak{D}_{L^1}) = \exp(it\mathbb{T}^i)$$

where $\mathbb{T}^i = \cup_{g\in\mathfrak{D}(\mathbb{R}^4)} T^i(g)$ is the union of $T^i(g)$ over the g:s. The rest of the proof goes as before, now with $\mathbb{T}^i$ replacing T^i.

This possibility will not be considered further.

4. COMPACTNESS OF FINITE-DIMENSIONAL INTERNAL SYMMETRY GROUPS

Given that the G_i:s are essentially self-adjoint operators on $\mathfrak{D}_L$ we consider the system $\{G_i; i \in I\}$ of these operators. The fundamental assumption (3.1) on the action of the G_i:s on $\mathfrak{A}$ turns out to be very restrictive. We have the following

Theorem 4.1. Assume that the T^i:s close under the Lie-product to a faithful representation of a symmetry algebra $\mathbf{g}$. Then the corresponding one-parameter groups close to a compact Lie group.

Lemma: If the T^i:s form a faithful representation of $\mathbf{g}$ we have $\dim \mathbf{g} \leq n^2$.

Proof: Since the T^i:s are Hermitian n-dimensional matrices, there are at most n^2 such ones. Thus if order $I > n^2$ all G_i:s are not linearly independent. Thus by a linear change of basis, all G_i:s with $i > n^2$ will have T^i:s = 0. This contradicts the assumption that the T^i:s form a faithful representation of $\mathbf{g}$. The G_i:s thus span a Lie-algebra $\mathbf{g}$ with $\dim \mathbf{g}$ = order $I \leq n^2$.

Proof of Theorem 4.1. Assume that the T^i:s close to a representation of $\mathbf{g}$ according to

$$\sum_m T^i_{lm} T^j_{mn} - \sum_m T^j_{lm} T^i_{mn} = \sum_k c^{ji}_k T^k_{ln} \tag{4.1}$$

where the c^{ij}_k:s are the structure constants of $\mathbf{g}$. Then, by using the Jacobi identity

$$\text{cycl.}\,[G_i, [G_j, \phi_k(g)]] = 0 \tag{4.2}$$

applied to Ω we get by simple computation that the G_i:s satisfy

$$[G_i, G_j]|_{V_g} = c^{ij}_k G^k|_{V_g} \tag{4.3}$$

The Lie-algebra $\mathbf{g}$ can thus be integrated to a unitary group matrix on V_g, the closure of which is compact. The procedure of passing from a representation on V_g to a representation on $\mathcal{H}$ goes as before. Since the G_i:s are represented by the T^i:s on V_g we obviously get a finite-dimensional, faithful, Hermitian representation of $\mathbf{g}$. As is very well known, this implies that $\mathbf{g}$ is compact.

Remark. The integrability of $\mathbf{g}$ can also be demonstrated by verifying the commutation relations directly on $\mathfrak{D}_L$. The compactness of $\mathbf{g}$ is demonstrated as above. The fact that we have a dense domain of entire vectors on which the G_i:s can be simultaneously integrated does not by itself exclude the possibility of having g noncompact and nilpotent.[10] The compactness of the internal symmetry group has also been discussed under different assumptions by Ref. 11 and 12.

5. CONCLUSIONS, CRITISISM AND SUGGESTIONS

Our main results are contained in theorems 3.1 and 4.1.

From the physical point of view, Theorem 3.1 gives a satisfactory connection between time independent charge-operators and observables in local quantum field theory.

More interesting is perhaps Theorem 4.1 which shows that space-time decoupled symmetry algebras, acting in the generating finite set of irreducible field operators according to (3.1), are of compact type. This means that if we want to have a non-compact symmetry algebra we must modify the basic assumptions. This will be discussed later on.

Actually it seems as if most of the field theoretical models considered in the literature are covered by our assumptions. This depends of course upon that assumption (3.1) is directly built in into these models. It is in fact not clear what one should mean with a symmetry in a field theory, if this is not the case.

Since we have restricted ourselves to a linear action of the G_i:s with respect to $\mathfrak{A}$, our results do not cover those models, where $SU(2) \otimes SU(2)$ is realized non-linearly on the field algebra. However, in the case of massive pions, $SU(2) \otimes SU(2)$ is not a symmetry of these models, and the case of mass-less pions is excluded by assumption. Other types of non-linear realization of the symmetry must face the difficulties already with the distribution theoretical definition of a product of field operators at the same point in space-time.

The symmetry algebras covered by our assumptions turn out to be integrable to group representations. Since this is a quite strong result we want to pass some comments on local representations here.

Firstly, all our symmetries are space-time decoupled by assumption (3.1). Hence our result does not apply to recent attempts[13] to consider local representations of non-trivial unifications of the Poincaré algebra with some internal symmetry algebra.

Secondly, even in the case of rotation group, $SU(2)$, we know from experience with the study of complex angular

momentum that local representations are of physical importance.

In passing it should be noticed, that the demonstration of Noether's Theorem utilises only the infinitesimal action of the symmetry group. Thus there is *a priori* no reason at all why the Lie-algebra of the generators should be integrable to a group representation in $\mathcal{H}$.

The automatic integrability for space–time decoupled symmetry algebras therefore seems to be directly connected with the fact, that in this case the internal parameter spaces do not have any physical interpretation. Hence, contrary to what is the case for space–time symmetries, the boundary conditions which distinguish integrable vectors from non-integrable ones, do not have physical significance.

The criticism of our basic assumptions can be summed up as follows:

(a) We work with finitely many field operators.

(b) We consider only space–time decoupled symmetries.

(c) We consider only linear actions of the $G_i: s$ on $\mathfrak{A}$.

(d) We lack in the Wightman framework a principle like Noether's Theorem, to give us the conserved currents corresponding to a given symmetry.

If point (d) were solved we would of course immediately know the action of the $G_i: s$ in $\mathfrak{A}$. On the other hand, as we mentioned in the introduction, the specification of the action of the $G_i: s$ in $\mathfrak{A}$ determines the corresponding equivalence classes of the local, locally conserved currents. In this approach it still remains to be investigated, whether fields satisfying assumption (3.1) always can be chosen. To get away from the compact case we must obviously consider modifications of assumption (3.1) in accordance with the criticism in points (a), (b), and (c).

On the other hand it is not at all clear that the space–time coupled symmetries always will generate conserved currents. Thus point (b) will eventually turn out to be disconnected to Noether's Theorem and therefore to it's converse.

The technical problems met in connection with point (c) has already been discussed. It should however be evident that Theorems 3.1 and 4.1 are quite loosely coupled to the structure described in Section 2. This means that the integrability holds for any symmetric operator that fulfills the conditions of the theorem independent of whether it is a charge operator or not. It also means that the results will be applicable to non-relativistic quantum field theory, in particular since the locality property is used only in section 2.

It is a pleasure to thank Professor M. Flato for many fruitful and constructive discussions on the subject of this paper.

I also would like to thank Professor B. Nagel for encouragement and helpful comments on earlier versions of the manuscript.

---

[1]E. Wigner, Ann. Math. (2) **40**, 149 (1939).
[2]E. Noether, Nachr. Akad. Wiss. Göttingen Math.-Phys. Kl., 235 (1918). For a treatment applicable to quantum field theory see K. Nishijima, *Fields and Particles* (Benjamin, New York, 1969).
[3]C. A. Orzalesi, Rev. Mod. Phys. **42**, 381 (1970).
[4]D . W. Robinson, in *Symmetry Principles and Fundamental Particles* (Freeman, San Francisco, 1966).
[5]L. Schwartz, *Théorie des distributions* (Hermann, Paris, 1957), Vols. 1 and 2.
[6]D. Kastler, D. W. Robinson, and A. Swieca, Commun. Math. Phys. **2**, 108 (1966).
[7]B. Schroer and P. Stichel, Commun. Math. Phys. **3**, 258 (1966).
[8]E. Nelson, Lecture notes, Princeton, 1964-1965.
[9]E. Nelson, Ann. Math. **70**, 572 (1959).
[10]See R. Goodman, Trans. Am. Math. Soc. **143**, 55 (1969).
[11]J. Lopuszanski, in *Springer Tracts in Modern Physics* (Springer-Verlag, Berlin, Heidelberg, 1970), Vol. 52, pp. 201–214.
[12]S. Doplicher, R. Haag, and J. E. Roberts, Commun. Math. Phys. **13**, 1 (1969).
[13]M. Flato and D. Sternheimer, Phys. Rev. Lett. **16**, 1185 (1969) and Commun. Math. Phys. **12**, 296 (1969).

Structure of Ališauskas–Jucys form of the 9j coefficients*

Alfred C. T. Wu

Department of Physics, University of Michigan, Ann Arbor, Michigan 48104
(Received 15 January 1973)

From the Ališauakas–Jucys triple summation expression, the Wigner 9j coefficients may be visualized as boundary values of a new generalized hypergeometric function $\phi^{(3)}(\alpha_{ik};\beta_k,\gamma_m;w_k)$ in three variables. Integral representations are given both for $\phi^{(3)}$ in general and its boundary values as the 9j coefficients. The Radon structure is discussed. It is seen that $\phi^{(3)}$ and the 9j coefficients in general do not belong to the class of hypergeometric functions whose Radon transforms are products of linear forms.

I. INTRODUCTION

In a previous paper,[1] the structure of the Wigner 9j coefficients was analyzed from the Bargmann approach. The generating function was derived, the 72-element symmetry was manifest, and a sixfold summation expression for the 9j coefficient was obtained. Contrary to the situation of the lower-order coupling and recoupling coefficients where the Clebsch–Gordan coefficent may be visualized as a ${}_3F_2$ function at $x=-1$ and the Racah coefficient as a ${}_4F_3$ function at $x=1$,[2] the 9j coefficient is seen not to belong to the ${}_pF_q$ family of functions.

However, the question was unanswered as to whether the 9j coefficient may be regarded as a boundary value of a member of some other class of generalized hypergeometric functions. A particularly interesting class of generalized hypergeometric functions is the Gel'fand type[3], being the Radon transforms of products of *linear* forms. Does the 9j coefficient satisfy the Gel'fand criterion? The sixfold summation expression derived in Ref. 1, having a rather complicated Radon transform, was not suited to answer this question.

Ališauskas and Jucys[4] have derived a remarkable triple summation expression for the 9j coefficients. This triple summation expression, while lacking in the manifest symmetry of the 9j coefficients, permits a definition of a *new* generalized hypergeometric function in three variables $\phi^{(3)}(\alpha_{kl};\beta_i,\gamma_m;w_k)$ [Eq. (11) below] of which the 9j coefficient is evaluated at $w_k=1$ together with the special values of the coefficients $\alpha,\beta,$ and γ.

A ninefold integral representation is given for $\phi^{(3)}$. When restricted to the case of 9j coefficients, a sixfold integral representation is obtained. It is seen that, in general, neither $\phi^{(3)}$ nor the 9j coefficient satisfy the Gel'fand criterion.

II. ALIŠAUSKAS-JUCYS TRIPLE SUM EXPRESSION OF 9j COEFFICIENTS

The 9j coefficient in the Ališauskas–Jucys triple summation form may be written as follows:

$$\begin{Bmatrix} j_{11} & j_{12} & j_{13} \\ j_{21} & j_{22} & j_{23} \\ j_{31} & j_{32} & j_{33} \end{Bmatrix} = K \sum_{x_1,x_2,x_3} \frac{\prod_{k=1}^{3}\prod_{l=1}^{4}(a_{kl})_{x_k}}{\prod_{k=1}^{3}(b_k)_{x_k}\prod_{\substack{m=1\\(m,n,k\ cyclic)}}^{3}(c_m)_{x_n+x_k}} \cdot \frac{1}{\prod_{k=1}^{3}x_k!} \tag{1}$$

where K is a multiplicative factor [see (6) below],

$$(a)_x \equiv \Gamma(a+x)/\Gamma(a), \tag{2}$$

a_{kl}, b_k, and c_m are certain linear combinations of the j_{pq}'s, namely,

$$(a_{kl}) \equiv \begin{pmatrix} 1+j_{21}+j_{31}-j_{11} & 1+j_{12}+j_{13}-j_{11} & j_{21}-j_{31}-j_{11} & j_{13}-j_{12}-j_{11} \\ j_{32}-j_{12}-j_{22} & 1+j_{12}+j_{32}-j_{22} & j_{23}-j_{21}-j_{22} & -1-j_{21}-j_{22}-j_{23} \\ 1+j_{31}+j_{33}-j_{32} & j_{33}-j_{13}-j_{23} & 1+j_{23}+j_{33}-j_{13} & j_{33}-j_{31}-j_{32} \end{pmatrix}, \tag{3}$$

$$(b_k) \equiv \begin{pmatrix} -2j_{11} \\ -2j_{22} \\ 2+2j_{33} \end{pmatrix}, \tag{4}$$

$$(c_m) \equiv \begin{pmatrix} j_{33}-j_{13}-j_{21}-j_{22} \\ 1+j_{33}+j_{21}-j_{11}-j_{32} \\ 1+j_{32}+j_{13}-j_{11}-j_{22} \end{pmatrix}, \tag{5}$$

$$K \equiv (-1)^{a_{23}+a_{34}} K_1 K_2 / K_3, \tag{6}$$

$$K_1 \equiv \frac{\nabla(j_{21}j_{22}j_{23})\,\nabla(j_{31}j_{32}j_{33})\,\nabla(j_{13}j_{23}j_{33})}{\nabla(j_{11}j_{21}j_{31})\,\nabla(j_{12}j_{22}j_{32})\,\nabla(j_{11}j_{12}j_{13})}, \tag{7}$$

$$\nabla(abc) \equiv [\Gamma(2+a+b+c)\,\Gamma(1+a+b-c) \times \Gamma(1+c+a-b)/\Gamma(1+b+c-a)]^{\frac{1}{2}},$$

$$K_2 \equiv \prod_{k=1}^{3}\Gamma(a_{k4})\Gamma(a_{12})\Gamma(a_{31})\Gamma(1-b_1)\Gamma(1-b_2)\Gamma(1-c_1), \tag{8}$$

$$K_3 \equiv \Gamma(1-a_{13})\Gamma(1-a_{21})\Gamma(1-a_{23})\Gamma(1-a_{32}) \times \prod_{k=1}^{3}\Gamma(1-a_{k4})\,\Gamma(b_3)\Gamma(c_2)\Gamma(c_3). \tag{9}$$

The apparent lack of symmetry among the entries in a_{kl}, b_k, and c_m in (3)–(5) is perhaps mitigated by the summation simplicity of Eq. (1).

III. 9j COEFFICIENT AS BOUNDARY VALUE OF A NEW HYPERGEOMETRIC FUNCTION $\phi^{(3)}(\alpha_{kl};\beta_k,\gamma_m;w_k)$

Equation (1) immediately suggests that the 9j coefficients may be regarded as boundary values of a function in three variables at $w_k=1$, $k=1,2,3$, namely,

$$\{9j\} = K\phi^{(3)}(a_{kl};b_k,c_m;w_k=1) \tag{10}$$

with the a's, b's, and c's given by (3)–(5). The $\phi^{(3)}$ function is defined as follows:

Copyright © 1973 by the American Institute of Physics

$$\phi^{(3)}(\alpha_{kl};\beta_k,\gamma_m;w_k) \equiv \sum_{x_1,x_2,x_3} \frac{\prod_{k=1}^{3}\prod_{l=1}^{4}(\alpha_{kl})_{x_k}}{\prod_{k=1}^{3}(\beta_k)_{x_k}\prod_{\substack{m=1\\(m,n,k\ cyclic)}}^{3}(\gamma_m)_{x_n+x_k}} \prod_{k=1}^{3}\frac{w_k^{x_k}}{x_k!} \tag{11}$$

$\phi^{(3)}$ does not seem to be a known function. In the next section, we examine its integral representation.

IV. INTEGRAL REPRESENTATION FOR $\phi^{(3)}$

Using the identity

$$(\gamma_m)_{x_n+x_k} = (\gamma_m + x_k)_{x_n}(\gamma_m)_{x_k}, \tag{12}$$

we see that the triple sum in Eq. (11) may be viewed as a folded produce of three ${}_4F_3$ functions, namely

$$\begin{aligned}\phi^{(3)} = &\sum_{x_1}\frac{\prod_{l=1}^{4}(\alpha_{1l})_{x_1}}{(\beta_1)_{x_1}(\gamma_2)_{x_1}(\gamma_3)_{x_1}}\cdot\frac{w_1^{x_1}}{x_1!}\\ &\times\sum_{x_2}\frac{\prod_{l=1}^{4}(\alpha_{2l})_{x_2}}{(\beta_2)_{x_2}(\gamma_3+x_1)_{x_2}(\gamma_1)_{x_2}}\cdot\frac{w_2^{x_2}}{x_2!}\\ &\times\sum_{x_3}\frac{\prod_{l=1}^{4}(\alpha_{3l})_{x_3}}{(\beta_3)_{x_3}(\gamma_2+x_1)_{x_3}(\gamma_1+x_2)_{x_3}}\cdot\frac{w_3^{x_3}}{x_3!}.\end{aligned} \tag{13}$$

Equation (13) has an immediate integral representation by iterating the well-known representation for the ${}_4F_3$ function. The result is

$$\begin{aligned}\phi^{(3)} = &\frac{\prod_{i=1}^{3}\Gamma(\beta_i)\Gamma(\gamma_i)}{\prod_{i,k=1}^{3}\Gamma(\alpha_{ik})\prod_{i\geq k=1}^{3}\Gamma(\beta_{ik})}\int_0^1\cdots\int\prod_{i,k=1}^{3}dt_{ik}\\ &\times t_{ik}^{\alpha_{ik}-1}(1-t_{ik})^{\beta_{ik}-1}\prod_{k=1}^{3}(1-w_k\tau_k)^{-\alpha_{k4}},\end{aligned} \tag{14}$$

where

$$(\beta_{ik}) \equiv \begin{pmatrix}\beta_1-\alpha_{11} & \gamma_3-\alpha_{21}-\alpha_{12} & \gamma_2-\alpha_{31}-\alpha_{13}\\ \gamma_3-\alpha_{21} & \beta_2-\alpha_{22} & \gamma_1-\alpha_{32}-\alpha_{23}\\ \gamma_2-\alpha_{31} & \gamma_1-\alpha_{32} & \beta_3-\alpha_{33}\end{pmatrix}, \tag{15}$$

$$\tau_k \equiv \prod_{3\geq l>k}(1-t_{lk})\prod_{m=1}^{3}t_{km}. \tag{16}$$

V. INTEGRAL REPRESENTATION FOR THE 9j COEFFICIENTS

When the boundary values are taken according to Eq. (10), the matrix (β_{ik}) of (15) may become triangular on account of a set of unexpected identities which come about by a judicious arrangement of the elements a_{kl} as done in (3):

$$c_m - a_{ik} - a_{ki} = 0, \quad i,k,m \text{ cyclic}. \tag{17}$$

The net effect of this is to reduce from a general ninefold integral of (14) for $\phi^{(3)}$ to a sixfold integral representation for the 9j coefficient. Thus

$$\begin{aligned}\{9j\} = KK'\int_0^1\cdots\int\prod_{i\geq k=1}^{3}dt_{ik}\\ \times t_{ik}^{a_{ik}-1}(1-t_{ik})^{b_{ik}-1}\prod_{k=1}^{3}(1-\hat{\tau}_k)^{-a_{k4}},\end{aligned} \tag{18}$$

where

$$K' = K_4/K_5, \tag{19}$$

$$K_4 = \prod_{i=1}^{3}\Gamma(b_i)\Gamma(c_i),$$

$$K_5 = \prod_{i\geq k=1}^{3}\Gamma(a_{ik})\Gamma(b_{ik}),$$

$$(b_{ik}) \equiv \begin{pmatrix}b_1-a_{11} & 0 & 0\\ c_3-a_{21} & b_2-a_{22} & 0\\ c_2-a_{31} & c_1-a_{32} & b_3-a_{33}\end{pmatrix}, \tag{20}$$

$$\hat{\tau}_k \equiv \prod_{3\geq l>k}(1-t_{lk})\prod_{l\leq k}t_{kl}, \quad k=1,2,3. \tag{21}$$

VI. RADON STRUCTURE

From the integral representation (14), we see that the folded (multi-loop-like) products of integration variables appearing in (16) in general would not render the integrand of (14) to be products of linear forms even after appropriate change of variables. This is true even for the boundary values (18) as far as the nondegenerate cases are concerned. By degenerate cases we mean when any one (or more) of the sixteen parameters a_{ik}, b_{ik} $(i \geq k = 1,2,3)$ and a_{k4} vanishes. When that happens, the multiloop structure is broken, and we are back in the more familiar situation of satisfying the Gel'fand criterion.[5] In this regard, we recall an analogous situation in the Radon structure of the multiperipheral versus multiloop (nonplanar) Veneziano functions.[6]

*Work supported in part by U.S. Atomic Energy Commission.

[1]A. C. T. Wu, J. Math. Phys. **13**, 84 (1972).
[2]See, e.g., G. Racah, Phys. Rev. **62**, 438 (1942); M. E. Rose, *Multipole Fields* (Wiley, New York, 1955), Appendix B.
[3]I. M. Gel'fand, M. I. Graev, and N. Ya. Vilenkin, *Generalized Functions* (Academic, New York, 1966), Vol. V.
[4]S. J. Ališauskas and A. P. Jucys, J. Math. Phys. **12**, 594 (1971), esp. Appendix; Note Erratum, J. Math. Phys. **13**, 575 (1972).
[5]A. C. T. Wu, J. Math. Phys. **12**, 437 (1971).
[6]A. C. T. Wu, J. Math. Phys. **12**, 2035 (1971), especially the last statement in Sec. 1.

Canonical transformations applied to the free Landau electron

M. H. Boon

Battelle Advanced Studies Center, CH-1227 Carouge-Geneva, Switzerland

T. H. Seligman*

Institut für Theoretische Physik, Universität Tübingen, D-74 Tübingen, Germany
(Received 29 January 1973)

It is shown that the problem of a free eletron in a uniform magnetic field provides an interesting example for the use of recently developed canonical transformation methods in quantum mechanics. The Schrödinger equation is immediately solved by a unitary transformation corresponding to a finite linear symplectic transformation in phase space. The natural invariance and dynamical groups involve affine and not only linear symplectic transformations, and the Weyl group turns out to be an appropriate invariance group to account completely for the infinite degeneracies.

I. INTRODUCTION

There has recently been a good deal of interest in the use of canonical transformations in quantum mechanical problems; and linear canonical transformations, in particular, have been extensively investigated in this context by Moshinsky and Quesne[1] and applied by them and others[2,3] to the harmonic oscillator and Coulomb systems. We wish to show that the free Landau electron system–i.e., the quantum system of an electron in a uniform magnetic field–provides a simple and interesting illustration of the use of these methods. It turns out, in fact, that the problem can be solved immediately by means of the unitary transformation in state space that corresponds to a given finite linear canonical transformation in phase space. The relation with the usual form of solution[4] is then easily written down. When we consider the symmetry of the problem, however, we find one important difference from the systems discussed in the earlier papers.[1,2,3] Namely, the invariance and dynamical groups for the Landau system are to be identified with *inhomogeneous* symplectic groups or their subgroups, rather than the usual linear symplectic groups. These inhomogeneous groups are composed of affine transformations which combine translations with the linear operations in phase space; when the latter has $2N$ dimensions, the corresponding group is denoted by $ISp(2N)$. Some of its general properties are given in the Appendix; see also Moshinsky,[5] and Wolf and Garcia.[6] We shall be led, in particular, to $ISp(2)$ as an invariance group for the Hamiltonian, and we shall find that each eigenspace (or Landau level) carries the same infinite-dimensional irreducible unitary ray representation of $ISp(2)$. This ray represenation can be obtained fairly directly; its restriction to the linear symplectic subgroup $Sp(2)$ is the ray representation can be obtained fairly directly; its restriction restriction to the subgroup of translations can be identified with the standard representation of the Weyl group for a single pair of canonical variables. It is perhaps worth emphasizing that the ordinary representation of the Weyl group appears here as a ray representation of the group of translations in two-dimensional phase space.[7]

Various aspects of the symmetry and degeneracy of the free Landau electron have previously been discussed by Johnson and Lippmann,[8] and by Dulock and McIntosh[9] who concentrate mainly, however, on the classical system. Invariance under translations in phase space has also been discussed by Opechowski and Tam,[10] by Grossmann,[11] and by one of the present authors.[12]

II. SOLUTION BY CANONICAL TRANSFORMATION

The Hamiltonian for a single (spinless) electron in a uniform magnetic field is

$$H = \tfrac{1}{2}(\mathbf{p} + e\mathbf{A})^2,$$

where $\hbar = c = m = 1$ and $e > 0$ and where $\mathbf{A}$ is the vector potential for the field which we take in the Landau gauge:

$$\mathbf{A} \equiv (-By, 0, 0)$$

expressed in terms of Cartesian coordinates with the magnetic field of magnitude B directed along the positive z axis. The momentum $\mathbf{p}$ has components p_x, p_y, p_z with respect to these axes, and since the motion in the z direction is unaffected by the field, it can be separated off to give the Hamiltonian

$$\tilde{H} = H - \tfrac{1}{2}p_z^2 = \tfrac{1}{2}(p_x - \beta y)^2 + \tfrac{1}{2}p_y^2, \qquad \beta = eB,$$

for motion in the plane at right angles to the field. From now on, we shall be effectively concerned only with $\tilde{H}$. Its eigensolutions are the well-known infinitely degenerate Landau levels, with energy $\beta(n + \frac{1}{2})$ for the nth level $(n = 0, 1, \ldots, \infty)$.[4] Let us, however, see how the canonical transformation procedure is applied to solve the problem. We make a linear transformation from the canonical (operator) variables[13] (x, y, p_x, p_y) to (operator) variables[13] $(Q, \overline{Q}, P, \overline{P})$ where

$$Q = -(p_x - \beta y)/\beta, \qquad \overline{Q} = -(p_y - \beta x)/\beta, \qquad P = p_y, \qquad \overline{P} = p_x.$$

We can verify that

$$[Q, P] = [\overline{Q}, \overline{P}] = i, \qquad Q, P \text{ commute with } \overline{Q}, \overline{P},$$

so that the new variables are true canonical ones, or alternatively that the four-dimensional transformation matrix is symplectic.[14] In the new variables, the Hamiltonian has the explicit harmonic oscillator form

$$\tilde{H} = \tfrac{1}{2}(P^2 + \beta^2 Q^2).$$

The nth oscillator level is clearly infinitely degenerate, and its eigenspace is spanned by the functions

$$\Psi_n(Q)\,\Phi_\alpha(\overline{Q}),$$

where $\Psi_n(Q)$ is the nth Hermite function and the $\Phi_\alpha(\overline{Q})$ run

Copyright © 1973 by the American Institute of Physics

with α over any complete set of functions of the coordinate $\bar{Q}$. The eigenfunctions for all n and α span the whole state space. To obtain them in terms of x and y, we use the formula given by Moshinsky and Quesne[1] for the unitary transformations that implement the canonical transformations. In the present case we need the unitary transformation corresponding to the inverse of the canonical transformation above; this can be written

$$\psi(x,y) = \frac{\beta}{2\pi} \int\int_{-\infty}^{\infty} \exp[i\beta(Q\bar{Q} + xy - xQ - y\bar{Q}] \times \Psi(Q,\bar{Q})\,dQd\bar{Q}$$

for arbitrary $\Psi(Q,\bar{Q})$ with image $\psi(x,y)$. One possible choice of eigenfunctions is obtained by choosing for the Φ_α the free-wave functions $\exp(2\pi ik\bar{Q})$, where $-\infty < k < \infty$; substituting then $\Psi_n(Q)\exp(2\pi ik\bar{Q})$ in the integral, we easily find the corresponding expressions in the x and y variables:

$$\psi_{nk}(x,y) = \Psi_n[\beta^{1/2}(y - 2\pi k/\beta)]\exp(2\pi ikx).$$

These are the usual Landau functions for propagation in the x direction.[4,15] An alternative choice for the Φ_α is the Hermite functions $\Psi_l (l = 0, 1, \ldots, \infty)$; physically, these can be viewed as the eigenfunctions of the operator $r_0^2 \equiv (\bar{P}^2 + \beta^2\bar{Q}^2)/\beta^2$ which is, as can be seen by writing it in the original variables, the square of the radius vector for the orbit centre.[8] The $\Psi_n(Q)\Psi_l(\bar{Q})$ are then simultaneous eigenfunctions of $\tilde{H}$ and r_0^2, and the corresponding integral, which is a little more difficult to accomplish, provides us with the functions $\psi_{nl}(x,y)$ given by Johnson and Lippmann.[15]

The canonical transformation above is evidently not the only one that sends $\tilde{H}$ into harmonic oscillator form; it can be combined with any other linear canonical transformation under which $\tilde{H}$ is invariant. In this way Landau functions for propagation in any given direction of the x–y plane can be obtained; each such direction provides us with a complete set of eigenfunctions.

III. INVARIANCE AND DYNAMICAL GROUPS

We shall continue to consider the two-dimensional problem described by $\tilde{H}$, and our starting point in this context will be the group $ISp(4)$ of all real affine canonical transformations in four-dimensional phase space. What transformations of $ISp(4)$ leave $\tilde{H} = \frac{1}{2}(P^2 + \beta^2Q^2)$ invariant? Evidently, since neither $\bar{P}$ nor $\bar{Q}$ appear in $\tilde{H}$, it is left unchanged by the whole inhomogeneous symplectic group $\overline{ISp}(2)$ of transformations involving only $\bar{P}$ and $\bar{Q}$. Indeed, if $O_\beta(2)$ is the group of linear symplectic transformations in P and Q that leave the form $\tilde{H}$ invariant, the maximal invariance group of transformations in $ISp(4)$ is the direct product

$$O_\beta(2) \times \overline{ISp}(2).$$

Here, as later, the symbols for groups of transformations in just one pair of canonical variables are barred for $\bar{P}$ and $\bar{Q}$ and unbarred for P and Q. The group $\overline{ISp}(2)$ can be expressed as a semidirect product

$$\overline{ISp}(2) = \bar{T}(2) \wedge \overline{Sp}(2)$$

of the normal subgroup $\bar{T}(2)$ of translations in $\bar{P}$ and $\bar{Q}$ and the subgroup $\overline{Sp}(2)$ of linear symplectic transformations in $\bar{P}$ and $\bar{Q}$.

Let us ignore the factor $O_\beta(2)$ for the moment, and take $\overline{ISp}(2)$ as the invariance group for $\tilde{H}$. Let $D(\overline{ISp}(2))$ denote its unitary representation on some eigenspace of $\tilde{H}$. We know from Sec. II that each eigenspace is effectively the full space of functions of the coordinate $\bar{Q}$. A set of generators for the representation are the operators

$$\bar{P}^2, \bar{Q}^2, \bar{P}\bar{Q} + \bar{Q}\bar{P}, \bar{P}, \bar{Q},$$

where the first three are generators for the restriction $D(\overline{Sp}(2))$ of $D(\overline{ISp}(2))$ to $\overline{Sp}(2)$, and the last two are generators for the restriction $D(\bar{T}(2))$ to $\bar{T}(2)$. Since $[\bar{Q},\bar{P}] = i$ we are in fact dealing with a ray representation of $\bar{T}(2)$,[7] and hence of $\overline{ISp}(2)$. Furthermore, the ray representation $D(\bar{T}(2))$ can be immediately reinterpreted as an ordinary (unitary) representation of the Weyl group $\bar{W}(2)$ for the single canonical pair $\bar{P},\bar{Q}$. This is the standard quantum-mechanical representation of $\bar{W}(2)$, and is well known to be irreducible and unique.[16] In other words $D(\bar{T}(2))$ is an irreducible unitary ray representation of $\bar{T}(2)$, unique for the given factor system. The ray representation $D(\overline{Sp}(2))$ is a direct multiple of a single irreducible ray representation of $\overline{Sp}(2)$; this representation is known,[1] and the multiple is two. Altogether, therefore, we have reached the result that $D(\overline{ISp}(2))$ is an irreducible ray representation, and the same for all eigenspaces. Furthermore, we observe that since the restriction $D(\bar{T}(2))$ remains irreducible, the translation group $\bar{T}(2)$, or equivalently the Weyl group $\bar{W}(2)$, is itself an adequate invariance group for the system and accounts completely for the degeneracy of the energy levels.

Including $O_\beta(2)$ (which is Abelian) again in the invariance group has only trivial consequences, since its generator in the state space is $\tilde{H}$ itself, and we need effectively retain only $\overline{ISp}(2)$. It is of interest to note, however, that the cylindrical symmetry around the magnetic field direction is included only in the larger invariance group.

Passing on to the question of dynamical groups, we now seek the largest subgroup of $ISp(4)$ that has $\overline{ISp}(2)$ as a direct factor: this gives us the beginning of a chain of groups

$$ISp(4) \supset ISp(2) \times \overline{ISp}(2).$$

As we know $Sp(2)$ to be a dynamical group of the one-dimensional oscillator,[1] we may complete the chain by

$$ISp(2) \supset Sp(2) \supset O_\beta(2),$$
$$\overline{ISp}(2) \supset \bar{T}(2).$$

One appropriate choice of dynamical group for our system would therefore be

$$Sp(2) \times \bar{T}(2).$$

The factor $\bar{T}(2)$ accounts fully for the degeneracies through its ray representation $\bar{W}(2)$ as we saw above; and $Sp(2)$ combines the half of the spectrum for even oscillator quantum number in one irreducible ray representation, and the half for odd oscillator quantum number in another equivalent one.[1,5] It follows, again, from our earlier discussion that this spectrum splitting would disappear if we replaced $Sp(2)$ by $ISp(2)$—or by its subgroup $T(2) \wedge O_\beta(2)$ which has generators $2\tilde{H} = P^2 + \beta^2Q^2, P, Q$.

Another choice of dynamical group is evidently $ISp(4)$ itself, for which the entire spectrum would occur in the single irreducible representation carried by the whole state space; this group contains in addition the trans-

formations used in Sec. II to solve the problem. There are, of course, other possibilities; in particular we may ask whether an appropriate choice can be made with semisimple groups only. For this purpose, the chain of groups would be

$$Sp(4) \supset Sp(2) \times \overline{Sp}(2) \supset O_\beta(2) \times \overline{Sp}(2),$$

with the last member as maximal semisimple invariance group. The consequences of taking one or other of the first two members as dynamical group follow largely from the discussion of the last paragraph. In short, for $Sp(4)$ the spectrum would be combined into two parts, and for $Sp(2) \times \overline{Sp}(2)$ into four.[1,5] The disadvantage here, however, is that in both cases degenerate states are associated with different parts, in such a way that there is always a twofold accidental degeneracy. This is unavoidable unless we allow non-semisimple groups and it will disappear only if translations, or equivalently the Weyl group, are introduced.

We finish the section with some remarks on the relationship between the various choices of invariance group and basis functions for eigenspaces. The Landau functions written down in Sec. II, and which span the nth eigenspace, evidently provide a basis for $D(\overline{T}(2))$, i.e., for the standard irreducible representation of the Weyl group $\overline{W}(2)$. They are eigenfunctions of $\overline{P}$, the generator of a noncompact one-parameter subgroup of $\overline{W}(2)$, and are uniquely identified by the corresponding wavenumber label $k(-\infty < k < \infty)$ since there are no degeneracies with respect to this subgroup. On the other hand in the semisimple case we have $\overline{Sp}(2)$ instead of $\overline{W}(2)$, and we have the noncompact one-parameter subgroup generated by $\overline{P}^2$ instead of that generated by $\overline{P}$. The eigenfunctions of $\overline{P}^2$ are still the Landau functions, but there is now a twofold accidental degeneracy between the functions with wavenumbers $\pm k$. This degeneracy is in accordance with the splitting of $D(\overline{Sp}(2))$ into two irreducible parts of opposite parity. The reduction of irreducible (ray) representations of $Sp(2)$—whose Lie algebra is also that of $O(2,1)$ and $SU(1,1)$—with respect to non-compact one-parameter subgroups is examined by Barut and Phillips.[17] In the case of a one-parameter subgroup that is compact, the corresponding basis will consist of localizable (i.e., square-integrable) functions. For example, the group generated by $r_0^2 = (\overline{P}^2 + \beta^2\overline{Q}^2)/\beta^2$, which belongs to the Lie algebra of $\overline{Sp}(2)$, yields the basis of functions $\psi_{nl}(x,y)$ mentioned at the end of Sec. II.

IV. CONCLUDING SUMMARY

The application of canonical transformations to the quantum mechanical problem of the free Landau electron has proved interesting in a number of ways. It provides a method of solving the Schrödinger equation with a linear symplectic transformation in $Sp(4)$, where the eigenfunctions are obtained by the use of the unitary operator for this transformation as given by Moshinsky and Quesne.[1] Furthermore, we could show that to explain the degeneracy completely it was necessary to have an invariance group involving affine transformations in phase space. In fact the smallest invariance group adequate in this respect turned out to be the group of translations in phase space, whose ray representation in the state space is just the standard irreducible representation of the Weyl group. On the other hand, if we retain only semisimple groups, a twofold accidental degeneracy appears which cannot be removed by any transformation in $Sp(4)$; it can only be lifted by reintroducing the translations and consequently the Weyl group.

ACKNOWLEDGMENT

One of us (T.H.S.) is grateful for the kind hospitality and support he received from the Battelle Institute.

APPENDIX

Much of the discussion of inhomogeneous symplectic groups in the text applies generally to the group $ISp(2N)$ of real affine canonical transformations in $2N$-dimensional phase space. The group can be expressed as a semidirect product

$$ISp(2N) = T(2N) \wedge Sp(2N)$$

of its normal subgroup of translations $T(2N)$ and its subgroup of linear canonical transformations $Sp(2N)$. Its Lie algebra is $N(2N+3)$-dimensional with generators

$$P_iP_j, Q_iQ_j, Q_iP_j + Q_jP_i, Q_iP_j - Q_jP_i \ (i \neq j), P_i, Q_i$$

where the Q_i and P_i $(i = 1 \cdots N)$ are canonical coordinates and momenta. The unitary ray representation $D(ISp(2N))$ of $ISp(2N)$ in the space of states is generated by taking the Q_i as multiplicative operators and $P_i = -i\partial/\partial Q_i$. The restriction of $D(ISp(2N))$ to $Sp(2N)$, generated by the $N(2N+1)$ bilinear operators, is the ray representation discussed by Moshinsky and Quesne and decomposes into two equivalent irreducible parts.[1] The restriction of $D(ISp(2N))$ to $T(2N)$ is generated by the P_i and Q_i where $[Q_i, P_j] = i\delta_{ij}$, and can therefore be reinterpreted as the (unique) irreducible unitary representation of the Weyl group[16] $W(2N)$ for N pairs of canonical variables. This implies, in particular, that $D(ISp(2N))$ is irreducible.

An arbitrary member of $ISp(2N)$ can be written

$$\begin{pmatrix}\mathbf{Q}'\\ \mathbf{P}'\end{pmatrix} = \begin{pmatrix}A & B\\ C & D\end{pmatrix}\begin{pmatrix}\mathbf{Q}\\ \mathbf{P}\end{pmatrix} + \begin{pmatrix}\mathbf{u}\\ \mathbf{v}\end{pmatrix},$$

where the bold face quantities $\mathbf{Q}$, etc., are real N-dimensional column vectors with components Q_i, $i = 1, \ldots, N$, etc., respectively and the entries in the matrix are real $N \times N$ submatrices. For the linear part, we can take any real symplectic matrix and for the translation part, any real vectors $\mathbf{u}, \mathbf{v}$. The unitary operator on the state space corresponding to the above transformation can be conveniently expressed as an integral operator

$$\Psi(Q') = \int K(Q', Q)\psi(Q)\,dQ$$

transforming an arbitrary wavefunction $\psi(Q)$ of the coordinates $Q \equiv \{Q_1, Q_2 \cdots Q_N\}$ into its image $\Psi(Q')$ with coordinates $Q' \equiv \{Q_1', Q_2' \cdots Q_N'\}$. Here $dQ = \prod_{i=1}^{N} dQ_i$ and each Q_i is integrated from $-\infty$ to $+\infty$. The kernel satisfies the unitarity condition

$$\int K^*(Q', Q'')K(Q', Q)\,dQ' = \delta(Q'' - Q),$$

where the right-hand side is a product of δ functions for each i and where the star means complex conjugate. The kernel is determined to within a constant phase which remains arbitrary because $D(ISp(2N))$ is a ray representation. There are different expressions for $K(Q', Q)$ according as the determinant of B is zero or

not. Omitting the details of the calculation, we find:

(i) For $\det B \neq 0$,

$$K(Q', Q) = (2\pi)^{-N/2} \, |\det B|^{-1/2} \exp[\tfrac{1}{2} i S(Q', Q)],$$

$$S(Q', Q) = (\mathbf{Q}, B^{-1}A\mathbf{Q}) + (\mathbf{Q}', DB^{-1}\mathbf{Q}') - 2(\mathbf{Q}, B^{-1}\mathbf{Q}') + 2(B^{-1}\mathbf{u}, \mathbf{Q}) - 2(DB^{-1}\mathbf{u} - \mathbf{v}, \mathbf{Q}').$$

The terms in $S(Q', Q)$ are scalar products of row vectors with column vectors.

(ii) For $\det B = 0$, whence $\det D \neq 0$,

$$K(Q', Q) = |\det D|^{-1/2} \, R(Q', Q) \times \exp\{i[\tfrac{1}{2}(\mathbf{Q}, D^{-1}C\mathbf{Q}) + (D^{-1}\mathbf{v}, \mathbf{Q})]\},$$

$$R(Q', Q) = \prod_{i=1}^{p} \delta(Y_i) \prod_{i=p+1}^{N} |\Lambda_i|^{-1} \times \exp\{i[-\tfrac{1}{4}\pi \operatorname{sgn}\Lambda_i + \tfrac{1}{2} Y_i^2 \Lambda_i^{-1}]\},$$

where

$$\mathbf{Y} = O(\mathbf{Q}' - (D^T)^{-1}\mathbf{Q} + BD^{-1}\mathbf{v} - \mathbf{u}), \qquad O^T BD^{-1} O = \Lambda$$

and O is a real orthogonal matrix with transpose O^T and Λ a diagonal matrix with diagonal elements Λ_i ($i = 1, 2 \cdots N$) of which the first p ($1 \leq p \leq N$) are zero and the remaining ones nonzero (note that BD^{-1} is always real symmetric). Also, $\operatorname{sgn}\Lambda_i \equiv \Lambda_i / |\Lambda_i|$ for $i > p$.

We should remark that the unitary operators for $ISp(2)$ have previously been calculated in terms of a Hermite function basis.[5]

When $\mathbf{u} = \mathbf{v} = \mathbf{0}$, the above expressions reduce to those given by Moshinsky and Quesne[1] for the ray representation of $Sp(2N)$. On the other hand, when the matrix is unity (pure translation) so that $B = C = 0, A = D = I$ (the unit matrix), we have

$$K(Q', Q) = \exp[i(\mathbf{v}, \mathbf{Q})] \, \delta(\mathbf{Q}' - \mathbf{Q} - \mathbf{u}).$$

It is easy to see that the unitary transformation with this kernel can be expressed as an operator

$$\exp[-i(\mathbf{u}, \mathbf{P})] \exp[i(\mathbf{v}, \mathbf{Q})]$$

(the components of $\mathbf{P}$ and $\mathbf{Q}$ being now operators). We thus obtain the usual representation of the Weyl group.

---

*Fellow of the Deutsche Forschungsgemeinschaft.

[1] M. Moshinsky and C. Quesne, J. Math. Phys. **12**, 1772 (1971).
[2] C. Quesne and M. Moshinsky, J. Math. Phys. **12**, 1780 (1971).
[3] M. Moshinsky, T. H. Seligman, and K. B. Wolf, J. Math. Phys. **13**, 901 (1971).
[4] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics* (Pergamon, London, 1959), Sec. 125.
[5] M. Moshinsky and C. Quesne, *Proceedings of XV Solvay Conference in Physics*, Brussels, 1970 (to be published).
[6] K. B. Wolf and A. Garcia, Comunicaciones Tecnicas C.I.M.A.S.S., U.N.A.M., Mexico D.F., Vol. 3, No. 9 (1972).
[7] On ray representations see, e.g., M. Hamermesh, *Group Theory* (Addison-Wesley, Reading, Mass., 1962), Chap. 12.
[8] M. H. Johnson and B. A. Lippmann, Phys. Rev. **76**, 828 (1949).
[9] V. A. Dulock and H. V. McIntosh, J. Math. Phys. **7**, 1401 (1966).
[10] W. Opechowski and W. G. Tam, Physica (Utr.) **42**, 525 (1969).
[11] A. Grossmann, C.N.R.S. Report No. 71/P. 409 (Marseille, 1971).
[12] M. H. Boon, J. Math. Phys. **13**, 1268 (1972).

[13] These variables are always assumed to be operators, except where coordinates x, y or $Q, \dot{Q}$ appear as arguments of functions. The context should always be clear.

[14] It then satisfies relations (2.1) of Ref. 1.
[15] With the z dependence omitted.
[16] J. von Neumann, Math. Ann. **104**, 570 (1931).
[17] A. O. Barut and E. C. Phillips, Commun. Math. Phys. **8**, 52 (1968).

Ergodic properties of simple model system with collisions*

Sheldon Goldstein†

Belfer Graduate School of Science, Yeshiva University, New York, New York

Oscar E. Lanford III‡

Department of Mathematics, University of California, Berkeley, California

Joel L. Lebowitz

Belfer Graduate School of Science, Yeshiva University, New York, New York

(Received 14 March 1973)

We investigate the ergodic properties of the discrete time evolution of a particle in a two-dimensional torus with velocity in the unit square. The dynamics consists of free motion for a unit time interval followed by a baker's transformation of the velocity.

1. INTRODUCTION

We are interested in the ergodic properties of dilute gas systems. These may be thought of as Hamiltonian dynamical systems in which the particles move freely except during binary "collisions". In a collision the velocities of the colliding particles undergo a transformation with "good" mixing properties (cf. Sinai's study of the billiard problem[1]). To gain an understanding of such systems we have studied the following simple discrete time model: The system consists of a single particle with coordinate $\underline{r} = (x, y)$ in a two-dimensional torus with sides of length (L_x, L_y), and "velocity" $\underline{v} = (v_x, v_y)$, in the unit square $v_x \in [0, 1], v_y \in [0, 1]$. The phase space Γ is thus a direct product of the torus and the unit square. The transformation T which takes the system from a dynamical state $(\underline{r}, \underline{v})$ at "time" j to a new dynamical state $T(\underline{r}, \underline{v})$ at time $j + 1$ may be pictured as resulting from the particle moving freely during the unit time interval between j and $j + 1$ and then undergoing a "collision" in which its velocity changes according to the baker's transformation, i.e.,

$$T(\underline{r}, \underline{v}) = (\underline{r} + \underline{v}, B\underline{v}), \tag{1.1}$$

with

$$B(v_x, v_y) = \begin{cases} (2v_x, \tfrac{1}{2}v_y), & 0 \le v_x < \tfrac{1}{2} \\ (2v_x - 1, \tfrac{1}{2}v_y + \tfrac{1}{2}), & \tfrac{1}{2} < v_x \le 1. \end{cases} \tag{1.2}$$

The normalized Lebesgue measure $d\mu = dxdydv_xdv_y/L_xL_y = d\underline{r}\,d\underline{v}/L_xL_y$ in Γ is left invariant by T. We call U_T the unitary transformation induced by T on $L^2(d\mu)$, $U_T\varphi = \varphi \circ T$. Our interest lies then in the ergodic properties of T and in the spectrum of U_T.

We note first that the transformation B on the velocities is, when taken by itself as a transformation of the unit square with measure $d\underline{v}$, well known to be isomorphic to a Bernoulli shift. It therefore has very good mixing properties. The isomorphism is obtained by setting

$$v_x = \sum_{j=1}^{\infty} 2^{-j}u_j, \qquad v_y = \sum_{j=1}^{\infty} 2^{-j}u_{1-j}, \tag{1.3}$$

with the u_j independent random variables taking the values 0 and 1 each with probability $\frac{1}{2}$. We then have

$$\begin{aligned} (B\underline{v})_x &= \sum_{j=1}^{\infty} 2^{-j}u_{j+1} = 2v_x - u_1, \\ (B\underline{v})_y &= \sum_{j=1}^{\infty} 2^{-j}u_{2-j} = \tfrac{1}{2}v_y + \tfrac{1}{2}u_1. \end{aligned} \tag{1.4}$$

2. ERGODIC PROPERTIES

The ergodic properties of our system which combines B with free motion turn out to depend on whether L_x^{-1} and L_y^{-1} satisfy the independence condition (I),

$$n_xL_x^{-1} + n_yL_y^{-1} \notin Z \text{ for } n_x \text{ and } n_y \text{ integers unless } n_x = n_y = 0. \tag{I}$$

Theorem 1: When (I) holds, the spectrum of U_T, on the complement of the one-dimensional subspace generated by the constants, is absolutely continuous with respect to Lebesgue measure and has infinite multiplicity.

It follows from Theorem 1 that when (I) holds the dynamical system (Γ, T, μ) is at least mixing. We do not know at present whether it is also a Bernoulli shift or at least a K system.

Theorem 2: When (I) does not hold the system (Γ, T, μ) is not ergodic.

The proof of Theorem 1 has two parts: a general characterization of unitary operators with Lebesgue spectrum and a set of estimates.

Lemma: Let U by a unitary operatory on a Hilbert space h, with spectral representation $U = \int_0^{2\pi} e^{i\theta}\underline{P}(d\theta)$. Assume that there exists a total set of vectors $\{\varphi_i\}$ such that $\sum_{n=1}^{\infty} |(U^n\varphi_i | \varphi_i)| < \infty$ for all i. (A set of vectors is said to be *total* if the finite linear span of this set of vectors is dense.) Then the spectral measure $\underline{P}(d\theta)$ is absolutely continuous with respect to Lebesgue measure, i.e., if E is a Borel set of Lebesgue measure 0, then $\underline{P}(E) = 0$.

Proof: We have

$(U^n\varphi_i | \varphi_i) = \int e^{in\theta}(\underline{P}(d\theta)\varphi_i | \varphi_i)$, i.e., the function $n \to (U^n\varphi_i | \varphi_i)$ is the Fourier transform of the measure $(\underline{P}(d\theta)\varphi_i | \varphi_i)$. On the other hand, $\sum_n |(U^n\varphi_i | \varphi_i)| < \infty$, so we can compute its inverse Fourier transform in the elementary way. By the uniqueness of the Fourier transform, we get:

$$(\underline{P}(d\theta)\varphi_i | \varphi_i) = \frac{d\theta}{2\pi} \cdot \sum_{n=-\infty}^{\infty} e^{-in\theta}(U^n\varphi_i | \varphi_i),$$

so the numerical measure $(\underline{P}(d\theta)\varphi_i | \varphi_i)$ is absolutely continuous with respect to Lebesgue measure. If E is a Borel setof Lebesgue measure 0,

$$\|\underline{P}(E)\varphi_i\|^2 = (\underline{P}(E)\varphi_i | \varphi_i) = 0, \quad \text{so } \underline{P}(E)\varphi_i = 0 \text{ for all } \varphi_i.$$

Copyright © 1973 by the American Institute of Physics

But the vectors $\{\varphi_i\}$ form a total set, so $\underline{P}(E) = 0$ as desired.

Now the estimates: Let $\chi(1) = 1$, $\chi(0) = -1$. For each finite subset X of Z, we define

$$\chi_X(\underline{v}) = \prod_{j\in X} \chi(u_j).$$

The χ_X form an orthonormal basis for $L^2(d\underline{v})$. Similarly, the functions $\exp(ik\cdot r)$; $\{k = (k_x, k_y), k_x = 2\pi n_x L_x, k_y = 2\pi n_y L_y, n_x$ and n_y integers$\}$, form an orthonormal basis for $L^2(d\underline{r})$. Thus, the functions $\varphi_{X,\underline{k}} = \exp(i\underline{k}\cdot\underline{r})\cdot\chi_X(\underline{v})$ form an orthonormal basis for $L^2(d\mu)$. We will prove that

$$\sum_{n=1}^{\infty} |(U_T^n \varphi_{X_1,\underline{k}_1} | \varphi_{X_2,\underline{k}_2})| < \infty \quad \text{unless } \underline{k}_1 = \underline{k}_2 = 0, \quad X_1 = X_2 = 0.$$

By straightforward computation,

$$U_T^n \varphi_{X_1,\underline{k}_1} = \varphi_{X_1+n}(\underline{v}) \exp(i\underline{k}\cdot\underline{r}) \times \exp[i\underline{k}\cdot(\underline{v} + B\underline{v} + \cdots + B^{n-1}\underline{v})].$$

Thus

$$\int d\underline{r}(U_T^n \varphi_{X_1,\underline{k}_1})\bar{\varphi}_{X_2,\underline{k}_2} = 0 \quad \text{unless } \underline{k}_1 = \underline{k}_2 (= \underline{k}),$$

so we assume $\underline{k}_1 = \underline{k}_2 = \underline{k}$. Also,

$$\int d\underline{v}(U_T^n \varphi_{X_1,0})\bar{\varphi}_{X_2,0} = 0 \quad \text{unless } X_2 = X_1 + n,$$

so the result is trivally true for $\underline{k} = 0$. We therefore assume $\underline{k} \neq 0$.

Now

$$(L_x L_y)^{-1} \int d\underline{r}d\underline{v}(U_T^n \varphi_{X_1,\underline{k}})\bar{\varphi}_{X_2,\underline{k}} = \int d\underline{v}\chi_{X_1}(B^n\underline{v})\chi_{X_2}(\underline{v}) \exp[i\underline{k}(\underline{v} + B\underline{v} + \cdots + B^{n-1}\underline{v})],$$

$$(B^j\underline{v})_x = \sum_{i=1}^{\infty} u_{j+i} 2^{-i},$$

$$\sum_{j=0}^{n-1} (B^j\underline{v})_x = \sum_{j=0}^{n-1} \sum_{i=1}^{\infty} u_{j+i} 2^{-i} = \sum_{l=1}^{\infty} u_l \sum_{i=1\vee(l-n+1)}^{l} 2^{-i} = \sum_{l=1}^{\infty} u_l \alpha_l^n$$

(where this equation defines α_l^n),

$$(B^j\underline{v})_y = \sum_{i=1}^{\infty} 2^{-i} u_{j+1-i},$$

$$\sum_{j=0}^{n-1} (B^j\underline{v})_y = \sum_{j=0}^{n-1} \sum_{i=1}^{\infty} 2^{-i} u_{j+1-i} = \sum_{l=-\infty}^{n-1} u_l \sum_{i=1\vee(-l+1)}^{n-l} 2^{-i} = \sum_{l=-\infty}^{\infty} u_l \beta_l^n.$$

Now let $l_2 = 1 \vee \max\{X_2\}$, $l_1 = \inf\{X_1\} \wedge 0$.

Then

$$U_T^n \varphi_{X_1,\underline{k}} \cdot \varphi_{X_2,\underline{k}} = \prod_{l=l_2+1}^{n+l_1-1} \exp[i(\alpha_l^n k_x + \beta_l^n k_y)u_l] \times [fn \text{ of the } u_l\text{'s for } l \notin (l_2, n + l_1)].$$

By independence, the integral of the product on the right is the product of the integrals, and the unspecified function of the u_l's, $l \notin (l_2, n + l_1)$ is no greater than one in absolute value, so

$$(L_x L_y)^{-1} |\int d\underline{v}d\underline{r}\, U_T^n \varphi_{X_1,\underline{k}} \cdot \varphi_{X_2,\underline{k}}| \leq \prod_{l=l_2+1}^{n+l_1-1} |\tfrac{1}{2}[\exp(i\alpha_l^n k_x + \beta_l^n k_y) + 1]|.$$

For l's within the limits of the product, we have

$$\alpha_l^n = \sum_{i=1}^{l} 2^{-i} = 1 - 2^{-l},$$

$$\beta_l^n = \sum_{i=1}^{n-1} 2^{-i} = 1 - 2^{-(n-l)}.$$

Thus, for most of the terms in the product, $\alpha_l^n \approx \beta_l^n \approx 1$, and the number of terms is $n -$ const for large n. In particular, if we put

$$\gamma = \tfrac{1}{2}|\exp[i(k_x + k_y)] + 1| < 1$$

(by our fundamental assumption),

$$|(U_T^n \varphi_{X_1,\underline{k}} | \varphi_{X_2,\underline{k}})| < \gamma^{n/2} \quad \text{for all sufficiently large } n,$$

we have

$$\sum_{n=1}^{\infty} |(U_T^n \varphi_{X_1,\underline{k}} | \varphi_{X_2,\underline{k}})| < \infty$$

as desired.

The fact that the multiplicity is infinite is trivial. We have $L^2(d\underline{v}) \subset L^2(d\underline{r}d\underline{v})$, and we already know that the spectrum of U_T restricted to $L^2(d\underline{v})$ has infinite multiplicity.

To obtain a proof of Theorem 2, we note that ergodicity is equivalent to

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int d\mu\, (U_T^n \varphi)\bar{\Psi} = (\int d\mu\, \varphi)(\int d\mu\, \bar{\Psi}), \qquad \varphi, \Psi \in L^2(d\mu).$$

For φ or Ψ orthogonal to the constants we must then have Cesaro convergence to zero when the system is ergodic. We prove that the system is nonergodic by finding φ or Ψ orthogonal to the constants such that the above integral converges (strictly) to a nonzero number.

Let n_x, n_y be such that $n_x/L_x + n_y/L_y \in Z$ and n_x and n_y are not both 0, and let $k_x = 2\pi n_x/L_x$, $k_y = 2\pi n_y/L_y$. We set $\varphi = \Psi = \varphi_{0,\underline{k}}$ and compute as before the relevant integrals:

$$I_n = \int d\mu\, (U_T^n \varphi_{0,\underline{k}})\bar{\varphi}_{0,\underline{k}} = \int d\underline{v} \exp\left[i\underline{k}\cdot\left(\sum_{j=0}^{n-1} B^j\underline{v}\right)\right] = \int d\underline{v} \prod_{l=\infty}^{\infty} \exp[i(k_x\alpha_l^n + k_y\beta_l^n)u_l] = \prod_{l=-\infty}^{\infty} \tfrac{1}{2}[1 + \exp(i\alpha_l^n k_x + \beta_l^n k_y)].$$

Here

$$\alpha_l^n = \sum_{i=1\vee(l-n+1)}^{l} 2^{-i} = 2^{-l} \sum_{m=0}^{(n-1)\wedge(l-1)} 2^m = 2^{-l}(2^{n\wedge l} - 1)$$

for $l > 0$ and vanishes for $l \leq 0$, and

$$\beta_l^n = \sum_{i=1\vee(-l+1)}^{n-l} 2^{-i} = 2^{l-1} \sum_{m=0\vee l}^{n-1} 2^{-m} = 2^{0\wedge l} - 2^{l-n}$$

for $l < n$ and vanishes for $l \geq n$.

We thus have found that

$$I_n = \prod_{l=-\infty}^{0} \tfrac{1}{2}\{1 + \exp[i(2^l - 2^{l-n})k_y]\}$$
$$\times \prod_{l=1}^{n-1} \tfrac{1}{2}(1 + \exp\{i[(1 - 2^{-l})k_x + (1 - 2^{-(n-l)})k_y]\})$$
$$\times \prod_{l-n}^{\infty} \tfrac{1}{2}\{1 + \exp[ik_x(2^{-(l-n)} - 2^{-l})]\}$$
$$= F_n^1(\underline{k})F_n^2(\underline{k})F_n^3(\underline{k})$$

with

$$F_n^1(\underline{k}) = F_n^1(k_y) = \prod_{m=0}^{\infty} \tfrac{1}{2}\{1 + \exp[ik_y(2^{-m} - 2^{-(m+n)})]\},$$
$$F_n^3(\underline{k}) = F_n^3(k_x) = F_n^1(k_x),$$
$$F_n^2(\underline{k}) = \prod_{l=1}^{n-1} \tfrac{1}{2}(1 + \exp\{i[(1 - 2^{-l})k_x + (1 - 2^{-(n-l)})k_y]\}).$$

Since $k_x + k_y \in 2\pi Z$, we have

$$F_n^2(\underline{k}) = \prod_{l=1}^{n-1} \tfrac{1}{2}\{1 + \exp[-i(k_x 2^{-l} + k_y 2^{-(n-l)})]\}.$$

We now assert that (for $k_x + k_y \in \pi Z$)

$$\lim_{n\to\infty} F_n^i(\underline{k}) = \alpha^i \neq 0, \qquad i = 1, 2, 3.$$

This is verified by observing that the log $F_n^i(\underline{k})$ coverage to a finite limit, thus completing the proof.

(If k_x and k_y are such that some of the terms at the beginning of the series which one obtains from the log $F_n^i(\underline{k})$ are singular, one easily removes the difficulty by an appropriate change in the functions φ and Ψ introduced at the beginning of the proof of Theorem 2. We also note that for the case where L_x/L_y is rational we can find explicitly a nonconstant function f which is left invariant by U_T. From the fact that $U_B(v_x + 2v_y) = 2v_x + v_y$ it follows that $f(x - y - v_x - 2v_y)$ is invariant if f is doubly periodic with periods L_x and L_y, so that we can construct an infinite family of orthonormal invariant functions $f_n : f_n = \exp\{(i2\pi n/L)(x - y - v_x - 2v_y)\}$ with $L_x/r = L_y/s = L$, r and s integers.)

---

*Supported in part by USAFOSR Grant No. 73-2430.

†National Science Foundation Fellow.

‡Alfred P. Sloan Foundation Fellow; also supported in part by NSF Grant GP-15735.

[1] Ya. Sinai, Russ. Math. Surv. 25, 137 (1970).

[2] Giovanni Gallavotti, *Modern Theory of the Billiard. An Introduction* (to appear).

Construction of reproducing kernels for analytic Hilbert spaces

R. E. Cutkosky

Carnegie-Mellon University, Pittsburgh, Pennsylvania
(Received 3 January 1973)

The reproducing kernels of suitably chosen Hilbert spaces provide a simple but powerful way to generate approximations to functions on which linear data is given. These data may consist either of exact values (for the problem of interpolation) or of values subject to experimental error. For applications, it is important to be able to construct reproducing kernels which embody the information which is available from general considerations about the physics of the problem. Explicit methods are given here for construction of reproducing kernels for Hilbert spaces of analytic functions, these functions being characterized by various types of behavior on the boundary of their domain of holomorphy.

Methods for exploiting to best effect general analytic information about functions to be fit to experimental data have recently undergone considerable development. A description of some of these methods, and further references, may be found in some recent papers by Ciulli and Nenciu.[1] Another class of techniques of especially great generality is based on the introduction of a Bayesian probability in a function space.[2] This function space is to be restricted in some way to functions which possess properties which are reasonable from the point of view of the physics of the problem, but the space is still to be large enough to allow sufficient flexibility for fitting data without undue bias. General discussions of this method as well as applications of some of its features are contained in Refs. 2-16.

The general problem of fitting data by analytic functions has been of especial interest in particle physics, where the information about domains of regularity comes from dispersion relations and their various generalizations. In many particle physics applications, information about boundary values on distant branch cuts is also obtained from general theoretical considerations applied to related data. It is hoped, however, that many features of these new approaches to data analysis will prove to be useful in other branches of physics. Even analyticity is not really a requirement for some of the ideas considered here; it is sufficient that the function space admit a reproducing kernel.

The primary way in which the work described in Refs. 2–16 differs from that of Ciulli and his co-workers[1] is through the use of Bayesian analysis. Bayesian techniques are usually considered to be very controversial by physicists. There is often good reason for being suspicious of too naive an introduction of prior probabilities. Properly considered, however, our use of prior probabilities is merely a device for representing quantitatively, in a logically self-consistent manner, subjective opinions about the functions under consideration. Since subjective opinions can never be entirely eliminated, it is only an advantage to use a formalism which requires such opinions to be more explicit and thus more open to correction.

In the original paper the natural connection between prior distributions and spaces with reproducing kernels was not fully recognized[2]; this has become clear in the course of subsequent work[3-16] as well as through discussions among the authors cited. A secondary purpose of this paper is thus to introduce the reader to some of the mathematical ideas which have been found to be useful. The primary purpose of the paper is to provide some methods and formulas for coping with a specific technical problem often encountered in this probabilistic method of data analysis. This is the problem of how to characterize, through their prior distributions, spaces of analytic functions which possess suitable behavior on the boundary of a domain of holomorphy.[14,16]

Brief summaries are given of the mathematical topics which are used later but which also are of interest in their own right. These include, first, some properties of spaces with reproducing kernels, and second, the definition of a probability measure in these spaces. Most of this discussion is adapted from the mathematical literature, but parts of Sec. 2 are new or are a refinement of material introduced in Ref. 2. In the main part of the paper, a specific class of "hypergeometric" spaces will first be introduced. Then some further generalizations will be introduced which allow the incorporation of a great variety of boundary behaviors. Finally, some suggestions will be given about generalizations to functions of several complex variables as well as about other unsolved problems.

1. REPRODUCING KERNELS

Reproducing kernels have been given a general definition by Aronszajn[17]; a survey containing many additional results is given by Meschkowski[18] (this is cited as M hereafter). Results of special interest to the interpolation problem have been described also by Davis.[19] For some recent work on the interpolation problem see Richter[20] and Mansfield.[21] An approach to the data analysis problem which is similar to that of Refs. 2–16 has also been given by Miller.[22]

Let A be a Hilbert space comprised of functions $f(x)$ defined for x in some domain D. The reproducing kernel is a function $H(x,y)$, which for fixed $y \in D$ lies in A, and which satisfies

$$(H(\cdot,y),f) = f(y); \tag{1.1}$$

the dot replaces the dummy variable with respect to which the inner product is defined. From (1.1) several results follow immediately:

$$(H(\cdot,x),H(\cdot,y)) = H(x,y), \tag{1.2}$$

$$H(x,y) = H(y,x)^*, \tag{1.3}$$

$$H(x,x) \geq 0, \tag{1.4}$$

$$|H(x,y)|^2 \leq H(x,x)H(y,y). \tag{1.5}$$

Inequality (1.5) follows from application of the Schwartz inequality to (1.2).

Copyright © 1973 by the American Institute of Physics

Reproducing kernels do not exist for all function spaces; spaces for which they do exist have much nicer properties. In particular, strong convergence of a sequence of functions $f_n(x)$ implies uniform convergence in a closed subdomain of D in which $H(x, x)$ is bounded. This is proved as follows (M, p. 45):

$$|f_n(x) - f(x)|^2 = |(H(\cdot, x), f_n - f)|^2. \tag{1.6}$$

By use of the Schwartz inequality, we have

$$\begin{aligned} |f_n(x) - f(x)|^2 &\le \|f_n - f\|^2 (H(\cdot, x), H(\cdot, x)) \\ &= \|f_n - f\|^2 H(x, x). \end{aligned} \tag{1.7}$$

Pointwise (but not necessarily uniform) convergence also follows from weak convergence, because

$$|f_n(x) - f_m(x)| = (H(\cdot, x), f_n - f_m). \tag{1.8}$$

Our subsequent calculations will be devoted to various generalizations of a Hilbert space A_0, which contains functions $f(u)$ which are analytic for $|u| < 1$, with

$$\|f\|^2 = \frac{1}{2\pi} \int |f(u)|^2 |du|. \tag{1.9}$$

The reproducing kernel for A_0 is[2,18,19,23]

$$H_0(u, v) = 1/(1 - uv^*). \tag{1.10}$$

The space A_0 can be immediately generalized to include a positive weight function $w(u)$ in (1.9): with $\phi(u)$ being the unique (up to a constant phase) nonvanishing analytic function for $|u| < 1$ which satisfies $|\phi(u)| = w(u)^{-1/2}$ for $|u| = 1$, we have

$$H(u, v) = \phi(u) H_0(u, v) \phi(v)^*. \tag{1.11}$$

The generalization to other domains D of holomorphy with a square-integral norm on the boundary of D is obtained by conformal transformation. Specifically, if the mapping to a circle is given by $u = u(x)$, let

$$\psi(x) = \left(\frac{du}{dx}\right)^{1/2}. \tag{1.12}$$

Then, if there is no weight function in the integral over $|dx|$, we have

$$H(x, y) = \psi(x) H_0(u(x), u(y)) \psi(y)^*. \tag{1.13}$$

In a conformal transformation of the circle into itself, H_0 is invariant.

As in Ref. (2), we denote certain linear functionals of $f \in A$ by

$$f_k \equiv I_k f \equiv \int_D m_k(x) f(x) dV_x, \tag{1.14}$$

where dV_x is a volume element in D. Here $m_k(x)$ may be any distribution defined in D for which a bound M_k exists, in the sense that

$$|f_k| \le M_k \|f\|. \tag{1.15}$$

We define a function

$$H_k(x) = H(x, \cdot) I_k(\cdot)^+ \equiv (I_k(\cdot) H(\cdot, x))^*. \tag{1.16}$$

Then we have

$$\begin{aligned} (H_k, f) &= (H(\cdot, \cdot) I_k(\cdot)^+, f) \\ &= I_k(\cdot)(H(\cdot, \cdot), f) \\ &= I_k(\cdot) f(\cdot) = f_k. \end{aligned} \tag{1.17}$$

Thus $H_k(x)$ is the function which is associated with I_k in the sense of the Riesz theorem (M, p. 25). We also define $H_{km} = I_k(\cdot) H_m(\cdot)$.

A generalization of (1.4) is that

$$H_{kk} \ge 0 \tag{1.18}$$

for all I_k constructed as in (1.12). It can be shown that to every function $H(x, y)$ satisfying both (1.3) and (1.18) there is a corresponding Hilbert space A (M, p. 96). This space consists of functions which can be represented in the form of a convergent sum

$$f(x) = \sum \alpha_k H_k(x). \tag{1.19}$$

The problem of interpolation is, given $f_k = I_k f$ for $k = 1, \ldots, N$, to find a suitable function $f^{(N)}$ such that $I_k f^{(N)} = f_k, k = 1, \ldots, N$. This $f^{(N)}(x)$ is a linear approximation to $f(x)$ if it can be written in the form

$$f^{(N)}(x) = \sum_{k=1}^{N} h_{k,N}(x) f_k, \tag{1.20}$$

where the $h_{k,N}$ do not depend on the f_k. The I_k may be said to be complete in a subdomain $d \subset D$ if there is some set of functions $h_{k,N}$ for $k \le N$ with $N \to \infty$ such that (1.20) converges to $f(x)$ for $x \in d$ and all $f \in A$. For each N, the linear approximation of smallest norm, $f^{(N)}(x)$, is given by choosing

$$h_{k,N}(x) = \sum_{m=1}^{N} H_m(x) L_{mk}, \tag{1.21}$$

where L_{mk} is the inverse of the $N \times N$ matrix H_{mk}. If A is a space of analytic functions, the I_k which are complete in any $d \subset D$ are automatically complete in all closed subregions $d \subset D$ (cf. Ref. 2).

If A is a separable space, in particular if it consists of functions analytic in some domain, then one can represent $H(x, y)$ by a convergent sum of orthonormal functions $\psi_n(x)$:

$$H(x, y) = \sum_n \psi_n(x) \psi_n(y)^*. \tag{1.22}$$

These $\psi_n(x)$ may be obtained, for example, by orthonormalizing any set of complete $H_k(x)$. For A_0, we may take $\psi_n(u) = u^n$.

In the remaining sections of this paper, we shall make extensive use of the following theorem[24]; (M, p. 97). Let A_1 and A_2 be two spaces with reproducing kernels $H_1(x, y)$ and $H_2(x, y)$; then

$$H(x, y) = H_1(x, y) + H_2(x, y) \tag{1.23}$$

is the reproducing kernel of a certain Hilbert space A. Here $f_1(x) \in A_1$ and $f_2(x) \in A_2$ may, for example, be functions analytic in domains D_1 and D_2, respectively, with $D_1 \cap D_2 = D_{12} \neq 0$, for which A_1 and A_2 also have a nontrivial intersection A_{12} containing functions analytic in $D_1 \cup D_2$. Let us call A the conjunction of A_1 and A_2. To obtain the norm in A we proceed as follows: Given a function $f(x)$ defined in D_{12}, we define a decomposition

$$f(x) = f_1(x) + f_2(x); \tag{1.24}$$

this is not unique because we may add $f_{12}(x)$ to $f_1(x)$ and subtract it from $f_2(x)$, where $f_{12} \in A_{12}$. We define

$$\|f\|^2 = \min (\|f_1\|_1^2 + \|f_2\|_2^2), \tag{1.25}$$

where the minimum is taken with respect to f_{12}. It can be shown that this norm (1.25) satisfies the usual axioms.

If $f_1(x)$ and $f_2(x)$ are the minimal decomposition of $f(x)$, that is, the decomposition for which the minimum in (1.25) occurs, then $(f_{12},f_1)_1 = (f_{12},f_2)_2$ for any $f_{12} \in A_{12}$. Using the reproducing property of H_1 and H_2, it follows that (1.23) is the minimal decomposition of H, and furthermore, for any I_k,

$$H_k(x) = H_{1k}(x) + H_{2k}(x) \tag{1.26}$$

is the minimal decomposition of $H_k(x)$.

The generalization of (1.23)–(1.26) to any finite number of components is trivial.

In the remaining work we shall be concerned mostly with spaces of real analytic functions. It is convenient to choose the I_k to be real functionals and the $H_k(x)$ are then also real analytic. Furthermore, we may choose a real analytic basis, and the coefficients are then real numbers.

2. GAUSSIAN DISTRIBUTION OF FUNCTIONS

Let $f \in A$ be represented in the form

$$f(x) = \sum a_n \psi_n(x). \tag{2.1}$$

We have

$$\|f\|^2 = \sum |a_n|^2. \tag{2.2}$$

We define a probability measure in A by taking the a_n to be independent Gaussian variables with a mean of zero[3]

$$\langle a_n \rangle = 0. \tag{2.3}$$

The variance is unity:

$$\langle a_n a_m^* \rangle = \delta_{nm}. \tag{2.4}$$

More specifically, if A is a real analytic space, the a_n are purely real; otherwise the real and imaginary parts are to be independent and have the same variance. In any other orthonormal basis (2.3) and (2.4) also hold. In Ref. 2 an equivalent (but more complicated) definition of the probability distribution was given in terms of the distribution of boundary values for functions in the space A_0 [Eq. (1.9)] and its trivial generalizations. The use of an equivalent *a priori* probability in Hilbert spaces of real functions has recently been introduced by Backus[25] in connection with data analysis problems of geophysics.

Now consider a random function (1) with the coefficients a_n chosen in sequence according to the prescription just given. It is clear that the norm (2) will almost certainly diverge; i.e., the $f(x)$ so generated almost certainly lies outside of A. On the other hand, the distribution of matrix elements (g,f), where $g \in A$ and f is a random function (1), has a Guassian distribution with a finite variance given by $\|g\|^2$. Furthermore, as indicated by (2.6) below, the functional values $f(x)$ also have a finite variance. In other words, theprobability distribution has really been defined over $\bar{A}$, the weak closure of A.

From (2.3), (2.4), and (1.22) we obtain

$$\langle f(x) \rangle = 0, \tag{2.5}$$

$$\langle f(x) f(y)^* \rangle = H(x,y). \tag{2.6}$$

Thus, for this Gaussian distribution in $\bar{A}$, all statistical properties of the distribution can be obtained directly from the reproducing kernel of A.

This definition of the probability measure can be criticized[14] on the grounds that the probability is zero that a random function lies in the space originally introduced. In defense of our definition, three points may be noted. First, Eq. (2.4) is the only possibility which is invariant. Second, spaces with reproducing kernels are simpler than general Hilbert spaces; in these the weak closure is tractable, as is demonstrated by (2.6) and (1.8). Third, when the distribution is used for statistical inference, any possible bias is slightly reduced by considering a slightly enlarged space of functions.

For the simple space A_0, we may take $\psi_n(u) = u^n$. Properties of random Taylor series

$$f(u) = \sum b_n u^n \tag{2.7}$$

have been studied extensively[26]; it is known that almost all functions in $\bar{A}_0$ have a natural boundary on $|u| = 1$. The same is of course true for all the trivial generalizations of A_0 described in Sec. 1. It is characteristic of random Taylor series such as (2.7) that they almost always have a circular natural boundary (Ryll-Nardzewski theorem; cf. Ref. 26). This fact does not depend on the a_n having a Gaussian distribution or on (2.4), but if the distribution of the b_n is not symmetrical around $b_n = 0$, there may be additional singularities inside the circular natural boundary. (An exception is that if the variance of the b_n drops sufficiently rapidly, the radius of the natural boundary may recede to ∞.)

More specific information about the boundary behaviour of the series (2.7) is also given by Kahane.[26] Suppose that the a_n are Gaussian with $\langle a_n \rangle = 0$ and

$$\limsup \langle b_n^2 \rangle^{1/n} = 1, \tag{2.8}$$

so the radius of convergence (and natural boundary) is at $|u| = 1$. Suppose also that, with $0 \le \alpha \le 1$,

$$\sum n^{2\alpha+\epsilon} \langle b_n^2 \rangle < \infty. \tag{2.9}$$

Let

$$F(\theta) = \lim_{\rho \to 1} f(\rho e^{i\theta}). \tag{2.10}$$

Then $F(\theta)$ satisfies the Lipschitz condition of order α, that is,

$$|F(\theta \pm h) - F(\theta)| = 0(h^\alpha). \tag{2.11}$$

On the other hand, if

$$\liminf n^{2\alpha+1} \langle b_n^2 \rangle > 0, \tag{2.12}$$

then, for all θ,

$$\limsup h^{-\alpha} |F(\theta \pm h) - F(\theta)| > 0. \tag{2.13}$$

(Kahane also gives estimates of logarithmic terms, but the results above are sufficient for our later work.)

These results may be generalized by replacing α by $\alpha + p$ in (2.9) and (2.12). Then, if p is a positive integer, (2.11) and (2.13) hold with (2.10) replaced by

$$F(\theta) = \lim_{u \to e^{i\theta}} \left(\frac{d}{du}\right)^p F(u). \tag{2.14}$$

Negative p can be included by integrating $|p|$ times; specifically, for $p = -1$:

$$F(\theta) = \lim_{\rho \to 1} \int_0^{\rho e^{i\theta}} f(u) du. \tag{2.15}$$

The consideration of random functions from $\overline{A}$ gives a new perspective to the problem of interpolation introduced in Sec. 1.[2] We now examine the average properties, with respect to the prior distribution over $\overline{A}$, of the interpolation $f^{(N)}(x)$ to N linear functionals f_k, $k = 1,\ldots,N$. Let us consider first the average value of all functions $f(x) \in \overline{A}$ which yield the N given values f_k. This average is equal to the minimum-norm interpolation given by Eqs. (1.20)–(1.21) [cf. (2.2)]:

$$\langle f(x)\rangle_N = F^{(N)}(x). \tag{2.16}$$

Furthermore, we may examine the variance of the deviation from the interpolated value, when averaged over all functions in $\overline{A}$:

$$\delta^2(x) = \langle |f(x) - f^{(N)}(x)|^2\rangle. \tag{2.17}$$

This averaged squared error is a minimum when $f^{(N)}(x)$ is the average value $F^{(N)}(x)$. Explicit formulas for the minimum $\delta^2(x)$ are given in Ref. 2. Note that even though the space of functions to be interpolated has been enlarged from A to $\overline{A}$, the best interpolation to any finite set of values still lies in A.

In the remaining sections we shall need information about smoothness properties of boundary values for the general series (2.1) analogous to those described for (2.7). This problem does not appear to have been studied in the literature, but we can easily obtain enough information about the boundary values of the reproducing kernel to suit our purposes.

Let $0 < \alpha \leq 1$, and let $H(x,x)$ be a finite constant on the boundary B of the domain of analyticity D. Then almost all $f \in \overline{A}$ have finite boundary values. The following two statements are equivalent:

(1) Almost all $f \in \overline{A}$ satisfy a Lipschitz condition of order α, but fail to satisfy a Lipschitz conditions of order $\alpha + \epsilon$, at almost all boundary points.

(2) If x and $x + h$ are boundary points,

$$H(x,x) - \mathrm{Re}\, H(x+h,x) - R(h^2) = 0(|h|^{2\alpha}) \tag{2.18}$$

$$\neq 0(|h|^{2\alpha+\epsilon}), \tag{2.19}$$

where $\epsilon > 0$, and $R(h^2)$ is a regular function of h^2. Equation (2.19) follows from the identity

$$\langle |f(x+h) - f(x)|^2\rangle = H(x+h,x+h) + H(x,x) - H(x+h,x) - H(x,x+h). \tag{2.20}$$

If $H(x,x)$ is not constant on B, but possesses a piecewise smooth first derivative with respect to arc length on the boundary, construct the unique (apart from a constant phase) nonvanishing analytic function $\phi(x)$ satisfying

$$|\phi(x)| = H(x,x)^{1/2}, \quad x \in B. \tag{2.21}$$

Then (2.19) holds for the reduced function

$$\overline{H}(x,y) = H(x,y)/(\phi(x)\phi(y)^*). \tag{2.22}$$

To generalize to $\alpha + p$, consider the functions obtained by differentiating p times ($p > 0$) or integrating $|p|$ times ($p < 0$).

Finally, if $n \geq p + \alpha$ the mean square value of $d^n f(x)/dx^n$ is unbounded as $x \to x_0$, x_0 being any boundary point. It follows that for almost all $f \in \overline{A}$, B is a natural boundary of $f(x)$. However, an approximation constructed from a finite amount of data will only have branch points on B, with use of the kernels constructed below.

3. HYPERGEOMETRIC REPRODUCING KERNELS

A. Isotropic circular region

Let the real analytic Hilbert space A be comprised of functions analytic for $|u| < 1$:

$$f(u) = \sum_0^\infty b_n u^n \tag{3.1}$$

with the norm

$$\|f\|^2 = \sum_0^\infty |b_n|^2 R_n, \tag{3.2}$$

where $R_n > 0$; let us choose $R_0 = 1$. The reproducing kernel $H(u,v)$ is then

$$H(u,v) = \sum_0^\infty z^n/R_n, \tag{3.3}$$

where $z = uv^*$. Over the Gaussian ensemble introduced in Sec. 2, the b_n are independent with

$$\langle b_n^2\rangle = R_n^{-1}. \tag{3.4}$$

Now suppose that A is to be characterized by a specific degree of local smoothness on $|u| = 1$. Let us suppose that the leading singularities on $|u| = 1$ of the functions to be approximated have the form $|u - u_0|^\nu$, possibly multiplied by powers of $\log(u - u_0)$, where ν is an arbitrary real number. It is then natural to assume that ν is the "critical Lipschitz order" for A. In other words, with $0 < \alpha \leq 1$, for $f \in A$ (2.11) holds with $\alpha + p < \nu$, while for almost all $f \in \overline{A}$, (2.11) fails to hold at some boundary points with $\alpha + p > \nu$. This is achieved (2.8), (2.12) if

$$R_n \sim n^{2\nu+1}. \tag{3.5}$$

A convenient family of reproducing kernels is given by

$$H(u,u') = F(a,b,c;z) \tag{3.6}$$

for which

$$R_n = \frac{\Gamma(a)\Gamma(b)\Gamma(c+n)\Gamma(1+n)}{\Gamma(c)\Gamma(a+n)\Gamma(b+n)} \tag{3.7}$$

and for which condition (1.18) will be satisfied if a, b, and c are positive.

We choose a = 1 so that the integral representation will be especially convenient for numerical calculations. The condition (3.5) then gives

$$2\nu = c - b - 1 \tag{3.8}$$

and we have

$$H(u,v) = G(z), \quad z = uv^*,$$

$$G(z) = \sum \frac{\Gamma(2\nu+b+1)\Gamma(b+n)}{\Gamma(2\nu+b+n+1)\Gamma(b)} z^n = \frac{1}{B(b,2\nu+1)} \int_0^1 \frac{t^{b-1}(1-t)^{2\nu} dt}{1-tz}. \tag{3.9}$$

It is easy to check, either directly from the integral in (3.9) or from the continuation formula for hypergeometric functions, that $H(u,v)$ satisfies condition (2.19).

The second parameter b evidently determines the value n_0 at which the transition occurs between the small n behavior of the coefficients ($R_n \approx 1$) and their large n behavior [$R_n = 0(n^{2\nu+1})$]. This parameter b (or n_0) also determines how the functional values are correlated on or near $|u| = 1$, for functions chosen randomly from A.

If $\nu > 0$, the mean square value on the boundary is finite, and the correlation function for boundary values is obtained from

$$h(\theta - \theta') = H(e^{i\theta}, e^{i\theta'}). \tag{3.10}$$

Noting that

$$\frac{1}{2\pi i} \oint \frac{du}{u} H(u, v) = 1, \tag{3.11}$$

we have

$$\frac{1}{2\pi} \int_0^{2\pi} h(\theta) d\theta = 1. \tag{3.12}$$

If we write, as a rough approximation near $\theta = 0$,

$$h(\theta) = i/(\theta + i\Delta), \tag{3.13}$$

we see that the correlation distance Δ can be estimated as

$$\Delta \approx H(1, 1)^{-1} = 2\nu/(2\nu + b). \tag{3.14}$$

However, if ν is small, the expression (3.13) is not accurate and we can better take

$$\Delta \approx n_0^{-1} \approx b^{-1}. \tag{3.15}$$

Another approach to the hypergeometric kernel (3.9) which shows how the formula can be generalized for other domains is suggested by the theorem contained in (1.23)–(1.26). First scale the primitive kernel H_0 of Eq. (1.10) to apply to $|u| < R, R \geq 1$: with the norm

$$\|f\|_R^2 = \frac{1}{2\pi R g(R)} \int_{|u|=R} |f|^2 \, |du|, \tag{3.16}$$

we have

$$H_R(u, v) = g(R) R^2/(R^2 - z). \tag{3.17}$$

We obtain the reproducing kernel for the conjunction of the Hilbert spaces A_R by summing over R; more generally, we may take

$$H(u, v) = \int_1^\infty H_R(u, v) dR \tag{3.18}$$

provided the integral is uniformly convergent. With $t = R^{-2}$ and an appropriate choice for $g(R)$, we obtain (3.9).

B. Conformal transformation of the circle

The space A constructed above is isotropic in the unit circle, in the sense that the width Δ of the boundary value correlation function is independent of the angle $\theta = \arg(u)$. This constancy of the width of the correlation function may not be in accord with the structure of the function which is to be estimated from data. Now, as pointed out following (1.13), a mapping of the circle into itself leaves H_0 invariant provided the weight function for the norm is scaled appropriately. This is no longer true of H as given by (3.9).

Let

$$u = \frac{x - \eta}{1 - \eta x} \quad (\eta \text{ real}) \tag{3.19}$$

and let $\phi(x)$ be nonvanishing and analytic in a region containing $|x| \leq 1$.

Then

$$\tilde{H}(x, y) = \phi(x) H(u(x), u(y), u(y)) \phi(y)^*, \tag{3.20}$$

has the following properties: (1) If $\nu > 0$, the variance of functional values on the circle is proportional to $|\phi(x)|^2$. (2) At each point $x = \exp(i\theta)$, the critical Lipschitz index for the space $\tilde{A}$ is ν. (3) The width of the boundary value correlation function at $x = \exp(i\theta)$ is

$$\Delta(\theta) = \Delta \frac{1 + \eta^2 - 2\eta \cos\theta}{1 - \eta^2}, \tag{3.21}$$

where Δ is the width characteristic of H.

C. Half plane Im $x > 0$

Consider first the Hilbert space containing functions $f(x)$ analytic for Im $x > 0$, satisfying $|f(x)| \to 0$ as $|x| \to \infty$ uniformly, with the norm

$$\|f\|^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} |f(x)|^2 dx. \tag{3.22}$$

The reproducing kernel for this space is

$$H(x, y) = i/(x - y^*). \tag{3.23}$$

To impose Lipschitz conditions on Im $x = 0$, we may either imitate the construction in Eqs. (3.16)–(3.18) or take a limit of the transformation (3.19)–(3.20). We obtain

$$H(x, y) = G(z), \tag{3.24}$$

where $G(z)$ is given by (3.9), but

$$z = 1 + i(x - y^*). \tag{3.25}$$

On the real axis, the boundary value correlation function ($\nu > 0$) has a correlation distance independent of x.

4. CONSTRUCTION OF GENERAL KERNELS

A. Simply connected regions

Let D be a simply connected region and B its piecewise-analytic boundary. The problem is to construct a reproducing kernel for a Hilbert space of functions analytic within D, for which the variance of boundary values, $v(s)$, and the correlation length for boundary values, $\Delta(s)$, are given functions of the arc length s along B. In applications it may also be desirable to allow the critical Lipschitz order ν to depend on s, but we shall first treat the case where ν is constant.

Let t be a continuous parameter, $0 \leq t \leq 1$, and for each t construct a domain $D_{(t)}$ with $D_{(t_1)} \supset D_{(t_2)}$ if $t_1 < t_2$, and with $D_{(1)} \equiv D$ and $D_{(0)}$ being the entire complex plane. Let $H_t(x, y)$ be the reproducing kernel corresponding to a square-integral norm over B_t, the boundary of $D_{(t)}$; in principle, this kernel is given by (1.13) if the mapping of $D_{(t)}$ onto a unit circle is known explicitly. Let $u(x)$ denote the mapping function for D onto the unit circle $|u| < 1$. Let x_0 be a point on an analytic arc of B; we shall examine $H_t(x, y)$ for t near 1 and x and y near to x_0.

Let $u_0 = u(x_0)$ and $\psi_0 = \psi(x_0)$ as given by (1.12). To first order in $(x - x_0)$ and $(y - x_0)$ we have

$$\begin{aligned} u &= u_0 + \psi_0^2 (x - x_0), \\ v &= u_0 + \psi_0^2 (y - x_0), \end{aligned} \tag{4.1}$$

$$1 - uv^* = -u_0^* \psi_0^2 (x - x_0) - u_0 \psi_0^{*2} (y - y_0). \tag{4.2}$$

Note that $u_0^* \psi_0 / \psi_0^*$ is a phase factor which rotates the direction of the boundary at x_0 into the vertical. With

$$\begin{aligned} i\xi &= u_0^* (x - x_0) \psi_0 / \psi_0^*, \\ i\eta &= u_0^* (y - x_0) \psi_0 / \psi_0^*. \end{aligned} \tag{4.3}$$

We have from 1.13, to leading order in ξ and η,

$$H_1(x,y) = i/(\xi - \eta^*) \tag{4.4}$$

as should be expected from (3.23).

Let s be the arc length up to the point x_0, and let $R(s,t)$ be the distance from x_0 to the nearest point of $B_{(t)}$; this is to have the form

$$R(s,t) = r(s)(1-t) + 0((1-t)^2). \tag{4.5}$$

Here $r(s)$ is an analytic function of s, at least where B is analytic. Then, for t near 1 and for points near x_0, we have

$$H_t(x,y) \approx \frac{i}{\xi - \eta^* + 2ir(s)(1-t)}. \tag{4.6}$$

Let $g(t)$ be a nonnegative normalized weighting function

$$\int_0^1 g(t)dt = 1, \tag{4.7}$$

which behaves near $t = 1$ as

$$g(t) \sim (1-t)^{2\nu}. \tag{4.8}$$

We have now constructed a set of Hilbert spaces $A(t)$; we shall let $g(t)H_t(x,y)$ be the reproducing kernel for $A(t)$ and then form the conjunction of these spaces. Thus, we construct

$$\tilde{H}(x,y) = \int_0 g(t)dt H_t(x,y) \tag{4.9}$$

and let

$$\tilde{H}(s) = \tilde{H}(x_0,x_0), \quad x_0 \in B. \tag{4.10}$$

We have, approximately,

$$\tilde{H}(s) \sim \frac{1}{2r(s)} \int_0^1 g(t) \frac{dt}{1-t}. \tag{4.11}$$

The approximation (4.11) will be most accurate when

$$\Delta = 2\left(\int_0 g(t) \frac{dt}{1-t}\right)^{-1} \tag{4.12}$$

is small. In any case, it is clearly consistent to identify $\Delta(s)$ with $\Delta r(s)$.

On eliminating $H(s)$ as in (2.21)–(2.22), it is clear that condition (2.19) is satisfied. The kernel which satisfies

$$H(x_0,x_0) = v(s) \tag{4.13}$$

is given by

$$H(x,y) = \phi(x)\tilde{H}(x,y)\phi(y), \tag{4.14}$$

where

$$(\phi(x_0))^2 = v(s)/\tilde{H}(s), \quad x_0 \in B. \tag{4.15}$$

B. Further generalizations

An N-fold connected domain D is the intersection of N singly connected domains $D_k, k = 1, \ldots, D_N$, of which $N-1$ consist of the region outside of their boundary B_k. Let us suppose that $H_k(x,y)$ is the reproducing kernel for the space A_k, of functions analytic in D_k, satisfying suitable boundary conditions on B_k; this kernel $H_k(x,y)$ is constructed as in part A of this section. Then the reproducing kernel for the space A of functions analytic in D, satisfying the same boundary conditions on each of the B_k, is

$$H(x,y) = \sum_{k=1}^{N} H_k(x,y). \tag{4.16}$$

It is clear that the Lipschitz order ν_k may be different on each of the B_k. This suggests a simple way to construct kernels for which the order ν is different on different parts of the same boundary curve. Let us consider a simply connected region D_0 with boundary B_0. On part of B_0 we have $\nu = \nu_1$, on the remainder $\nu = \nu_2$. Construct any suitable domains $D_i \supset D_0$, having boundaries B_i, such that $B_i \cap B_0$ is the part of B_0 on which the index is ν_i. Then take

$$H(x,y) = H_1(x,y) + H_2(x,y),$$

where $H_i(x,y)$ is defined in D_i with the order ν_i. The value of $H(x_0,x_0)$ for $x_0 \in B$, is apportioned between H_1 and H_2; the value of H_2 could be chosen first, somewhat arbitrarily, and then the value of H_1 chosen to make up the difference.

It is clear that there are many hidden parameters involved in the constructions outlined in this section. However, the kernel $H(x,y)$ at interior points is not very sensitive to these ambiguities, because in going from the boundary to the interior, variations in a function are smoothed over. Thus the ambiguities can usually be resolved on the grounds of simplicity.

5. EXAMPLES

A. Strip centered on real axis

Consider first the sapce of functions $f(\xi)$ analytic in the strip $|\mathrm{Im}(\xi)| < \pi/(4t)$ which vanish in the strip uniformly as $|\xi| \to \infty$, with the norm

$$\|f\|_t^2 = \frac{t}{2\pi} \int_{-\infty}^{\infty} d\xi \left[\left|f\left(\xi - \frac{i\pi}{4t}\right)\right|^2 + \left|f\left(\xi + \frac{i\pi}{4t}\right)\right|^2\right]. \tag{5.1}$$

The reproducing kernel is

$$H_t(\xi,\eta) = 1/\cosh t(\xi - \eta^*). \tag{5.2}$$

To impose Lipschitz conditions on $|\mathrm{Im}(\xi)| = \pi/4$, we integrate (5.2) with a weight function:

$$H(\xi,\eta) = \int_0^1 \frac{g(t)dt}{\cosh t(\xi - \eta^*)}. \tag{5.3}$$

The width of the correlation function on the boundary is a constant.

B. Elliptical region

Let B be a unifocal ellipse with semi-axes a and b, let $r = a + b$, and let D be the region inside B. The function

$$z = \tfrac{1}{2}(\zeta + \zeta^{-1}) \tag{5.4}$$

maps the annulus $1 \le \zeta < r$ onto D. These domains arise as a result of conformal maps which maximize the rate of convergence of polynomial approximations to data on a given segment, which becomes here $-1 < z < 1$[2,27,28]

A suitable primitive kernel is

$$H_0(z,w) = \sum b_n T_n(z) T_n(w)^*, \tag{5.5}$$

where

$$T_n(z) = \tfrac{1}{2}(\zeta^n + \zeta^{-n}) \tag{5.6}$$

are Chebyshef polynomials, and

$$b_n = (r^{2n} + r^{-2n} + 2\delta_{On})^{-1}. \tag{5.7}$$

The norm here is a simple square integral over $|\zeta| = r$, or a square integral with the weight $|z^2 - 1|^{-1/2}$ over B.

One method to impose Lipschitz conditions on B is to replace b_n by $R_n b_n$, where R_n is given by (3.7), (3.9). For large n, $b_n \sim r^{-2n}$ and $T_n(z) \sim \frac{1}{2}\zeta^n$, so the properties of power series described in Sec. 2 apply here.

A second way is to consider nested unifocal ellipses, with $r(t) = r/\sqrt{t}$, and integrate with the same weight function $g(t)$ as in (3.9). Then we have

$$H(z, w) = \sum B_n T_n(z) T_n(w)^*, \tag{5.8}$$

where

$$B_n = \int_0^1 \frac{g(t)dt}{r^{2n}/t^n + t^n/r^{2n} + 2\delta_{On}} . \tag{5.9}$$

For large n, we obtain

$$B_n \sim b_n R_n \tag{5.10}$$

showing that the two approaches are essentially equivalent. In the annular domain, the width of the correlation function is independent of ζ, but on the ellipse it is proportional to $|z^2 - 1|^{1/2}$.

C. Cut plane

In many applications the domain of analyticity C of functions $f(s)$ is the entire s plane except for two cuts along the real axis: $s < s$, and $s > s_2$. We show here how to apply the previous results to obtain a variety of kernels for this domain. The most complicated technical problem to be faced in applying the general procedure of Sec. 4 is that of constructing maps for arbitrary domains; it is therefore useful to illustrate what can be done with elementary maps.

For the first example, map C onto $|u(s)| < 1$ and use the kernel derived in Sec. 3. Then, for large s, the boundary value correlation distance $\Delta(s)$ has the behavior

$$\Delta(s) = 0(s^2) \qquad (|s| \to \infty). \tag{5.11}$$

Another way to put (5.11) is to say that the structure of the functions $f(s)$ on the cuts is characterized by a constant amount of variation in constant intervals of u; for large s, this is equivalent to constant intervals of $(1/s)$.

Second, one may map C onto the strip $|\mathrm{Im}(\xi)| < \pi/4$ and use (5.3) as the kernel. Then, for $s \to \infty$,

$$\Delta(s) = 0(s). \tag{5.12}$$

In other words, there is a constant amount of structure in constant intervals of $\xi \sim \log s$.

Finally, we may consider C as the union of C_L and C_R, where these are the entire plane except for only the left- and right-hand cuts, respectively. Then we have, as shown in Sec. 4B, a simple way to impose different Lipschitz conditions on the two cuts. If we map each of C_L and C_R onto a circle, we again obtain (5.11). However, we may also map either of them into a half plane; then there is a constant degree of structure in constant intervals of $s^{1/2}$, for large s, or

$$\Delta(s) = 0(s^{3/2}). \tag{5.13}$$

In all of these examples the regions near s_1 and s_2 are expanded greatly. In physical examples where s is an energy variable, the effect is that, for small momenta, a constant degree of structure appears in equal intervals of momentum. This effect is in accord with general physical intuition.

6. SEVERAL COMPLEX VARIABLES

Reproducing kernels for functions of several complex variables have been widely used for furthering the general theory of analytic functions.[18,29] The norm of the corresponding Hilbert spaces, however, has been defined as an integral over the entire domain (Bergman kernels). Hilbert spaces in which the norm is defined by a boundary integral appear to be more useful for the kinds of physical applications envisaged here but have been studied less extensively. There is, however, a general rule which can be often used: if the domain of analyticity D is a product domain, for example, $D = D_1 \times D_2$, where D_i is the domain of analyticity in the single variable x_i, then

$$H(x_1, x_2, y_1, y_2) = H_1(x_1, y_1) H_2(x_2, y_2) \tag{6.1}$$

is the reproducing kernel for the direct product $A = A_1 \otimes A_2$ (M, p. 105). This allows the immediate application of our previous formulas to many interesting situations.

An application similar to the one that is described below has already been treated by Shih[8].

Consider, for example, functions which satisfy Mandelstam's representation[30]; first let

$$f(s, t) = \int_{s_0}^{\infty} \int_{t_0}^{\infty} \frac{\rho(s', t')ds'dt'}{(s' - s)(t' - t)} . \tag{6.2}$$

Map each of the cut s and t planes to a circle or half plane and use a hypergeometric kernel. There are usually higher two-body thresholds, which suggests taking $\nu = \frac{1}{2}$. Then we write

$$H_{st}(s, t, s', t') = H_s(s, s') H_t(t, t'), \tag{6.3}$$

where H_x is given by a prescription from Sec. 3. For the general Mandelstam representation we have to add two more terms:

$$H = H_{st}(s, t, s', t') + H_{su}(s, u, s', u') + H_{tu}(t, u, t', u'). \tag{6.4}$$

(where $s + t + u = \text{const}$). For greater generality, envelope factors may be supplied to each term in (6.4); specifically, the first term may be replaced by

$$H_{st} = \phi(s, t) H_s(s, s') H_t(t, t') \phi(s', t')^*, \tag{6.5}$$

where $\log \phi$ satisfies a representation similar to (6.2).

As a more specific example, consider the $\pi^0 - \pi^0$ elastic scattering amplitude. Crossing symmetry requires that H_{st}, H_{su}, and H_{tu} be the same function H_{00}. In this case we may distinguish between the terms in which the 2π and 4π thresholds are lowest as follows:

$$H_{00} = H_2(s, s') H_4(t, t') + H_4(s, s') H_2(t, t'). \tag{6.6}$$

7. SUMMARY AND DISCUSSION

In fitting data, one necessarily uses functions drawn from a certain preselected class, and the choice of this class will influence the results. It has been shown that the definition of the class of functions to be used can be conveniently made in terms of the reproducing kernel

of a function space. Formulas have been given which will allow the incorporation in this kernel of general information about the behavior of the functions on the boundary of a domain of analyticity

The availability of kernels with a great deal of flexibility emphasizes anew the importance of using sound physical and mathematical arguments to establish the proper class of functions to use in a specific situation. Much work remains to be done on this problem. The approach through reproducing kernels, especially when the kernel is interpreted as the covariance matrix for random functional values (Eq. 2.6) suggests new ways to formulate and study the mathematical problem of characterizing the general behavior of functions suitable for use for a concrete physical problem

An example of the kind of problem for which the solution is unknown but would be of interest is presented by the scattering amplitudes which in the nonrelativistic wave mechanics arise from a superposition of Yukawa potentials. Here one might define a space of Yukawa potential functions through its reproducing kernel; the question is then, what is the kernel for the space to which the scattering amplitude belongs. This question could refer to the amplitude with t, s, or the angular momentum fixed, or more generally as a joint function of s and t as in Sec. 6.

In the relativistic theory a wide class of unsolved problems will present themselves to the reader. For applications it is especially important to find ways to characterize theoretically the envelope factor ϕ which appears in Eqs. (1.11), (2.21), (3.20), (4.14), (6.5). Approximate empirical estimations would also be useful.[31]

There are many directions in which the material of this paper needs to be extended. First, there is the problem of constructing kernels for spaces in which there is separate information about the real and imaginary part of boundary values. Work on this problem has already been done by Ross[4] and by Shepard and Shih[13] for spaces in which the norm is given by separately weighted square integrals over the real and imaginary parts. Another problem is that of imposing nonlinear constraints, such as given by unitarity, on the functions to be used for interpolation. An approximate linearization can be resorted to, but it is often cumbersome, and the bias which it introduces has not been studied. Finally, only the most simple kind of construction has been used in Sec. 6 for kernels for functions of more than one variable. It would beof interest to see whether more general techniques, based perhaps on the Bergman–Oka–Weil integral, could be used to pratical advantage in constructing kernels.

ACKNOWLEDGMENTS

I wish to thank Professors V. Mizell and C. C. Shih for interesting and helpful discussions. This work was supported, in part, by the U.S. Atomic Energy Commission. The manuscript was prepared at the Aspen Center for Physics.

[1]S. Ciulli and G. Nenciu, Nuovo Cimento **8,** 735 (1972); Commun. Math. Phys. **26,** 237 (1972); J. Math. Phys. (to be published).
[2]R. E. Cutkosky, Ann. Phys. (N.Y.) **54,** 110 (1969).
[3]P. Prešnajder and J. Pišut, Nuovo Cimento A **3,** 603 (1970).
[4]G. G. Ross, Nucl. Phys. B **31,** 113 (1971).
[5]Y. A. Chao, Phys. Rev. Lett. **25,** 309 (1970).
[6]R. E. Cutkosky and C. C. Shih, Phys. Rev. D **4,** 2751 (1971).
[7]C. C. Shih, Phys. Rev. D **4,** 3293 (1971).
[8]C. C. Shih, CMU preprint (1971), and private communication.
[9]Y. A. Chao and E. Pietarinen, Phys. Rev. Lett. **26,** 1060 (1971).
[10]R. C. Miller, T. B. Novey, A. Yokosawa, R. E. Cutkosky, H. R. Hicks, R. L. Kelly, C. C. Shih, and G. Burleson, Nucl. Phys. B **37,** 401 (1972).
[11]P. Lichard and P. Prešnajder, Nucl. Phys. B **33,** 605 (1971).
[12]R. L. Kelly, Phys. Lett. B **39,** 635 (1972).
[13]H. K. Shepard and C. C. Shih, Nucl. Phys. B **42,** 397 (1972).
[14]E. Pietarinen, Nordita preprint, 1972.
[15]C. C. Shih and H. K. Shepard, preprint, 1972.
[16]S. C. Cheung, CMU preprint, COO-3066-1, 1972.
[17]N. Aronszajn, Trans. Am. Math. Soc. **68,** 337 (1950).
[18]H. Meschkowski, *Hilbertsche Räume Mit Kernfunktion* (Springer, Berlin, 1962).
[19]P. J. Davis, *Interpolation and Approximation* (Blaisdell, New York, 1963).
[20]N. Richter, SIAM J. Num. Anal. **7,** 67 (1970); SIAM J. Num. Anal. **8,** 467 (1971); SIAM J. Num. Anal. **8,** 583 (1971).
[21]L. E. Mansfield, SIAM J. Num. Anal. **8,** 115 (1971).
[22]K. Miller, SIAM J. Math. Anal. **1,** 52 (1970).
[23]J. L. Walsh, *Interpolation and Approximation by Rational Functions in the Complex Domain* (American Mathematical Society, Providence, Rhode Island, 1956), 2nd ed., Vol. 20.
[24]In ignorance of the mathematical literature, a proof was included in Ref. 2.
[25]G. Backus, Proc. Natl. Acad. Sci. USA **65,** 1 (1970); Proc. Natl. Acad. Sci. USA **65,** 281 (1970); Proc. Natl. Acad. Sci. USA **67,** 282 (1970).
[26] J. P. Kahane, *Some Random Series of Functions* (Heath, Lexington, Massachusetts, 1968).
[27]R. E. Cutkosky and B. B. Deo, Phys. Rev. **174,** 1859 (1968).
[28]S. Ciulli, Nuovo Cimento A **61,** 787 (1969).
[29]S. Bochner and W. T. Martin, *Several Complex Variables* (Princeton U. P., Princeton, 1948).
[30]R. J. Eden, P. V. Landshoff, D. I. Olive, and J. C. Polkinghorne, *The Analytic S-Matrix* (Cambridge U. P., Cambridge, 1966).
[31]R. E. Cutkosky, Nucl. Phys. B **13,** 351 (1969).

The gravitational influence of a beam of light. I*

Raymond W. Nackoney

Physics Department, University of Chicago, Chicago, Illinois 60637
(Received 27 March 1972)

In this paper an exact solution is provided for the Einstein field equations of a beam of light. The beam is circular in cross section, infinite in length and duration, and considered in the geometrical limit with the flux being a function of only the radial coordinate. Null and timelike trajectories of this metric are considered in two classes: those with and without angular motion around the circular beam. The elements of the beam are initially without angular or radial motion and are found to follow a set of null trajectories which allows the beam to be in a state of equilibrium. A null trajectory, without angular motion, which is initially inclined slightly to the direction of the beam will curve away from the beam. Such a trajectory will continue to curve away until it is directed perpendicular to and then opposite to the beam direction, thereupon coming back into the beam upon the mirrored outgoing trajectory. Such trajectories are discussed with emphasis upon Lorentz transformations along the direction of the beam. Trajectories with angular motion include orbital cases. Beams of sufficient energy flux may allow closed null orbitals and regions of stable and unstable closed timelike orbits. Null helical trajectories which spiral back along the length of the beam can always occur; however, forward spiraling trajectories are restricted to large fluxes and are discussed in the light of Lorentz transformations.

INTRODUCTION

The propagation of light acted as a cornerstone in the formulation of the principle of special relativity. The property was incorporated that light travels through the flat space–time of special relativity with a unique speed c and without a change in its direction. In general relativity, special relativity is valid in two senses. First, it is valid locally, even though there may exist an over-all curvature due to the energy–momentum tensor of matter. Hence light may accelerate over a nonlocal region of a gravitational field. Second, special relativity is valid throughout all space–time in the case where no matter whatsoever is present. In this case, transmission of information is usually obtained by using light. But electromagnetic radiation itself has a nonzero energy–momentum tensor and, therefore, is active in modifying the geometry of space–time. The question arises as to whether a beam of light, with no other matter present, will interact upon itself gravitationally to focus or retard its own propagation.

This problem was first approached in the 1931 work of Tolman, Ehrenfest, and Podolsky[1,2] in which the effect of a pulse of light, and also a finite stationary beam of light, was obtained as a first-order perturbation on Minkowski space as expanded in the strength of the gravitational field. The bending of light in a particle's gravitational field was well known at that time, and the emphasis of the 1931 paper was directed to the converse problem of a particle in light's gravitational field. One result which was obtained was that to first order a null trajectory traveling parallel to the beam was neither deflected nor retarded. The present paper treats the exact field of an infinite beam and will prove to maintain the conclusion of Tolman, Ehrenfest, and Podolsky.

The metric we use is a case of Vaidya's Newtonian metric.[3] This metric has been applied to spherical systems, cylindrical systems, and ones which are asymptotically flat at infinity.[4–9] Whereas previous studies of systems with cylindrical symmetry dealt with a radial flux of gravitational radiation (Bondi and Sachs), this report deals with an axial flux of electromagnetic (or neutrino) radiation. The previous electromagnetic cases examined earlier are all spherical.

STATIONARY BEAM METRIC

We consider the problem in general relativity of an infinite beam of light with circular cross section, of radius R, which propagates along the z axis of a cylindrical coordinate system in the positive z direction. The beam's energy density ρ is measured in a local Minkowski frame and is without azimuthal, axial, or temporal dependencies. Hence, ρ is a function of the radius r only. We shall ignore all diffraction effects at the edge of the beam and all nongravitational interactions; and, hence, we initially have the flux solely in the positive z direction. We wish to consider the gravitational field produced by such a system in the geometrical limit, that is, where the wavelength of the light is made to go to zero. To do so, we must find a metric which is a solution of the field equations for the beam. Our major demand on the metric is that it will contain a retarded time which parametrizes successive cross sections of the beam.

Retarded time arises from the following considerations: The invariant interval in Minkowski space has its description in cylindrical coordinates given by

$$ds^2 = d\bar{t}^2 - d\bar{r}^2 - \bar{r}^2 d\bar{\theta}^2 - d\bar{z}^2. \tag{1}$$

A retarded time metric (r, t, θ, z) is obtained by the transformation

$$\bar{t} = t + z, \quad \bar{r} = r, \quad \bar{\theta} = \theta, \quad \bar{z} = z. \tag{2}$$

Using this transformation, we obtain the Vaidya form

$$ds^2 = dt^2 + 2dtdz - dr^2 - r^2 d\theta^2, \tag{3}$$

which has no diagonal term for the z component. To see the retarded time property, consider a test ray moving axially ($dr = 0, d\theta = 0$). Since light has the property that $ds = 0$, Eq. (3) becomes

$$0 = dt^2 + 2dtdz. \tag{4}$$

There are two solutions for Eq. (4), namely,

$$\frac{dz}{dt} = -\frac{1}{2}$$
and
$$dt = 0. \tag{5}$$

Copyright © 1973 by the American Institute of Physics

The negative sign of the first solution indicates a ray moving parallel to the z axis in the negative z direction. The second solution is compatible with light moving parallel to the axis in the positive direction. In this case $dt = 0$ or $t = \text{const}$, which states that t is the retarded time parametrizing the light ray.

A beam of light is composed of elements, each of which may be considered as a test ray; hence, a beam which is propagating along the z axis in a positive direction has its cross sections labeled by the retarded time t if the metric is of the form of Eq. (3). In our case we must assume a curvilinear metric. Hence we have the metric tensor, as defined by

$$ds^2 = g_{ik}dx^i dx^k \tag{6}$$

to be given quite generally by

$$\begin{aligned} &g_{00} = f(r), \quad g_{11} = -e^{2\lambda}, \quad g_{22} = -r^2, \\ &g_{03} = e^{\alpha}, \quad g_{33} = 0, \end{aligned} \tag{7}$$

where f, α, and λ are unknown functions of r; all other components of g_{ik} are zero. The contravarient metric tensor is found to be

$$\begin{aligned} &g^{00} = 0, \quad g^{11} = -e^{-2\lambda}, \quad g^{22} = -r^{-2}, \\ &g^{03} = e^{-\alpha}, \quad g^{33} = -fe^{-2\alpha}, \end{aligned} \tag{8}$$

all other components being zero.

As in our Minkowski case, we can see that the retarded time property holds. Let the 4-velocity of the beam, i.e., its elements, be given by $u^i(r)$. Our beam of light is axially directed, hence

$$u^1 = 0, \quad u^2 = 0. \tag{9}$$

The condition on the remaining two components of u^i is

$$0 = g_{ik}u^i u^k \tag{10}$$

or

$$0 = f(r)u^0 u^0 + 2e^{\alpha}u^0 u^3. \tag{11}$$

The two solutions of Eq. (11) are

$$u^3 = -\tfrac{1}{2}fe^{-2\alpha}u^0$$

and

$$u^0 = 0. \tag{12}$$

The first is the anti-parallel solution. Parallel rays and the beam itself are given by the second solution

$$u^0 = 0, \tag{13}$$

and a retarded time parametrizes cross sections or elements of the beam.

The Christoffel symbols are determined from the relation

$$\Gamma^i_{jk} = \tfrac{1}{2}g^{il}(g_{lj,k} + g_{lk,j} - g_{jk,l}). \tag{14}$$

For the metric of Eqs. (7) we obtain

$$\begin{aligned} &\Gamma^0_{01} = \frac{1}{2}\frac{d\alpha}{dr}, \\ &\Gamma^1_{00} = \frac{1}{2}\frac{df}{dr}e^{-2\lambda}, \quad \Gamma^1_{11} = \frac{d\lambda}{dr}, \quad \Gamma^1_{22} = -re^{-2\lambda}, \\ &\Gamma^1_{03} = \frac{1}{2}\frac{d\alpha}{dr}e^{\alpha-2\lambda}, \quad \Gamma^2_{12} = r^{-1}, \\ &\Gamma^3_{01} = \frac{1}{2}\left(\frac{df}{dr} - f\frac{d\alpha}{dr}\right)e^{-\alpha}, \quad \Gamma^3_{13} = \frac{1}{2}\frac{d\alpha}{dr}. \end{aligned} \tag{15}$$

The Ricci tensor is generated from the Christoffel symbols

$$R_{ik} = \Gamma^l_{kl,i} - \Gamma^l_{ik,l} + \Gamma^l_{mi}\Gamma^m_{kl} - \Gamma^m_{lm}\Gamma^l_{ik}. \tag{16}$$

A calculation gives the nonzero components:

$$\begin{aligned} &R_{22} = re^{-2\lambda}\frac{d}{dr}(\alpha - \lambda), \\ &R_{00} = -\tfrac{1}{2}e^{-2\lambda}\left[\frac{d^2f}{dr^2} + \frac{1}{r}\frac{df}{dr} - \frac{df}{dr}\frac{d(\alpha+\lambda)}{dr} + f\left(\frac{d\alpha}{dr}\right)^2\right], \\ &R_{03} = -\tfrac{1}{2}e^{-2\lambda}\left[\frac{d^2\alpha}{dr^2} + \frac{1}{r}\frac{d\alpha}{dr} - \frac{d\alpha}{dr}\frac{d\lambda}{dr} + \left(\frac{d\alpha}{dr}\right)^2\right], \\ &R_{11} = \frac{d^2\alpha}{dr^2} + \frac{1}{2}\left(\frac{d\alpha}{dr}\right)^2 - \frac{d\alpha}{dr}\frac{d\lambda}{dr} - \frac{1}{r}\frac{d\lambda}{dr}. \end{aligned} \tag{17}$$

THE ENERGY-MOMENTUM TENSOR

The energy–momentum tensor is proportional to the quadratic form of the light's 4-velocity u^i

$$T_{ik} = C(g_{il}u^l)(g_{km}u^m), \tag{18}$$

where C is proportional to the energy density. Let us consider an observer who is situated at rest in our metric. Such an observer has a 4-velocity $\bar{u}^i$ with

$$\bar{u}^1 = 0, \quad \bar{u}^2 = 0, \quad \bar{u}^3 = 0. \tag{19}$$

The condition on the magnitude of the 4-velocity is

$$g_{ik}\bar{u}^i\bar{u}^k = 1 \tag{20}$$

or for our observer,

$$g_{00}\bar{u}^0\bar{u}^0 = 1 \tag{21}$$

and, therefore,

$$\bar{u}^0 = (g_{00})^{-1/2}. \tag{22}$$

The energy–momentum tensor must be such that an observer who is at rest with respect to the global metric measures the energy density to be ρ in his local Minkowski frame. This condition is satisfied by the contraction

$$T_{ik}\bar{u}^i\bar{u}^k = \rho \tag{23}$$

or

$$T_{00}\bar{u}^0\bar{u}^0 = C(g_{0l}u^l)(g_{0m}u^m)/g_{00} = \rho;$$

therefore,

$$C = \rho[g_{00}/(g_{0l}u^l)(g_{0m}u^m)]. \tag{24}$$

Hence the energy–momentum tensor of the beam is given by

$$T_{ik} = \rho[g_{00}/(g_{0l}u^l)(g_{0m}u^m)]g_{is}u^s g_{kt}u^t. \tag{25}$$

The 4-velocity of our beam of light has three of its four components given by Eqs. (9) and (13):

$$u^0 = 0, \quad u^1 = 0, \quad u^2 = 0.$$

as a result of which Eq. (25) reduces to

$$T_{ik} = \rho g_{00}(g_{03})^{-2}g_{i3}g_{k3}. \tag{26}$$

The only nonzero component of T_{ik} is

$$T_{00} = \rho f. \tag{27}$$

The trace $T^i{}_i = 0$ as expected for electromagnetic (or neutrino) radiation.

FIELD EQUATIONS

Since the trace of the energy–momentum tensor is zero, the Einstein field equations with the cosmological constant set to zero are given by

$$R_{ik} = -(8\pi G/c^4)T_{ik}. \tag{28}$$

On writing Eqs. (28) explicitly in terms of Eqs. (17) and (27), we obtain

$$e^{-2\lambda}\left[\frac{d^2f}{dr^2} + \frac{1}{r}\frac{df}{dr} - \frac{df}{dr}\left(\frac{d\alpha}{dr} + \frac{d\lambda}{dr}\right) + f\left(\frac{d\alpha}{dr}\right)^2\right] = \frac{16\pi G}{c^4}\rho f, \tag{29}$$

$$e^{-2\lambda} r \frac{d(\lambda - \alpha)}{dr} = 0, \tag{30}$$

$$\tfrac{1}{2} e^{\alpha - 2\lambda}\left[\frac{d^2\alpha}{dr^2} + \frac{1}{r}\frac{d\alpha}{dr} - \frac{d\alpha}{dr}\frac{d\lambda}{dr} + \left(\frac{d\alpha}{dr}\right)^2\right] = 0, \tag{31}$$

$$\frac{d^2\alpha}{dr^2} + \frac{1}{2}\left(\frac{d\alpha}{dr}\right)^2 - \frac{d\alpha}{dr}\frac{d\lambda}{dr} - \frac{1}{r}\frac{d\lambda}{dr} = 0. \tag{32}$$

Equation (30) is immediately solvable; we find

$$\lambda = \alpha + \text{const.} \tag{33}$$

Substituting this solution into Eq. (32) we have

$$\frac{d^2\alpha}{dr^2} - \frac{1}{r}\frac{d\alpha}{dr} - \frac{1}{2}\left(\frac{d\alpha}{dr}\right)^2 = 0. \tag{34}$$

The solution of Eq. (34) is

$$e^{\alpha} = (A + Br^2)^{-2}. \tag{35}$$

Substituting Eq. (35) into Eq. (31) we find

$$-8AB(A + Br^2)^{-2} = 0. \tag{36}$$

Hence we must have either $A = 0$ or $B = 0$. If $A = 0$, we find

$$e^{\alpha} = B^{-2}r^{-4}. \tag{37}$$

Implying that e^{α}, and therefore e^{λ} as well, becomes singular at $r = 0$ as a fourth-order pole. These singularities produce an essential singularity for f at $r = 0$ for realistic energy distributions, e.g., $\rho =$ const. We, therefore, disregard this solution and take $B = 0$. Thus, our solution will be taken as

$$e^{\alpha} = D \text{ (a const)} \tag{38}$$

and

$$e^{\lambda} = E \quad \text{(a const)}. \tag{39}$$

Substituting Eqs. (38) and (39) into Eq. (29) we have

$$\frac{d^2f}{dr^2} + \frac{1}{r}\frac{df}{dr} - \frac{16\pi G}{c^4}E^2\rho f = 0. \tag{40}$$

The metric is now given by

$$ds^2 = f(r)dt^2 + 2Ddtdz - E^2dr^2 - r^2d\theta^2. \tag{41}$$

By rescaling z we can make $D = 1$. At any radius r, the circumference is $2\pi r$. Since space–time is locally Minkowskian, the circle near $r = 0$, given by $r =$ const, is small enough to be locally measurable, and hence we demand that the ratio of the proper circumference to the proper radius is to be 2π; that is, $E = 1$. Similarly, by scaling t, it is reasonable for

$$f = 1 \quad \text{at} \quad r = 0. \tag{42}$$

Equation (40) becomes

$$\frac{d^2f}{dr^2} + \frac{1}{r}\frac{df}{dr} - 4mf = 0, \tag{43}$$

where

$$m(r) = (4\pi G/c^4)\rho(r). \tag{44}$$

For $r > R$, $m = 0$, and we obtain as a solution to Eq. (43)

$$f = A + B\log(r/R). \tag{45}$$

Inside, the beam f varies with the distribution $m(r)$. If the energy density distribution is constant, the interior solution is a linear combination of zero-order Bessel functions of imaginary argument of the first and second kinds:

$$f = aI_0(2m^{1/2}r) + bK_0(2m^{1/2}r).$$

The modified Bessel function of the second kind is singular at $r = 0$; hence we set $b = 0$ and take the solution to be

$$f = I_0(2m^{1/2}r), \quad m = \text{const}, \tag{46}$$

which is consistent with Eq. (42). The exterior solution matches smoothly by choosing the integration constants so as to make

$$f = I_0(2m^{1/2}R) + 2m^{1/2}RI_1(2m^{1/2}R)\log(r/R). \tag{47}$$

In general, our metric is specified by a single function $f(r)$;

$$ds^2 = f(r)dt^2 + 2dtdz - dr^2 - r^2d\theta^2, \tag{48}$$

where $f(r)$ satisfies the equation

$$\frac{d^2f}{dr^2} + \frac{1}{r}\frac{df}{dr} - 4m(r)f = 0 \tag{49}$$

and

$$m(r) = (4\pi G/c^4)\rho(r). \tag{50}$$

Solution of the Einstein field equations for an infinite beam of light is not limited to beams with circular cross section. If the cross section is rectangular, infinite in extent in the y direction and finite in the x direction whose density is a function of x only, we find that

$$ds^2 = f(x)dt^2 + 2dtdz - dx^2 - dy^2, \tag{51}$$

where

$$\frac{d^2f}{dx^2} - 4m(x)f = 0. \tag{52}$$

Outside of the beam, $f = A + Bx$. If $m = \text{const}$, then inside of the beam, $f = \cosh(2m^{1/2}x)$.

If the beam is elliptical, we use confocal coordinates ν and μ, $(-1 < \nu < 1, \mu > 1)$. Let $D = (\frac{1}{2}$ focal distance); then the metric for a beam whose density m is a function of μ only is

$$ds^2 = f(\mu,\nu)dt^2 + 2dtdz - \frac{D^2(\mu^2-\nu^2)}{\mu^2-1}d\mu^2 - \frac{D^2(\mu^2-\nu^2)}{1-\nu^2}d\nu^2, \quad (53)$$

where f is a function of ν as well as μ. Dependence on ν is inserted as a mathematical device, in order that we may obtain a solution in terms of an eigenexpansion. We obtain the equation governing f:

$$(\mu^2-1)\frac{\partial^2 f}{\partial\mu^2} + \mu\frac{\partial f}{\partial\mu} - 4D^2\mu^2 m(\mu)f + (1-\nu^2)\frac{\partial^2 f}{\partial\nu^2} - \nu\frac{\partial f}{\partial\nu} + 4D^2\nu^2 m(\mu)f = 0. \quad (54)$$

Outside of the beam, $f(\mu,\nu) = A + B\log(\mu + (\mu^2-1)^{1/2})$, independent of ν. Inside, if $m(\mu)$ is a constant, by using the constant b, we can separate the equation and obtain the same equation for both variables μ and ν, e.g., our equation for μ becomes

$$(\mu^2-1)\frac{d^2 f}{d\mu^2} + \mu\frac{df}{d\mu} - (4D^2 m\mu^2 + b)f = 0, \quad (55)$$

which is Mathieu's equation with a scaling factor $i2Dm^{1/2}$, that is imaginary. For the interior solution there is no eigensolution for ν which is constant; hence, a summation over the eigenspace is necessary. The same expansion coefficients will then apply to the hyperbolic solution for μ, and the solution f will be independent of ν.

In all of these cases the basic form of the metric is the same. As a consequence, the geodesics of each of the beams will be similar.

GEODESIC EQUATIONS

The solution for the metric from the preceding section yields four nonvanishing Christoffel symbols

$$\Gamma^1_{00} = \frac{1}{2}\frac{df}{dr}, \quad \Gamma^3_{01} = \frac{1}{2}\frac{df}{dr}, \quad \Gamma^1_{22} = -r, \quad \Gamma^2_{12} = r^{-1}. \quad (56)$$

The geodesic equations are given by the expression

$$\frac{dv^k}{ds} + \Gamma^k_{il}v^i v^l = 0, \quad (57)$$

where v^k is the 4-velocity of a test ray or particle. Combining Eqs. (56) and (57) we obtain the four geodesic equations

$$\frac{dv^1}{ds} - rv^2v^2 + \frac{1}{2}\frac{df}{dr}v^0v^0 = 0, \quad (58)$$

$$\frac{dv^2}{ds} + \frac{2}{r}v^1v^2 = 0, \quad (59)$$

$$\frac{dv^3}{ds} + \frac{df}{dr}v^0v^1 = 0, \quad (60)$$

$$\frac{dv^0}{ds} = 0. \quad (61)$$

Solutions for Eqs. (61) and (59) are immediate; they are

$$v^0 = A, \quad (62)$$

$$v^2 = hr^{-2}, \quad (63)$$

where A and h are constants. Inserting these results in the remaining two equations, we may write

$$\frac{dv^1}{ds} - \frac{h^2}{r^3} + \frac{A^2}{2}\frac{df}{dr} = 0 \quad (64)$$

and

$$\frac{dv^3}{ds} + A\frac{df}{dr}v^1 = 0. \quad (65)$$

After multiplying Eq. (64) by v^1, we integrate both equations to obtain

$$v^1 = \pm\,[C - (h^2/r^2) - A^2 f]^{1/2} \quad (66)$$

and

$$v^3 = D - Af, \quad (67)$$

where C and D are constants. In addition, we have the line element

$$B = f(v^0)^2 + 2v^0v^3 - (v^1)^2 - r^2(v^2)^2, \quad (68)$$

where either $B = 0$, if the test particle is null, or $B = 1$, if the test particle is timelike.

In summary, the geodesics equations are given by

$$v^0 = A, \quad (69)$$

$$v^2 = h/r^2, \quad (70)$$

$$v^3 = D - Af, \quad (71)$$

$$v^1 = \pm\,(C - h^2/r^2 - A^2 f)^{1/2}, \quad (72)$$

and the line element Eq. (68) becomes

$$B = 2AD - C, \quad (73)$$

where A, C, D, and h are constants, and B equals zero or one depending on whether the geodesic is null or nonnull, respectively. In the following work we will assume that a light ray or particle may pass through the beam without interaction, save the gravitational interaction. Hence all geodesics will be continued indefinitely.

TRAJECTORIES IN THE R-Z PLANE ($h = 0$)

Null trajectories in r-z plane

We first consider null trajectories ($B = 0$). By scaling the affine parameter such that $D = 1$, we may characterize the geodesics by the constant A so as to rewrite the Eqs. (69)–(72) in the form

$$v^0 = A, \quad (74)$$

$$v^2 = 0, \quad (75)$$

$$v^1 = \pm\,(2A - A^2 f)^{1/2}, \quad (76)$$

$$v^3 = 1 - Af. \quad (77)$$

Equation (76), along with the fact that $f \geq 1$, implies that

$$0 \leq A \leq 2.$$

There are five cases to be considered:

(i) $A = 2$. By Eq. (76) the radicand of v^1 is $4(1-f)$. Since $f \geq 1$, such a trajectory can exist only for $r = 0$ ($f = 1$), in which case the ray propagates along the z axis in the direction opposite to that of the beam. It is neither deflected nor retarded.
(ii) $1 < A < 2$. Consider a ray which starts at $r = 0$, then v^1 is initially positive. For $A > 1, v^3 < 0$ for all r. Hence our ray is inclined towards the negative z direction. As r increases, f increases and v^1 decreases smoothly, becoming zero at $f = 2/A$, where v^3 reaches its largest absolute magnitude $v^3 = -1$. At this point $dv^1/ds \neq 0$, v^1 assumes the negative root of its solution, and the trajectory returns to $r = 0$ along the mirror image of its outgoing path. The complete trajectory is an infinite series of negatively z directed undulations alternating on either side of the z axis, with maximum radius given by $f = 2/A$.
(iii) $A = 1$. At $r = 0$, $f = 1$ and hence $v^3 = 0$; the ray is approaching or departing from the axis radially. As we follow such a ray out, v^3 becomes negative, and the ray is swept back along the negative z axis. It reaches a maximum radius at $f = 2$ and then returns to the z axis along the mirror image path of the outgoing ray. Again, the complete trajectory is a series of such undulations.
(iv) $0 < A < 1$. At $r = 0$, $v^3 > 0$, and the ray is inclined along the beam direction. As the ray goes out, v^1 decreases smoothly to zero at $f = 2/A$, as before. But now v^3 changes sign at $f = 1/A$, and the ray begins to be inclined toward the negative z axis. Hence the outward directed trajectory turns around upon itself (see Fig. 1). The trajectory is a set of lobe-shaped loops which alternate about the z axis.

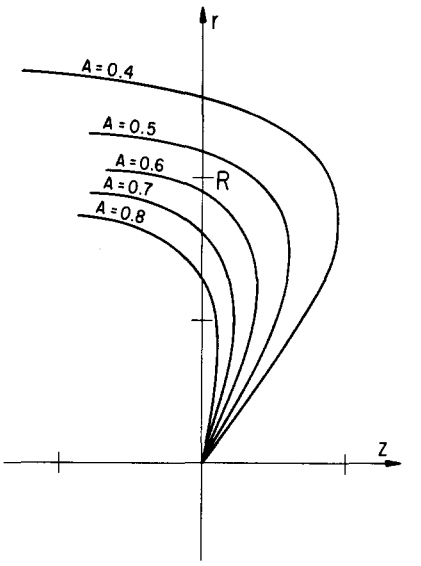


FIG. 1. Outward directed trajectories in the r-z plane for $mR^2 = 1.5$.

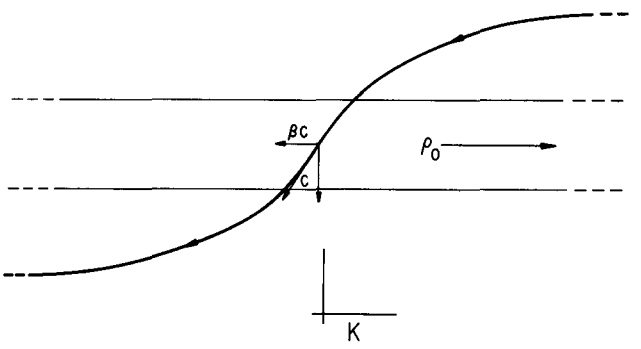


FIG. 2. A null trajectory in the r-z plane as seen in frame K ($\rho = \rho_0$).

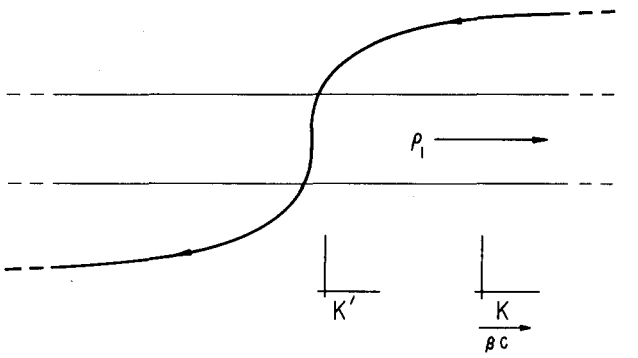


FIG. 3. A null trajectory in the r-z plane as seen in frame K' ($\rho = \rho_1$) moving opposite to the direction of the beam with speed βc with respect to frame K.

Let us consider a test ray which is initially traveling anti-parallel to the beam of density ρ_0 at some nonzero radius. Assuming that the density of the beam is comparatively small, we obtain an undulating trajectory, case (ii). The result is similar to that which one would expect for a particle in a cylindrical Newtonian potential well, as indicated in Fig. 2. If we increase ρ to $\rho_1 > \rho_0$, the corresponding angular deflection at a given radius increases. At a specific ρ_1, the angular deflection at $r = 0$ reaches $\pi/2$; hence, the ray travels perpendicular to the beam axis, and case (iii) applies (see Fig. 3). As the density increases to $\rho_2 > \rho_1$, case (iv) applies. The test ray will reverse upon itself twice in crossing the beam, once on each side of the beam, attaining a maximum deflection from its initial direction at $r = 0$ (see Fig. 4).

This reversing property of the obtained solutions which gives rise to the lobe-shaped loops can be explained in terms of special relativity and the "potential well" solution of case (ii). Consider again the case where $\rho = \rho_0$, i.e., we measure ρ as ρ_0 in some local Minkowski frame K. At the instant the ray crosses the z axis the local speed of the ray must be "c." Hence the component of the test ray's velocity in the direction opposite to the beam must be less than c, say βc (see Fig. 2). Now according to special relativity, there exists no preferred frame for observing the propagation of light. In particular, we may observe the propagation in a frame K' which is moving with respect to the K frame at a velocity βc in the direction opposite to the beam's flux. Such a frame K' exists since $\beta c < c$. In K' the z component of the ray's velocity is zero. Hence an observer in the K' frame sees the test ray traveling perpendicular to the beam at $r = 0$. In addition, the density of the beam is Lorentz transformed in K' relative to K, and we obtain a density $\rho > \rho_0$; namely, $\rho = \rho_1$, as expected from case (iii) (see Fig. 3). It is similarly

TABLE I. The z coordinate at the maximum radial position for null trajectories in the $r-z$ plane as a function of beam flux (mR^2) and maximum radius (r/R). For those cases where the maximum radial position is at positive z, a loop will occur in the extended trajectory. Since, for a given beam, these values of z go through a maximum, only for those beams where $mR^2 \geq 2.57$ does z become positive.

r/R	z				
	$mR^2 = 2.20$	$mR^2 = 2.50$	$mR^2 = 2.57$	$mR^2 = 2.60$	$mR^2 = 3.00$
0.97.....	−.1668	−.0860	−.0672	−.0600	+.0415
1.02.....	−.1191	−.0300	−.0094	−.0013	+.1132
1.03.....	−.1158	−.0260	−.0052	+.0036	+.1201
1.04.....	−.1139	−.0226	−.0014	+.0069	+.1244
1.05.....	−.1134	−.0207	−.0002	+.0087	+.1280
1.06.....	−.1143	−.0212	+.0006	+.0091	+.1300
1.07.....	−.1155	−.0221	−.0001	+.0090	+.1296
1.08.....	−.1181	−.0234	−.0021	+.0075	+.1286
1.09.....	−.1219	−.0261	−.0045	+.0046	+.1272
1.10.....	−.1261	−.0301	−.0083	+.0014	+.1243
1.15.....	−.1564	−.0604	−.0385	−.0293	+.0955

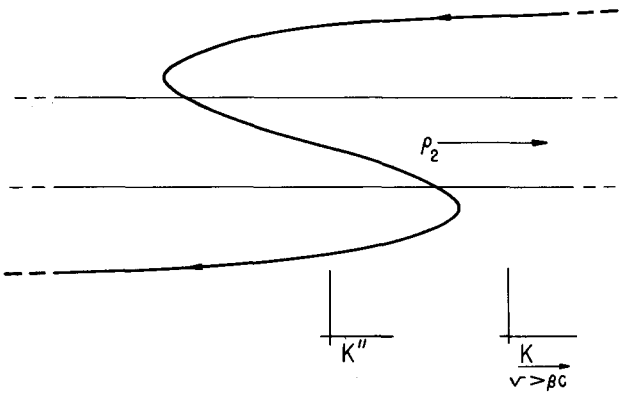


FIG. 4. A null trajectory in the $r-z$ plane as seen in frame K'' ($\rho = \rho_2$) moving opposite to the direction of the beam with speed greater than βc with respect to the frame K.

true that the case where the trajectory turns around, shown in Fig. 4, can be generated by transforming to a frame K'' moving at a speed greater than βc with respect to K in the negative z direction. The density of the beam is then Lorentz transformed to ρ_2.

(v) $A = 0$. In this case $v^1 = 0$ and $v^3 = 1$, the ray travels parallel to and in the same direction as the beam. The ray is not deflected by the beam. Since the beam is itself composed of such rays, the ray is in equilibrium and will neither focus itself nor retard its own propagation. This conclusion holds for the elliptical and rectangular cross sections as well. Thus we answer the question raised in the Introduction. The exact solution we have produced is in agreement with the approximate solution of Tolman, Ehrenfest, and Podolsky.

With the occurrence of the lobe-shaped loops of case (iv), one may ask if it is possible to have a complete loop; that is, a closed null trajectory in the $r-z$ plane. Table I gives a locus of maximum radii positions (maximum r, z) for trajectories which start at $(r, z) = (0, 0)$ for several different beams. As can be seen from Table I and from Fig. 1, for a given beam the range of values of z at the maximum radius positions has a maximum. For our cases, $mR^2 = 2.2$ and 2.5, this value is at $r_{\max} = 1.05R$, and for the remaining three choices this value is at $r_{\max} = 1.06R$. In particular, at $mR^2 = 2.57$ the maximum value of z is zero. Since the trajectory is mirrored about the maximum r position, this trajectory will come back to the axis at $(r, z) = (0, 0)$; a closed loop will result. Closed loops can occur for $mR^2 > 2.57$ since the

TABLE II. Considering the set of null trajectories in the $r-z$ plane which start from $(r, z) = (0, 0)$, the times and radial coordinates at which these trajectories reintersect $z = 0$ are tabulated. For $mR^2 = 3.5$, an observer at $(r, z) = (1.8R, 0)$, say, will see the effect of radiation scattered from $(r, z) = (0, 0)$ before an observer who is situated at $(r, z) = (0.9R, 0)$ will, even though the second observer is between the source and the first observer.

r/R	T		
	$mR^2 = 1.5$	$mR^2 = 2.5$	$mR^2 = 3.5$
0.30.....	0.298	0.285	0.281
0.40.....	0.384	0.373	0.366
0.50.....	0.469	0.451	0.430
0.60.....	0.545	0.515	0.483
0.70.....	0.615	0.567	0.518
0.80.....	0.680	0.604	0.537
0.90.....	0.730	0.629	0.541
1.00.....	0.769	0.640	0.530
1.10.....	0.798	0.642	0.517
1.20.....	0.827	0.645	0.510
1.30.....	0.850	0.650	0.507
1.40.....	0.875	0.657	0.508
1.50.....	0.899	0.666	0.511
1.60.....	0.923	0.677	0.516
1.70.....	0.947	0.688	0.522
1.80.....	0.972	0.701	0.529
1.90.....	0.997	0.714	0.537
2.00.....	1.021	0.728	0.545
2.10.....	1.044	0.742	0.554
2.20.....	1.069	0.756	0.563
2.30.....	1.095	0.770	0.573
2.40.....	1.120	0.785	0.583
2.50.....	1.145	0.800	0.593

locus of the positions goes to positive z. For a given mR^2 we will call $\bar{r}$ the radius at which z, at maximum r, takes its maximum value. Then for $mR^2 > 2.57$, there exists two radii r_1 and r_2, with $r_1 < \bar{r}$ and $\bar{r} < r_2$ such that closed loops occur; that is, an infinitely repeatable figure eight trajectory in the $r-z$ plane centered on $(r, z) = (0, 0)$. For maximum r such that $r_1 < r_{\max} < r_2$, the trajectories will intersect themselves before coming back to $r = 0$. In these cases, succeeding loops will occur at greater z and the trajectory will progress in the direction of the beam.

Other effects are connected with the region centered about $\bar{r}$. Consider Table II in which we tabulate the times of signal reception at $(r, z) = (r, 0)$ for rays which started at $(0, 0)$. There is a plateau in reception time for $mR^2 < 2.57$ and an actual dip in reception time for $mR^2 > 2.57$. Two observers, O_1 and O_2, situated at $(r_1, 0)$ and $(r_2, 0)$, respectively, with $r_2 > r_1$ can observe a burst of scattered light which originates at some inner radius $(r, 0)$. It is possible that

O_2 will receive information of this burst before O_1 does, even though O_1 is between O_2 and the source. The drastic curvatures of the looping null rays account for this effect.

Particle trajectories in the r-z plane

In the case of particle trajectories, the quantity B in Eqs. (69) to (73) is now equal to one. Since we do not have an affine parameter to scale, the constant D is free. With an extra degree of freedom, which amounts to picking a speed for our test particle, we can ensure a figure eight loop for a trajectory which starts on the axis for all values of mR^2 and all values of the constant A. In addition, all observers can now throw a projectile which they can catch without moving; that is, looping is possible for all radii and not just for those centered about the $\bar{r}$ of the previous section. All other possible trajectories can be obtained by an appropriate Lorentz transformation. This includes trajectories which have cusps at their maximum radii alternating about the beam and progressing in the direction of the beam's flux; that is, the case of an object dropped into the beam from rest.

TRAJECTORIES WITH ROTATION ($h \neq 0$)

Helical null trajectories

In the Schwarzschild solution for a spherically symmetric source, there exists a region near the Schwarzschild radius at which null orbital trajectories are possible. The question may be asked as to whether similar null orbitals are possible for the light beam. At the moment we shall restrict our concern to circular orbits and those which spiral around the beam in a helix of constant radius. We previously derived the geodesic equations (69)–(73). At the moment we will use Eqs. (69) and (70) and the differentiated forms of Eqs. (71) and (72), viz., Eqs. (58) and (60):

$$v^0 = A, \tag{69}$$

$$v^2 = \frac{h}{r^2}, \tag{70}$$

$$\frac{dv^3}{ds} = -A^2 \frac{df}{dr} v^1, \tag{58'}$$

$$\frac{dv^1}{ds} = \frac{h^2}{r^3} - \frac{A^2}{2} \frac{df}{dr}. \tag{60'}$$

Since, for trajectories of constant radius, $v^1 = 0$, Eq. (60) becomes

$$\frac{h^2}{r^3} = \frac{A^2}{2} \frac{df}{dr}; \tag{78}$$

similarly Eq. (58') yields

$$\frac{dv^3}{ds} = 0$$

or

$$v^3 = Q, \tag{79}$$

a constant. The condition that v^1 be null requires that $v_i v^i = 0$; hence

$$0 = fA^2 - h^2/r^2 + 2QA.$$

By virtue of Eq. (78) this equation becomes

$$0 = A\left[\left(f - \frac{r}{2}\frac{df}{dr}\right)A + 2Q\right]. \tag{80}$$

Equation (80) has two solutions. The first, $A = 0$, requires that $h = 0$ and describes a forward directed light ray which does not orbit [see Eqs. (74)–(77) and case (v) of that section]. The second solution is

$$Q = -\frac{1}{2}A\left(f - \frac{1}{2}r\frac{df}{dr}\right). \tag{81}$$

Closed circular orbits can exist only for $Q = 0$. Since we have assumed that $A \neq 0$, the vanishing of Q implies the following condition on f:

$$f = \frac{1}{2}r\frac{df}{dr}. \tag{82}$$

As a specific example we consider the interior and exterior solutions for a beam of constant m.

INTERIOR: $r < R$. The interior solution is specified by

$$f = I_0(2m^{1/2}r), \tag{46}$$

$$\frac{df}{dr} = 2m^{1/2}I_1(2m^{1/2}r), \tag{83}$$

which yields, as the form of Eq. (81),

$$Q = -\tfrac{1}{2}A[I_0(2m^{1/2}r) - m^{1/2}rI_1(2m^{1/2}r)]. \tag{84}$$

For the closed circular orbits, Eq. (82) requires

$$I_0(2m^{1/2}r) = m^{1/2}rI_1(2m^{1/2}r). \tag{85}$$

This condition can be satisfied for the r', where

$$m(r')^2 = 1.669. \tag{86}$$

For r such that $mr^2 < 1.669$, $Q < 0$ and the orbit spirals toward the rear; on the other hand for $mr^2 > 1.669$, $Q > 0$ and the orbit spirals forward.

EXTERIOR: $r > R$. We recall that the exterior solution is given by

$$f(r) = I_0(2m^{1/2}R) + 2m^{1/2}RI_1(2m^{1/2}R)\log(r/R), \tag{47}$$

$$\frac{df}{dr} = \frac{2m^{1/2}R}{r} I_1(2m^{1/2}R), \tag{87}$$

so that Q has the form

$$Q = -\tfrac{1}{2}A[W + 2m^{1/2}RI_1(2m^{1/2}R)\log(r/R)], \tag{88}$$

where

$$W = I_0(2m^{1/2}R) - m^{1/2}RI_1(2m^{1/2}R).$$

Now the $\log(r/R)$ for $r > R$ is always positive. The quantity W in Eq. (88) is the same as the quantity previously discussed in the interior case, Eq. (84). Hence if $mR^2 = 1.669$, then $Q = 0$ for $r = R$ and a circular orbit exists at the beam's radius R. If $mR^2 < 1.669$, then the quantity W is positive, Q is negative for all $r > R$, and all orbits spiral to the rear. If $mR^2 > 1.669$, the quantity W is negative and there exists an r^* such that if $r = r^*$, $Q = 0$ and a circular orbit exists at $r = r^*$; if $r < r^*$, $Q > 0$ and

the orbit spirals forward; or if $r > r^*$, $Q < 0$ and the orbit spirals to the rear.

The results may be summarized:

(1) $mR^2 < 1.669$.
$Q < 0$ for all r.
The helical orbits all spiral to the rear.
No closed circular orbits exist.

(2) $mR^2 = 1.669$.
Circular orbits exist at $r = R$ only.
For $r \neq R$, $Q < 0$ and the orbits spiral to the rear.

(3) $mR^2 > 1.669$.
There exist two radii at which circular orbits are possible.
The first is $r' < R$, such that $m(r')^2 = 1.669$.
The second is $r^* > R$, given by $Q = 0$ in Eq. (88).
For $r < r'$, $Q < 0$ and the orbits spiral to the rear.
For $r' < r < r^*$, $Q > 0$ and the orbits spiral forward.
For $r^* < r$, $Q < 0$ and the orbits spiral to the rear.

Note that for all values of mR^2 there are helical orbits spiraling backward. For $mR^2 > 1.669$ there is a region situated about $r = R$ where helical orbits spiraling forward exists.

These results are consistent with special relativity. Were forward spiraling helices to exist regardless of the density of the beam, there would exist a new Minkowski frame moving with respect to the initial frame in the direction of propagation of the beam in which the energy density of the beam is even less than it was in the initial frame. Furthermore, by picking the new Minkowski frame properly, the orbit would close up to become a circle. Thus we would obtain the absurd result that it is possible for light to orbit around a beam of light produced by two flashlight batteries.

Particle orbits

We shall consider circular timelike orbits. For a circular orbit $d^2r/ds^2 = 0$; hence, Eq. (64) becomes

$$-\frac{1}{2}A^2\frac{df}{dr} + \frac{h^2}{r^3} = 0. \tag{89}$$

The quantities v^1 and v^3 are also zero. Combining Eq. (89) with Eqs. (71) and (72) and the line element condition (73) (with $B = 1$), we obtain

$$h^2 = \left(r^3\frac{df}{dr}\right)\Big/\left(2f - r\frac{df}{dr}\right), \tag{90}$$

$$A^2 = 2\Big/\left(2f - r\frac{df}{dr}\right). \tag{91}$$

The quantity h^2 becomes infinite at r', given by

$$f(r') = \frac{r'}{2}\frac{df(r')}{dr}, \tag{92}$$

which is the condition previously derived for circular null geodesics, Eq. (82). If there exists both an r'_1 and an r'_2, $r'_2 > r'_1$, such that Eq. (92) is satisfied, then h^2 is negative for all r such that $r'_1 < r < r'_2$. In this region no circular orbits will exist.

Stability of circular orbits

Differentiation of h^2 with respect to r leads to

$$\frac{dh^2}{dr} = \left[-4r^3\Big/\left(2f - r\frac{df}{dr}\right)^2\right]\left[\left(\frac{df}{dr}\right)^2 - \frac{1}{r}f\frac{df}{dr} - 2mf^2\right]. \tag{93}$$

Outside the beam, dh^2/dr changes sign at $\tilde{r}$, where

$$f(\tilde{r}) = \tilde{r}\frac{df(\tilde{r})}{dr}. \tag{94}$$

We note here for later reference that inside the beam, one may show that dh^2/dr is always positive.

In order to calculate the stability of a circular orbit or radius r, we will go to a neighboring orbit, at a radius $r + \Delta r$, which has the same h and which also has dz/ds and dr/ds initially zero. By calculating whether the initial acceleration $d^2\Delta r/ds^2$ is toward or away from the original circular orbit, we determine its stability. From Eqs. (72) and (71)

$$\left(\frac{dr}{ds}\right)^2 = A^2f_I + \frac{h^2}{r_I^2} - A^2f - \frac{h^2}{r^2}, \tag{95}$$

$$\frac{dz}{ds} = Af_I - Af, \tag{96}$$

where "I" indicates initial values for a given orbit. By the line element (73)

$$A^2 = \frac{1 + h^2/r_I^2}{f_I}; \tag{97}$$

this is more general than Eq. (91) which holds for circular orbits only. In going from the circular orbit at r to the new orbit at $r + \Delta r$, A^2 changes from its value in the circular orbit A_0^2, since r_I and f_I change

$$A^2 = A_0^2 + \Delta A^2,$$

where

$$\Delta A = -\frac{A_0^2}{f_I}\frac{df_I}{dr}\Delta r - \frac{2h^2}{f_I r_I^3}\Delta r,$$

or using the circular orbit values for A_0 and h, Eqs. (90) and (91),

$$\Delta A = -4\left[f\left(2f - r\frac{df}{dr}\right)\right]^{-1}\frac{df}{dr}\Delta r. \tag{98}$$

Equation (64), d^2r/ds^2 suffers a change

$$\frac{d^2\Delta r}{ds^2} = -\frac{\Delta A^2}{2}\frac{df}{dr} - \frac{A^2}{2}\frac{d^2f}{dr^2}\Delta r - \frac{3h^2}{r^4}\Delta r$$

or

$$\frac{d^2\Delta r}{ds^2} = 2\left[f\left(2f - r\frac{df}{dr}\right)\right]^{-1}\left[\left(\frac{df}{dr}\right)^2 - \frac{f}{r}\frac{df}{dr} - 2mf^2\right]\Delta r.$$

But dh^2/dr, as given by Eq. (93), implies

$$\frac{d^2\Delta r}{ds^2} = \left[\left(2f - r\frac{df}{dr}\right)\Big/2r^2f\right]\frac{dh^2}{dr}\Delta r. \tag{99}$$

Therefore, wherever circular orbits exist, the orbits are stable for $dh^2/dr > 0$ and unstable for $dh^2/dr < 0$.

We plot the three possible cases of h^2 versus r, where R is the beam's radius.

(i) The quantity $f(df/dr) < f$ for all r and Graph A of Fig. 5 applies. With h^2 and dh^2/dr always positive, circular orbits are possible at all r, and all such orbits are stable. Light will not be able to occupy a circular orbit in this case. If $m(r)$ were a constant, this case would be described by

$$mR^2 < 0.645.$$

(ii) If there exists a region such that $r(df/dr) > f$, but $r(df/dr) < 2f$ for all r, we get Graph B as seen in Fig. 5. Circular orbits are possible for all r. But for r between R and the value of r when $f = r(df/dr)$, $dh^2/dr < 0$ and the circular orbits are unstable. If $m(r)$ were constant, this case would be described by

$$0.645 < mR^2 < 1.669.$$

(iii) If there exists a region where $r(df/dr) > 2f$, we obtain Graph C as seen in Fig. 5. The region bounding R where $r(df/dr)$ is greater than $2f$ has h^2 negative and, hence, has no circular orbits. This region is bounded by two radii where null orbits are possible. Outside of the beam the null orbital is encompassed by a region where $dh^2/dr < 0$ and the orbits are unstable. All other circular orbits are stable. If $m(r)$ were a constant, this case would be described by

$$mR^2 > 1.669.$$

Noncircular orbits, which start outside of the beam and enter into the region of instability, spiral down until they reach the inner region of stability. Here they reach their minimum radii and then spiral out once again. In this case there are two orbits inside of the beam which have the same minimum radius and the same angular velocity. One is unstable and will spiral out; the other is stable and remains comparatively close to its minimum radius. The feature distinguishing the two is the z velocity associated with the orbit. If both started with $v^3 = 0$ at their maximum radius, the unstable orbit will attain the larger velocity parallel to the z axis as compared to that attained by the stable orbit.

One may naturally ask as to the practical nature of a gravitational field due to a beam of light. The energy density ρ is scaled in the field equations in the form mR^2, i.e., $10^{-48}\rho R^2$ (*CGS*) with R being the radius of the beam. The field of the beam would be measurable if this scaling factor were on the order of unity. The factor of 10^{-48} is, however, a formidable one to overcome. For example, if we could construct a laser, of cross section 1 cm^2, whose output equaled the solar constant, i.e., 4×10^{33} ergs/sec, we would need enough such lasers so as to fill out a beam of fifty million kilometers diameter. This, unfortunately, is not practical. Astronomically we may do a bit better if we assume that a beam of light could have the same energy flux and size as that of the plasma jet associated with *M*87. The total energy in particles and field for the jet[10] is about 2×10^{55} ergs and assuming that the jet propagated from *M*87 at its velocity of turbulence, 1500 km/sec, we get an energy output of 10^{42} ergs/sec. Ascribing this output to a beam of light, our scaling factor would still

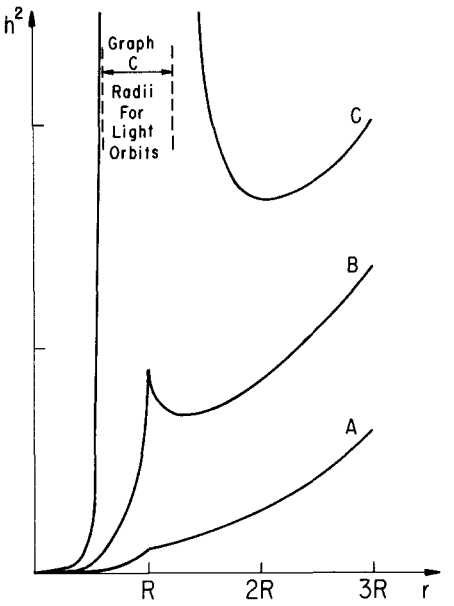


FIG. 5. The square of the angular momentum per unit mass h^2 for a circular orbit situated at radius r. Graph A: $mR^2 = 0.25$. Graph B: $mR^2 = 1.0$. Graph C: $mR^2 = 4.0$. Stable timelike orbits are allowed for dh^2/dr positive. For dh^2/dr negative, the orbits are unstable. In Graph C we have two radii at which null orbits are possible. Between these two positions no circular orbits can occur.

be on the order of 10^{-16}. Although one may not derive any comfort from such magnitudes, or lack of them, the solution and its simplicity in terms of special relativity is rather satisfying.

A subsequent paper will treat first order perturbations of the beam and neighboring solutions.

ACKNOWLEDGMENTS

In the course of writing this paper, many individuals were of assistance and unfortunately too many to mention adequately. I wish to primarily thank William Barletta, Philip Greenberg, Sol Krasner, and Roland Winston for their kind support. I also wish to thank my thesis committee consisting of Roland Winston, S. Chandrasekhar, Peter Freund, and Melba Phillips for their assistance.

---

*Submitted as a thesis to the Department of Physics, The University of Chicago, in partial fulfillment of the requirements for the Ph.D. degree.

[1]R. C. Tolman, P. Ehrenfest, and B. Podolsky, Phys. Rev. 37, 602 (1931).
[2]R. C. Tolman, *Relativity, Thermodynamics and Cosmology* (Clarendon, Oxford, 1934), pp. 272-88.
[3]P. C. Vaidya, Nature 171, 260 (1953).
[4]H. Bondi, Nature 186, 535 (1960).
[5]H. Bondi, M. G. van der Burg, and A. W. K. Metzner, Proc. Roy. Soc. (London) 269A, 21 (1962).
[6]R. Sachs, Proc. Roy. Soc. (London) 270A, 103 (1962).
[7]H. Bondi, Proc. Roy. Soc. (London) 281A, 39 (1964).
[8]R. W. Lindquist, R. A. Schwartz, and C. W. Misner, Phys. Rev. 137, B1364 (1965).
[9]W. J. Kaufmann III, Astrophys. J. 153, 849 (1968).
[10]G. Burbridge, Astrophys. J. 124, 416 (1956).

The Clebsch–Gordan coefficients of $SU(3)$ and the orthogonalization problem

S. Datta Majumdar and Biplab Kumar Basu

Department of Physics, Indian Institute of Technology, Kharagpur, India
(Received 14 July 1971; revised manuscript received 18 February 1972)

The Clebsch–Gordan coefficients of the group $SU(3)$ are determined by integrating the product of three matrix elements of finite transformations belonging to three irreducible representations of the group. Compact expressions involving a single or a double sum over products of 3-j and 6-j symbols of $SU(2)$ are obtained for several different classes of coefficients by suitably restricting the initial states but keeping the final states of the matrix elements arbitrary. To orthogonalize the CG coefficients, a linear combination of several integrals with the same final but different initial states is taken. The coefficients of the linear combination are determined by the Schmidt procedure and are found to be expressible in terms of integrals of the same type.

1. INTRODUCTION

If the representations of finite transformations of a compact Lie group are known, the Clebsch–Gordan coefficients (CGC) can be obtained from the relation[1]

$$I\begin{pmatrix}\rho_1 & \rho_2 & \rho\\ \nu_1\lambda_1 & \nu_2\lambda_2 & \nu\lambda\end{pmatrix} \equiv \int d\Omega D^{\rho_1}_{\nu_1\lambda_1}(\alpha)D^{\rho_2}_{\nu_2\lambda_2}(\alpha)D^{\rho *}_{\nu\lambda}(\alpha)$$
$$= \frac{\int d\Omega}{\mathfrak{N}}\sum_{\gamma}\begin{pmatrix}\rho_1 & \rho_2 & \rho_\gamma\\ \nu_1 & \nu_2 & \nu\end{pmatrix}\begin{pmatrix}\rho_1 & \rho_2 & \rho_\gamma\\ \lambda_1 & \lambda_2 & \lambda\end{pmatrix} \quad (1)$$

by integrating the product of three matrix elements (m.e.) of finite transformations belonging to three irreducible representations (IR's) of the group. Here, ρ_1, ρ_2, ρ characterize the IR's, $\mathfrak{N}$ is the dimensionality of ρ, and $\lambda_1\lambda_2\lambda$, $\nu_1\nu_2\nu$ denote the initial and the final states of the m.e. Although relation (1) is given in standard text books, it does not appear to have been used for the actual calculation of the CGC. However, the use of this relation not only gives simple and convenient formulas for the CGC but also leads to a satisfactory solution of the orthogonalization problem in the case when a particular IR appears more than once in the reduction of a direct product.

For finding the exact form of the representation matrices we need a set of parameters for the group. In the case of the group $SU(3)$, with which alone we are concerned here, a set of parameters was obtained by Nelson[2] by writing a general element $\mathcal{S}$ in the form

$$\mathcal{S} = e^{-i\beta Y}e^{-i\alpha_3 T_3}e^{-i\alpha_2 T_2}e^{-i\gamma T_3}e^{-i\nu N}e^{i\gamma' T_3}e^{i\alpha_2' T_2}e^{i\alpha_3' T_3}, \quad (2)$$

where

$$T = \tfrac{1}{2}\begin{pmatrix}\sigma & & \\ & & \\ & & 0\end{pmatrix}, \quad N = \begin{pmatrix}0&0&1\\0&0&0\\1&0&0\end{pmatrix}, \quad Y = \tfrac{1}{3}\begin{pmatrix}1&0&0\\0&1&0\\0&0&-2\end{pmatrix}.$$

These parameters are directly related to but are more convenient than those given a long time ago by Murnaghan,[3] and form the basis of the present investigations. Since the basic states of an IR of $SU(3)$ are simultaneous eigenstates of T^2, T_3, Y with the eigenvalues,[4] $j(j+1)$, $m, 2\delta - \frac{2}{3}(p-q)$, respectively, the m.e. of $\mathcal{S}$ must have the form

$$\langle j\mu\delta\,|\,\mathcal{S}\,|\,j'\mu'\delta'\rangle = \sum_{\bar{j},m} e^{-i\beta Y}\mathfrak{D}^{j}_{\mu m}(-\alpha_3,-\alpha_2,-\gamma)$$
$$\times\langle \bar{j}\bar{m}\bar{\delta}\,|\,K\,|\,jm\delta\rangle(i)^{\bar{m}'-\bar{m}}D^{\bar{j}}_{\bar{m}\bar{m}'}(-2\nu)$$
$$\times\langle \bar{j}\bar{m}'\bar{\delta}\,|\,K\,|\,j'm'\delta'\rangle\mathfrak{D}^{j'}_{m'\mu'}(\gamma',\alpha_2',\alpha_3'), \quad (3)$$

where $K = ({}^{1}{}_{i\sigma_2})$, and $\mathfrak{D}^{j}_{m',m}(\alpha,\beta,\gamma) = e^{-im'\alpha}D^{j}_{m',m}(\beta)\,e^{-im\gamma}$ are rotation matrices. The m.e. of K, the only unfamiliar quantities occurring in this expression, can be determined by applying the transformation K to a basic state and expanding the result in a series of basic states. The expansion coefficients, which are identical with the m.e., are found to take the form[5]

$$\langle \bar{j}\bar{m}\bar{\delta}\,|K|\,jm\delta\rangle = (-)^{2\delta}(2\bar{j}+1)^{1/2}(2j+1)^{1/2}$$
$$\times\begin{Bmatrix}\bar{j} & \tfrac{1}{2}(p-2\delta) & \tfrac{1}{2}(q+\delta+m)\\ j & \tfrac{1}{2}(p-2\bar{\delta}) & \tfrac{1}{2}p\end{Bmatrix}, \quad (4)$$

where the quantity within the double braces is a 6-j symbol of $SU(2)$. The m.e. of K and of $\exp(-i\nu N)$ have nonzero values only when

$$\begin{aligned} 2\bar{m} &= -p+q+3\delta+m,\\ 2\bar{\delta} &= p-q-\delta+m,\\ \delta-m &= \delta'-m'. \end{aligned} \quad (5)$$

Combining Eqs. (3) and (4), we get the general form of the m.e. of finite transformations in an arbitrary IR of the group. Substituting this in Eq. (1) and performing the integrations, we get the value of the "scalar product" of the CGC on the rhs of the equation. In the nondegenerate case, the CGC $\left(\begin{smallmatrix}\rho_1\rho_2\rho\\ \nu_1\nu_2\nu\end{smallmatrix}\right)$ carrying the labels of the final states is given by the ratio $I(\nu_1\lambda_1, \nu_2\lambda_2, \nu\lambda)/I(\lambda_1\lambda_1, \lambda_2\lambda_2, \lambda\lambda)^{1/2}$. But, if degeneracy is present, then the integrals, as such, give us only a set of nonorthogonal CGC. To orthogonalize them, we have to form suitable linear combinations of integrals with the same final but different initial states. If orthogonalization is carried out by the Schmidt process, then the coefficients of the linear combinations also involve integrals of the same type. Thus, the problem of calculating the CGC of a compact group reduces to the evaluation of certain integrals. In the present paper we evaluate these integrals for the group $SU(3)$, keeping the final states arbitrary but choosing initial states for which the results assume particularly simple forms. The orthogonalization problem is discussed in Sec. 4 of the paper.

First, we choose initial states for which $j_1' = j_2' = j' = 0$. This choice leads to a very simple formula for the CGC if Nelson's form of the special m.e. $\langle j\mu\delta\,|\,\mathcal{S}\,|\,000\rangle$ is used in the calculation. However, the simplicity is gained at the cost of generality. Since p_1q_1, p_2q_2, pq are connected by the relation

$$3(\delta-\delta_1-\delta_2) = (p-p_1-p_2)-(q-q_1-q_2), \quad (6)$$

Copyright © 1973 by the American Institute of Physics

the simultaneous vanishing of the three δ's implies that p and q can take only the values $p = p_1 + p_2 - \lambda, q = q_1 + q_2 - \lambda$. Coefficients of this kind will be called "semistretched" CGC for all admissible values of λ.

Next, we choose initial states for which $j' = \delta_1'$, $j_2' = \delta_2'$, $j' = \delta'$, and use the compact form of the m.e. of $\exp(-i\nu N)$ given in Sec. 3 of II. No loss of generality is incurred by this choice if the j's are kept otherwise arbitrary. If, however, they are given the special values $j_1' = p_1/2$, $j_2' = p_2/2$, $j' = p/2$, then the ${}_3F_2$ series in the expressions for the m.e. reduce to unity giving again a simple formula for the CGC. As in the previous case, the choice of the initial states puts a restriction on the p, q values of the product representation. From Eq. (6) it follows that p and q can now take only the values $p = p_1 + p_2 - 2r$, $q = q_1 + q_2 + r$. The domain of permissible values of p and q can be extended considerably by using the symmetry properties of the CGC and by taking $j_1' = \frac{1}{2}(p_1 - n_1)$, $j_2' = \frac{1}{2}(p_2 - n_2)$, $j' = \frac{1}{2}p$, where n_1, n_2 are integers.

We see that convenient expressions for the CGC (or, rather, the integrals) can be obtained by suitably restricting the initial states and using special forms of the m.e. of finite transformations. However, for obtaining CGC of all kinds and for carrying out the orthogonalization program, we may also require the general expression with no restriction imposed on the initial and final states. The expression is easily derived by the procedure outlined in I.

In course of these investigations[6] it has been frequently necessary to transform a Saalschützian ${}_4F_3(1)$ series into a 6-j symbol. This question is discussed in Appendix A and the conditions under which such a transformation is possible are clearly formulated. Another result, which has also been used, is the expansion of $(\csc\nu)^{\lambda-2}D^{j}_{\mu',\mu}(-2\nu)$ in a series of $D^{J}_{m',m}(-2\nu)$ (for $\mu' + \mu = m' + m$). This expansion is worked out in Appendix B. It is found to take an elegant form with 6-j symbols occurring as coefficients of $D^{J}_{m',m}(-2\nu)$.

2. THE 'SEMISTRETCHED' CG COEFFICIENTS

We now evaluate the integral in the "semistretched" case using Nelson's[2] form of the m.e. $\langle jm\delta|\exp(-i\nu N)|000\rangle$. For $j_1' = j_2' = j' = 0$ the integral contains only the five variables, $\beta, \alpha_3, \alpha_2, \gamma$, and ν. As in I, integration over the first four variables yields a product of two CGC of $SU(2)$. But, the integral over ν now takes the form

$$\int D^{\frac{1}{2}(p_1+q_1+1)}_{\frac{1}{2}(q_1-p_1)+\delta_1+j_1+\frac{1}{2},\frac{1}{2}(q_1-p_1)+\delta_1-j_1-\frac{1}{2}}(-2\nu) \times \left(\csc\nu D^{\frac{1}{2}(p_2+q_2+1)}_{\frac{1}{2}(q_2-p_2)+\delta_2+j_2+\frac{1}{2},\frac{1}{2}(q_2-p_2)+\delta_2-j_2-\frac{1}{2}}(-2\nu)\right) \times D^{\frac{1}{2}(p+q+1)}_{\frac{1}{2}(q-p)+\delta+j+\frac{1}{2},\frac{1}{2}(q-p)+\delta-j-\frac{1}{2}}(-2\nu)\sin 2\nu d\nu. \tag{7}$$

This integral can be evaluated by expanding the function inside the square brackets in (7) in a series of the form

$$\sum_J G_J D^{J}_{\frac{1}{2}(q_2-p_2)+\delta_2+j-j_1,\frac{1}{2}(q_2-p_2)+\delta_2-j+j_1}(-2\nu), \tag{8}$$

using the relation

$$\tfrac{1}{2}(2j+1)\int_0^{\pi} D^{j_1}_{m_1m_1'}(\alpha)D^{j_2}_{m_2m_2'}(\alpha)D^{j}_{mm'}(\alpha)\sin\alpha d\alpha = \begin{Bmatrix} j_1 & j_2 & j \\ m_1 & m_2 & m \end{Bmatrix}\begin{Bmatrix} j_1 & j_2 & j \\ m_1' & m_2' & m' \end{Bmatrix}. \tag{9}$$

The expansion coefficients are given by Eq. (B7) of Appendix B. In the present case they reduce to

$$G_J = (-)^{J+\frac{1}{2}(p_2+q_2)+j-j_1-j_2+1}(2J+1)\begin{Bmatrix} \frac{1}{2}(q_2+\delta_2+j-j_1) & J & \frac{1}{2}(p_2-\delta_2-j+j_1) \\ \frac{1}{2}(p_2-\delta_2+j-j_1) & j_2 & \frac{1}{2}(q_2+\delta_2-j+j_1) \end{Bmatrix}. \tag{10}$$

Using this expression for G_J and integrating the series of triple products of D functions term by term, we have

$$\sum_\gamma \begin{pmatrix} p_1q_1 & p_2q_2 & pq_\gamma \\ j_1\delta_1 & j_2\delta_2 & j\delta \end{pmatrix}\begin{pmatrix} p_1q_1 & p_2q_2 & pq_\gamma \\ 00 & 00 & 00 \end{pmatrix} = \sum_J (-)^{\frac{1}{2}(p_2+q_2)+2j+J}(2J+1) \times \left(\frac{(2j_1+1)(2j_2+1)(p+1)(q+1)}{(2j+1)(p_1+1)(q_1+1)(p_2+1)(q_2+1)}\right)^{1/2}\begin{Bmatrix} j_1 & j_2 & j \\ \delta_1 & \delta_2 & \delta \end{Bmatrix}\begin{Bmatrix} \frac{1}{2}(q_2+\delta_2+j-j_1) & J & \frac{1}{2}(p_2-\delta_2-j+j_1) \\ \frac{1}{2}(p_2-\delta_2+j-j_1) & j_2 & \frac{1}{2}(q_2+\delta_2-j+j_1) \end{Bmatrix} \times \begin{Bmatrix} \frac{1}{2}(p_1+q_1+1) & J & \frac{1}{2}(p+q+1) \\ \frac{1}{2}(q_1-p_1+1)+j_1+\delta_1 & \frac{1}{2}(q_2-p_2)+j-j_1+\delta_2 & \frac{1}{2}(q-p+1)+j+\delta \end{Bmatrix} \times \begin{Bmatrix} \frac{1}{2}(p_1+q_1+1) & J & \frac{1}{2}(p+q+1) \\ \frac{1}{2}(q_1-p_1-1)-j_1+\delta_1 & \frac{1}{2}(q_2-p_2)-j+j_1+\delta_2 & \frac{1}{2}(q-p-1)-j+\delta \end{Bmatrix}. \tag{11}$$

The expression gives only CGC of a particular class for which $p = p_1 + p_2 - \lambda, q = q_1 + q_2 - \lambda$.

3. CG COEFFICIENTS OF OTHER KINDS

Although after the evaluation of the integral (1) for arbitrary initial and final states the problem of calculating the CGC[6-13] of $SU(3)$ may be regarded as solved, the inherent complexity of the resulting expression makes any such calculation interminably long and cumbersome. To obtain simpler expressions, we keep the final states of the m.e. arbitrary but choose initial states for which $j_1' = \delta_1'$, $j_2' = \delta_2'$, $j' = \delta' = p/2$. This can be done without any loss of generality if no other restriction is imposed on j_1' and j_2'. For $j_1' = \delta_1'$, $j_2' = \delta_2'$, $j' = \delta'$, three of the 6-j symbols representing the m.e. of K become "stretched" reducing the number of implicit summations in the expression by 3. If, further, δ' is taken to be equal to $p/2$, then all the summations disappear from the m.e. of $e^{-i\nu N}$ with the state $|\frac{1}{2}p\ m'\ \frac{1}{2}p\rangle$ on the rhs. This becomes evident when one writes the m.e. in the form

$$\langle jm\delta|e^{-i\nu N}|j'm'\delta'\rangle = \sum_{\bar{j}}(-)^{\delta'-\delta}(2\bar{j}+1)[(2j+1)(2j'+1)]^{1/2}$$

$$\times \begin{Bmatrix} \bar{j} & \frac{1}{2}(p-2\delta) & \frac{1}{2}(q+\delta+m) \\ j & \frac{1}{2}(q+\delta-m) & \frac{1}{2}p \end{Bmatrix}$$

$$\times D^{\bar{j}}_{\bar{m},\bar{m}'}(-2\nu) \begin{Bmatrix} \bar{j} & \frac{1}{2}(p-2\delta') & \frac{1}{2}(q+\delta'+m') \\ j' & \frac{1}{2}(q+\delta'-m') & \frac{1}{2}p \end{Bmatrix}$$

with

$$2\bar{m} = -p+q+3\delta+m, \quad 2\bar{m}' = -p+q+3\delta'+m'. \quad (12)$$

For $\delta' = p/2$ one of the j's in the 6-j symbol on the rhs of $D^{\bar{j}}_{\bar{m},\bar{m}'}(-2\nu)$ vanishes, giving the relations

$$\bar{j} = \tfrac{1}{2}(q+p/2+m') = \tfrac{1}{2}(q+p-\delta+m) = \bar{m}'$$

and

$$\bar{j} - \tfrac{1}{2}(p-2\delta) = \tfrac{1}{2}(q+\delta+m).$$

As a consequence, both the 6-j symbols in (12) become "stretched" and the summation over $\bar{j}$ disappears. The m.e., thus, takes the simple form

$$\langle jm\delta | e^{-i\nu N} | \frac{p}{2} m' \frac{p}{2} \rangle$$

$$= (-)^{p/2-\delta} \left(\frac{(q+\delta+m)!(p-2\delta)!(j+\delta)!}{(p+q-\delta+m)!(\delta-m)!} \frac{(j-m)!(p+q+1)!q!(p-\delta+m)!(2j+1)}{(p-j-\delta)!(p+j-\delta+1)!(q-j+\delta)!(q+j+\delta+1)!(j-\delta)!(j+m)!} \right)^{1/2}$$

$$\times D^{\frac{1}{2}(p+q-\delta+m)}_{\frac{1}{2}(p+q-\delta+m),\frac{1}{2}(q-p+3\delta+m)}(-2\nu). \quad (13)$$

The same result is obtained from Eq. (22) of II if the powers of $\sin\nu$ and $\cos\nu$ are replaced by the functions $D^{\bar{j}}_{\bar{j}\bar{m}}(-2\nu)$ with appropriate values of $\bar{j}, \bar{m}$. Using these results, we can obtain an expression for the isoscalar factor (ISF) containing only four explicit and five implicit summations. However, the difficulties of handling the multiple sums persist, and we have to content ourselves with the discussion of special cases only.

From the foregoing analysis it is clear that the ISF will take the simplest form when $j_1' = \delta_1' = p_1/2$, $j_2' = \delta_2' = p_2/2$, $j' = \delta' = p/2$. Since the m.e. of $\exp(-i\nu N)$ with the state $|\frac{1}{2}p\, m'\, \frac{1}{2}p\rangle$ on the rhs has been shown to be a multiple of the function $D^{\bar{j}}_{\bar{j}\bar{m}}(-2\nu)$, the ISF is obtained by evaluating the integral

$$\int D^{\bar{j}_1}_{\bar{j}_1\bar{m}_1}(-2\nu) D^{\bar{j}_2}_{\bar{j}_2\bar{m}_2}(-2\nu) D^{\bar{j}}_{\bar{j}\bar{m}}(-2\nu) \sin^2\nu \sin 2\nu \, d2\nu.$$

The integral is seen to have the value

$$\left(\frac{(\bar{j}_1 - \bar{m}_1 + 1)(\bar{j} - \bar{m} + 1)}{(2\bar{j}_1+1)(2\bar{j}+1)} \right)^{1/2} \times \frac{1}{2\bar{j}+2} \begin{Bmatrix} \bar{j}_1 + \frac{1}{2} & \bar{j}_2 & \bar{j} + \frac{1}{2} \\ \bar{m}_1 - \frac{1}{2} & \bar{m}_2 & \bar{m} - \frac{1}{2} \end{Bmatrix}$$

when the extra factor $\sin^2\nu$ in the volume element is absorbed in the D functions and use is made of the relation (9). Multiplying this by the CGC of $SU(2)$ arising from integration over the other variables, we have

$$\sum_\gamma \begin{pmatrix} p_1 q_1 & p_2 q_2 & pq \\ j_1\delta_1 & j_2\delta_2 & j\delta \end{pmatrix}_\gamma \begin{pmatrix} p_1 q_1 & p_2 q_2 & pq \\ \frac{1}{2}p_1 \frac{1}{2}p_1 & \frac{1}{2}p_2 \frac{1}{2}p_2 & \frac{1}{2}p \frac{1}{2}p \end{pmatrix}_\gamma = (-)^{p-2\delta} \frac{(q+1)(p+q+2)(p-2\delta+1)}{(2j+1)}$$

$$\times \left(\frac{(j_1+\delta_1)!(p_1+q_1+1)!q_1!(2j_1+1)(j_2+\delta_2)!(p_2+q_2+1)!q_2!(2j_2+1)(j+\delta)!(p+q+1)!q!(2j+1)}{(p_1-j_1-\delta_1)!(p_1+j_1-\delta_1+1)!(q_1-j_1+\delta_1)!(q_1+j_1+\delta_1+1)!(j_1-\delta_1)!(p_2-j_2-\delta_2)!(p_2+j_2-\delta_2+1)!(j_2-\delta_2)!} \right.$$

$$\left. \times \frac{1}{(q_2-j_2+\delta_2)!(q_2+j_2+\delta_2+1)!(p-j-\delta)!(p+j-\delta+1)!(q-j+\delta)!(q+j+\delta+1)!(j-\delta)!} \right)^{1/2}$$

$$\times \sum_{m_1,m_2} \left(\frac{(j_1-m_1)!(p_1-\delta_1+m_1)!(j_2-m_2)!(p_2-\delta_2+m_2)!(j-m)!(p-\delta+m)!}{(\delta_1-m_1)!(j_1+m_1)!(\delta_2-m_2)!(j_2+m_2)!(\delta-m)!(j+m)!} \right)^{1/2}$$

$$\times \frac{(q+\delta+m)!}{(p+q-\delta+m+2)!} \begin{Bmatrix} j_1 & j_2 & j \\ m_1 & m_2 & m \end{Bmatrix} \begin{Bmatrix} \frac{1}{2}p_1 & \frac{1}{2}p_2 & \frac{1}{2}p \\ m_1' & m_2' & m' \end{Bmatrix}. \quad (14)$$

This expression possesses the desired features, but gives only a class of ISF for which

$$p = p_1 + p_2 - 2r, \quad q = q_1 + q_2 + r.$$

The restriction on the values of p, q arises from Eq. (6) which, in the present case, takes the form

$$\tfrac{1}{2}(p - p_1 - p_2) = q_1 + q_2 - q.$$

The set of admissible values of p, q can be enlarged by taking one or more of the initial δ's to be equal to $-q/2$. Thus, for $\delta_1' = p_1/2$, $\delta_2' = -q_2/2$, $\delta' = p/2$, the possible values of p, q are

$$p = p_1 - q_2 - 2r, \quad q = q_1 - p_2 + r.$$

The ISF for this case can be obtained without a fresh calculation by making certain simple changes in the expression (14). This is a consequence of the symmetry

$$D^{pq}_{j\mu\delta,j'\mu'\delta'}(\alpha) = D^{qp*}_{j-\mu-\delta,j'-\mu'-\delta'}(\alpha),$$

which permits us to write the integral for (14) in the alternative form

$$\int D^{pq}_{j\mu\delta,\frac{1}{2}p\,\mu'\,\frac{1}{2}p}(\alpha) D^{p_1 q_1 *}_{j_1\mu_1\delta_1,\frac{1}{2}p_1\mu_1'\frac{1}{2}p_1}(\alpha) \times D^{q_2 p_2}_{j_2-\mu_2-\delta_2,\frac{1}{2}p_2,-\mu_2'-\frac{1}{2}p_2}(\alpha) d\Omega.$$

From this the integral for the ISF in the second case can be obtained by making the interchanges $p_1 \leftrightarrow p, q_1 \leftrightarrow q$, $\mu \leftrightarrow \mu_1, \delta \leftrightarrow \delta_1, \mu' \leftrightarrow \mu'_1, p_2 \leftrightarrow q_2$, and changing the sign of $\mu_2\delta_2\mu'_2$. Besides these, there are four other cases in which the ISF simplifies and takes a form similar to (14). The values of $\delta'_1, \delta'_2, \delta'$ in these cases and the corresponding values of p, q are tabulated below.

$2\delta'_1$	$2\delta'_2$	$2\delta'$	p	q
$-q_1$	$-q_2$	$-q$	p_1+p_2+r	q_1+q_2-2r
$-q_1$	p_2	$-q$	p_1-q_2+r	q_1-p_2-2r
$-q_1$	p_2	p	p_2-q_1-2r	q_2-p_1+r
p_1	$-q_2$	$-q$	p_2-q_1+r	q_2-p_1-2r

As already indicated, the integral on the lhs of (1) can be looked upon as an unnormalized CGC carrying the labels of the final (or, the initial) states of the m.e. The normalization factor for the CGC carrying the labels of the final states is

$$N = I(\lambda_1\lambda_1, \lambda_2\lambda_2, \lambda\lambda)^{-1/2}.$$

In the case of the expression (14) the normalization factor is obtained by setting $j_1 = \delta_1 = p_1/2$, $j_2 = \delta_2 = p_2/2$, $j = \delta = p/2$, $m_1 = m'_1$, $m_2 = m'_2$, $m = m'$. The expression then reduces to

$$I(\lambda_1\lambda_1, \lambda_2\lambda_2, \lambda\lambda) = \frac{(q+1)(p+q+2)}{(p+1)} \times \sum_m \frac{1}{(q+\tfrac{1}{2}p+m+1)(q+\tfrac{1}{2}p+m+2)}.$$

This is clearly seen to have the value 1. We can, therefore, replace the sum on the lhs of (14) by the single term

$$\begin{pmatrix} p_1 q_1 & p_2 q_2 & pq \\ j_1\delta_1 & j_2\delta_2 & j\delta \end{pmatrix}.$$

4. THE ORTHOGONALIZATION PROBLEM

According to our previous analysis an orthogonalized CGC,[14,15] the ϵth member of a degenerate set, can be written as

$$\begin{pmatrix} \rho_1 & \rho_2 & \rho_\epsilon \\ \nu_1 & \nu_2 & \nu \end{pmatrix} = \sum_{ijk} A^{\epsilon}_{ijk} \int D^{\rho_1}_{\nu_1 i}(\alpha) D^{\rho_2}_{\nu_2 j}(\alpha) D^{\rho *}_{\nu k}(\alpha) d\Omega \quad (15)$$

with suitable values of A^{ϵ}_{ijk}, the number of the set (ijk) of the initial states in the sum being equal to the multiplicity of ρ. Multiplying this by another member of the set and using the orthogonality conditions, we have

$$\sum_{\nu_1\nu_2} \begin{pmatrix} \rho_1 & \rho_2 & \rho_\epsilon \\ \nu_1 & \nu_2 & \nu \end{pmatrix} \begin{pmatrix} \rho_1 & \rho_2 & \rho_{\epsilon'} \\ \nu_1 & \nu_2 & \nu \end{pmatrix} = \sum_{ijk, i'j'k', \gamma} A^{\epsilon}_{ijk} A^{\epsilon'}_{i'j'k'} \begin{pmatrix} \rho_1 & \rho_2 & \rho_\gamma \\ i & j & k \end{pmatrix} \begin{pmatrix} \rho_1 & \rho_2 & \rho_\gamma \\ i' & j' & k' \end{pmatrix}. \quad (16)$$

Orthogonality of the CGC carrying the labels of the final states on the lhs of (16) is, thus, seen to imply the orthogonality of the CGC on the rhs looked upon as vectors in the "γ-space." Since the scalar products of these vectors are given by the integrals on the lhs of (1), one can apply Schmidt's method and express the coefficients A^{ϵ}_{ijk} in terms of integrals of the same type. However, a good deal of arbitrariness remains in the choice of the coefficients. This arbitrariness is inherent in the problem and cannot be removed.

As an illustration of the orthogonalization scheme we consider CGC of the type $\begin{pmatrix} p_1 q_1 & 1\,1 & pq \\ \nu_1 & \nu_2\ \nu & \end{pmatrix}$. In this case all the product representations are nondegenerate with the exception of the one with $p = p_1, q = q_1$. As this representation is double degenerate, the orthogonalization program can be carried out with two integrals of the type

$$I_1 = I(\nu_1 0, \nu_2 0, \nu 0), \qquad I_2 = I(\nu_1\lambda_1, \nu_2 0, \nu\lambda_1)$$
$$\text{with } \lambda_1 = |\tfrac{1}{2}p_1\ \mu'_1\tfrac{1}{2}p_1\rangle.$$

Writing the orthogonalized CGC as

$$(\mathrm{CGC})_1 = \alpha^1_1 I_1 + \alpha^1_2 I_2, \qquad (\mathrm{CGC})_2 = \alpha^2_1 I_1 + \alpha^2_2 I_2$$

and using Schmidt's method, we have

$$\alpha^1_1 = [I(00,00,00)]^{-1/2}, \alpha^1_2 = 0,$$
$$\alpha^2_2 = -\frac{I(00,00,00)}{I(\lambda_1 0, 00, \lambda_1 0)}\alpha^2_1 = \left(I(\lambda_1\lambda_1, 00, \lambda_1\lambda_1) - \frac{I(\lambda_1 0, 00, \lambda_1 0)^2}{I(00,00,00)}\right)^{-1/2}. \quad (17)$$

When the integrals are evaluated and combined as in (17), the coefficients α^i_j are found to take the values

$$\alpha^1_1 = \left[\frac{(p_1+q_1+1)(p_1+q_1+3)}{p_1^2+q_1^2-p_1q_1+p_1+q_1}\right]^{1/2}, \qquad \alpha^1_2 = 0,$$

$$\alpha^2_1 = (-)^{p_1+1}\left(\frac{(q_1+2)}{3p_1q_1(p_1+2)(p_1^2+q_1^2-p_1q_1+p_1+q_1)}\right)^{1/2} \times (2q_1-p_1)(p_1+q_1+1),$$

$$\alpha^2_2 = \left(\frac{4(q_1+2)(p_1^2+q_1^2-p_1q_1+p_1+q_1)}{3p_1q_1(p_1+2)}\right)^{1/2}.$$

The CGC obtained in this way agree with those given by Kuriyan *et al.* after an orthogonal transformation in the "γ-space."

In conclusion, we wish to emphasize once again that the methods, in their broad outline, are applicable to any compact group. For calculating the CGC of a compact group we require three things: (i) a set of parameters for labeling the elements of the group, (ii) the invariant volume element in the parameter space, and (iii) an analytic expression for the m.e. of finite transformations. When these three things are known, the calculation of the coefficients becomes an exercise in elementary algebra.

APPENDIX A

In this section we investigate the conditions under which a Saalschützian ${}_4F_3(1)$ series can be regarded as equivalent to a 6-j symbol. The ${}_4F_3(1)$ series which have turned up in the present calculations could all be transformed into series with purely negative parameters. We, therefore, consider a Saalschützian series

$${}_4F_3(-a,-b,-c,-d;-u,-v,-w;1) \quad (A1)$$

containing only negative parameters, and try to transform it into a 6-j symbol noting the constraints put on

the parameters at various stages of the process. For this we require an identity of the type

$$ {}_2F_1(-a,-b;-u;z)\,{}_2F_1(1+v-d,1+w-d;1+c-d;z) = {}_2F_1(-u+a,-u+b;-u;z) \times {}_2F_1(-v+c,-w+c;1+c-d;z) \quad \text{(A2)} $$

which holds when $u \geq a,b; c \geq d$. Equating the coefficients of Z^d on the two sides of the identity, we get a relation connecting the series (A1) with another series whose form depends on the relative magnitudes of v,w,c,d. If the second series is to assume the standard form of the 6-j symbol, then all the factorials containing r on the rhs of (A2), with the exception of $(u-r)!$, must occur in the denominator. This is ensured by taking $v,w \geq c$. The relation then takes the form

$$ {}_4F_3(-a,-b,-c,-d;-u,-v,-w;1) = \frac{c!d!(v-d)!(w-d)!(u-a)!(u-b)!(v-c)!(w-c)!}{u!v!w!} \times \sum_r (-)^r \frac{(u-r)!}{r!(u-a-r)!(u-b-r)!(v-c-d+r)!(w-c-d+r)!(c-r)!(d-r)!}. \quad \text{(A3)} $$

Comparing this with the expression for the 6-j symbol and solving the algebraic equations for the j's, we have

$$ {}_4F_3(-a,-b,-c,-d;-u,-v,-w;1) = (-)^{1-u}\left(\frac{a!b!c!d!(v-d)!(w-d)!}{(b-w-1)!(a-v-1)!(a-w-1)!}\,\frac{(u-a)!(u-b)!(v-c)!(w-c)!(u-c)!(u-d)!}{(b-v-1)!}\right)^{1/2}\frac{1}{u!v!w!} \times \begin{Bmatrix} \tfrac{1}{2}(u+v-a-c) & \tfrac{1}{2}(a+b-v-w-2) & \tfrac{1}{2}(u+w-b-c) \\ \tfrac{1}{2}(u+v-b-d) & \tfrac{1}{2}(v+w-c-d) & \tfrac{1}{2}(u+w-a-d) \end{Bmatrix}. \quad \text{(A4)} $$

This is one of the possible forms to which the ${}_4F_3(1)$ series reduces when the above inequalities are satisfied. This form is obtained by taking one particular solution of the equations for the j's. Nothing essentially new is obtained from the other solutions of the equations. Since Eq. (A4) is symmetrical in c and d, the restriction $c \geq d$ can be dropped. However, additional restrictions arise from the nonnegative character of the arguments of the factorials, and the conditions for the validity of the relation (A4) take the form

$$ u \geq a,b > v,w \geq c,d \quad \text{(A5)} $$

With these restrictions on the parameters the j's take only physical values consistent with the triangular inequalities.

APPENDIX B

We shall now evaluate the coefficients in the expansion

$$ (\csc\nu)^{2\lambda} D^j_{\mu',\mu}(-2\nu) = \sum_J G_J(\mu',\mu,m',m) D^J_{m',m}(-2\nu) \quad \text{(B1)} $$

for $\mu'+\mu = m'+m$, and, thus, fill a gap in the derivation of the expression (11) for the "semistretched" CGC. For $\mu'-\mu \geq 0, m'-m \geq 0$, Eq. (B1) can be written as

$$ (\sin^2\nu)^{\mu'-m'-\lambda}\,{}_2F_1(-j+\mu',j+\mu'+1;\mu'-\mu+1;\sin^2\nu) = \sum_J G_J (-)^{-\mu'+\mu+m'-m}\left(\frac{(J+m')!(J-m)!(j-\mu')!(j+\mu)!}{(J-m')!(J+m)!(j+\mu')!(j-\mu)!}\right)^{1/2}\frac{(\mu'-\mu)!}{(m'-m)!}\,{}_2F_1(-J+m',J+m'+1;m'-m+1;\sin^2\nu). \quad \text{(B2)} $$

To determine G_J, we express each power of $\sin\nu$ on the lhs as a linear combination of Jacobi polynomials by means of the formula

$$ z^p = \sum_n \frac{(\alpha+\rho)!(2n+\alpha+\beta+1)(n+\alpha+\beta)!(-\rho)_n}{(n+\alpha+\beta+\rho+1)!(n+\alpha)!} \times P_n^{\alpha,\beta}(1-2z). \quad \text{(B3)} $$

For $n = j-m'$, $\alpha = m'-m$, $\beta = m'+m$, the Jacobi polynomials reduce to multiples of the hypergeometric functions on the rhs of the equation. Equating coefficients of the same polynomial, we then have

$$ G_J = (-)^{J-m-\mu'+\mu}\left(\frac{(j+\mu')!(j-\mu)!(J+m')!(J+m)!}{(j-\mu')!(j+\mu)!(J-m')!(J-m)!}\right)^{1/2} \times \frac{(2J+1)(\mu'-m-\lambda)!(\mu'-m'-\lambda)!}{(\mu'-\mu)!(\mu'-\lambda+J+1)!(\mu'-\lambda-J)!} \times {}_4F_3(-j+\mu',j+\mu'+1,\mu'-m-\lambda+1, \mu'-m'-\lambda+1;\mu'-\mu+1, \mu'-\lambda+J+2,\mu'-\lambda-J+1;1). \quad \text{(B4)} $$

This expression is derived under the restrictions

$$ \mu'-\mu \geq 0, \quad \text{(B5a)} $$

$$ m'-m \geq 0, \quad \text{(B5b)} $$

$$ m' \geq 0, \quad \text{(B5c)} $$

$$ \mu'-\mu-m'+m-2\lambda \geq 0, \quad \text{(B5d)} $$

the last three restrictions arising from the use of the expansion (B3). We shall now put it into a more convenient form by using the results of the preceding section.

By a reversal of the series the expression (B4) can be written as

$$ G_J = (-)^{J-m+j-2\mu'+\mu} \times \left(\frac{(J+m')!(J+m)!}{(j+\mu')!(j+\mu)!(j-\mu')!(j-\mu)!(J-m')!(J-m)!}\right)^{1/2} \times \frac{(2J+1)(2j)!(j-m-\lambda)!(j-m'-\lambda)!}{(J+j+1-\lambda)!(j-J-\lambda)!} $$

$$\times {}_4F_3(\mu'-j, \lambda-j-J-1, \lambda+J-j, \mu-j; \\ -2j, \lambda+m-j, \lambda+m'-j; 1). \quad \text{(B6)}$$

The ${}_4F_3(1)$ series occurring in this expression is of the Saalschütztian type and all its parameters are negative. To see if the conditions (A5) are also satisfied, we write

$$a = J+j-\lambda+1, \quad b = j-\mu, \quad c = j-J-\lambda, \\ d = j-\mu', \quad u = 2j, \quad v = j-\lambda-m, \quad w = j-\lambda-m',$$

and examine the signs of $u-a$, $u-b$, $b-v$, etc. Since the values of the summation index range from 0 to $j-\mu'$, the highest power of $\sin^2\nu$ on the lhs of (B2) is $j-m'-\lambda$. This determines the range of the parameter n of the Jacobi polynomials and, hence, also the range of J. It is easily seen that the maximum and minimum values of J are $j-\lambda$ and m', respectively. If $u-a$ is not to be negative for any value of J within this range then the minimum value λ can take is $\frac{1}{2}$. With λ thus restricted, the inequalities (A5) are all satisfied and the ${}_4F_3(1)$ series reduces to a multiple of the 6-j symbol

$$\begin{Bmatrix} \frac{1}{2}(j-J-1+\mu'-m) & \frac{1}{2}(J+\mu'+\lambda-1) \\ \frac{1}{2}(j+\mu+J-m) & \frac{1}{2}(J-\mu-\lambda) \\ & \frac{1}{2}(j+m-\lambda) \\ & \frac{1}{2}(j-m'+\lambda-1) \end{Bmatrix}.$$

By using Regge symmetry and Eqs. (A4) [and (B6)], the expansion (B1) can now be written as[16]

$$(\csc\nu)^{2\lambda} D^j_{\mu'\mu}(-2\nu) = \sum_J (-)^{J-m+j+\mu+1} \left(\frac{(j+J+\lambda)!(j-J+\lambda-1)!(\mu'-m'-\lambda)!(\mu'-m-\lambda)!}{(J+j+1-\lambda)!(j-J-\lambda)!(\mu'-m'+\lambda-1)!(\mu'-m+\lambda-1)!} \right)^{1/2} \\ \times (2J+1) \begin{Bmatrix} \frac{1}{2}(j+m'+\lambda-1) & J & \frac{1}{2}(j-m'+\lambda-1) \\ \frac{1}{2}(j-m-\lambda) & \frac{1}{2}(\mu'-\mu-1) & \frac{1}{2}(j+m-\lambda) \end{Bmatrix} D^J_{m',m}(-2\nu). \quad \text{(B7)}$$

Equation (B7) possesses certain obvious symmetries and remains valid even when the restrictions (B5b) and (B5c) are removed. To see this, we make the interchanges $\mu' \leftrightarrow -\mu$, $m' \leftrightarrow -m$, everywhere in Eq. (B2). After the interchanges Eq. (B2) and the inequalities (B5a), (B5b), (B5d) continue to hold, but the inequality (B5c) gets replaced by $m \geq 0$. The last condition is automatically fulfilled if $m'-m \geq 0$ and $m' \leq 0$. Equation (B1) now takes the form

$$(\csc\nu)^{2\lambda} D^j_{\mu'\mu}(-2\nu) = \sum_J G_J(-\mu, -\mu', -m, -m') \times D^J_{m',m}(-2\nu).$$

But, since $G_J(-\mu, -\mu', -m, -m') = G_J(\mu', \mu, m', m)$, the same expansion holds for both positive and negative values of m' provided the other conditions (B5a), (B5b), (B5d) are satisfied. In the same way it can be shown that the expansion remains valid for $m'-m \leq 0$, if $\mu'-\mu \geq 0$, and $\mu'-\mu-m+m'-2\lambda \geq 0$. This is a consequence of the symmetry $G_J(\mu', \mu, m, m') = (-)^{m'-m} G_J(\mu', \mu, m', m)$.

[1] S. D. Majumdar, Prog. Theor. Phys. **41**, 1382 (1969). This paper is referred to as I.
[2] T. J. Nelson, J. Math. Phys. **8**, 857 (1967).
[3] F. D. Murnaghan, *The Unitary and Rotation Groups* (Spartan, Washington, D.C., 1962).
[4] S. D. Majumdar, J. Phys. A **1**, 203 (1968).
[5] S. D. Majumdar and B. K. Basu, Ann. Phys. (N.Y.) **56**, 464 (1970). This paper is referred to as II.
[6] M. Moshinsky, Rev. Mod. Phys. **34**, 813 (1962).
[7] L. K. Pandit and N. Mukunda, **6**, 1547 (1965).
[8] K. T. Hecht, Nucl. Phys. **62**, 1 (1965).
[9] J. G. Kuriyan, D. Lurié, and A. J. MacFarlane, J. Math. Phys. **6**, 722 (1965).
[10] R. T. Sharp and H. von Baeyer, J. Math. Phys. **7**, 1105 (1966).
[11] M. Resnikoff, J. Math. Phys. **8**, 63 (1967).
[12] L. C. Biedenharn, A. Giovannini, and J. D. Louck, J. Math. Phys. **8**, 691 (1967).
[13] R. M. Asherova and Yu. F. Smirnov, Nucl. Phys. B **4**, 399 (1968).
[14] C. K. Chew and R. T. Sharp, Can. J. Phys. **44**, 2789 (1966).
[15] S. D. Majumdar, Prog. Theor. Phys. **42**, 1475 (1969).

[16] The result is obtained directly from Eq. (A3) (without using Regge symmetry) by taking another solution of the equations for the j's.

Matter symmetries in general relativistic kinetic theory

R. Berezdivin[*†]

University of California, Berkeley, California
Universidad Simón Bolívar, Caracas, Venezuela

R. K. Sachs*

University of California, Berkeley, California
(Received 29 November 1972)

It is shown that a surface forming matter symmetry for collision-free general relativistic kinetic theory gives rise to an isometry on the space-time. Specific solutions are given, and it is shown that limiting the surface forming conditions to the mass subbundle does not alter the results. A possible method for solving more general cases is given.

I. INTRODUCTION

Matter symmetries in general relativistic kinetic theory[1] have been previously defined[2] and the extent to which they induce isometries on the spacetime has been questioned.[2] A brief summary of the relevant concepts is here presented as well as further results.

II. TANGENT BUNDLE AND KINETIC THEORY

The tangent bundle[3] TM over the space–time M is an eight-dimensional differentiable manifold. If $\{x^a\}$ is a system of local coordinates on M then $\{x^a, p^a\}$ is one on TM, with p^a the components of a tangent vector at the point x^a of M. Only the submanifold of TM with p^a future-directed and non-spacelike will concern us: this is the relativistic one-particle phase space, with x^a a point in the space–time and p^a the momentum of a particle at that point; this restriction, however, alters none of the results as will later be seen, and we can thus simply work on TM. With the above coordinates on TM the coordinate basis for vectorfields over TM, that is, for operators on functions of x^a and p^a, is $\{\partial/\partial x^a, \partial/\partial p^a\}$. This basis is awkward for calculations, for the components of vectors in TM, under transformation of coordinates in M, $\bar{x}^a = \bar{x}^a(x^b)$, which induces a transformation of coordinates on TM given by $\{\bar{x}^a(x^b), [\partial \bar{x}^a(x^b)/\partial x^c]p^c\}$, do not transform as vectors over M. If, however, we go to the connection basis, $\{\nabla_a, \partial/\partial p^a\}$, where $\nabla_a = \partial/\partial x^a - \Gamma^c_{ab} p^b \partial/\partial p^c$, and write a vector field over TM as $Q = Q^a \nabla_a + Q^{a+4}\partial/\partial p^a$ and now carry out the coordinate transformation, both Q^a and Q^{a+4} will transform as vector fields over M. This can best be seen by realizing that with the dual connection basis $\{dx^a, Dp^a\}$, with $Dp^a = dp^a + \Gamma^a_{bc} p^b dx^c$, since $dx^a(\nabla_a) + Dp^a(\partial/\partial p^a)$ is a scalar on TM, and since dx^a/ds and Dp^a/ds are vector fields on M, so will Q^a and Q^{a+4}.

In the connection basis the Liouville operator is $L = p^a \nabla_a$, and if the distribution function of matter is f, a differentiable function on TM, then the Liouville equation is $Lf = 0$. The Einstein equations have on the right the stress energy tensor defined in terms of an integral over the fibre composed of all future-directed non-spacelike tangent vectors at the point x of M,

$$T^{ab} = \int f(x,p) p^a p^b d^*p \tag{1}$$

with d^*p the natural volume element on the fibre. The Einstein–Liouville equations[4] apply to collisionless systems with no macroscopic electromagnetic field.

III. MATTER SYMMETRIES

Roughly, a matter symmetry will exist if there exists a vector field over TM that leaves f unchanged. This vector field will connect the points in the phase space where the distribution of matter is the same. Formally, a matter symmetry from (x, F) to (x, F') exists if an observer at x with a local Lorentz frame F measures f on the fibre above x and another observer at x' with a local Lorentz frame F' measures f' on the fibre at x' to be the same as f. A one-parameter group of matter symmetries exists if we can define a matter symmetry from any point to any other point along a curve in M. Letting φ^a be the tangent to this curve and allowing that the momenta must change as under a Lorentz transformation, in going from x^a to $x^a + dx^a$ we must have, to first order in ϵ, $p'^a = p^a + \epsilon A^a{}_b p^b$ with $p'^2 = p^2$. Thus

$$p^2 = p^2 + \epsilon[g_{ab,c}\varphi^c + 2A_{ab}]p^a p^b,$$

so that $g_{ab,c}\varphi^c + 2A_{(ab)} = 0$. We must pick the transformation properties of A_{ab} correctly so that the above equation be coordinate independent. To do so, we define

$$A'^a{}_b = A^a{}_b + \Gamma^a_{bc}\varphi^c,$$

so that the above equation will now give us $A'_{(ab)} = 0$. We may require A'_{ab} to transform as a cotensor on M, for if we compute the change if f under the action of the matter symmetry, with $ds = \epsilon$,

$$0 = \frac{df}{ds} = \frac{\partial f}{\partial x^a}\varphi^a + \frac{\partial f}{\partial p^a}A^a{}_b p^b = \varphi^a \frac{\partial f}{\partial x^a} + \frac{\partial f}{\partial p^a}[A'^a{}_b p^b - \Gamma^a_{bc} p^b \varphi^c] = \varphi^a \nabla_a f + A'^a{}_b p^b \frac{\partial f}{\partial p^a}.$$

We have seen that this equation is coordinate independent if φ^a and $A'^a{}_b p^b$ transform as vector fields over M, so that A'_{ab} should transform as a cotensor. The equation for a one-parameter group of matter symmetries may then be written, dropping primes, as $Wf = 0$, or equivalently, $\mathcal{L}_W f = 0$, the Lie derivative $\mathcal{L}$, with

$$W = \varphi^a \nabla_a + A^a{}_b p^b \frac{\partial}{\partial p^a}, \tag{2a}$$

and

$$A_{(ab)} = 0. \tag{2b}$$

A one-parameter group of matter symmetries will simply be called a matter symmetry. Various properties of matter symmetries can be shown:

(A) They carry fibres into fibres linearly and isometrically: linearly because φ^a is independent of p and $A^a{}_b p^b$ is linear in p and isometrically because they leave p^2 unchanged.

(B) If W is a matter symmetry, then so is $\psi(x)W$.

(C) If $A_{ab} = 0$, called a horizontal matter symmetry (the connection Γ^a_{bc} uniquely splits any vectorfield over TM into horizontal and vertical coordinate independent projections[5] the momenta are parallely propagated along φ^a. This holds for any horizontal vector field over TM. For notice that a displacement along W can be expressed as

$$\left(\frac{dx^a}{ds}, \frac{dp^a}{ds}\right) ds = (\varphi^a, A^a{}_b p^b - \Gamma^a_{bc} p^b \varphi^c) ds$$

Copyright © 1973 by the American Institute of Physics

or as

$$\left(\frac{dx^a}{ds}, \frac{Dp^a}{ds}\right) ds = (\varphi^a, A^a{}_b p^b) ds . \tag{3}$$

If $A_{ab} = 0$ we must have $Dp^a/ds = 0$, that is, $p^a{}_{;b}\varphi^b = 0$.

(D) If $A_{ab} = \varphi_{a;b}$, so that W is called the natural lift of the Killing vector field φ^a, denoted by (NAT LIFT φ), the momenta are Lie transported along φ^a. For

$$\frac{dp^a}{ds} = \frac{Dp^a}{ds} - \Gamma^a_{bc} p^b \varphi^c = \varphi^a{}_{;b} p^b - \Gamma^a_{bc} p^b \varphi^c$$

from (3). This immediately gives $\mathcal{L}_\varphi p^a = 0$. This is what we expect an isometry to do. (NAT LIFT φ) $f = 0$ thus imposes a limitation on f that could reflect the isometry on the space-time. See below.

(E) There are conditions imposed on the stress energy tensor derived from a distribution function subject to a matter symmetry. Indeed, from (1), integrating by parts and assuming a proper behavior for f at ∞ on any fibre, it can be shown[6] that

$$T^{ab}{}_{;c} = \int (\nabla_c f) p^a p^b d^* p .$$

From this, $Lf = 0$ can be seen to imply $T^{ab}{}_{;b} = 0$; however, we need the vanishing of $T^{ab}{}_{;b}$ and all the higher moments to give us Liouville's equation. We can again integrate by parts and get, if φ is a Killing vector field,

$$\int [(\text{NAT LIFT } \varphi) f] p^a p^b d^* p = 0, \tag{4}$$

and

$$\int (Wf) p^a p^b d^* p = \mathcal{L}_\varphi T^{ab} - (A^a{}_d - \varphi^a{}_{;d}) T^{ba} - (A^b{}_d - \varphi^b{}_{;d}) T^{ad} . \tag{5}$$

Hence, for φ giving rise to an isometry, the second moment of (NAT LIFT φ) f is 0 but (NAT LIFT φ) f may oscillate so it does not follow that it is 0. However, (NAT LIFT φ) is usually taken to be the matter symmetry induced by a Killing vector field φ.[7]

(F) If the stress energy tensor from (1) is that of a perfect fluid, we can get from (5) that $Wf = 0$ implies $\mathcal{L}_\varphi \rho = \mathcal{L}_\varphi p = 0$, p the pressure, and $\mathcal{L}_\varphi u^a = (A^a{}_c - \varphi^a{}_{;c}) u^c$, for ρ, p, and u^a either the kinematically or the dynamically defined quantities.[8] Thus, even in hydrodynamics there exists the question of whether a matter symmetry gives rise to an isometry. However, a Killing vector field here gives $\mathcal{L}_\varphi \rho = \mathcal{L}_\varphi p = \mathcal{L}_\varphi u^a = 0$, which is in hydrodynamics as much a "matter symmetry" as we ever can have. Therefore, the problem of getting a matter symmetry from an isometry, trivial in hydrodynamics, is, in kinetic theory, not as interesting as the inverse problem, and we content ourselves with (4).

(G) In classical Newtonian theory a matter symmetry can only be due to the Galilean group, translations, rotations, and velocity transformations, giving,

$$\frac{dx^i}{ds} = \varphi^i, \quad \text{and} \quad \frac{d\mathbf{p}}{ds} = \mathbf{A} \times \mathbf{p} + m\mathbf{v}, \quad \varphi_{(i,j)} = 0.$$

By using the fact that V, the Newtonian potential, is a Poisson integral, $\mathcal{L}_\varphi V$ can be written in terms of integrals of f, which integrated by parts, and using $\mathcal{L}_\varphi \rho = 0$, easily gotten from $Wf = 0$, will give us the conclusion that $\mathcal{L}_\varphi V = 0$. This is the isometry here. If instead of the Galilean group, the complete Heckman-Schücking group is allowed, the situation gets extremely complicated since V is not then a Poisson integral.

(H) An example of matter symmetries is provided by an isotropic distribution of velocities. Specifically, one has a three-parameter group of matter symmetries with closed two dimensional orbits. The matter symmetries are each vertical. Ehlers, Geren, and Sachs[9] showed that it necessarily follows, for nonstationary space-times and nonzero rest mass particles, that M must be a Robertson-Walker space-time.

IV. ONE-PARAMETER GROUP OF MATTER SYMMETRIES

Consider the space spanned by W and L at each point of TM. It is a two-dimensional subspace of TTM. Call $D = D(x,p)$ the distribution which assigns such a space to each (x,p) in TM. Now form the sequence of brackets

$$[W,L], [W,[W,L]], [L,[W,L]], [W,[W,[W,L]]], \cdots$$

and at each step take the new space $\overline{D}$ spanned by W, L, and the brackets up to that step, and continue taking brackets until further brackets do not increase the dimension of $\overline{D}$, which is at most eight dimensional since it is contained in TTM. This will happen in a finitely different number of ways, as some of the brackets being linearly dependent on the previous ones will imply that some of the following ones also are (the number of ways is not one as we may bracket with either W or L at each step). If the dimension of $\overline{D}$ does not vary over TM, hence a distribution, it will be involutive.[3] By Frobenius' theorem, $\overline{D}$ is then completely integrable. f will not vary on its integral submanifolds of dimension equal to the dimension of $\overline{D}$. That is, we keep on adding equations until the system is completely integrable. And since D is involutive, some of the brackets will be linearly dependent on the others; this will be very useful, along with the Einstein-Liouville equations, in searching for a possible isometry induced by W.

The solution to the problem was attempted in the following manner: Assume $\overline{D}$ of dimension n, n from 1 to 8, and solve each case using the linear independence of only n of the brackets in $\overline{D}$. The $n = 1$ case was very easy and is simply a subcase of the $n = 2$ case, which was solved. The $n > 2$ cases were not attempted. The $n = 2$ case corresponds to $\overline{D} = D$, a surface, and W and L are said to be surface forming. The necessary and sufficient condition is that $[W, L] \in D$. As for degeneracy, from $n = 2$ on any eight dimensional region of TM it was shown that it necessarily gives $n = 2$ on all TM. Similarly, the same results and solutions as were gotten on TM were also gotten on the mass subbundle.

We have thus a possible, although difficult, method for solving the problem completely. We need to do some calculations first.

V. COMMUTATORS

We use the easily shown fact[6] that if we know $Z(t_a p^a)$ for all one forms $t_a(x)$, Z, a vector field over TM, is completely determined. We also need that for a one form $t_a(x)$, and a vector field

$$Z = H^a \nabla_a + V^a \frac{\partial}{\partial p^a},$$

we have

$$Z(t_a p^a) = H^a \left(\frac{\partial}{\partial x^a}(t_c p^c) - \Gamma^d_{ba} p^b \frac{\partial}{\partial p^d}(t_c p^c) \right) + V^a t_a ,$$

so that

$$Z(t_a p^a) = H^a p^c t_{c;a} + t_a V^a.$$

Then, for any $t_a(x)$, from (2), $L = p^a \nabla_a$, and the above, we get

$$L(t_a p^a) = t_{a;b} p^a p^b \tag{6}$$

and

$$W(t_a p^a) = \varphi^a p^c t_{c;a} + A^a{}_b t_a p^b. \tag{7}$$

We therefore get for the Lie bracket or commutator of W and L, acting on $t_a p^a$,

$$\begin{aligned}[W, L](t_a p^a) &= W(t_{a;b} p^a p^b) - (t_{a;b}\varphi^b)_{;d} p^d p^a - (A^a{}_c t_a)_{;d} p^d p^c \\ &= t_{a;b;c} p^a p^b \varphi^c + t_{a;b}(p^a A^b{}_c p^c + p^b A^a{}_c p^c) \\ &\quad - t_{a;c;b} p^a p^b \varphi^c - p^a p^d t_{a;b} \varphi^b{}_{;c} - t_{a;d} p^d p^c A^a{}_c \\ &\quad - A^a{}_{c;d} p^d p^c t_a \\ &= t_{a;b} p^a p^d (A^b{}_d - \varphi^b{}_{;d}) + p^a p^b (2 t_{a;[b;c]} \varphi^c - A^c{}_{a;b} t_c),\end{aligned}$$

where we operated on $t_{a;b}$ by taking it to be $l_a k_b$, so that we have

$$[W, L] = (A^b{}_d - \varphi^b{}_{;d}) p^d \nabla_b + p^a p^b [\varphi^c R^d{}_{abc} - A^d{}_{a;b}] \frac{\partial}{\partial p^d}. \tag{8}$$

We can now easily treat the preliminary case when W and L commute. From (8) we get $A_{ab} = \varphi_{a;b}$ and thus φ^a is Killing and W its natural lift. Conversely, if W is the natural lift of the Killing vector field φ^a, we immediately get that the horizontal part of $[W, L]$ is zero. As for the vertical part, using the definition of the curvature tensor, its index symmetries, and the antisymmetry of $A_{ab} = \varphi_{a;b}$, we get

$$\varphi^c R_{dabc} = \varphi^c R_{cbad} = \varphi_{b;a;d} - \varphi_{b;d;a},$$

so that the vertical part of $[W, L]$ is

$$(\varphi_{b;a;d} - \varphi_{b;d;a} - \varphi_{d;a;b}) p^a p^b = (-\varphi_{d;a;b} + \varphi_{d;b;a}) p^a p^b = 0.$$

This commutativity applies equally well to the flows[3] of W and L. We may thus travel certain parameter distances along either integral curve first but still reach the same point. This is as expected because translations along an isometry should change nothing.

VI. SURFACE FORMING MATTER SYMMETRY

We now treat the nondegenerate $n = 2$ case. From (8) and the surface forming conditions, with α and β functions on TM,

$$\begin{aligned}(A^a{}_b - \varphi^a{}_{;b}) p^b \nabla_a + (R^d{}_{abc}\varphi^c - A^d{}_{a;b}) p^a p^b \frac{\partial}{\partial p^d} \\ = (\alpha\varphi^a + \beta p^a)\nabla_a + \alpha A^a{}_b p^b \frac{\partial}{\partial p^a}.\end{aligned} \tag{9}$$

Equation (9) holds throughout TM so that we may differentiate it with respect to the arbitrary momenta to get, using commas to indicate the differentiation, for the horizontal part,

$$A^a{}_b - \varphi^a{}_{;b} = \alpha_{,b}\varphi^a + \beta_{,b} p^a + \beta\delta^a_b. \tag{10}$$

Differentiating again, we obtain

$$\alpha_{,bd}\varphi^a + \beta_{,bd} p^a + \beta_{,b}\delta^a_d + \beta_{,d}\delta^a_b = 0. \tag{11}$$

If we now let t^a be any non-lightlike vector on M orthogonal to both φ^a and p^a and contracting Eq. (10) with it, we get

$$t_{(d}\beta_{,b)} = 0.$$

This gives $\beta_{,d} = 0$ and then equation (11) gives $\alpha_{,bd} = 0$ so that $\alpha = f_c(x) p^c$. We thus have, also differentiating the vertical part of (9),

$$\varphi_{a;c} - A_{ac} = -\beta g_{ac} - f_c \varphi_a, \tag{12}$$

and

$$R_{c(ba)d}\varphi^c - A_{d(a;b)} - A_{d(b} f_{a)} = 0, \tag{13}$$

where everything is now an object on M and commas will now indicate ordinary differentiation on M. Note that the previous differentiation could have been performed had the surface forming conditions held on any open region of eight dimensions in TM and that from (12) and (13) it follows that if X is the projection of that region under the projection map $\pi : TM \to M$, then the conditions will hold throughout TX: We can go from (12) and (13) to Eq. (9) which now would hold on TX. Covariantly differentiating (12), we now get

$$\varphi_{a;c;d} - A_{ac;d} = -\beta_{,d} g_{ac} - f_{c;d}\varphi_a - f_c[A_{ad} - \beta g_{ad} - f_d \varphi_a].$$

If we now antisymmetrize with respect to c and d and then symmetrize with respect to a and c, we get, using the definition of the Riemann tensor,

$$\begin{aligned}\tfrac{1}{2} R^b{}_{(ac)d}\varphi_b = \tfrac{1}{2}(A_{(ac);d} - \underbrace{A_{ad;c}}) - \tfrac{1}{2}(\beta_{,d} g_{ac} - \beta_{,(c} g_{a)d}) \\ - \tfrac{1}{2}[f_{[c;d]}\varphi_a + f_{[a;d]}\varphi_c] + \tfrac{1}{2}\beta[f_{(c} g_{a)d} - f_d g_{ac}] \\ + f_{[c}\underbrace{f_{d]} f_a} - \tfrac{1}{2}(\underbrace{A_{ad}} f_c - A_{(ac)} f_{d)})\end{aligned}$$

where $\underbrace{\quad}$ also denotes symmetrization with respect to the indices so connected. Then, since $A_{(ab)} = 0$ and $f_{[c} f_{d]} = 0$, noticing that (13) will cause the terms $A_{ad} f_c$ and $A_{ad;c}$ to cancel out with the Riemann tensor term, we may obtain

$$\begin{aligned}-\beta_{,d} g_{ac} + \beta_{,(c} g_{a)d} - [f_{[c;d]}\varphi_a + f_{[a;d]}\varphi_c] \\ + \beta[f_{(c} g_{a)d} - f_d g_{ac}] = 0.\end{aligned} \tag{14}$$

This and (12) are equivalent to (13) and (12).

We first assume $\beta \neq 0$ and define $\overline{W} = W/\beta$, $\bar{f}_c = f_c + \beta_{,c}/\beta$, and get from (14)

$$\bar{f}_{(c} g_{a)d} - g_{ac}\bar{f}_d - [\bar{f}_{[c;d]}\overline{\varphi}_a + \bar{f}_{[a;d]}\overline{\varphi}_c] = 0.$$

This will imply, again contracting with a non-lightlike vector, that $\bar{f}_d = 0$; placing hats on (12), with $\bar{\beta} = 1$, will give

$$\overline{\varphi}_{a;c} = \overline{A}_{ac} - g_{ac}. \tag{15}$$

This is now our only surface forming condition. We drop the bars and in a coordinate system adapted to φ^a, with $\varphi^a = \delta^a_0$,

$$Wf = \varphi^a \frac{\partial f}{\partial x^a} + (\varphi^a{}_{,b} + \delta^a_b) p^b \frac{\partial f}{\partial p^b}$$

or

$$f = f(x^i, e^{-x^0} p^a), \qquad i = 1, 2, 3. \tag{16}$$

(15) now gives

$$0 + [a, oc] = A_{ac} - g_{ac},$$

or

$$g_{ac} = e^{-2x^0} h_{ac}(x^i).$$

This conformal relation gives $G_{ab} = G'_{ab} + H_{ab}(x^i)$, with G'_{ab} the Einstein tensor for h_{ab}. Then G^{ab} is proportional to e^{4x^0}, since it is proportional to $(g^{ab})^2$. Then from the Einstein equations and (16),

$$-G^{ab} = T^{ab} = \int f(x,p)p^a p^b d^*p$$
$$= \int f(x^i, e^{-x^0}p^a)p^a p^b d^*p = \int f(x^i, p'^a)e^{2x^0}p'^a p'^b d^*p',$$

where we define $p'^a = e^{-x^0}p^a$ and d^*p remains unchanged under such a transformation since it is the invariant volume element on the fibre. Then, with T'^{ab} independent of x^0,

$$-G^{ab} = e^{2x^0}T'^{ab}.$$

This is inconsistent with the previous result that G^{ab} be proportional to e^{4x^0}, so we must have made an inconsistent assumption in taking $\beta \neq 0$.

If $\beta = 0$, from (14), $f_{[c;d]}\varphi_a + f_{[a;d]}\varphi_c = 0$ which gives $f_{[c;d]} = 0$ so that $f_a = \sigma_{,a}$ and defining $\overline{\varphi}_a = e^\sigma \varphi_a$ and differentiating, $\overline{\varphi}_{a;c} = (e^\sigma)_{,c}\phi_a + e^\sigma \varphi_{a;c} = \overline{A}_{ac}$ with $\overline{W} = e^\sigma W$ the natural lift of the Killing vectorfield $\overline{\varphi}^a$. Thus, a properly scaled matter symmetry that is surface forming with the Liouville operator is the natural lift of a Killing vector field on M.

VII. MASS SUBBUNDLE

The question is whether a constraint on the surface forming conditions to the mass subbundle might weaken the restrictions on solutions and thus not give us a Killing vector field. Equation (9) is of the form, with α and β functions on TM, and B_{abc} a tensor field on M,

$$B_{abc}p^a p^b = \alpha A_{cd}p^d, \tag{17}$$

and

$$A^a{}_b p^b = \alpha\varphi^a + \beta p^a. \tag{18}$$

In $T_m M$, where $p^2 = m^2 =$ a constant $\neq 0$, defining for $\bar{p}^2 > 0$,

$$\bar{\alpha}(x,\bar{p}^a) = [(\bar{p}^2)^{1/2}/m]\alpha(x, \bar{p}^a m/(\bar{p}^2)^{1/2}),$$

and

$$\bar{\beta}(x,\bar{p}^a) = \beta(x, \bar{p}^a m/(\bar{p}^2)^{1/2}),$$

(18) gives, for $\bar{p}^a = p^a(\bar{p}^2)^{1/2}/m$ so that $p^2 = m^2$,

$$A^a{}_b \bar{p}^b m/(\bar{p}^2)^{1/2} = [m/(\bar{p}^2)^{1/2}][\bar{\alpha}(x,\bar{p})\varphi^a + \bar{\beta}(x,\bar{p})\bar{p}^a].$$

The same may be done for Eq. (17) so that it and (18) hold for all $\bar{p}^a$ such that $\bar{p}^2 \neq 0$ even though $\bar{p}^2 \neq m^2$. This is a smooth extension to a four-dimensional region in the fibre M_x near $p^2 = m^2$ and to an eight-dimensional region in TM. This can be done without ever touching the points $p^2 = 0$. Therefore, as we noted in the previous section, we will obtain the same results throughout TM and thus the same solutions as on M but for a fixed mass now. For zero rest mass particles the treatment is more complicated and can be found elsewhere; we still obtain the same results.[6]

VIII. HORIZONTAL SUBCASE

Solutions were already found for horizontal surface forming matter symmetries.[2] When the horizontal component was a lightlike vector, the general solution was given. For the spacelike case a general class of solutions were given. In the timelike case no nontrivial solution exists.

To construct a specific spacelike example, we need only a three-dimensional example.[6] With $i = 1, 2$ and $a = 1, 2, 3$ now, in

$$d\sigma^2 = dt^2 - R^2(t)[dx^2 + dy^2], \qquad h = h(x^i, p^a),$$

$Lh = 0$ gives $h = h(p^t R)$, $R^i{}_i = 0$ gives $R(t) = (at + b)^{2/3}$, and the Einstein equations are satisfied when we can find an h such that

$$\frac{1}{R^5}\int \frac{(p^x)^2 h dp^x dp^y}{[(p^x)^2 + (p^y)^2]^{1/2}} = \frac{1}{R^4}[\dot{R}^2 + \dot{R}\ddot{R}].$$

This will be satisfied if $R[\dot{R}^2 + R\ddot{R}] =$ a constant $= K$.

This holds indeed with $K = \frac{2}{9}a^2$. We can fix t such that $R = t^{2/3}$. Our solution is then,

$$ds^2 = dt^2 - dz^2 - t^{4/3}[dx^2 + dy^2], \tag{19}$$

$$f = \delta(p^a p_a)\delta(p^z)h(t^{4/3}[(p^x)^2 + (p^y)^2]^{1/2}), \tag{20}$$

with h any function of its argument.

The stress energy tensor for (20) is, from (19) and the Einstein equations,

$$T^{ab} = \frac{4}{9}\begin{pmatrix} t^{-2} & 0 & 0 & 0 \\ 0 & \frac{1}{2}t^{-10/3} & 0 & 0 \\ 0 & 0 & \frac{1}{2}t^{-10/3} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

This can be interpreted as zero mass particles traveling isotropically in the xy plane and uniformly distributed in the z direction. It is homogeneous, anisotropic, and of Bianchi type I.

IX. CONCLUSION

We have presented matter symmetries and discussed some of the problems involved. We have shown that a surface forming matter symmetry gives rise to an isometry on the spacetime. We have found some specific solutions and have shown that limiting the surface forming conditions to the mass subbundle alters nothing. We have indicated a possible method for solving more general cases, by using the linear dependence of a sequence of Lie brackets with higher dimensional space forming conditions. We expect the conclusion to remain valid under less strong conditions than surface forming.

ACKNOWLEDGMENTS

One of us (RB) would like to thank Romualdo Tabensky, Douglas Eardly, and Edison Liang for various enlightening discussions.

---

*Partially supported by the National Science Foundation under Grant No. GP-21495. Parts of this work were submitted as a Ph.D. thesis at the University of California at Berkeley in partial fulfillment of the requirements for a Doctor of Philosophy degree.

†Present address: Universidad Simon Bolivar, Caracas, Venezuela.

[1] R. K. Sachs and J. Ehlers, in *Brandeis University Summer Institute in Theoretical Physics*, edited by M. Chretien, S. Deser, and J. Goldstein (Gordon and Breach, New York, 1969).

[2] R. Berezdivin and R. K. Sachs, in *Relativity*, edited by M. Carmeli, S. I. Fickler, and L. Witten (Plenum, New York, 1970).

[3] R. L. Bishop and S. I. Goldberg, *Tensor Analysis on Manifolds* (Macmillan, New York, 1968).

[4] J. R. Ipser and K. S. Thorne, Astrophys. J. **154**, 262 (1968).

[5] N. J. Hicks, *Notes on Differential Geometry* (Van Nostrand, Princeton, N. J., 1965).

[6] R. Berezdivin, Ph.D. thesis (University of California, Berkeley, 1972).

[7] J. Ehlers, 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).

[8] J. Ehlers, in *Relativistic Fluid Dynamics*, Centro Internazionale Matematico Estivo (Edizionni Cremonese, Rome, 1971).

[9] J. Ehlers, P. Geren, and R. K. Sachs, J. Math. Phys. **9**, 1344 (1968).

Exact parastatistics for one-dimensional lattice space

R. B. McQuistan

Department of Physics and Laboratory for Surface Studies, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201

(Received 21 July 1972)

Exact parastatistics are developed which describe the occupation of a one-dimensional lattice space by any arbitrary distribution of particles $\{g_\lambda\}$, $(\lambda = 1,2,3,\ldots,)$, where λ is the particle correlation distance, i.e., the number of contiguous lattice sites occupied by a particle.

1. INTRODUCTION

Two previous papers have considered the exact occupation statistics for single types of particles on a one-dimensional lattice space; the first of these[1] dealt with dumbbells and the second[2] considered λ-bells (where, e.g., $\lambda = 2$ for dumbbells). The present paper treats a more general problem which considers an arbitrary distribution of a number of different types of particles on a one-dimensional lattice space. The word parastatistics has been used here as it has been used in the literature[3] to denote those statistics applicable in situations involving more than one kind of particle. In the present paper, these particles may include any number of simple particles (occupying a single lattice site), dumbbells (occupying two adjacent lattice sites), and λ-bells (occupying λ contiguous lattice sites where $\lambda = 1, 2, 3, \ldots$). We shall state and prove three theorems dealing with the distribution of such particles.

(I) $A[\{q_\lambda\}, N]$ the occupational degeneracy for any arbitrary set $\{q_\lambda\}$ of particles on a one-dimensional lattice space of N compartments.

(II) $N_p[\{q_\lambda\}, N]$, the number of groups of contiguous vacancies of length p created when $\{q_\lambda\}$ is placed on a one-dimensional lattice space of N compartments in all possible ways.

(III) $N_r[\{q_\lambda\}, N]$, the number of groups of r contiguous particles arising when $\{q_\lambda\}$ is placed in all possible ways on a one-dimensional lattice space of N equivalent sites.

2. OCCUPATIONAL DEGENERACY

Given a set of particles $\{q_\lambda\}$, where $\lambda(= 1, 2, 3\ldots)$ is the number of contiguous lattice sites occupied by each different type of particle. The particles of one species are indistinguishable from other particles of the same species. This set of particles is distributed on a one-dimensional lattice space consisting of N lattice sites and we assume that the occupation of a lattice site is either 0 or 1.

Theorem 1: $A[\{q_\lambda\}, N]$, the number of independent ways of arranging the set of particles $\{q_\lambda\}$, on a one-dimensional lattice space of N compartments is given by (see Fig. 1)

$$A[\{q_\lambda\}, N] = \frac{\left(N - \sum_\lambda (\lambda - 1)q_\lambda\right)!}{\prod_\lambda (q_\lambda!)\left(N - \sum_\lambda \lambda q_\lambda\right)!}. \tag{1}$$

In Fig. 1 we see, for example, if $q_1 = 1$, $q_2 = 2$, and $q_3 = 2$ and $N = 14$ the number of arrangements *for a particular sequence of particles* is 56; there are $5!/1!2!2! = 30$ of such independent sequences so that there are a total of (30) (56) = 1680 arrangements as given by Eq. 1.

Proof: There are a total of $[N - \sum_\lambda (\lambda - 1)q_\lambda]$ individuals to be permuted; $[N - \sum_\lambda \lambda q_\lambda]$ of these are vacancies, q_1 are simple particles, q_2 are dumbbells, $\cdots q_\lambda$ are λ-bells. The number of independent ways of arranging all these individuals, subject to the constraint of the indistinguishability of the particles (and vacancies) within a particular species, is then given by the multinomial coefficient expressed in Eq. 1.

3. THE NUMBER OF GROUPS OF p CONTIGUOUS VACANT SITES

Theorem 2: $N_p[\{q_\lambda\}, N]$, the number of groups of p $(p = 0, 1, 2, \cdots N - \sum_\lambda \lambda q_\lambda)$ contiguous vacant compartments created when $\{q_\lambda\}$ is placed on a one-dimensional lattice space of N compartments in all possible ways is given by

$$N_p[\{q_\lambda\}, N] = \frac{\left(\sum_\lambda q_\lambda\right)!}{\prod_\lambda (q_\lambda!)} [1 + \sum_\lambda q_\lambda] \times \binom{N - \sum_\lambda (\lambda - 1)q_\lambda - p - 1}{\sum_\lambda q_\lambda - 1}. \tag{2}$$

In Fig. 1 we see, for example, that if $q_1 = 1$, $q_2 = 2$, and $q_3 = 2$ and $N = 14$, N_2 the number of groups consisting of exactly two adjacent vacancies arising when a particular sequence of the $\{q_\lambda\}$ is arranged in all possible ways is 30; all the possible arrangements are shown for one of these sequences in Fig. 1. There are $5!/1!2!2! = 30$ of such independent sequences so that there are a total of 900 groups of exactly two vacancies as given by Eq. 2.

Proof: For a particular sequence of particles we consider the group of p contiguous vacant compartments *and* the particle which terminates it as a unit; then there are $[N - p - 1 - \sum_\lambda (\lambda - 1)q_\lambda]$ remaining individuals consisting of $\sum_\lambda (q_\lambda) - 1$ particles and $N - p - \sum_\lambda \lambda q_\lambda$ vacancies. If we assume initially that the particles are not distinguishable among themselves but are distinguishable from the vacancies then the remaining $[N - p - 1 - \sum_\lambda (\lambda - 1)q_\lambda]$ individuals may be arranged in

$$\frac{\left(N - p - 1 - \sum_\lambda (\lambda - 1)q_\lambda\right)!}{\left(\sum_\lambda (q_\lambda) - 1\right]![N - p - \sum_\lambda \lambda q_\lambda)!} =$$

Copyright © 1973 by the American Institute of Physics

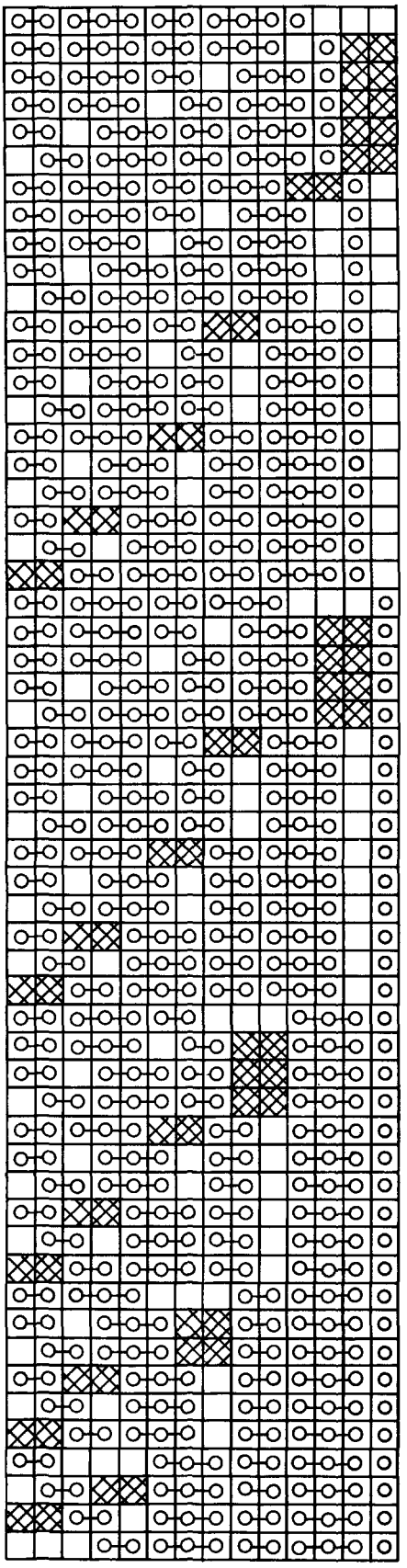


FIG. 1. All the possible independent arrangements of a particular sequence composed of the set of particles $\{q_\lambda\} = \{q_1 = 1, q_2 = 2$ and $q_3 = 3\}$ on a one-dimensional lattice space having $N = 14$ sites. There are 56 independent arrangements possible for this particular sequence. The number of groups consisting of exactly 2 adjacent vacancies created when this particular sequence of particles is arranged in all possible ways is shown to be 30 [see cross-hatched compartments]

$$\times \binom{N - \sum_\lambda (\lambda - 1)q_\lambda - p - 1}{\sum_\lambda (q_\lambda) - 1}$$

ways. However, the unit consisting of the p contiguous vacancies plus the terminating particle can be placed between the $\sum_\lambda q_\lambda$ particles in $(1 + \sum_\lambda q_\lambda)$ ways. Now, the particles characterized by a particular value of λ are distinguishable from particles having a different value of λ so that the $\sum_\lambda q_\lambda$ particles can be arranged in $(\sum_\lambda q_\lambda)!/\prod_\lambda (q_\lambda!)$ ways.

4. THE NUMBER OF GROUPS OF r CONTIGUOUS PARTICLES

Theorem 3: $N_r[\{q_\lambda\}, N]$, the number of groups of r $(r = 0, 1, 2, \ldots, \sum q_\lambda)$ contiguous particles arising when $\{q_\lambda\}$ is placed on a one-dimensional lattice space of N compartments in all possible ways is given by

$$N_r[\{q_\lambda\}, N] = \frac{\left(\sum_\lambda q_\lambda\right)!}{\prod_\lambda (q_\lambda!)} \left(N + 1 - \sum_\lambda \lambda q_\lambda\right) \times \binom{N - \sum_\lambda (\lambda - 1)q_\lambda - r - 1}{\sum_\lambda q_\lambda - r} . \quad (3)$$

If $q_1 = 1$, $q_2 = 2$, and $q_3 = 2$ and $N = 14$ as shown in Fig. 2, N_2 the number of groups of exactly two adjacent particles arising when *a particular sequence* of $\{q_\lambda\}$ is arranged in all possible ways is 40. All the possible arrangements are shown for one of these sequences in Fig. 2. There are $5!/1!2!2! = 30$ of such independent sequences so that there are a total of 1200 groups of exactly two particles, in accordance with Eq. 3.

Proof: For a particular sequence of particles we consider the group consisting of a sum of r adjacent particles *and* the vacancy which terminates the sum as a unit. There are a total of $N - \sum_\lambda (\lambda - 1)q_\lambda - r - 1$ remaining individuals consisting of $N - \sum_\lambda \lambda q_\lambda - 1$ vacancies and $\sum_\lambda (q_\lambda) - r$ particles. These individuals may be permuted in

$$\frac{\left[N - \sum_\lambda (\lambda - 1)q_\lambda - r - 1\right]!}{\left(\sum_\lambda (q_\lambda) - r\right)!\left(N - \sum_\lambda \lambda q_\lambda - 1\right)!} = \times \binom{N - \sum_\lambda (\lambda - 1)q_\lambda - r - 1}{\sum_\lambda (q_\lambda) - r}$$

ways. The unit consisting of r contiguous particles plus the terminating vacancy can be placed between the $N - \sum_\lambda \lambda q_\lambda$ vacancies in $N + 1 - \sum_\lambda \lambda q_\lambda$ ways. As before, a particle characterised by a particular value of λ is distinguishable from particles having a different value of λ but not from particles having the

same value of λ. Thus the $\sum_\lambda q_\lambda$ particles can be arranged in $(\sum_\lambda q_\lambda)!/\prod_\lambda (q_\lambda!)$ independent ways.

5. STATISTICAL PROBABILITY OF SUCCESS

For a given $\{q_\lambda\}$ and $N, S_d[\{q_\lambda\}, N]$ the statistical probability of success for a particle occupying d contiguous sites is defined to be the ratio of the number of ways of placing the particle on the $A[\{q_\lambda\}, N]$ arrays, to the total number of spatially independent attempts which can be made to place the particle of length d on that set of arrays.

As we have pointed out previously[2] when one is treating particles for which $\lambda \geq 2$ there is a profound difference in the statistical and kinetic probability of success. In this section we calculate $S_d[\{q_\lambda\}, N]$ the statistical probability of success for the situation under discussion.

Given a $\{q_\lambda\}$, arranged on a one-dimensional lattice space of N sites, what is the ensemble average probability of placing an additional particle of length d when the attempts to place the particle on the lattice space are performed in a spatially random manner?

To answer this question we have recourse to Eq. 2 and the following reasoning: A group of p contiguous vacancies will accommodate a particle occupying d contiguous lattice sites in $[p - d + 1]$ ways if $d \leq p$. If there are $N_p[\{q_\lambda\}, N]$ of such groups then $[p - d + 1]N_p[\{q_\lambda\}, N]$ is the number of ways of placing the d-bell on all of the groups of p contiguous vacancies. The total number of successes is then obtained by summing over all appropriate values of p, i.e., $d \leq p \leq N - \sum_\lambda \lambda q_\lambda$. The total number of attempts is $(N - d + 1)A[\{q_\lambda\}, N]$ because $(N - d + 1)$ attempts can be made to place the particles of length d on each of the $A[\{q_\lambda\}, N]$ arrangements, i.e.,

$$S_d[\{q_\lambda\}, N] = \frac{\prod_\lambda (q_\lambda!)\,(N - \sum_\lambda \lambda q_\lambda)!(1 + \sum_\lambda q_\lambda)!}{[N - d + 1][N - \sum_\lambda (\lambda - 1)q_\lambda]!\prod_\lambda (q_\lambda!)}$$

$$\sum_{p=d}^{N-\sum_\lambda q_\lambda} (p - d + 1) \binom{N - p - 1 - \sum_\lambda (\lambda - 1)q_\lambda}{\sum_\lambda q_\lambda - 1}$$

$$= \frac{(N - \sum_\lambda \lambda q_\lambda)!(1 + \sum_\lambda q_\lambda)!}{[N - d + 1][N - \sum_\lambda (\lambda - 1)q_\lambda]!}$$

$$\sum_{p=d}^{N-\sum_\lambda \lambda q_\lambda} (p - d + 1) \binom{N - p - 1 - \sum_\lambda (\lambda - 1)q_\lambda}{\sum_\lambda q_\lambda - 1}$$

$$= \frac{(N - \sum_\lambda \lambda q_\lambda)![N - \sum_\lambda (\lambda - 1)q_\lambda - d + 1]!}{(N - d + 1)[N - \sum_\lambda (\lambda - 1)q_\lambda]!(N - \sum_\lambda \lambda q_\lambda - d)!}$$

$$= \frac{[N - \sum_\lambda (\lambda - 1)q_\lambda - d + 1]}{(N - d + 1)} \frac{\binom{N - \sum_\lambda (\lambda)q_\lambda - d}{d}}{\binom{N - \sum (\lambda - 1)q_\lambda}{d}}. \quad (4)$$

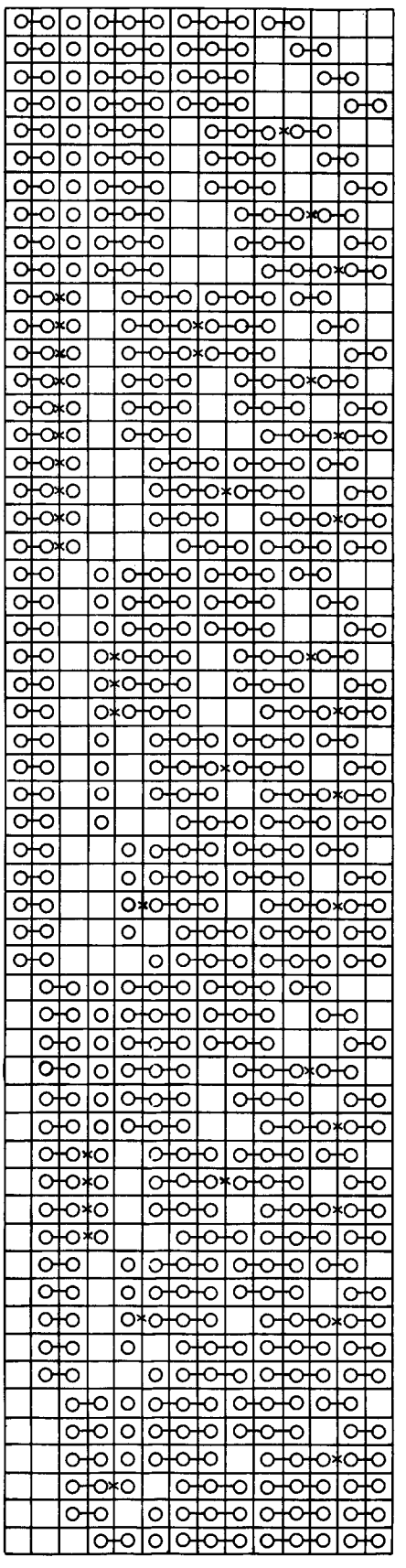


FIG. 2. For the same sequence of particles shown in Fig. 1, the number of groups consisting of exactly two adjacent particles (irrespective of the kind of particle) created when this particular sequence of particles is arranged in all possible ways is shown to be 40 (see crosses between compartments which separate the two particles).

Thus if $\{q_\lambda\}$ reduces to q_1, i.e., only simple particles are involved, and $d = 1$, then Eq. 4 yields

$$\lim_{N\to\infty} S[q_1, N] = 1 - \theta_1, \tag{5}$$

where $\theta_1 \equiv q_1/N$; if $\{q_\lambda\} = q_2$, i.e., only dumbbells are considered, then Eq. 4 yields

$$\lim_{N\to\infty} S[q_2, N] = 2(1 - \theta_2)^2/(2 - \theta_2) \tag{6}$$

where $\theta_2 \equiv 2q_2/N$. Equation (6) is in agreement with the results of the Bethe approximation[3,4,5] and previously published exact results!

ACKNOWLEDGMENT

The author wishes to express his appreciation to Dr. Ralf Vanselow for critically reading the manuscript of this paper.

[1]D. Lichtman and R. B. McQuistan, J. Math. Phys. **8**, 2441 (1969).
[2]R. B. McQuistan, Nuovo Cimento B **58**, 86 (1968).
[3]H. A. Bethe, Proc. R. Soc. A **150**, 59 (1935).
[4]J. K. Roberts, Proc. R. Soc. A **161**, 141 (1937).
[5]J. K. Roberts and A. R. Miller, Proc. Camb. Philos. Soc. **35**, 293 (1939).

Projection operator techniques for compact groups

Joseph Y. Shapiro

Department of Physics, Fordham University, Bronx, New York 10458
(Received 6 September 1972)

Projection operator techniques for reducing the Kronecker product of N irreducible representations of a compact group are developed. Among the results are a generalized Clebsch–Gordan series and generalized Wigner–Eckart theorems. The methods developed here are easily applied to any compact group provided one can (a) compute the matrix elements, between product states, of the projector onto the identity representation, and (b) select a subset of product states such that their projections form a basis for the subspace belonging to the identity representation. For the case of angular momentum projection, methods are given for solving both these problems.

1. INTRODUCTION

The reduction of a Kronecker product of N irreducible representations (IR's) of a group to a direct sum of IR's can be treated in many ways. The conventional approach[1] is to perform successive binary couplings of the IR's, using the appropriate Clebsch–Gordan coefficients at each step. Although this method is satisfactory for N small, it becomes complicated as N increases, because of the large number of couplings required. An alternative technique is to project out of a Kronecker product state that component which transforms according to any desired IR. This projection operator method has been used with considerable success in dealing with the angular momentum coupling problem[2–5] [i.e., the three dimensional rotation group $R(3)$ and its covering group $SU(2)$], especially for nuclear structure calculations.[6] The present paper is an attempt to extend the usefulness of the method in two directions: first, to include any compact group and second, to develop the projection operator equivalent of the "Racah algebra" for the reduction of matrix elements containing tensor operators.[1,7]

Section 2 contains a summary of well-known properties of the projection operators. The matrix elements of the projection operators between Kronecker product states are studied in Sec. 3. Of particular importance is Eq. (3.9), which shows that a matrix element of the projector onto any row of any IR is proportional to a matrix element, between product states with one additional factor, of P^0, the projector onto the identity representation. In Secs. 4 and 5 a "Racah algebra" using projection operators is developed. The expression for the reduction of a product of N matrix elements of IR's to a sum of matrix elements (generalized Clebsch–Gordan series), together with other relations involving products of representation matrix elements, is given in Sec. 4. Section 5 contains some generalized Wigner–Eckart theorems. The results of both sections involve coefficients containing matrix elements of P^0 between product states. The connection between the "Racah algebra" using projection operators and that of the conventional binary coupling method is discussed in both sections.

In order for the results derived in this paper to be useful for a specific compact group, two problems remain to be solved, one major, one minor. The major problem is the evaluation of the P^0 matrix elements: they replace the $3j$, $6j$, etc. symbols of the conventional coupling method. For $SU(2)$, methods of evaluating these matrix elements have been given by the author in an earlier paper,[8] hereafter referred to as I. A much simpler method is outlined in part A of sec. 6. The minor problem is to select a subset of the Kronecker product space basis such that its projection by P^0 forms a basis for the subspace corresponding to the identity representation. This problem is minor because it can always be solved by trial and error. For the case of $SU(2)$, a rule for selecting the product states is proven in part B of Sec. 6, for the more general problem of projecting onto any row of any IR.

2. PROPERTIES OF THE PROJECTION OPERATORS

Let $R, S, \cdots$ be the elements of a compact group G. The matrix elements of the IR's of G will be written as $D^j_{\lambda\mu}(R)$, where j (or some other lower case Latin letter) stands for a set of symbols specifying the IR and λ or μ (or other lower case Greek letter) each stand for a set of symbols specifying a particular basic state (row) of the IR D^j. For example, for the group $SU(3)$ of particle physics, j is a pair of integers (p, q) and λ is the set (I, I_3, Y). Since G is compact, we may (and shall) take the matrices of the IR, to be unitary, so that the orthogonality condition becomes[1]

$$\frac{[j]}{V}\int dR D^j_{\lambda\mu}(R)^* D^k_{\rho\sigma}(R) = \delta_{jk}\delta_{\lambda\rho}\delta_{\mu\sigma}. \tag{2.1}$$

Here $[j]$ is the dimension of D^j and $V = \int dR$, the volume of the group parameter space using the invariant integration of G. Each δ function in Eq. (2.1) is actually a product of a set of δ functions. In an arbitrary (in general reducible) representation of G, let O_R be the operator corresponding to the group element R. Then, in this representation, we can define a set of operators

$$P^j_{\lambda\mu} = \frac{[j]}{V}\int dR D^j_{\lambda\mu}(R)^* O_R. \tag{2.2}$$

The properties of these operators are well known,[1] so that we merely summarize them here. It follows directly from Eqs. (2.1) and (2.2) that

$$O_R P^j_{\lambda\mu} = \sum_\nu D^j_{\nu\lambda}(R) P^j_{\nu\mu}, \tag{2.3}$$

$$P^j_{\lambda\mu} O_R = \sum_\nu D^j_{\mu\nu}(R) P^j_{\lambda\nu}, \tag{2.4}$$

$$P^j_{\lambda\mu} P^k_{\rho\sigma} = \delta_{jk}\delta_{\mu\rho} P^j_{\lambda\sigma}, \tag{2.5}$$

$$P^j_{\lambda\mu} O_R P^k_{\rho\sigma} = \delta_{jk} D^j_{\mu\rho}(R) P^j_{\lambda\sigma}, \tag{2.6}$$

and that

$$P^j_{\lambda\mu}\psi^k_\nu = \delta_{jk}\delta_{\mu\nu}\psi^k_\lambda, \tag{2.7}$$

where $\psi^k_\nu, \psi^k_\lambda$ are basic vectors ν and λ, respectively, of an IR D^k. Since G is compact, our representation is

Copyright © 1973 by the American Institute of Physics

completely reducible, i.e., every vector in the representation space is a linear combination of basic vectors of IR's. Then it follows from Eq. (2.7) that the adjoint of Eq. (2.2) is given by

$$P^{j\dagger}_{\lambda\mu} = P^{j}_{\mu\lambda} \tag{2.8}$$

and that

$$\sum_{j,\lambda} P^{j}_{\lambda\lambda} = I, \tag{2.9}$$

where I is the identity operator of the representation.

As a special case, Eq. (2.5) gives

$$P^{j}_{\lambda\lambda} P^{k}_{\mu\mu} = \delta_{jk}\delta_{\lambda\mu} P^{j}_{\lambda\lambda}, \tag{2.10}$$

from which it follows that the "diagonal" operators $P^{j}_{\lambda\lambda}$ are idempotent and Eq. (2.9) is a resolution of the identity into projectors.

Let ψ be an arbitrary vector in the representation space. Then Eq. (2.3) states that the vectors $P^{j}_{\lambda\mu}\psi$, for λ varying but j and μ fixed, either all vanish or form a basis for the IR D^j. Different values of μ will, in general, lead to different bases. It follows that the component $P^{j}_{\lambda\lambda}$ in Eq. (2.9) is the projector onto the entire subspace which transforms like row λ of D^j (D^j can appear more than once in the decomposition). Furthermore, since Eq. (2.5) gives as a special case

$$P^{j}_{\lambda\mu} = P^{j}_{\lambda\mu} P^{j}_{\mu\mu},$$

one can think of the general operator $P^{j}_{\lambda\mu}$ as acting in two steps: it first projects out of ψ its component which transforms as row μ of D^j and then changes this projection into its partner $P^{j}_{\lambda\mu}$ belonging to row λ of D^j [recall Eq. (2.7)]. The operators $P^{j}_{\lambda\mu}$, for $\lambda \neq \mu$, are sometimes called "transfer" operators.

The Hermitian operators

$$P^{j} = \sum_{\lambda} P^{j}_{\lambda\lambda} = \frac{[j]}{V}\int dR\, \chi^{j}(R)^{*} O_R, \tag{2.11}$$

where χ^j is the character of D^j, are also projectors, onto the entire subspace transforming like vectors of D^j (not just onto a particular row). From Eqs. (2.9) and (2.10), they satisfy

$$P^{j}P^{k} = \delta_{jk}P^{j} \tag{2.12}$$

and

$$\sum_{j} P^{j} = I. \tag{2.13}$$

Equations (2.3) and (2.4) give

$$[O_R, P^j] = 0, \tag{2.14}$$

i.e., P^j is invariant under the group.

We shall be especially interested in the projector

$$P^0 = P^{0}_{00} = \frac{1}{V}\int dR\, O_R \tag{2.15}$$

onto the identity representation. From Eqs. (2.3) and (2.14), P^0 satisfies

$$P^0 = O_R P^0 = P^0 O_R \,. \tag{2.16}$$

3. MATRIX ELEMENTS OF $P^{j}_{\lambda\mu}$ BETWEEN PRODUCT STATES

We now consider the important case where the representation is a direct product of IR's of G. The basic vectors of the direct product representation of N factors will be denoted by

$$|\mathbf{j}_N \boldsymbol{\lambda}_N\rangle = |j_1\lambda_1\rangle|j_2\lambda_2\rangle\cdots|j_N\lambda_N\rangle, \tag{3.1}$$

where the factor $|j_i\lambda_i\rangle$ is a member of an orthonormal basis of the IR D^{j_i}. The effect of $P^{j}_{\lambda\mu}$ on a product state can be expressed in terms of the matrix representation of $P^{j}_{\lambda\mu}$ between these states. Orthonormality of the product states and Eq. (2.2) gives, for a general matrix element

$$\begin{aligned}
&\langle \mathbf{j}_N \boldsymbol{\lambda}_N | P^{j}_{\lambda\mu} | \mathbf{k}_N \boldsymbol{\mu}_N\rangle \\
&= \frac{[j]}{V}\int dR D^{j}_{\lambda\mu}(R)^{*}\langle \mathbf{j}_N\boldsymbol{\lambda}_N | O_R | \mathbf{k}_N\boldsymbol{\mu}_N\rangle \\
&= \sum_{\boldsymbol{\nu}_N}\frac{[j]}{V}\int dR D^{j}_{\lambda_1\mu_1}(R)^{*}\prod_{i=1}^{N}[D^{j_i}_{\nu_i\mu_i}(R)\delta_{j_ik_i}\delta_{\lambda_i\nu_i}] \\
&= \prod_{i=1}^{N}\delta_{j_ik_i}\cdot\frac{[j]}{V}\int dR D^{j}_{\lambda\mu}(R)^{*}\prod_{i=1}^{N} D^{j_i}_{\lambda_i\mu_i}(R),
\end{aligned} \tag{3.2}$$

where the notation $\sum_{\nu_N}$ means a sum over $\nu_1, \nu_2, \cdots, \nu_N$.

For the special case of projection onto the identity IR, Eq. (3.2) becomes

$$\langle \mathbf{j}_N\boldsymbol{\lambda}_N | P^0 | \mathbf{k}_N\boldsymbol{\mu}_N\rangle = \prod_{i=1}^{N}\delta_{j_ik_i}\cdot\frac{1}{V}\int dR\prod_{i=1}^{N} D^{j_i}_{\lambda_i\mu_i}(R). \tag{3.3}$$

For unitary representations, the complex conjugate and adjoint representations are identical. In this case, one usually selects a basis $|\bar{j}\bar{\lambda}\rangle$ for the IR $D^{\bar{j}}$ adjoint to D^j such that

$$D^{j}_{\lambda\mu}(R)^{*} = \eta(j\lambda\mu)D^{\bar{j}}_{\bar{\lambda}\bar{\mu}}(R), \tag{3.4}$$

where $\eta(j\lambda\mu)$ is a phase factor satisfying

$$\eta(j\lambda\mu)\eta(j\lambda\mu)^{*} = 1, \tag{3.5}$$

$$\eta(\bar{j}\bar{\lambda}\bar{\mu}) = \eta(j\lambda\mu), \tag{3.6}$$

$$\eta(j\lambda\mu)\eta(j\mu\nu) = \eta(j\lambda\nu). \tag{3.7}$$

Equation (3.6) follows from the requirement that the adjoint of the adjoint give back the original IR; Equation (3.7) ensures that the matrices $D^{\bar{j}}$ also form a representation. As an example, the usual choice of phases for $SU(2)$ gives[9]

$$D^{j\,*}_{\lambda\mu} = (-1)^{\lambda-\mu}D^{j}_{-\lambda-\mu}, \tag{3.8}$$

so that $\bar{j} = j$ (all IR's are self-adjoint), $\bar{\lambda} = -\lambda$, $\bar{\mu} = -\mu$, and $\eta = (-1)^{\lambda-\mu}$.

Assuming Eq. (3.4), substitution into the integral on the right of Eq. (3.2) shows that it is of the same form as the integral in Eq. (3.3), but with $N+1$, instead of N factors. One obtains

$$\langle \mathbf{j}_N\boldsymbol{\lambda}_N | P^{j}_{\lambda\mu} | \mathbf{k}_N\boldsymbol{\mu}_N\rangle = \eta(j\lambda\mu)[j]\langle \mathbf{j}_N\boldsymbol{\lambda}_N, \bar{j}\bar{\lambda} | P^0 | \mathbf{k}_N\boldsymbol{\mu}_N, \bar{j}\bar{\mu}\rangle, \tag{3.9}$$

where

$$|\mathbf{j}_N\boldsymbol{\lambda}_N, \bar{j}\bar{\lambda}\rangle = |\mathbf{j}_N\boldsymbol{\lambda}_N\rangle|\bar{j}\bar{\lambda}\rangle \tag{3.10}$$

is a direct product of $N+1$ factors, the extra term being a basic state for the adjoint representation [selected so that Eq. (3.4) holds]. Eq. (3.9) shows that projecting (by $P^{j}_{\lambda\mu}$) a direct product of N factors is

equivalent to projecting a direct produce of $N+1$ factors onto the identity representation. All results of the projection operator method can be expressed in terms of matrix elements of P^0, rather than the more complicated matrix elements of $P^j_{\lambda\mu}$. For the group $SU(2)$, Eq. (3.9) has been derived in I (after a long calculation) for the special case of the "diagonal" operators $P^j_{\lambda\lambda}$.

We note the following properties of the matrix elements of P^0.

1. Equation (3.3) shows that they are diagonal in the individual sets of quantum numbers j_i, i.e., projecting a direct product does not change the IR's of its components. From Eq. (3.2), this is also true of matrix elements of $P^j_{\lambda\mu}$. Henceforth, we shall only consider elements diagonal in the j_i.

2. Equation (3.3) also shows that the matrix elements are unchanged if the same permutation of the factors (of the indices i) is performed on both sides. In other words, any permutation T of the factors in the direct product commutes with P^0

$$[P^0, T] = 0. \tag{3.11}$$

3. It follows from Eqs. (3.3) and (3.4) and the fact that P^0 is Hermitian that

$$\langle \mathbf{j}_N \boldsymbol{\lambda}_N | P^0 | \mathbf{j}_N \boldsymbol{\mu}_N \rangle^* = \langle \mathbf{j}_N \boldsymbol{\mu}_N | P^0 | \mathbf{j}_N \boldsymbol{\lambda}_N \rangle = \prod_{i=1}^N \eta(j_i \lambda_i \mu_i) \langle \bar{\mathbf{j}}_N \bar{\boldsymbol{\lambda}}_N | P^0 | \bar{\mathbf{j}}_N \bar{\boldsymbol{\mu}}_N \rangle. \tag{3.12}$$

As a special case, diagonal matrix elements ($\lambda_i = \mu_i$) are real and equal to the matrix elements in which each factor is replaced by the corresponding factor in the adjoint representation. [The phases vanish by Eq. (3.7), which implies $\eta(j\lambda\lambda) = 1$].

4. Some of the quantum numbers specifying the row in an IR are additive [e.g., I_3 and Y for $SU(3)$]. An additive quantum number is an eigenvalue of a generator Ω satisfying

$$\Omega = \sum_{i=1}^N \Omega_i, \tag{3.13}$$

where

$$\Omega_i | j_i \lambda_i \rangle = \omega_i | j_i \lambda_i \rangle. \tag{3.14}$$

Using Eq. (2.16) with $O_R = \exp(i\theta\Omega)$ in a matrix element shows that the matrix elements of P^0 vanish unless the additive quantum numbers sum to zero on each side.

For particular groups, other relationships between matrix elements may be derived by using Eq. (2.16) [or its matrix representation, Eq. (4.4) of the next section] for selected operators O_R. An example of this will be given in Sec. 6 for the group $SU(2)$.

4. RELATIONS INVOLVING PRODUCTS OF MATRIX ELEMENTS OF IRREDUCIBLE REPRESENTATIONS

Complete reducibility of a direct product representation leads to an expansion of the form

$$\prod_{i=1}^N D^{j_i}_{\lambda_i \mu_i}(R) = \sum_{j,\lambda,\mu} c^j_{\lambda\mu} D^j_{\lambda\mu}(R)^*.$$

The orthogonality relation (2.1) and Eq. (3.3) determine the coefficients, giving

$$\prod_{i=1}^N D^{j_i}_{\lambda_i \mu_i}(R) = \sum_{j,\lambda,\mu} [j] \langle \mathbf{j}_N \boldsymbol{\lambda}_N, j\lambda | P^0 | \mathbf{j}_N \boldsymbol{\mu}_N, j\mu \rangle D^j_{\lambda\mu}(R)^*, \tag{4.1}$$

which expresses the product of an arbitrary number of D's as a sum of D's. This is a generalization of the usual Clebsch–Gordan series for reducing the product of two D's.

Another expression involving products of D's is obtained from the matrix element, between product states, of Eq. (2.4). This gives

$$\sum_\nu D^j_{\mu\nu}(R) \langle \mathbf{j}_N \boldsymbol{\lambda}_N | P^j_{\lambda\nu} | \mathbf{j}_N \boldsymbol{\nu}_N \rangle = \sum_{\mu_N} \langle \mathbf{j}_N \boldsymbol{\lambda}_N | P^j_{\lambda\mu} | \mathbf{j}_N \boldsymbol{\mu}_N \rangle \prod_{i=1}^N D^{j_i}_{\mu_i \nu_i}(R). \tag{4.2}$$

Replacing D^j by its adjoint representation and using Eqs. (3.4) to (3.7) and Eq. (3.9), one obtains an equation in terms of matrix elements of P^0,

$$\sum_\nu D^j_{\mu\nu}(R) \langle \mathbf{j}_N \boldsymbol{\lambda}_N, j\lambda | P^0 | \mathbf{j}_N \boldsymbol{\nu}_N, j\nu \rangle = \sum_{\mu_N} \langle \mathbf{j}_N \boldsymbol{\lambda}_N, j\lambda | P^0 | \mathbf{j}_N \boldsymbol{\mu}_N, j\mu \rangle \prod_{i=1}^N D^{j_i}_{\mu_i \nu_i}(R). \tag{4.3}$$

Note that the phase factors have cancelled out in Eq. (4.3).

For a group such as $SU(2)$, where all quantum numbers in the set ν are additive, the left-hand sides of Eqs. (4.2) and (4.3) reduce to the single term $\nu = -\sum_i \nu_i$, so that these equations express a single D as a sum of products of D's.

Letting $j = 0$ (identity IR) in Eq. (4.2) or Eq. (4.3) gives

$$\sum_{\mu_N} \langle \mathbf{j}_N \boldsymbol{\lambda}_N | P^0 | \mathbf{j}_N \boldsymbol{\mu}_N \rangle \prod_{i=1}^N D^{j_i}_{\mu_i \nu_i}(R) = \langle \mathbf{j}_N \boldsymbol{\lambda}_N | P^0 | \mathbf{j}_N \boldsymbol{\nu}_N \rangle. \tag{4.4}$$

This relation also follows from the matrix element of Eq. (2.16).

The matrix element of Eq. (2.6), with $k = j$, leads to

$$D^j_{\mu\rho}(R) \langle \mathbf{j}_N \boldsymbol{\lambda}_N | P^j_{\lambda\sigma} | \mathbf{j}_N \boldsymbol{\sigma}_N \rangle = \sum_{\mu_N, \rho_N} \langle \mathbf{j}_N \boldsymbol{\lambda}_N | P^j_{\lambda\mu} | \mathbf{j}_N \boldsymbol{\mu}_N \rangle \prod_{i=1}^N D^{j_i}_{\mu_i \rho_i}(R) \times \langle \mathbf{j}_N \boldsymbol{\rho}_N | P^j_{\rho\sigma} | \mathbf{j}_N \boldsymbol{\sigma}_N \rangle. \tag{4.5}$$

Replacing D^j by its adjoint and using Eqs. (3.4) to (3.7) and Eq. (3.9) leads to

$$D^j_{\mu\rho}(R)^* \langle \mathbf{j}_N \boldsymbol{\lambda}_N, j\lambda | P^0 | \mathbf{j}_N \boldsymbol{\sigma}_N, j\sigma \rangle = [j] \sum_{\mu_N, \rho_N} \langle \mathbf{j}_N \boldsymbol{\lambda}_N, j\lambda | P^0 | \mathbf{j}_N \boldsymbol{\mu}_N, j\mu \rangle \times \prod_{i=1}^N D^{j_i}_{\mu_i \rho_i}(R) \langle \mathbf{j}_N \boldsymbol{\rho}_N, j\rho | P^0 | \mathbf{j}_N \boldsymbol{\sigma}_N, j\sigma \rangle. \tag{4.6}$$

Again the phase factors cancel. Eq. (4.6) gives, for any compact group, a single D as a sum of products of D's.

Other relations involving products of D's follow from Eqs. (4.1) to (4.6) by using the unitarity of the representation matrices. We omit the details.

Equations (4.1), (4.3), (4.4), and (4.6) are generalizations of known results,[1] involving products of two or three D's, which arise out of the conventional binary coupling method. To see the relationship note that, acting in the direct product space $D^{j_1} \times D^{j_2} \times \cdots \times D^{j_N}$, one can write

$$P^0 = \sum_\alpha | \mathbf{j}_N \alpha \rangle \langle \mathbf{j}_N \alpha |, \tag{4.7}$$

where the vectors $|\mathbf{j}_N\alpha\rangle$ are a complete orthonormal set of vectors arising from the coupling of the product to give the identity representation. The label α distinguishes different members of this set. For the matrix elements, one has

$$\langle \mathbf{j}_N\boldsymbol{\lambda}_N|P^0|\mathbf{j}_N\boldsymbol{\mu}_N\rangle = \sum_\alpha \langle \mathbf{j}_N\alpha|\mathbf{j}_N\boldsymbol{\lambda}_N\rangle^*\langle \mathbf{j}_N\alpha|\mathbf{j}_N\boldsymbol{\mu}_N\rangle. \quad (4.8)$$

For the special case of the direct product of three factors, the recoupling coefficients on the right of Eq. (4.8) are 3-j symbols

$$\langle j_1j_2j_3\alpha|j_1\mu_1,j_2\mu_2,j_3\mu_3\rangle = \begin{pmatrix} j_1j_2j_3 & \alpha \\ \mu_1\mu_2\mu_3 & \end{pmatrix} \quad (4.9)$$

and Eq. (4.8) becomes

$$\langle j_1\lambda_1,j_2\lambda_2,j_3\lambda_3|P^0|j_1\mu_1,j_2\mu_2,j_3\mu_3\rangle = \sum_\alpha \begin{pmatrix} j_1j_2j_3\alpha \\ \lambda_1\lambda_2\lambda_3 \end{pmatrix}^* \begin{pmatrix} j_1\,j_2\,j_3\alpha \\ \mu_1\mu_2\mu_3 \end{pmatrix}. \quad (4.10)$$

For a simply reducible group, such as $SU(2)$, the sum in Eq. (4.10) reduces to a single term.

Eq. (4.8) enables us to express our results in terms of recoupling coefficients, rather than P^0 matrix elements. Substitution into Eqs. (3.3) and (4.1) gives

$$\frac{1}{V}\int dR \prod_{i=1}^N D^{j_i}_{\lambda_i\mu_i}(R) = \sum_\alpha \langle \mathbf{j}_N\alpha|\mathbf{j}_N\boldsymbol{\lambda}_N\rangle^*\langle \mathbf{j}_N\alpha|\mathbf{j}_N\boldsymbol{\mu}_N\rangle \quad (4.11)$$

and

$$\prod_{i=1}^N D^{j_i}_{\lambda_i\mu_i}(R) = \sum_{j,\lambda,\mu,\alpha} [j]\langle \mathbf{j}_N j\alpha|\mathbf{j}_N\boldsymbol{\lambda}_N,j\lambda\rangle^* \times D^j_{\lambda\mu}(R)^*\langle \mathbf{j}_N j\alpha|\mathbf{j}_N\boldsymbol{\mu}_N,j\mu\rangle. \quad (4.12)$$

Upon substitution of Eq. (4.9), one obtains the known results in terms of 3-j symbols from Eq. (4.11) with $N = 3$ and Eq. (4.12) with $N = 2$.

Direct substitution of Eq. (4.8) into Eqs. (4.3) and (4.6) gives overly complicated results, so we proceed somewhat differently. Recall that these equations are matrix representations of the operator equations (2.4) and (2.6), a fact that is partially obscured by the change to matrix elements of P^0. Consider Eq. (4.3). The P^0 matrix elements are diagonal in the individual j_i, so that Eq. (4.3) is valid for an arbitrary bra vector $\langle \mathbf{j}_N\mu_N,j\lambda|$ and thus equivalent to the vector equation

$$\sum_\nu P^0|\mathbf{j}_N\nu_N,j\nu\rangle D^j_{\mu\nu}(R)^* = \sum_{\mu_N} P^0|\mathbf{j}_N\mu_N,j\mu\rangle \prod_{i=1}^N D^{j_i}_{\mu_i\nu_i}(R). \quad (4.13)$$

Taking the scalar product with any state coupled to give the identity representation leads to a relation involving recoupling coefficients, since P^0 acting to the left on such a state leaves it unchanged. For example, using one of the states introduced in Eq. (4.7) gives

$$\sum_\nu \langle \mathbf{j}_N j\alpha|\mathbf{j}_N\boldsymbol{\nu}_N,j\nu\rangle D^j_{\mu\nu}(R)^* = \sum_{\mu_N}\langle \mathbf{j}_N j\alpha|\mathbf{j}_N\boldsymbol{\mu}_N,j\mu\rangle \prod_{i=1}^N D^{j_i}_{\mu_i\nu_i}(R). \quad (4.14)$$

Setting $j = 0$ gives the analog of Eq. (4.4)

$$\sum_{\mu_N}\langle \mathbf{j}_N\alpha|\mathbf{j}_N\boldsymbol{\mu}_N\rangle \prod_{i=1}^N D^{j_i}_{\mu_i\nu_i}(R) = \langle \mathbf{j}_N\alpha|\mathbf{j}_N\boldsymbol{\nu}_N\rangle. \quad (4.15)$$

Equation (4.6) is equivalent to the operator equation

$$D^j_{\mu\rho}(R)^*P^0 = [j]\sum_{\mu_N,\rho_N} P^0|\mathbf{j}_N\boldsymbol{\mu}_N,j\mu\rangle \times \prod_{i=1}^N D^{j_i}_{\mu_i\rho_i}(R)\cdot\langle \mathbf{j}_N\boldsymbol{\rho}_N,j\rho|P^0 \quad (4.16)$$

leading to

$$D^j_{\mu\rho}(R)^*\delta_{\alpha\beta} = [j]\sum_{\mu_N,\rho_N}\langle \mathbf{j}_N j\alpha|\mathbf{j}_N\boldsymbol{\mu}_N,j\mu\rangle \times \prod_{i=1}^N D^{j_i}_{\mu_i\rho_i}(R)\cdot\langle \mathbf{j}_N j\beta|\mathbf{j}_N\boldsymbol{\rho}_N,j\rho\rangle. \quad (4.17)$$

Setting $N = 2$ in Eqs. (4.14) and (4.17) and $N = 3$ in Eq. (4.15) leads to known results in terms of 3-j symbols.[1]

In summary, we have generalized the known results involving products of two or three D's to the case of N D's. The results have been given in two forms—the first, involving matrix elements of P^0 between product states and the second, recoupling coefficients between product states and any complete orthonormal set of states coupled to give the identity representation. Which form is to be preferred depends upon the relative difficulty of computing these two sets of coefficients. For $SU(2)$, results similar to ours with recoupling coefficients have been obtained by Kumar.[10] His coefficients differ from ours. Rather than using an orthonormal set of vector coupled states, as here, Kumar has constructed a set of symmetrically coupled states which, though complete, are not independent. He has given both a generating function and an explicit formula for his coefficients. Formulas for calculating the P^0 matrix elements for $SU(2)$ are given in I. An alternate derivation is given in Sec. 6. Extensive tables of P^0 matrix elements for $SU(2)$ have been computed by Miranda and Reith.[11] For $SU(3)$, polynomial expressions for P^0 in terms of the group generators have been obtained by Asherova and Smirnov[12] and by Noz.[13] The latter author has used her expression to calculate P^0 matrix elements for some cases of interest in particle physics. More generally, Schott[14] has obtained polynomial expressions for P^0 in terms of the generators for $SU(n)$. These could be used to evaluated matrix elements. Experience with the unitary groups suggests that the expressions with projection operators are more useful than those with recoupling coefficients, except possibly for those cases ($N = 2$ or $N = 3$) where 3-j symbols are involved.

Finally, other things being equal, Eqs. (3.3) and (4.1) (projection operators) are preferable to Eqs. (4.11) and (4.12) (recoupling coefficients) because the former involve one less sum than the latter.

5. WIGNER-ECKART TYPE THEOREMS

In this section we consider the problem of reducing (i.e., of factoring out the dependence on the sets of indices specifying the rows of IR's of our group G) matrix elements containing tensor operators. We define a tensor operator $T(k)$ belonging to an IR D^k to be a set of $[k]$ operators $T(k,\sigma)$ which transform as

$$O_R T(k,\sigma)O_R^{-1} = \sum_\tau D^k_{\tau\sigma}(R)T(k,\tau). \quad (5.1)$$

Two types of matrix elements will be examined. First, we consider matrix elements which differ from those of the conventional Wigner–Eckart theorem only by having more than three factors, i.e., they contain a product of an arbitrary number of tensor operators be-

tween product states with an arbitrary number of factors. We call these "conventional matrix elements." Second, we look at matrix elements which contain projection operators as well as tensor operators. These arise in any calculation in which one constructs wave functions using projection operators, rather than conventional binary coupling techniques.

A. Conventional matrix elements

Consider matrix elements of the form

$$\langle \mathbf{n}'_S \mathbf{j}'_S \mu'_S \,|\, \prod_{q=1}^{R} T_q(k_q, \sigma_q) \,|\, \mathbf{n}_N \mathbf{j}_N \mu_N \rangle, \tag{5.2}$$

i.e., of a product of R tensor operators between two product states, of S and N factors, respectively. The individual factors in the product states, and the individual tensor operators, each belong to rows of IR's of G. The sets $\mathbf{n}'_S, \mathbf{n}'_N$ stand for other quantum numbers, not associated with the group G, required to specify the factors in the product states. These individual factors could themselves be vector coupled states. We require, of course, that both product states in Eq. (5.2) are vectors in the same space and that the product of tensor operators acts in this space. Note that, for the special case of no tensor operators (identity operator in the space), Eq. (5.2) includes the recoupling coefficients introduced in the last section. It also includes, for $SU(2)$, such quantities as the transformation matrix from LS to jj coupling.

From Eq. (5.1), the product of tensor operators in Eq. (5.2) transforms as

$$\prod_{q=1}^{R} T_q(k_q, \sigma_q) = O_R^{-1} \sum_{\tau_R} \prod_{q=1}^{R} D^{k_q}_{\tau_q \sigma_q}(R) T_q(k_q, \tau_q) \cdot O_R .$$

Substituting into Eq. (5.2) and letting O_R^{-1} and O_R operate on the product states gives

$$\begin{aligned}
&\langle \mathbf{n}'_S \mathbf{j}'_S \mu'_S \,|\, \prod_{q=1}^{R} T_q(k_q, \sigma_q) \,|\, \mathbf{n}_N \mathbf{j}_N \mu_N \rangle \\
&\quad = \sum_{\boldsymbol{\nu}'_S, \tau_R, \boldsymbol{\nu}_N} \prod_{i=1}^{S} D^{j_i j'_i}_{\nu_i \mu'_i}(R)^* \prod_{q=1}^{R} D^{k_q}_{\tau_q \sigma_q}(R) \prod_{p=1}^{N} D^{j_p}_{\nu_p \mu_p}(R) \\
&\qquad \times \langle \mathbf{n}'_S \mathbf{j}'_S \boldsymbol{\nu}'_S \,|\, \prod_{q=1}^{R} T_q(k_q, \tau_q) \,|\, \mathbf{n}_N \mathbf{j}_N \boldsymbol{\nu}_N \rangle .
\end{aligned}$$

Integrating over the group volume and using Eqs. (3.3) and (3.4) gives

$$\begin{aligned}
&\langle \mathbf{n}'_S \mathbf{j}'_S \mu'_S \,|\, \prod_{q=1}^{R} T_q(k_q, \sigma_q) \,|\, \mathbf{n}_N \mathbf{j}_N \mu_N \rangle \\
&\quad = \sum_{\boldsymbol{\nu}'_S, \tau_R, \boldsymbol{\nu}_N} \left[\prod_{i=1}^{S} \eta(j'_i \nu'_i \mu'_i) \right] \langle \bar{\mathbf{j}}'_S \bar{\boldsymbol{\nu}}'_S, \mathbf{k}_R, \tau_R, \mathbf{j}_N \boldsymbol{\nu}_N \,|\, P^0 \,| \\
&\qquad \times \bar{\mathbf{j}}'_S \bar{\mu}'_S, \mathbf{k}_R \sigma_R, \mathbf{j}_N \mu_N \rangle \\
&\qquad \times \langle \mathbf{n}'_S \mathbf{j}'_S \boldsymbol{\nu}'_S \,|\, \prod_{q=1}^{R} T_q(k_q, \tau_q) \,|\, \mathbf{n}_N \mathbf{j}_N \boldsymbol{\nu}_N \rangle,
\end{aligned} \tag{5.3}$$

where

$$|\bar{\mathbf{j}}'_S \bar{\mu}'_S, \mathbf{k}_R \sigma_R, \mathbf{j}_N \mu_N \rangle = |\bar{\mathbf{j}}'_S \bar{\mu}'_S \rangle \, |\mathbf{k}_R \sigma_R \rangle \, |\mathbf{j}_N \mu_N \rangle \tag{5.4}$$

is a product state with $S + R + N$ factors.

To get a Wigner–Eckart type theorem one must factor the P^0 matrix element into a sum of terms, each of which is a product of two factors, one independent of the set $\{\boldsymbol{\nu}'_S, \tau_R, \boldsymbol{\nu}_N\}$, the other independent of $\{\boldsymbol{\mu}'_S, \sigma_R, \boldsymbol{\mu}_N\}$. The former factors come outside the sums over $\boldsymbol{\nu}'_S, \tau_R$ and $\boldsymbol{\nu}_N$. These sums are then independent of all the quantum numbers specifying rows of IR's, and become reduced matrix elements. We have assumed that the μ'_i dependence of the phase factors in Eq. (5.3) can also be factored out and taken outside the sum.

The obvious way of factoring the P^0 matrix element in Eq. (5.3) is to use Eq. (4.8), leading directly to a Wigner–Eckart type theorem involving recoupling coefficients

$$\begin{aligned}
&\langle \mathbf{n}'_S \mathbf{j}'_S \mu'_S \,|\, \prod_{q=1}^{R} T_q(k_q, \sigma_q) | \mathbf{n}_N \mathbf{j}_N \mu_N \rangle \\
&\quad = \delta(\mathbf{j}'_S, \mu'_S) \sum_{\alpha} \langle \bar{\mathbf{j}}'_S \mathbf{k}_R \mathbf{j}_N \alpha \,|\, \bar{\mathbf{j}}'_S \bar{\mu}'_S, \mathbf{k}_R \sigma_R, \mathbf{j}_N \mu_N \rangle S_\alpha ,
\end{aligned} \tag{5.5}$$

where $\delta(\mathbf{j}'_S, \mu'_S)$ is a phase factor arising from the product of phases in Eq. (5.3) and the S_α are reduced matrix elements, independent of the sets $\boldsymbol{\mu}'_S, \sigma_R$, and $\boldsymbol{\mu}_N$. In the special case of a single tensor operator between single "vector-coupled" states, the recoupling coefficient multiplying S_α becomes a 3-j symbol and one obtains the usual form of the Wigner–Eckart theorem. Apart from those cases where the recoupling coefficients reduce to 3-j symbols, Eq. (5.5) is of limited usefulness because of the difficulties involved in evaluating the recoupling coefficients.

One can derive a more useful generalization of the Wigner–Eckart theorem, involving P^0 matrix elements rather than recoupling coefficients. In Eq. (4.7) we introduced an orthonormal set of states $|\mathbf{j}_N \alpha\rangle$ arising from coupling the product states of $D^{j_1} \times D^{j_2} \times \cdots \times D^{j_N}$ to give the identity representation. The number of states in this set (number of values taken by α) equals the number of times the identity representation appears in the decomposition of the above Kronecker product representation. One can also get states in the identity representation by projecting product states. Consider the set of projections $P^0 |\mathbf{j}_N \boldsymbol{\lambda}_N\rangle$ for a fixed set $\mathbf{j}_N = (j_1, j_2, \ldots, j_N)$, but for all sets $\boldsymbol{\lambda}_N$ leading to nonzero vectors. This set spans the same space as the set $|\mathbf{j}_N \alpha\rangle$, but is in general linearly dependent. Nevertheless, it is always possible to select a subset of product states such that their projections are complete and independent, although not normalized or orthogonal. We denote such a subset of product states by

$$|\alpha\rangle = |\mathbf{j}_N \boldsymbol{\lambda}_N(\alpha)\rangle . \tag{5.6}$$

The number of product states in this subset (number of values for α) obviously equals the number of states $|\mathbf{j}_N \alpha\rangle$. For $SU(2)$, a rule for selecting the set $|\alpha\rangle$ is proven in the next section. If such rules cannot be found for other groups, one can always proceed by trial and error, i.e., by projecting products and testing for independence until one finds the required number.

Having found a set $|\alpha\rangle$, one could orthonormalize its projections and obtain an expansion of the form of Eq. (4.7). However, it is more convenient not to do this, but to proceed as follows. Define a matrix M by

$$M_{\alpha\beta} = \langle P^0 \alpha | P^0 \beta \rangle = \langle \alpha \,|\, P^0 \,|\, \beta \rangle , \tag{5.7}$$

where $|\alpha\rangle$ and $|\beta\rangle$ are any two members of the subset of product states of Eq. (5.6). Since the projections of this subset are independent, M is nonsingular. Then, one easily proves the expansion

$$P^0 = \sum_{\alpha, \beta} P^0 |\alpha\rangle (M^{-1})_{\alpha\beta} \langle \beta | P^0 , \tag{5.8}$$

where the sums go over the subset of Eq. (5.6). Acting on a product state, Eq. (5.8) gives

$$P^0 |\mathbf{j}_N \boldsymbol{\lambda}_N\rangle = \sum_{\alpha} b(\alpha \,|\, \mathbf{j}_N \boldsymbol{\lambda}_N) P^0 |\alpha\rangle , \tag{5.9}$$

where

$$b(\alpha|\mathbf{j}_N\boldsymbol{\lambda}_N) = \sum_\beta (M^{-1})_{\alpha\beta}\langle\beta|P^0|\mathbf{j}_N\boldsymbol{\lambda}_N\rangle. \quad (5.10)$$

Equations (5.9) and (5.10) give the projection of an arbitrary product state in terms of a basis $P^0|\alpha\rangle$ of independent projections. For the special case $|\mathbf{j}_N\boldsymbol{\lambda}_N\rangle = |\gamma\rangle$, one of the subset of Eq. (5.6), Eq. (5.10) gives the simple result

$$b(\alpha|\gamma) = \delta_{\alpha\gamma} \quad (5.11)$$

and Eq. (5.9) becomes an obvious identity. Since the set of projections of the states $|\gamma\rangle$ is complete, this proves Eq. (5.8).

The coefficients $b(\alpha|\mathbf{j}_N\boldsymbol{\lambda}_N)$ play an important role in calculations involving the projection operator method. Equation (5.10) shows that their evaluation does not require the entire matrix of P^0, but only that part between an arbitrary product and our subset leading to independent projections.

To obtain a Wigner–Eckart type theorem, one substitutes Eq. (5.9) into Eq. (5.3) giving

$$\langle\mathbf{n}'_S\mathbf{j}'_S\boldsymbol{\mu}'_S|\prod_{q=1}^R T_q(k_q,\sigma_q)|\mathbf{n}_N\mathbf{j}_N\boldsymbol{\mu}_N\rangle = \delta(\mathbf{j}'_S,\boldsymbol{\mu}'_S)\sum_\alpha b(\alpha|\bar{\mathbf{j}}'_S\bar{\boldsymbol{\mu}}'_S,\mathbf{k}_R\sigma_R,\mathbf{j}_N\boldsymbol{\mu}_N)R_\alpha. \quad (5.12)$$

Here δ is the same phase factor that appeared in Eq. (5.5). The reduced matrix element R_β is determined by evaluating Eq. (5.12) for the special set $\boldsymbol{\mu}'_S(\beta), \sigma_R(\beta), \boldsymbol{\mu}_N(\beta)$ corresponding to one of our independent projections. Using Eq. (5.11), this gives

$$R_\alpha = \delta(\mathbf{j}'_S,\boldsymbol{\mu}'_S(\alpha))^*\langle\mathbf{n}'_S\mathbf{j}'_S\boldsymbol{\mu}'_S(\alpha)|\prod_{q=1}^R T_q(k_q,\sigma_q(\alpha))|\mathbf{n}_N\mathbf{j}_N\boldsymbol{\mu}_N(\alpha)\rangle, \quad (5.13)$$

i.e., the reduced matrix elements are (up to a phase) just the original matrix elements for the special sets $\boldsymbol{\mu}'_S(\alpha), \sigma_q(\alpha), \boldsymbol{\mu}_N(\alpha)$ leading to independent projections. Substitution of Eq. (5.13) into Eq. (5.12) gives an extremely useful form of Wigner–Eckart theorem

$$\langle\mathbf{n}'_S\mathbf{j}'_S\boldsymbol{\mu}'_S|\prod_{q=1}^R T_q(k_q\sigma_q)|\mathbf{n}_N\mathbf{j}_N\boldsymbol{\mu}_N\rangle = \delta(\mathbf{j}'_S\boldsymbol{\mu}'_S)\sum_\alpha \delta(\mathbf{j}'_S,\boldsymbol{\mu}'_S(\alpha))^* b(\alpha|\bar{\mathbf{j}}'_S\bar{\boldsymbol{\mu}}'_S,\mathbf{k}_R\sigma_R,\mathbf{j}_N\boldsymbol{\mu}_N) \times\langle\mathbf{n}'_S\mathbf{j}'_S\boldsymbol{\mu}'_S(\alpha)|\prod_{q=1}^R T_q(k_q,\sigma_q(\alpha))|\mathbf{n}_N\mathbf{j}_N\boldsymbol{\mu}_N(\alpha)\rangle. \quad (5.14)$$

In most applications, the set $|\alpha\rangle$ is small, so that the problem of calculating the geometrical factors by Eq. (5.10) is relatively simple, provided the P^0 matrix elements are known.

B. Matrix elements containing projectors

In performing calculations using projected wave functions, one usually obtains matrix elements containing the projectors $P^j_{\lambda\mu}$ themselves. One then can do the reduction by a simple step-by-step procedure, using a generalization of Eq. (5.9). From (3.9) and (5.9), one has

$$\begin{aligned}\langle\mathbf{j}_N\boldsymbol{\lambda}_N|P^j_{\lambda\mu}|\mathbf{j}_N\boldsymbol{\mu}_N\rangle &= \eta(j\lambda\mu)[j]\langle\mathbf{j}_N\boldsymbol{\lambda}_N,\bar{j}\bar{\lambda}|P^0|\mathbf{j}_N\boldsymbol{\mu}_N\bar{j}\bar{\mu}\rangle\\ &= \eta(j\lambda\mu)[j]\sum_\alpha b(\alpha|\mathbf{j}_N\boldsymbol{\mu}_N,\bar{j}\bar{\mu})\langle\mathbf{j}_N\boldsymbol{\lambda}_N,\bar{j}\bar{\lambda}|P^0|\alpha\rangle\\ &= \sum_\alpha \eta(j\mu(\alpha)\mu)b(\alpha|\mathbf{j}_N\boldsymbol{\mu}_N,\bar{j}\bar{\mu})\langle\mathbf{j}_N\boldsymbol{\lambda}_N|\\ &\quad\times P^j_{\lambda\mu(\alpha)}|\mathbf{j}_N\boldsymbol{\mu}_N(\alpha)\rangle. \end{aligned} \quad (5.15)$$

Note that here the states $|\alpha\rangle$ are those which give complete independent projections onto the identity representation from the $N+1$ factor Kronecker product $D^{j_1}\times D^{j_2}\times\ldots\times D^{j_N}\times D^{\bar{j}}$. In deriving (5.15) we have used

$$\eta(j\lambda\mu) = \eta(j\mu\lambda)^*, \quad (5.16)$$

which follows from Eqs. (3.5) and (3.7). Equation (5.15) is equivalent to the vector equation

$$P^j_{\lambda\mu}|\mathbf{n}_N\mathbf{j}_N\boldsymbol{\mu}_N\rangle = \sum_\alpha \eta(j\mu(\alpha)\mu)b(\alpha|\mathbf{j}_N\boldsymbol{\mu}_N,\bar{j}\bar{\mu}) \times P^j_{\lambda\mu(\alpha)}|\mathbf{n}_N\mathbf{j}_N\boldsymbol{\mu}_N(\alpha)\rangle, \quad (5.17)$$

where we have explicitly included the other quantum numbers $\mathbf{n}_N$ required to uniquely specify the product state. Equation (5.17) gives the reduction of the product state and of one index of the projection operator. Its adjoint is

$$\langle\mathbf{n}_N\mathbf{j}_N\boldsymbol{\lambda}_N|P^j_{\lambda\mu} = \sum_\alpha \eta(j\lambda\lambda(\alpha))b(\alpha|\mathbf{j}_N\boldsymbol{\lambda}_N,\bar{j}\bar{\lambda})^* \times\langle\mathbf{n}_N\mathbf{j}_N\boldsymbol{\lambda}_N(\alpha)|P^j_{\lambda(\alpha)\mu}. \quad (5.18)$$

Equations (5.17) and (5.18) enable one to completely reduce matrix elements containing projectors and tensor operators betwen product states. The most important cases are

$$\begin{aligned}\langle\mathbf{n}'_S\mathbf{j}'_S\boldsymbol{\mu}'_S|P^j_{\mu'\mu}\prod_{q=1}^R T_q(k_q,\sigma_q)|\mathbf{n}_N\mathbf{j}_N\boldsymbol{\mu}_N\rangle &= \sum_{\alpha,\beta}\eta(j\mu'\mu'(\alpha))\eta(j\mu(\beta)\mu)\\ &\times b(\alpha|\mathbf{j}'_S\boldsymbol{\mu}'_S,\bar{j}\bar{\mu}')^* b(\beta|\mathbf{k}_R\sigma_R,\mathbf{j}_N\boldsymbol{\mu}_N,\bar{j}\bar{\mu})\\ &\times\langle\mathbf{n}'_S\mathbf{j}'_S\boldsymbol{\mu}'_S(\alpha)|P^j_{\mu'(\alpha)\mu(\beta)}\prod_{q=1}^R T_q(k_q,\sigma_q(\beta))|\mathbf{n}_N\mathbf{j}_N\boldsymbol{\mu}_N(\beta)\rangle\end{aligned} \quad (5.19)$$

and

$$\begin{aligned}\langle\mathbf{n}'_S\mathbf{j}'_S\boldsymbol{\mu}'_S|P^{j'}_{\mu'\lambda'}\prod_{q=1}^R T_q(k_q\sigma_q)P^j_{\lambda\mu}|\mathbf{n}_N\mathbf{j}_N\boldsymbol{\mu}_N\rangle &= \sum_{\alpha,\beta,\gamma}\eta(j'\mu'\mu'(\alpha))\eta(j'\lambda'(\beta)\lambda')\eta(j\mu(\gamma)\mu)\\ &\times b(\alpha|\mathbf{j}'_S\boldsymbol{\mu}'_S,\bar{j}'\bar{\mu}')^* b(\beta|\mathbf{k}_R\sigma_R,j\lambda,\bar{j}'\bar{\lambda}')b(\gamma|\mathbf{j}_N\boldsymbol{\mu}_N,\bar{j}\bar{\mu})\\ &\times\langle\mathbf{n}'_S\mathbf{j}'_S\boldsymbol{\mu}'_S(\alpha)|P^{j'}_{\mu'(\alpha)\lambda'(\beta)}\prod_{q=1}^R T_q(k_q\sigma_q(\beta))\\ &\times P^j_{\lambda(\beta)\mu(\gamma)}|\mathbf{n}_N\mathbf{j}_N\boldsymbol{\mu}_N(\gamma)\rangle.\end{aligned} \quad (5.20)$$

We have used the fact that in Eq. (5.19) $\Pi_q T_q(k_q,\sigma_q)\times|\mathbf{n}_N\mathbf{j}_N\boldsymbol{\mu}_N\rangle$ transforms as $|\mathbf{k}_R\sigma_R,\mathbf{j}_N\boldsymbol{\mu}_N\rangle$ and in Eq. (5.20) $\Pi_q T_q(k_q,\sigma_q)P^j_{\lambda\mu}|\rangle$ transforms as $|\mathbf{k}_R\sigma_R,j\lambda\rangle$.

6. ANGULAR MOMENTUM PROJECTION

A. Matrix elements of P^0

For the results of the previous sections to be useful, one must be able to evaluate the matrix elements of P^0. In the case of angular momentum, methods of doing this were given in I, starting with Löwdin's expansion of P^0 in terms of the raising and lowering operators J_+ and J_-.[5] Here we discuss a more direct method, using Eq. (3.3) with the known expressions[9] for the rotation matrices. In terms of Euler angles α,β,γ (β about y axis) one has

$$D^j_{mm'}(\alpha,\beta,\gamma) = e^{im\alpha}d^j_{mm'}(\beta)e^{im'\gamma}, \quad (6.1)$$

where

$$d^j_{mm'}(\beta) = D^j_{mm'}(0,\beta,0) \quad (6.2)$$

is the matrix corresponding to a rotation through β about the y axis. For angular momentum we use the symbols m and m' to label the rows of the IR's, rather than λ or μ. The possible values of j are $j = 0, 1/2, 1, 3/2, \cdots$, and m (or m') takes the $2j + 1$ possible values $m = -j, -j+1, \ldots, j-1, j$ for given j. Since we have parameterized by the Euler angles rather than the complete parameter space of $SU(2)$, the IR's corresponding to j half odd-integral are double valued. The volume element is

$$\frac{1}{V}\int dR = \frac{1}{8\pi^2}\int_0^{2\pi} d\alpha \int_0^{\pi} \sin\beta d\beta \int_0^{2\pi} d\gamma. \tag{6.3}$$

Performing the integrals over α and γ, Eq. (3.3) becomes

$$\langle \mathbf{j}_N \mathbf{m}_N | P^0 | \mathbf{j}_N \mathbf{m}'_N \rangle = \delta_{\Sigma m_i, 0}\delta_{\Sigma m'_i, 0} \cdot \frac{1}{2}\int_0^{\pi} \prod_{i=1}^{N} d^{j_i}_{m_i m'_i}(\beta) \sin\beta d\beta. \tag{6.4}$$

With the usual phase convention, the d^j are real, so that the P^0 matrix elements are real. The δ functions in Eq. (6.4) merely express the fact that J_Z is additive. Because of Eq. (3.8), Eq. (3.9) gives

$$\langle \mathbf{j}_N \mathbf{m}_N | P^j_{mm'} | \mathbf{j}_N \mathbf{m}'_N \rangle = (-1)^{m-m'}(2j+1)\langle \mathbf{j}_N \mathbf{m}_N, j-m | P^0 | \mathbf{j}_N \mathbf{m}'_N, j-m' \rangle \tag{6.5}$$

and Eq. (3.12) gives the symmetry relations

$$\langle \mathbf{j}_N \mathbf{m}_N | P^0 | \mathbf{j}_N \mathbf{m}'_N \rangle = \langle \mathbf{j}_N \mathbf{m}'_N | P^0 | \mathbf{j}_N \mathbf{m}_N \rangle = \langle \mathbf{j}_N - \mathbf{m}_N | P^0 | \mathbf{j}_N - \mathbf{m}'_N \rangle. \tag{6.6}$$

An additional symmetry may be obtained from Eq. (4.4) with O_R a rotation about the y axis through π. Recalling that[9]

$$D^j_{mm'}(0, \pi, 0) = d^j_{mm'}(\pi) = (-1)^{j-m}\delta_{m,-m'}, \tag{6.7}$$

one obtains

$$\langle \mathbf{j}_N \mathbf{m}_N | P^0 | \mathbf{j}_N \mathbf{m}'_N \rangle = (-1)^{\bar{J}}\langle \mathbf{j}_N - \mathbf{m}_N | P^0 | \mathbf{j}_N \mathbf{m}'_N \rangle = (-1)^{\bar{J}}\langle \mathbf{j}_N \mathbf{m}_N | P^0 | \mathbf{j}_N - \mathbf{m}'_N \rangle, \tag{6.8}$$

where

$$\bar{J} = \sum_{i=1}^{N} j_i. \tag{6.9}$$

Equation (6.8) is a new result which was not obtained in I. As a special case, if all j_i are integers ($j_i = l_i$) and if all $m'_i = 0$ (or all $m_i = 0$) one obtains

$$\langle \mathbf{l}_N \mathbf{m}_N | P^0 | \mathbf{l}_N 0 \rangle = (-1)^{L}\langle \mathbf{l}_N \mathbf{m}_N | P^0 | \mathbf{l}_N 0 \rangle,$$

i.e.,

$$\langle \mathbf{l}_N \mathbf{m}_N | P^0 | \mathbf{l}_N 0 \rangle = 0 \quad \text{if } L = \sum_i l_i \text{ is odd}, \tag{6.10}$$

an extension of the well-known parity rule for the Clebsch–Gordan coefficients.

We now outline how Eq. (6.4) may be used to evaluate the matrix elements. Complete details will be given in another paper, together with extensive numerical tables for the evaluation of the matrix elements. First we use the symmetry properties[9]

$$d^j_{mm'}(\beta) = (-1)^{m-m'}d^j_{-m-m'}(\beta) = (-1)^{m-m'}d^j_{m'm}(\beta) \tag{6.11}$$

to rewrite Eq. (6.4). We introduce new numbers p_i and q_i defined by

$$\begin{aligned} p_i + q_i &= |m_i + m'_i|, \\ p_i - q_i &= |m_i - m'_i| \end{aligned} \tag{6.12}$$

or, equivalently,

$$\begin{aligned} p_i &= \max(|m_i|, |m'_i|), \\ q_i &= |m_i + m'_i| - p_i = p_i - |m_i - m'_i|. \end{aligned} \tag{6.13}$$

We also will need integers K and L defined by

$$\begin{aligned} K &= \frac{1}{2}\sum_{i=1}^{N}(p_i + q_i) = \frac{1}{2}\sum_{i=1}^{N}|m_i + m'_i|, \\ \bar{L} &= \frac{1}{2}\sum_{i=1}^{N}(p_i - q_i) = \frac{1}{2}\sum_{i=1}^{N}|m_i - m'_i|. \end{aligned} \tag{6.14}$$

Using Eq. (6.11) one easily shows that Eq. (6.4) becomes

$$\langle \mathbf{j}_N \mathbf{m}_N | P^0 | \mathbf{j}_N \mathbf{m}'_N \rangle = \delta_{\Sigma_i m_i, 0}\delta_{\Sigma_i m'_i, 0} \times \frac{(-1)^{\bar{L}}}{2}\int_0^{\pi}\prod_{i=1}^{N} d^{j_i}_{p_i q_i}(\beta) \sin\beta\, d\beta. \tag{6.15}$$

From the definitions (6.12) one has

$$0 \leq |q_i| \leq p_i \leq j_i. \tag{6.16}$$

Rotation matrices satisfying Eq. (6.16) may be expressed in terms of Jacobi polynomials[9]

$$d^j_{pq}(\beta) = \left[\frac{(j+p)!(j-p)!}{(j+q)!(j-q)!}\right]^{1/2} (\cos\beta/2)^{p+q} (\sin\beta/2)^{p-q} \times P^{(p-q, p+q)}_{j-p}(\cos\beta). \tag{6.17}$$

Substituting Eq. (6.17) into (6.15), one obtains

$$\langle \mathbf{j}_N \mathbf{m}_N | P^0 | \mathbf{j}_N \mathbf{m}'_N \rangle = \delta_{\Sigma m_i, 0}\delta_{\Sigma m'_i, 0} \times \prod_{i=1}^{N}\left[\frac{(j_i + p_i)!(j_i - p_i)!}{(j_i + q_i)!(j_i - q_i)!}\right]^{1/2} \frac{(-1)^{L}}{2} \times \int_{-1}^{1}\left(\frac{1+x}{2}\right)^{K}\left(\frac{1-x}{2}\right)^{L}\prod_{i=1}^{N} P^{(p_i - q_i, p_i + q_i)}_{j_i - p_i}(x)\, dx. \tag{6.18}$$

The Jacobi polynomials in Eq. (6.18) can be expressed in terms of hypergeometric functions[15]

$$P^{(\alpha,\beta)}_n(x) = \binom{n+\alpha}{n} F\left(-n, n+\alpha+\beta+1, \alpha+1, \frac{1-x}{2}\right) = (-1)^n\binom{n+\beta}{n} F\left(-n, n+\alpha+\beta+1, \beta+1, \frac{1+x}{2}\right). \tag{6.19}$$

Using either of the expressions (6.19) in Eq. (6.18) and expanding the hypergeometric functions, the integral reduces to a sum of beta functions and can be evaluated. Alternatively, expanding a single hypergeometric function, one can get recurrence relations between the integrals for different matrix elements. The integrals in Eq. (6.18) depend upon a relatively small number of parameters; K, L and the set $\{N(jpq) \mid |q| \leq p < j\}$, where $N(jpq)$ is the number of individual angular momenta in the matrix element which lead to a given j, p and q. Note that this includes the four cases (m, m') equal to (p, q), (q, p), $(-p, -q)$, and $(-q, -p)$ respectively. Also cases with $p = j$ do not contribute to the

integral [they do contribute to the normalization factor in Eq. (6.18)] because the corresponding Jacobi polynomials in Eq. (6.18) are unity [Eq. (6.19) with $n = 0$].

B. Selecting a set of product states with complete, independent projections

In deriving the Wigner–Eckart type theorems in Sec. 5, we introduced a special set of product states $|\alpha\rangle$ which has complete, independent projections under P^0. Such a set can always be found by trial and error, but it is, of course, more convenient to have a rule which automatically gives the set. We here prove such a rule for angular momentum projection of a Kronecker product representation to give specified values j and m for the total angular momentum and its projection (we do not restrict ourselve to $j = 0$). The rule is a generalization of one given by Löwdin[16] for projection of a product of spin 1/2's. Our proof will also show that the functions of the projected set may be ordered in such a way that, after Schmidt orthonormalization, they agree with the functions obtained by standard binary coupling techniques for the sequential coupling

$$\mathbf{J}_1 = \mathbf{j}_1,$$
$$\mathbf{J}_{i+1} = \mathbf{J}_i + \mathbf{j}_{i+1}, \quad i = 1, 2, \ldots, N-1 \tag{6.20}$$

with $J_N = j$. Even the phases of the two sets of functions agree, provided that the usual phase convention[9] is used for the Clebsch–Gordan coefficients in the sequentially coupled states. We denote the sequentially coupled states [for the coupling of Eq. (6.20)] by

$$|(J_2 J_3 \cdots J_{N-1})jm\rangle. \tag{6.21}$$

In Eq. (6.21) we have omitted as labels the individual quantum numbers j_i since these are common to all members of the set. The allowed values of the J_i may be determined from the rule for combining two angular momenta j_1 and j_2, i.e., one gets each J value

$$J = |j_1 - j_2|, |j_1 - j_2| + 1, \ldots, j_1 + j_2 \tag{6.22}$$

once and only once.

For fixed values of j and m we order the states (6.21) by saying that the state specified by the set $\{J_i\}$ appears before the state specified by $\{J'_i\}$ if $J_k > J'_k$ for the first integer k for which $J_k \neq J'_k$.

For the product states $|\mathbf{j}_N \mathbf{m}_N\rangle$, we introduce the labels

$$M_k = \sum_{i=1}^{k} m_i, \quad k = 1, 2, \ldots, N, \tag{6.23}$$

i.e.,

$$m_1 = M_1,$$
$$m_i = M_i - M_{i-1}, \quad i = 1, 2, 3, \ldots, N, \tag{6.24}$$

and rewrite the states as

$$|M_1 M_2 \cdots M_N\rangle. \tag{6.25}$$

Again, we omit the labels j_i, which are common to all members of the set.

We first consider the "principal" case $m = j$.

Theorem: The projections by P^j_{jj} of the subset of product states

$$|j_1 J_2 J_3 \cdots J_{N-1} j\rangle, \tag{6.26}$$

where the sets $\{J_i\}$ are the same as for the sequentially coupled states (6.21), form a basis for the subspace with $J^2 = j(j+1)$ and $J_z = j$. Furthermore, if the product states of Eq. (6.26) are ordered according to the convention for J values stated above (for the sequentially coupled states) and their projections are orthonormalized by the Schmidt process, one has

$$(P^j_{jj}|j_1 J_2 \cdots J_{N-1} j\rangle)_S = |(J_2 J_3 \cdots J_{N-1})jj\rangle, \tag{6.27}$$

where the subscript S on the left hand side means "after Schmidt orthonormalization."

Proof: By their construction, one clearly has the correct number of states to form a basis. Now consider a particular product state (6.26). It has an expansion in terms of the complete orthonormal set of sequentially coupled states

$$|j_1 J_2 J_3 \cdots J_{N-1}\rangle = \sum_{\substack{J'_2, \ldots, J'_{N-1}, j' \\ (J'_i \geq J_i,\, j' \geq j)}} c(J_2 \cdots J_{N-1} j\,|\,J'_2 \cdots J'_{N-1} j')\, |(J'_2 \cdots J'_{N-1})j'j\rangle. \tag{6.28}$$

The restrictions on the values of J'_i and j' in Eq. (6.28) follow because J'_i is the intermediate total quantum number for the coupling of the first i angular momenta and $J_i = \sum_{k=1}^{i} m_k$ the corresponding azimuthal quantum number.

The effect of operating with P^j_{jj} on Eq. (6.28) is to pick out those terms with $j' = j$, i.e.

$$P^j_{jj}|j_1 J_2 \cdots J_{N-1} j\rangle = \sum_{\substack{J'_2, \ldots, J'_{N-1} \\ (J'_i \geq J_i)}} c(J_2 \cdots J_{N-1} j\,|\,J'_2 \cdots J'_{N-1} j)\, |(J'_2 \cdots J'_{N-1})jj\rangle. \tag{6.29}$$

Now consider the effect of orthonormalizing the states (6.29), using the ordering given above. By our convention, the first state must have $J'_2 \leq J_2$ for all J'_2. But Eq. (6.29) only contains terms with $J'_2 \geq J_2$. Thus, for the first projected product state, only the terms with $J'_2 = J_2$ survive in Eq. (6.29). For these terms, our ordering says that $J'_3 \leq J_3$ for all J'_3, so that only the terms with $J'_3 = J_3$ survive. Repeating this procedure, one sees that the expansion of the first projected product state reduces to the single term with $J'_i = J_i$, $i = 2, 3, \ldots, N-1$. The first step of the orthonormalization removes this term from the sums for all other projected products. Repeating the above argument for the second (partially orthonormalized) state, one sees that the sum again reduces to the single term with $J'_i = J_i$. Continuing this process, one gets the result

$$(P^j_{jj}|j_1 J_2 \cdots J_{N-1} j\rangle)_S = \frac{c(J_2 \cdots J_{N-1} j\,|\,J_2 \cdots J_{N-1} j)}{|c(J_2 \cdots J_{N-1} j\,|\,J_2 \cdots J_{N-1} j)|^{1/2}}\, |(J_2 \cdots J_{N-1})jj\rangle \tag{6.30}$$

for all the projected products in our set. The denominator in Eq. (6.30) is the result of normalizing at each stage.

The argument of the preceding paragraph is valid only if the coefficients c appearing in Eq. (6.30) are nonzero. To see that this is so, note that Eq. (6.29) gives

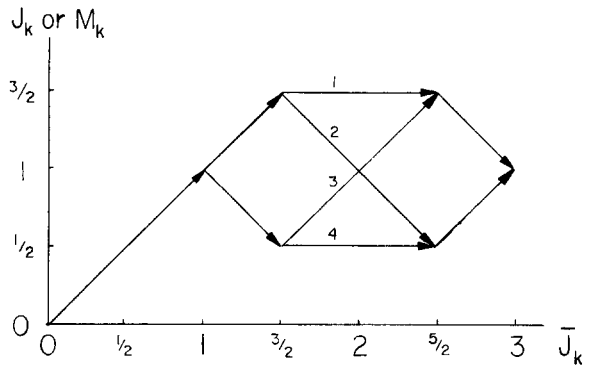


FIG. 1. With ordinate J_k, the paths 1–4 give the independent sets of intermediate angular momenta for the sequential coupling $[(1+\frac{1}{2})+1]+\frac{1}{2}$ to give $j=1$. With ordinates M_k [Eq. (6.23)], the paths specify, for the above angular momenta, product states such that their projections by P^1_{m1} are complete and independent.

$$c(J_2J_3\cdots J_{N-1}j\,|J_2J_3\cdots J_{N-1}j)$$
$$=\langle(J_2J_3\cdots J_{N-1})jj\,|P^j_{jj}|j_1J_2\cdots J_{N-1}j\rangle$$
$$=\langle(J_2J_3\cdots J_{N-1})jj\,|j_1J_2\cdots J_{N-1}j\rangle, \qquad (6.31)$$

where in the last step we have let the Hermitian operator P^j_{jj} operate to the left. Uncoupling the state to the left of Eq. (6.31) and using the orthonormality of the single particle states gives a product of Clebsch–Gordan coefficients $(j_1m_1j_2m_2\,|jm)$. With $J_1=j_1$ and $J_N=j$ we get

$$c(J_2J_3\cdots J_{N-1}j\,|J_2J_3\cdots J_{N-1}j)$$
$$=\prod_{i=1}^{N-1}(J_iJ_ij_{i+1}J_{i+1}|-J_i|J_{i+1}J_{i+1}). \qquad (6.32)$$

All Clebsch–Gordan coefficients in Eq. (6.32) are of the form $(j_1j_1j_2j-j_1\,|jj)$. For angular momenta satisfying the triangle rule (as ours do here) these coefficients are nonzero. With the usual phase convention, they also are all positive; one has[9]

$$(j_1j_1j_2j-j_2\,|jj)=\left[\frac{(2j_1)!(2j+1)!}{(j_1-j_2+j)!(j_1+j_2+j+1)!}\right]^{1/2}. \qquad (6.33)$$

Thus all c's appearing in Eq. (6.30) are not only nonzero, but they are also real and positive, so that Eq. (6.30) leads directly to Eq. (6.27).

The discussion of Sec. 2 shows that precisely the same set of product states leads to a basis for the subspace with the above j, but with an arbitrary m. One merely projects with the appropriate "transfer" operator P^j_{mj}, rather than the projector P^j_{jj}. Again, with our ordering, the orthonormalized projected products coincide with the sequentially coupled states.

The set of products given by the above theorem may be obtained simply from a diagram giving the allowed states for sequential coupling. We illustrate in Fig. 1 for the angular momentum coupling $[(1+\frac{1}{2})+1]+\frac{1}{2}\to 1$. The tips of the arrows give the possible values of the intermediate angular momentum J_k [obtained from Eq. (6.22) at each stage, with the restriction that one must wind up with a specified j] plotted against $\bar{J}_k=\sum_{i=1}^{k}j_i$, i.e., against the algebraic sum of the individual angular momenta up to that step. Each path corresponds to a vector coupled state. We have numbered the paths in Fig. 1 according to our convention for ordering states. The product states are obtained by setting $M_k=J_k$, i.e.the M_k values are just the ordinates of the arrow tips along the paths, and the m_k are [by Eq. (6.24)] just the differences of successive ordinates. In our example, the four product states are [written first as in Eq. (6.25) and then as $\Pi|j_im_i\rangle$]

$$\phi_1=|1\tfrac{3}{2}\tfrac{3}{2}1\rangle=|11\rangle\,|\tfrac{1}{2}\tfrac{1}{2}\rangle\,|10\rangle\,|\tfrac{1}{2}-\tfrac{1}{2}\rangle,$$
$$\phi_2=|1\tfrac{3}{2}\tfrac{1}{2}1\rangle=|11\rangle\,|\tfrac{1}{2}\tfrac{1}{2}\rangle\,|1-1\rangle\,|\tfrac{1}{2}\tfrac{1}{2}\rangle,$$
$$\phi_3=|1\tfrac{1}{2}\tfrac{3}{2}1\rangle=|11\rangle\,|\tfrac{1}{2}-\tfrac{1}{2}\rangle\,|11\rangle\,|\tfrac{1}{2}-\tfrac{1}{2}\rangle,$$
$$\phi_4=|1\tfrac{1}{2}\tfrac{1}{2}1\rangle=|11\rangle\,|\tfrac{1}{2}-\tfrac{1}{2}\rangle\,|10\rangle\,|\tfrac{1}{2}\tfrac{1}{2}\rangle.$$

ACKNOWLEDGMENT

The author would like to thank Dr. Liang-Shiun Chou for many stimulating discussions which helped to clarify some points in the paper.

[1] M. Hamermesh, *Group Theory and its Application to Physical Problems* (Addison-Wesley, Reading Massachusetts, 1962).
[2] P.-O. Löwdin, Phys. Rev. **97**, 1509 (1955).
[3] P.-O. Löwdin, Adv. Phys. **5**, 1 (1956).
[4] P.-O. Löwdin, Rev. Mod. Phys. **34**, 520 (1962).
[5] P.-O. Löwdin, Rev. Mod. Phys. **36**, 966 (1964).
[6] N. MacDonald, Adv. Phys. **19**, 371 (1970). This review article contains an extensive bibliography.
[7] W. J. Holman, III, and L. C. Biedenharn, in *Group Theory and its Applications*, edited by E. M. Loebl (Academic, New York, 1971), Vol. II.
[8] J. Shapiro, J. Math. Phys. **6**, 1680 (1965).
[9] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U. P., Princeton, 1957).
[10] K. Kumar, Aust. J. Phys. **19**, 719 (1966).
[11] T. Reith and M. D. Miranda, M.S. research paper, Fordham University (1968).
[12] R. M. Asherova and Y. F. Smirnov, Nucl, Phys. **4**, 399 (1970).
[13] M. E. Noz, Ph.D. disseration, Fordham University, 1969.
[14] J. Schott, Ph.D. dissertation, Fordham University, 1970.
[15] A. Erdelyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953).
[16] P.-O. Löwdin, "Spin Degeneracy Problem," technical note from the Quantum Chemistry Group of Uppsala University, 1957.

Quantum field theory in terms of sesquilinear forms

Walter Wyss*

Department of Mathematics and Physics, University of Colorado, Boulder, Colorado
(Received 7 February 1973)

We propose to describe a quantum field theory by continuous positive sesquilinear forms; using positive linear functionals would then be a special case. Our formalism still can accommodate a scattering theory.

INTRODUCTION

Wightman proposed to describe a quantum field theory by a positive linear functional on a topological *-algebra $\mathfrak{A}$, vanishing on a certain subspace.[1-4] This formalism allows the construction of field operators and a scattering matrix (Haag-Ruelle).[3,5] Every positive linear functional W on $\mathfrak{A}$ gives rise to a sesquilinear form by $B(f,g) = W(f^*g)$, $f,g \in \mathfrak{A}$, but not every continuous positive sesquilinear form on $\mathfrak{A}$ is derivable in this way from a positive linear functional.

Hence we propose to describe a quantum field theory by continuous positive sesquilinear forms on $\mathfrak{A}$. This formalism in general does not lead to field operators but still can accommodate a scattering theory, modeled after the Haag-Ruelle theory and given by a sesquilinear form. In this paper we are only concerned with partial aspects of the above description.

In Sec. 1 we summarize the properties of positive linear functionals of $\mathfrak{A}$. Section 2 then deals with the axioms for a neutral scalar field in terms of positive linear functionals. In Sec. 3 we study sesquilinear forms on $\mathfrak{A}$ and Sec. 4 gives a representation of all continuous positive sesquilinear forms. In Sec. 5 we list the axioms for a neutral scalar field in terms of continuous positive sesquilinear forms and in Sec. 6 we develop a scattering theory within this formalism.

1. POSITIVE LINEAR FUNCTIONALS ON THE FIELD ALGEBRA

The topological vector space

$$\mathfrak{A} = \bigoplus_{n=0}^{\infty} \mathcal{S}^n,$$

where

$$\mathcal{S}^0 = \mathbb{C}, \mathcal{S}^n = \mathcal{S}(\mathbb{R}^{4n}),$$

equipped with the multiplication

$$(fg)_n(p_1,\ldots,p_n) = \sum_{k=0}^{n} f_k(p_1,\ldots,p_k)g_{n-k}(p_{k+1},\ldots,p_n)$$

and the involution

$$(f^*)_n(p_1,\ldots,p_n) = \overline{f_n(-p_n,\ldots,-p_1)}, \quad (f^*)_0 = \overline{f_0},$$

is called the field algebra for a neutral scalar field.[1,4,6] The field algebra is thus a topological *-algebra, but can also be looked upon as a ordered topological vector space, ordered by the cone

$$K = \overline{\{f \in \mathfrak{A}; f = \textstyle\sum \lambda_j (f^j)^* f^j, \quad \lambda_j \geq 0, \quad f^j \in \mathfrak{A}\}}.$$

The topology of $\mathfrak{A}$ is the direct sum topology of an increasing family of nuclear Fréchet spaces, and hence described by an uncountable set of seminorms. By $\mathfrak{B}(\mathfrak{A})$ we denote the family of bounded sets of $\mathfrak{A}$.

Lemma 1.1.

(1) $\mathfrak{A}$ is a nuclear LF-space and thus complete, bornological, barreled, reflexive, and separable

(2) K is a closed, proper, strict $\mathfrak{B}$ cone.

Proof: See Refs. 4 and 6.

Let $\mathfrak{A}'$ be the topological dual of the field algebra $\mathfrak{A}$, and

$$K' = \{T \in \mathfrak{A}^*; \quad T(f) \geq 0, \quad \forall f \in K\}$$

the dual cone of K.

A linear functional on $\mathfrak{A}$ that takes positive values on K, is called a positive linear functional.

Lemma 1.2.

(1) $\mathfrak{A}'$ is nuclear, complete, reflexive and thus barreled.

(2) Every positive linear functional in continuous.

Proof: See Refs. 4 and 6.

We state now some well-known properties of a positive linear functional.

Lemma 1.3: A positive linear functional $W \in K'$ satisfies

(1) $W(f^*) = \overline{W(f)}$ (hermitian),

(2) $|W(f^*g)|^2 \leq W(f^*f)W(g^*g)$ (Cauchy-Schwarz inequality),

(3) $$W(f^*g) = \frac{1}{4}W((f+g)^*(f+g)) - \frac{1}{4}W((f-g)^*(f-g)) - \frac{i}{4}W((f+ig)^*(f+ig)) + \frac{i}{4}W((f-ig)^*(f-ig))$$ (polarization identity).

Lemma 1.4: Every positive linear functional W gives rise to a continuous seminorm on $\mathfrak{A}$ by $p(f) = \sqrt{W(f^*f)}$

Proof: The Cauchy-Schwarz inequality implies that $p(\cdot)$ is a seminorm. Since W is continuous and the product of two bounded set of $\mathfrak{A}$ is bounded, $p(\cdot)$ is bounded on bounded sets. The fact that $\mathfrak{A}$ is bornological implies then that $p(\cdot)$ is a continuous seminorm on $\mathfrak{A}$.

With every positive linear functional on $\mathfrak{A}$ one can associate a left module

Lemma 1.5: Let $W \in K'$. Then $I(W) = \{f \in \mathfrak{A}; W(f^*f) = 0\}$ is a closed left ideal of $\mathfrak{A}$, and $H(W) = \mathfrak{A}/I(W)$ is a positive definite inner product space and also a left module.

Proof: This is the famous Gel'fand-Segal construction. The inner product in $H(W)$ is given by

$$(\epsilon(f), \epsilon(g)) = W(f^*g),$$

Copyright © 1973 by the American Institute of Physics

where $\epsilon(f)$ stands for the equivalence class of f modulo $I(W)$. The left module structure then follows from

$$f\epsilon(g) = \epsilon(fg).$$

Note: A positive linear functional W on $\mathfrak{A}$ is called a state if $W(1) = 1$. In this paper, all positive linear functionals will be states.

2. POSITIVE LINEAR FUNCTIONALS–AXIOMS FOR A NEUTRAL SCALAR FIELD (WIGHTMAN)

First we introduce some maps of the field algebra into itself.

Definition 2.1: (1) The Fourier transform F: $\mathfrak{A} \to \mathfrak{A}$.

$$(Ff)_n(x_1, \ldots, x_n) = (2\pi)^{-2n} \int_e^{-i(p_1x_1+\cdots+p_nx_n)} f_n(p_1, \ldots, p_n)dp_1 \ldots dp_n,$$

$$(Ff)_0 = f_0.$$

F turns out to be an automorphism.

(2) Representation of the Poincaré group $\mathbb{R}^4 \wedge L_+^\uparrow$.

$$((a,\wedge)f)_n(p_1, \ldots, p_n) = e^{ia(p_1+\cdots+p_n)} f_n(\wedge^{-1}p_1, \ldots, \wedge^{-1}p_n), \qquad ((a,\wedge)f)_0 = f_0.$$

This representation is thus given by *-automorphisms.

(3) The TCP Operator θ: $\mathfrak{A} \to \mathfrak{A}$.

$$(\theta f)_n(p_1, \ldots, p_n) = \overline{f_n(p_1, \ldots, p_n)}, \qquad (\theta f)_0 = \overline{f_0}.$$

θ then is a *-antiautomorphism.

Secondly we need two subspaces of the field algebra.

Definition 2.2: (1) The spectrum ideal I_1:

$$I_1 = \{f \in \mathfrak{A}; f_0 = 0, f_n(p_1, \ldots, p_n) = 0 \text{ if } q_k \equiv \sum_{l=k}^{n} p_l \in \overline{V}_+, \forall\, 1 \le k \le n, \forall n\}.$$

I_1 turns out to be a left ideal.

(2) The locality ideal I_2:

I_2 is defined to be the two-sided ideal generated by elements of the form $(Ff)_n(x_1, \ldots, x_n) = (F\varphi)_n(x_1, \ldots, \underline{x}, \ldots, x_n) - (F\varphi)_n(x_1, \ldots, \underline{x}_\pi, \ldots, x_n)$, where $\underline{x} = \{x_l, x_{l+1}, \ldots, x_{l+k}\} \subset \{x_1, \ldots, x_n\}$ for any l and k, and π is a permutation of the elements in $\underline{x}$. Furthermore,

$$(F\varphi)_n(x_1, \ldots, \underline{x}, \ldots, x_n) = 0 \text{ if } (x_i - x_k)^2 > 0, \forall\, x_i, x_k \in \underline{x}, i \neq k.$$

Now we are in a position to state the axioms for a neutral scalar field.

Axioms (local Field Theory):

A local field theory for a neutral scalar field is given by a positive linear functional W on the field algebra, satisfying

(1) $W((a,\wedge)f) = W(f)$ (Lorentz covariance),

(2) $W(I_1) = 0$ (spectrum condition),

(3) $W(I_2) = 0$ (locality),

(4) $W(1) = 1$ (normalization).

Axioms (weakly local field theory):

A weakly local field theory for a neutral scalar field is given by a positive linear functional W on the field algebra, satisfying

(1) $W((a,\wedge)f) = W(f)$ (Lorentz covariance),

(2) $W(I_1) = 0$ (spectrum condition),

(3) $W(\theta f) = \overline{W(f)}$ (weak locality),

(4) $W(1) = 1$ (normalization).

The following example shows that the two axiom schemes are nonvoid.

Example: Let $\mathfrak{N}$ be a graded vector space and K a map

$$K: \mathfrak{A} \to \mathfrak{N},$$

satisfying

$$(Kfg)_n(p_1, \ldots, p_n) = \sum_{k=0}^{\infty} \int d\mu(q_1) \cdots d\mu(q_k) \frac{1}{k!} \sum_{l=0}^{n} \left\{ r_1, \sum \cdots, r_n \right\}$$

$$(Kf)_{k+l}(p_{r_1}, \ldots, p_{r_l}, -q_1, \ldots, -q_k)$$

$$(Kg)_{k+n-l}(p_{r_{l+1}}, \ldots, p_{r_n}, q_1, \ldots, q_k),$$

where $d\mu(q) = \theta(q)\delta(q^2 - m^2)dq$.

Then $W(f) \equiv (Kf)_0$ is a positive linear functional on the field algebra, and according to Ref. 4 there are maps K such that W satisfies all the above axioms.

3. SESQUILINEAR FORMS ON THE FIELD ALGEBRA

A bilinear form Φ on the field algebra $\mathfrak{A}$ is a mapping

$$\Phi: \mathfrak{A} \times \mathfrak{A} \to \mathbb{C},$$

such that Φ_f, Φ_g, defined by

$$\Phi_f(g) = \Phi(f,g), \qquad \Phi_g(f) = \Phi(f,g)$$

are linear functionals on $\mathfrak{A}$.

A sesquilinear form B on the field algebra $\mathfrak{A}$ is a mapping

$$B: \mathfrak{A} \times \mathfrak{A} \to \mathbb{C}$$

such that B_f, B_g, defined by

$$B_f(g) = B(f,g), \qquad B_g(f) = B(f^*,g)$$

are linear functionals on $\mathfrak{A}$.

Bilinear and sesquilinear forms on $\mathfrak{A}$ are in a one-to-one correspondence by $B(f,g) = \Phi(f^*,g)$.

Definition 3.1: Let Φ be a bilinear form on $\mathfrak{A}$. Then Φ is called

(1) separately continuous, if Φ_f, Φ_g are continuous linear functionals on $\mathfrak{A}$; the space of separately continuous bilinear forms is denoted by $B(\mathfrak{A}, \mathfrak{A})$;

(2) hypocontinuous, if $\{\Phi_f; f \in \mathfrak{B}(\mathfrak{A})\}$ and $\{\Phi_g; g \in \mathfrak{B}(\mathfrak{A})\}$ are equicontinuous subsets of $\mathfrak{A}'$; the space of hypocontinuous bilinear forms is denoted by $\mathfrak{K}(\mathfrak{A}, \mathfrak{A})$;

(3) jointly continuous, if Φ is continuous from the Cartesian product $\mathfrak{A} \times \mathfrak{A}$ into $\mathbb{C}$; the space of jointly continuous bilinear forms is denoted by $\mathfrak{B}(\mathfrak{A}, \mathfrak{A})$.

We clearly have the inclusions $\mathfrak{B}(\mathfrak{A}, \mathfrak{A}) \subset \mathfrak{K}(\mathfrak{A}, \mathfrak{A}) \subset B(\mathfrak{A}, \mathfrak{A})$.

Lemma 3.1: For the field algebra $\mathfrak{A}$ we have $\mathfrak{K}(\mathfrak{A}, \mathfrak{A}) = B(\mathfrak{A}, \mathfrak{A})$.

Proof: This follows immediately from the fact that $\mathfrak{A}$ is barreled.

On the space $\mathfrak{K}(\mathfrak{A}, \mathfrak{A})$ we put the topology of uniform convergence on products of bounded sets of $\mathfrak{A}$.

Lemma 3.2: Let $\mathfrak{L}(\mathfrak{A}, \mathfrak{A}')$ be the space of continuous linear maps from $\mathfrak{A}$ into $\mathfrak{A}'$, equipped with the topology of uniform convergence on bounded sets. Then $\mathfrak{K}(\mathfrak{A}, \mathfrak{A})$ is topologically isomorphic to $\mathfrak{L}(\mathfrak{A}, \mathfrak{A}')$ and thus complete.

Proof: Take $\Phi \in \mathfrak{K}(\mathfrak{A}, \mathfrak{A})$ and look at the map $\alpha(\Phi)$ defined by

$$(\alpha(\Phi)f)(g) = \Phi(f, g).$$

$\alpha(\Phi)$ then maps a bounded set of $\mathfrak{A}$ into a simply bounded set of $\mathfrak{A}'$; simply bounded sets of $\mathfrak{A}'$ however are bounded. Hence $\alpha(\Phi)$ is a bounded linear map from $\mathfrak{A}$ into $\mathfrak{A}'$ and due to the LF-structure of $\mathfrak{A}$, $\alpha(\Phi) \in \mathfrak{L}(\mathfrak{A}, \mathfrak{A}')$, i.e. $\alpha(\Phi)$ is continuous.

Take $\psi \in \mathfrak{L}(\mathfrak{A}, \mathfrak{A}')$ and look at the map $\beta(\psi)$ defined by

$$\beta(\psi)(f, g) = \psi(f)(g).$$

Since $\mathfrak{A}$ is barreled, $\beta(\psi)$ is a hypocontinuous bilinear form on $\mathfrak{A}$.

We have now

$$\alpha: \mathfrak{K}(\mathfrak{A}, \mathfrak{A}) \to \mathfrak{L}(\mathfrak{A}, \mathfrak{A}'),$$

$$\beta: \mathfrak{L}(\mathfrak{A}, \mathfrak{A}') \to \mathfrak{K}(\mathfrak{A}, \mathfrak{A}),$$

and $\alpha\beta = \beta\alpha = 1$.

We now prove the continuity of α and β. This will depend extensively on the barreledness of $\mathfrak{A}$ and $\mathfrak{A}'$.

The neighborhoods of zero in $\mathfrak{K}(\mathfrak{A}, \mathfrak{A})$ are given by

$$U(A, B, \epsilon) = \{\Phi \in \mathfrak{K}(\mathfrak{A}, \mathfrak{A}); |\Phi(A, B)| < \epsilon; A, B \mathfrak{B}(\mathfrak{A})\}.$$

The neighborhoods of zero in $\mathfrak{L}(\mathfrak{A}, \mathfrak{A}')$ are given by

$$U(A, V) = \{\psi \in \mathfrak{L}(\mathfrak{A}, \mathfrak{A}'); \psi(A) \subset V; V \in U(\mathfrak{A}')\}.$$

We thus have to show that $\beta U(A, V)$ is a neighborhood in $\mathfrak{K}(\mathfrak{A}, \mathfrak{A})$ and $\alpha U(A, B, \epsilon)$ is a neighborhood in $\mathfrak{L}(\mathfrak{A}, \mathfrak{A}')$.

For $\psi \in U(A, V)$ we get with

$$|\beta(\psi)(f, g)| = |\psi(f)(g)|$$

that $\beta U(A, V) = U(A, V^0, 1)$.

On the other hand

$$\alpha U(A, B, \epsilon) \supset U(A, \epsilon' B^0), \qquad \epsilon' < \epsilon.$$

Finally, $\mathfrak{L}(\mathfrak{A}, \mathfrak{A}')$ is complete because $\mathfrak{A}$ is an LF-space.

The space of hypocontinuous bilinear forms can also be characterized in terms of tensor products.[7,8] Since $\mathfrak{A}$ and $\mathfrak{A}'$ are both nuclear, we know that on $\mathfrak{A} \otimes \mathfrak{A}$ (resp. $\mathfrak{A}' \otimes \mathfrak{A}'$) the π-topology and the ϵ-topology coincide. The completion of $\mathfrak{A} \otimes \mathfrak{A}$ (resp. $\mathfrak{A}' \otimes \mathfrak{A}'$) in either of these topologies is denoted by $\mathfrak{A} \hat{\otimes} \mathfrak{A}$ (resp. $\mathfrak{A}' \hat{\otimes} \mathfrak{A}'$); $\mathfrak{A} \hat{\otimes} \mathfrak{A}$ (resp. $\mathfrak{A}' \otimes \mathfrak{A}'$) is again nuclear.

Lemma 3.3: $\mathfrak{K}(\mathfrak{A}, \mathfrak{A}) \cong \mathfrak{A}' \hat{\otimes} \mathfrak{A}' \cong \mathfrak{L}(\mathfrak{A}, \mathfrak{A}')$,
$\mathfrak{K}(\mathfrak{A}', \mathfrak{A}') \cong \mathfrak{A} \hat{\otimes} \mathfrak{A} \cong \mathfrak{L}(\mathfrak{A}', \mathfrak{A})$.

Proof: Since $\mathfrak{A}$ and $\mathfrak{A}'$ are both nuclear, barreled, and complete, we know from the theory of tensor products, Ref. 8, that $\mathfrak{A}' \hat{\otimes} \mathfrak{A}' \cong \mathfrak{L}(\mathfrak{A}, \mathfrak{A}')$ and $\mathfrak{A} \otimes \mathfrak{A} \cong \mathfrak{L}(\mathfrak{A}, \mathfrak{A}')$. The rest follows from Lemma 3.2.

From now on we will be interested in sesquilinear forms. The special structure of our field algebra leads to an interesting representation of jointly continuous sesquilinear forms.

Lemma 3.4 (Kernel theorem): Every $\Phi \in \mathfrak{B}(\mathfrak{A}, \mathfrak{A})$ has the representation

$$\Phi(f, g) = \sum \lambda_i A_i(f) B_i(g),$$

where $\{A_i\}$ and $\{B_i\}$ are bounded sets in $\mathfrak{A}'$ and $\sum |\lambda_i| < \infty$. A jointly continuous sesquilinear form thus has the representation

$$B(f, g) = \sum \lambda_i \overline{A_i(f)} B_i(g),$$

where $\{A_i\}$ and $\{B_i\}$ are bounded sets in $\mathfrak{A}'$ and $\sum |\lambda_i| < \infty$.

Proof: See Ref. 9. Since $\mathfrak{A}'$ is barreled, there is equality between equicontinuous and bounded sets in $\mathfrak{A}'$.

We state now two well-known facts about sesquilinear forms.

Lemma 3.5: Let B be a sesquilinear form on the field algebra $\mathfrak{A}$. $B(f, f)$ is called the quadratic form associated with B. Then we have the polarization identity

$$B(f, g) = \frac{1}{4} B(f + g, f + g) - \frac{1}{4} B(f - g, f - g) - \frac{i}{4} B(f + ig, f + ig) + \frac{i}{4} B(f - ig, f - ig).$$

Lemma 3.6: A sesquilinear form B is called positive if $B(f, f) \geq 0$.

A positive sesquilinear form satisfies

(1) $B(f, g) = \overline{B(g, f)}$ (Hermitian),

(2) $|B(f, g)|^2 \leq B(f, f) B(g, g)$ (Cauchy–Schwarz inequality).

Lemma 3.7: Let B be a separately continuous positive sesquilinear form on the field algebra. Then $p(f) = \sqrt{B(f, f)}$ is a continuous seminorm on $\mathfrak{A}$.

Proof: It follows immediately from the Cauchy–Schwarz inequality that $p(\cdot)$ is a seminorm. Since B is separately continuous it has to be hypocontinuous (Lemma 3.1) and then $p(\cdot)$ is bounded on bounded sets. This however implies the continuity of $p(\cdot)$ because $\mathfrak{A}$ is bornological.

Corollary 3.1: Every positive separately continuous sesquilinear form on the field algebra is jointly continuous.

Proof: From the Cauchy–Schwarz inequality we find

$$|B(f,g)| \leq \sqrt{B(f,f)}\,\sqrt{B(g,g)},$$

where the right-hand side is clearly a seminorm (Lemma 3.7) on $\mathfrak{A} \times \mathfrak{A}$.

Similarly to the Gel'fand–Segal construction we can associate an inner product space with every positive continuous sesquilinear form. This inner product space, however, is in general not a left module.

Lemma 3.8: Let B be a positive continuous sesquilinear form on the field algebra. The $J(B) = \{f \in \mathfrak{A};\ B(f,f) = 0\}$ is a closed subspace of $\mathfrak{A}$ and $K(B) = \mathfrak{A}/J(B)$ is a positive definite inner product space.

Proof: Let $\pi(f)$ represent the equivalence class of f modulo $J(B)$. Then the inner product in $K(B)$ is given by

$$(\pi(f), \pi(g)) = B(f,g).$$

The lemma then becomes trivial.

4. POSITIVE LINEAR FUNCTIONALS–POSITIVE SESQUILINEAR FORMS. RELATIONS AND REPRESENTATIONS

Let W be a positive linear functional on the field algebra such that $W(1) = 1$, i.e., W is a state. Then we can associate two sesquilinear forms to W, by

$$B(f,g) = W(f^*g) \quad \text{and} \quad C(f,g) = W(f^*g) - \overline{W(f)}W(g).$$

These sesquilinear forms are positive and continuous.

The Cauchy–Schwarz inequality for $B(f,g) = W(f^*g)$ is the well-known inequality

$$|W(f^*g)|^2 \leq W(f^*f)W(g^*g).$$

The Cauchy–Schwarz inequality for $C(f,g) = \overline{W(f)}W(g)$, however improves the previous inequality. Explicitly we get

$$\begin{aligned}|W(f^*g)|^2 - W(f^*f)W(g^*g) &\leq W(f^*g)W(f)\overline{W(g)} + \overline{W(f^*g)}\,\overline{W(f)}W(g) \\ &\quad - W(f^*f)|W(g)|^2 - W(g^*g)|W(f)|^2.\end{aligned}$$

Or if $f \in \ker W, f \notin I(W)$, the inequality reads

$$|W(f^*g)|^2 - W(f^*f)W(g^*g) \leq -W(f^*f)|W(g)|^2.$$

Lemma 4.1: Let W be a state on the field algebra. Then $B(f,g) = W(f^*g)$ defines a continuous sesquilinear form, satisfying $B(f,f) \geq 0$ and $B(f,g) = B(1,f^*g)$. Conversely every positive sesquilinear form on the field algebra, satisfying $B(f,g) = B(1,f^*g)$ defines a positive linear functional.

Proof: The first part of the lemma is obvious. Next let $W(f) = B(1,f)$. Then this is indeed a positive linear functional by $W(f^*f) = B(1,f^*f) = B(f,f) \geq 0$. Note that we do not have to require our positive sesquilinear form to be continuous.

Lemma 4.2: Let W be a state on the field algebra $\mathfrak{A}$. Then $C(f,g) = W(f^*g) - \overline{W(f)}W(g)$ defines a continuous positive sesquilinear form, satisfying $C(1,g) = 0$, $\forall g \in \mathfrak{A}$.

Proof: The positivity follows from $|W(f)|^2 \leq W(f^*f)$. The rest is obvious.

The nuclear structure of our field algebra leads us now to an interesting representation theorem for separately continuous positive sesquilinear forms.

Theorem 4.1: Let B be a separately continuous positive sesquilinear form on the field algebra. Then

$$B(f,g) = \sum \wedge_{ij} \overline{T_i(f)}\,T_j(g),$$

where $\wedge = \{\wedge_{ij}\}$ is a positive hermitian matrix, satisfying $\sum_{i,j}|\wedge_{ij}| < \infty$, and $|T_i(f)| < M(f)$, i.e., $T_i \in \mathfrak{A}'$ form a bounded set.

Proof: By Lemma 3.8, $K(B)$ is a positive inner product space, that we can complete in the norm $\|\pi(f)\|^2 = B(f,f)$ to a Hilbert space. The map π: $\mathfrak{A} \to K(B) \subset \overline{K(B)}$ $\|\cdot\|$ is continuous and hence by the nuclearity of $\mathfrak{A}$ has the form[9]

$$\pi(f) = \sum \lambda_i T_i(f) x_i,$$

where $\sum|\lambda_i| < \infty$, $\{T_i\} \subset \mathfrak{A}'$ is bounded, and $\|x_i\| \leq 1$. Then

$$B(f,g) = (\pi(f), \pi(g)) = \sum \overline{\lambda_i}\,\overline{T_i(f)}\,\lambda_j T_j(g)(x_i, x_j).$$

The proof then is completed by setting $\wedge_{ij} = \overline{\lambda_i}\,(x_i, x_j)\lambda_j$.

Note: (1) For a positive sesquilinear form satisfying $B(f,g) = B(1,f^*g)$, one observes from the proof of the above theorem that the vectors of the Gel'fand–Segal domain for the functional $W(f) = B(1,f)$ have an expansion in terms of Gel'fand–Segal Hilbert space vectors with summable coefficients. Not all Hilbert space vectors can have that property; remember that

$$l_1 \subset l_2.$$

(2) We get all separately continuous positive sesquilinear forms on the field algebra via the representation in Theorem 4.1, and this is in accordance with Lemma 3.4.

Corollary 4.1: Let B be a separately continuous positive sesquilinear form on the field algebra $\mathfrak{A}$, satisfying $B(1,1) = 1$. The B is uniquely characterized by the representation

$$B(f,g) = \sum \wedge_{ij} \overline{T_i(f)}\,T_j(g),$$

where $\wedge = \{\wedge_{ij}\}$ is a positive hermitian matrix, with $\sum_{i,j}|\wedge_{ij}| < \infty$; $\{T_i\} \subset \mathfrak{A}'$ is a bounded set of continuous linear functionals; and $\sum \wedge_{ij} \overline{T_i(1)}\,T_j(1) = 1$.

Proof: This is an immediate consequence of Theorem 4.1.

Remark: A positive linear functional W on the field algebra $\mathfrak{A}$ leads to a continuous positive sesquilinear form B, with $B(1,1) = 1$. However, a continuous positive sesquilinear form B, satisfying $B(1,1) = 1$ can not necessarily be written as $B(f,f) = W(f^*f)$, where W is a positive linear functional on $\mathfrak{A}$. One is thus tempted to investigate the possibility to describe quantum field theory in terms of sesquilinear forms and then axiomatize by abstraction.

5. POSITIVE SESQUILINEAR FORMS–AXIOMS FOR A NEUTRAL SCALAR FIELD

With the notations of Sec. 2, we like to propose axiom systems for a quantum field theory.

Axioms (local field theory)

A local Field Theory for a neutral scalar field is given by a continuous positive sesquilinear form B on the field algebra, satisfying

(1) $B((a,\wedge)f,(a,\wedge)g) = B(f,g)$ (Lorentz covariance),
(2) $B(f,I_1) = 0,\ \forall f \in \mathfrak{A}$ (spectrum condition),
(3) $B(f,I_2) = 0, \forall f \in \mathfrak{A}$ (locality),
(4) $B(1,1) = 1$ (normalization).

Axioms (weakly local field theory)

A weakly local field theory for a neutral scalar field is given by a continuous positive sesquilinear form B on the field algebra, satisfying

(1) $B((a,\wedge)f,(a,\wedge)g) = B(f,g)$ (Lorentz covariance),
(2) $B(f,I_1) = 0,\ \forall f \in \mathfrak{A}$ (spectrum condition),
(3) $B(\theta f,\theta g) = \overline{B(f,g)}$ (weak locality),
(4) $B(1,1) = 1$ (normalization).

The following example, actually the example in Sec. 2, shows the structure of a continuous positive sesquilinear form, satisfying the above axioms.

Example:

$$B(f,g) = \sum_{k=0}^{\infty} \int d\mu(q_1)\cdots d\mu(q_k) \times \frac{1}{k!}\overline{(Kf)_k(q_1,\ldots,q_k)}\,(Kg)_k(q_1,\ldots,q_k),$$

where K is the operator defined in Ref. (4). In terms of the graphs defining K, we see that B is gotton by integrating over internal free legs.

6. SCATTERING THEORY IN TERMS OF SESQUILINEAR FORMS

We want to describe partial aspects of a scattering theory associated with a continuous positive sesquilinear form. Our main objective is to show the existence of a scattering form.

Let's denote by W_0 the positive linear functional describing a neutral free scalar field of mass m. Furthermore, let $W_0(f,g) = W_0(f^*g)$.

Definition 6.1: Let B be a continuous positive sesquilinear form on the field algebra $\mathfrak{A}$ and $\alpha = \{\alpha(t)\}$ a one parameter family of continuous linear maps of $\mathfrak{A}$. B is called α-free if the two parameter family

$$\{B_{t_1,t_2}(f,g) = B(\alpha(t_1)f,\alpha(t_2)g)\}$$

of continuous sesquilinear forms has the following limit in $\mathfrak{B}(\mathfrak{A},\mathfrak{A})$

$$\lim_{t_1,t_2\to\infty} B_{t_1,t_2}(f,g) = \lim_{t_1,t_2\to-\infty} B_{t_1,t_2}(f,g) = W_0(f,g).$$

Theorem 6.1: Let B be an α-free continuous positive sesquilinear form. Then the following limit exists

$$\lim_{t_1,t_2\to\infty} B_{-t_1,t_2}(f,g) = S(f,g)$$

and is a jointly continuous sesquilinear form.

Proof: Let $t_1,t_2,t_3,t_4 > t_0$ and $f \in A$, $g \in B$ where $A, B \in \mathfrak{B}(\mathfrak{A})$.

Then

$$\begin{aligned}
|B_{-t_1,t_2}(f,g) - B_{-t_3,t_4}(f,g)| &= |B_{-t_1,t_2}(f,g) - B_{-t_3,t_2}(f,g) \\
&\quad + B_{-t_3,t_2}(f,g) - B_{-t_3,t_4}(f,g)| \\
&\le |B_{-t_1,t_2}(f,g) - B_{-t_3,t_2}(f,g)| \\
&\quad + |B_{-t_3t_2}(f,g) - B_{-t_3,t_4}(f,g)| \\
&\le |B(\{\alpha(-t_1) - \alpha(-t_3)\}f,\alpha(t_2)g| \\
&\quad + |B(\alpha(-t_3)f,\{\alpha(t_2) - \alpha(t_4)\}g)| \\
&\le [B(\alpha(t_2)g,\alpha(t_2)g)]^{1/2} \\
&\quad \times [B(\{\alpha(-t_1) - \alpha(-t_3)\}f,\{\alpha(-t_1) - \alpha(-t_3)\}f)]^{1/2} \\
&\quad + [B(\alpha(-t_3)f,\alpha(-t_3)f)]^{1/2} \\
&\quad \times [B(\{\alpha(t_2) - \alpha(t_4)\}g,\{\alpha(t_2) - \alpha(t_4)\}g)]^{1/2}
\end{aligned}$$

Since B is α-free we know the following estimates:

$$\begin{aligned}
&|B(\alpha(t_1)f,\alpha(t_2)g) - B(\alpha(t_3)f,\alpha(t_4)g)| < \epsilon_1, \\
&|B(\alpha(-t_1)f,\alpha(-t_2)g) - B(\alpha(-t_3)f,\alpha(-t_4)g)| < \epsilon_2, \\
&|W_0(f,g) - B(\alpha(t_1)f,\alpha(t_2)g)| < \epsilon_3, \\
&|W_0(f,g) - B(\alpha(-t_1)f,\alpha(-t_2)g)| < \epsilon_4,
\end{aligned}$$

provided $t_1,t_2,t_3,t_4 > t_0$ and f,g vary in bounded sets of $\mathfrak{A}$.

Then

$$\begin{aligned}
|B_{-t_1,t_2}(f,g) - B_{-t_3,t_4}(f,g)| &\le [W_0(g,g) + \epsilon]^{1/2} \\
&\quad \times [B_{-t_1,-t_1}(f,f) - B_{-t_1,-t_3}(f,f) - B_{-t_3,-t_1}(f,f) \\
&\quad + B_{-t_3,-t_3}(f,f)]^{1/2} + [W_0(f,f) + \epsilon]^{1/2}[B_{t_2,t_2}(g,g) \\
&\quad - B_{t_2,t_4}(g,g) - B_{t_4,t_2}(g,g) + B_{t_4,t_4}(g,g)]^{1/2}.
\end{aligned}$$

Using the above estimates then gives $|B_{-t_1,t_2}(f,g) - B_{-t_3,t_4}(f,g)| < \epsilon$, hence $\{B_{-t_1,t_2}\}$ is a Cauchy net in $\mathfrak{K}(\mathfrak{A},\mathfrak{A})$ and due to the completeness of $\mathfrak{K}(\mathfrak{A},\mathfrak{A})$ (Lemma 3.2), $\{B_{-t_1,t_2}\}$ converges to a hypocontinuous sesquilinear form $S(f,g)$. Because of

$$\begin{aligned}
&\lim_{t_1,t_2\to\infty} |B_{-t_1,t_2}(f,g)| \\
&\quad \le \lim_{t_1,t_2\to\infty} [B(\alpha(-t_1)f,\alpha(-t_1)f)B(\alpha(t_2)g,\alpha(t_2)g)]^{1/2} \\
&\quad \le [W_0(f,f)]^{1/2}[W_0(g,g)]^{1/2}
\end{aligned}$$

and the fact that $[W_0(f,f)]^{1/2}$ is a continuous seminorm (Lemma 3.7) on $\mathfrak{A}$ implies that $S(f,g)$ is actually jointly continuous.

Definition 6.2: Let B be an α-free continuous positive sesquilinear form. Then $S(f,g)$ is called the α-scattering form of B. S is again a jointly continuous sesquiiinear form.

Corollary 6.1: Let B be an α-free continuous positive sesquilinear form on the field algebra. Then B has an α-scattering form.

Proof: Obvious.

Definition 6.3: An α-free continuous sesquilinear form B is called α-trival, if $S(f,g) = W_0(f^*g)$.

Remarks: (1) Instead of working with $B(f,g)$ corresponding to $W(f^*g)$ one can as well axiomatize $C(f,g)$ corresponding to $W(f^*g) - \overline{W(f)}W(g)$. The same axioms as in Sec. 5 would apply except for the normalization; instead we would require $C(1,1) = 0$. An α-scattering form would then be expressed by

$$S(f,g) = W_0(f)\,W_0(g) + \lim_{t_1,t_2\to\infty} C(\alpha(-t_1)f, \alpha(t_2)g);$$

one has however to assume that $\alpha(t)1 = 1 \forall t$.

(2) Let B be a continuous positive sesquilinear form, satisfying the axioms in Sec. 5. The existence and essential uniqueness of a one parameter family $\alpha = \{\alpha(t)\}$ of continuous linear maps leading to a nontrivial α-scattering form will be the goal of future investigations.

CONCLUSION

We propose to describe quantum field theory in terms of sesquilinear forms. This formalism contains the well-known Wightman theory. Whereas the latter leads to field operators on an inner product space our theory only inherits the inner product space. Amongst the most general representations that we have to describe our formalism, the Wightman theory is characterized by a quadratic relation. Our formalism allows a scattering theory which means that asymptotically it becomes a Wightman theory.

ACKNOWLEDGMENTS

I would like to thank Professors W. E. Brittin, R. K. Goodrich, K. T. Mahanthappa, B. Misra, and A. S. Wightman for helpful remarks.

---

*Supported by National Science Foundation, GP 19479.

[1]H. J. Borchers, Nuovo Cimento **24**, 214 (1962).
[2]R. F. Streater and A. S. Wightman, *PCT, Spin, Statistics, and All That* (Benjamin, New York, 1964).
[3]R. Jost, *The General Theory of Quantized Fields* (American Mathematical Society, Providence, Rhode Island, 1965).
[4]W. Wyss, in *Lectures in Theoretical Physics*, edited by W. E. Britten (Gordon and Breach, New York, 1969), Vol. XI D.
[5]K. Hepp, *Brandeis University Summer Institute 1965* (Gordon and Breach, New York, 1966).
[6]W. Wyss, Commun. Math. Phys. **27**, 223 (1972).
[7]A. Grothendieck, "Produits tensoriels topologiques et espaces nucléaires," Mem. Am. Math. Soc. **16**, (1955).
[8]F. Trèves, *Topological Vector Spaces, Distribution and Kernels* (Academic, New York, 1967).
[9]H. H. Schaefer *Topological Vectorspaces* (Macmillan, New York, 1966).

The decay of the continuous spectrum for solutions of the Korteweg-deVries equation

Mark J. Ablowitz and Alan C. Newell

Department of Mathematics, Clarkson College of Technology Potsdam, New York 13676
(Received 15 December 1972)

The asymptotic behavior of the solution $u(x, t)$ of the Korteweg–deVries equation $u_t + uu_x + u_{xxx} = 0$ is investigated for the class of problems where the initial data does not give rise to an associated discrete spectrum. It is shown that the behavior is different in the three regions (i) $x \gg t^{1/3}$, (ii) $x = 0(t^{1/3})$, (iii) $x \ll -t^{1/3}$. Asymptotic solutions in each of these regions are found which match at their respective boundaries. One of the Berezin–Karpman similarity solutions is the asymptotic state in (ii) and $u(x, t)$ decays like $1/t^{2/3}$ in this region. For region (i) there is exponential decay, whereas in region (iii) the structure of $u(x, t)$ is highly oscillatory and the amplitudes of the oscillations decay as $1/t^{1/2}$ when $-x/t$ is of order unity. For x/t large and negative these amplitudes decay at least as $1/(-x/t)^{1/4}t^{1/2}$.

1. INTRODUCTION

In recent years there has been considerable interest in certain classes of nonlinear evolution equations which arise in a large number of physical contexts. In particular, special attention has been paid to the classical Korteweg-deVries (KdV) equation[1] which was first suggested in order to describe the development and propagation of moderately small amplitude shallow water waves. In a truly remarkable paper, Gardner, Green, Kruskal, and Miura[2] outlined a method by which one could obtain the general solution to the KdV equation once appropriate initial data is given.

Essentially, the ideas are the following. By judiciously considering the eigenvalues of the equation

$$\psi_{xx} + [\lambda + (u/6)]\psi = 0, \tag{1.1}$$

where the potential $u(x,t)$ satisfies the KdV equation

$$u_t + uu_x + u_{xxx} = 0 \tag{1.2}$$

on $-\infty < x < \infty$, it is possible to relate the solution of (1.2) to a linear integral equation so long as the initial data decays sufficiently rapidly as $|x| \to \infty$. The integral equation is the same one found by Gel'fand and Levitan[3] to solve the inverse scattering problem. Indeed, the solution to (1.2) is given by

$$u(x,t) = 12\frac{d}{dx}K(x,x,t), \tag{1.3}$$

where $K(x,y;t)$ for $y > x$ satisfies the Gel'fand–Levitan integral equation

$$K(x,y;t) + B(x+y;t) + \int_x^\infty K(x,z;t)B(y+z;t)dz = 0 \tag{1.4}$$

with

$$B(\xi;t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} b_0(k)e^{i(k\xi+8k^3t)}dk + \sum_n C_n^2(0)e^{-K_n\xi+8K_n^3t}. \tag{1.5}$$

In Eq. (1.5), $b_0(k)$ is the reflection coefficient corresponding to the initial condition

$$u(x,0) = g(x) \tag{1.6}$$

and $C_n(0)$ corresponds to the amplitudes of the discrete eigenfunctions of (1.1). The terms corresponding to the discrete eigenvalues ($\lambda = -K_n^2$) of (1.1) directly relate to a sequence of permanent traveling waves with wave speeds $4K_n^2$. For large times, these traveling waves or solitons appear as separate extremely stable waves having the form

$$u(x,t) = 12K_n^2\,\mathrm{sech}^2K_n(x - x_n - 4K_n^2t). \tag{1.7}$$

Neglect of the reflection term in (1.5) leads to an exact solution of (1.4) first found by Kay and Moses.[4] Gardner, Green, Kruskal, and Miura pointed out how these ideas could be used to predict the number of solitons which emerge from arbitrary initial conditions as well as the interaction properties. Indeed, these waves do not interact strongly in the sense that no radiation is produced; they pass through each other without change of shape and the only interaction memory is a phase shift.

To date, the important question of the decay of the initial data when no solitons are produced is yet unresolved. In this paper, we deduce the longtime structure for the solution of the KdV equation when the initial data does not give rise to discrete eigenvalues on (1.1), and decays sufficiently rapidly as $|x| \to \infty$. The KdV equation is Galilean invariant but this choice of boundary condition fixes the reference frame. A sufficient condition for the $C_n(0) = 0$ is that $g(x) < 0$ for $-\infty < x < \infty$.

Consider (1.5) with $C_n = 0$. A stationary phase analysis suggests that when t is large, and x/t is negative and finite (specifically $x \ll -t^{1/3}$), the structure of $B(x,t)$ and ultimately $u(x,t)$ for $x \ll -t^{1/3}$ is highly oscillatory. The oscillation amplitudes decay as $1/\sqrt{t}$. We also show that a "mean" contribution of $O(1/t)$ plays a crucial role. For x/t positive (specifically $x \gg t^{1/3}$)$B(x,t)$ and ultimately $u(x,t)$ for $x \gg t^{1/3}$ decays exponentially. In the middle regime the stationary points of the phase in the integral for $B(x,t)$ coalesce. In this region, the structure of the solution is deduced by recognizing that essentially $B(x,t)$ takes the form

$$B(x;t) \sim \frac{b_0(0)}{t^{1/3}}A\left(\frac{x}{t^{1/3}}\right) + \cdots, \tag{1.8}$$

where

$$A\left(\frac{x}{t^{1/3}}\right) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\exp\{i[k(x/t^{1/3}) + 8k^3]\}dk. \tag{1.9}$$

Expressions (1.8) and (1.9) suggest that in this regime a consistent solution to the Gel'fand–Levitan equation may be found by looking for a solution for $K(x,y,t)$ of the form

$$K(x,y,t) = (1/t^{1/3})K_s(X,Y) + \cdots, \qquad X = x/t^{1/3}, \quad Y = y/t^{1/3}. \tag{1.10}$$

Substitution of (1.10) and (1.8) into (1.4) results to leading order in a canonical Gel'fand–Levitan integral

Copyright © 1973 by the American Institute of Physics

equation for K_s with $x, y,$ and z replaced by $x/t^{1/3}$, $y/t^{1/3}$, and $z/t^{1/3}$, respectively. In this way the parameter t is eliminated. This implies that $u(x,t)$ must have the form

$$u(x,t) = (1/t^{2/3})F(x/t^{1/3}),$$

within $-t^{1/3} < x < t^{1/3}$. (1.11)

However, since the integral

$$\int_x^\infty K(x,z)B(y+z)dx \tag{1.12}$$

involves the behavior of the functions $K(x,z)$ and $B(y+z)$ in the region $z \gg O(t^{1/3})$ care must be exercised in obtaining (1.11). However, we will show that for $z \gg O(t^{1/3})$, $B(y+z)$ and $K(x,z)$ are exponentially decaying and thus the contribution of these functions to the integral is negligible. Nevertheless, we still emphasize that the preceding argument is merely meant to be suggestive. Confirmation that the choice (1.11) is correct relies on a successful matching of the asymptotic solutions of the three different regions.

Similarity solutions which are precisely in the form (1.11) have been examined by Berezin and Karpman.[5] They investigated the properties of the similarity solution and showed that there is a three parameter manifold of solutions. Only one solution, however, has the necessary properties which will match the exponential decay for large positive $x/t^{1/3}$ and the rapidly oscillating structure for large negative $x/t^{1/3}$.

The fact that, for large negative $(x/t^{1/3})$, $u(x,t)$ is rapidly oscillating suggests that the Whitham,[6] Kruskal and Zabusky[7] WKB approach may provide a relevant description in this region. Specifically, we examine the WKB theory and show that in a certain limit, the solutions obtained from this approach agree with the Berezin–Karpman similarity solution. Of interest is the appearance of a mean term which decays like $1/t$ and which is produced as the DC component of a quadratic self-interaction of the rapid oscillation.

The results obtained from this theory suggest the form in which an asymptotic (large t) solution can be obtained for the Gel'fand–Levitan equation in the $x \ll -t^{1/3}$ regime. We show that

$$K(x,y,t) \sim M\left(\frac{-x}{t}, \frac{-y}{t}, t\right) + \frac{1}{\sqrt{t}}\left\{h\left(\frac{-x}{t}, \frac{-y}{t}\right) \exp\left[i\theta\left(-\frac{x+y}{t}, t\right)\right] + (*)\right\} + \cdots, \tag{1.13}$$

where

$$\theta(-x,t) = \frac{2}{3\sqrt{24}}\left(\frac{-x}{t}\right)^{3/2} t + \theta_0$$

and $(*)$ denotes the complex conjugate, provides the required asymptotic representation of the solution in this regime. The first term plays the role of the mean.

The solution to the KdV equation therefore has the following structure as $t \to +\infty$. For $x/t^{1/3} \gg 1$,

$$u(x,t) \sim \frac{3^{1/4}}{\sqrt{\pi t}}\left(\frac{x}{t}\right)^{1/4} b_0\left(i\left(\frac{1}{12}\frac{x}{t}\right)^{1/2}\right) \times \exp\left[-\frac{2}{3\sqrt{3}}\left(\frac{x}{t}\right)^{3/2} t\right] + \cdots; \tag{1.14}$$

for $|x|/t^{1/3} = O(1)$,

$$u(x,t) \sim \frac{1}{t^{2/3}} f(\eta) \quad \text{where } \eta = \frac{-x}{t^{1/3}}; \tag{1.15}$$

and for $x/t^{1/3} \ll -1$,

$$u(x,t) \sim -\frac{3^{1/4}}{\sqrt{\pi t}}\left(\frac{-x}{t}\right)^{1/4} \frac{b_1([(1/12)(-x/t)]^{1/2})}{1 - \frac{1}{4}|b_1([(1/12)(-x/t)]^{1/2})|^2} \times \exp\left\{i\left[\frac{2}{3\sqrt{3}}\left(\frac{-x}{t}\right)^{3/2} t + \frac{\pi}{4}\right]\right\} + (*) - \frac{3^{1/2}}{\pi t(-x/t)^{1/2}} \times \left(\frac{b_1((1/12)(-x/t))}{1 - \frac{1}{4}|b_1([(1/12)(-x/t)]^{1/2})|^2}\right)^2 + \cdots, \tag{1.16}$$

where $b_1(k)$ is the reflection coefficient corresponding to the initial condition $u(x,0) = g(-x)$. The analysis in the region $x \ll -t^{1/3}$ is not straightforward and relies on using the reflection properties of solutions to the KdV equation in order to avoid having to know the behavior of $K(x,z)$ and $B(y+z)$ in the integral (1.12) for values of the arguments corresponding to the regimes $|z| = O(t^{1/3})$ and $z \gg t^{1/3}$. In addition, for $|z|/t^{1/3} = O(1)$ and $t \to \infty$, the quantity $h(-x/t, -y/t)$ in (1.13) undergoes a rapid transition as $y \to x$. The expression for the mean quantity $M(-x/t, -x/t)$ is then not readily differentiable. In order to find the correct asymptotic behavior for the mean component of $u(x,t) = 12(d/dx)K(x,x,t)$, the differentation must be carried out before the asymptotic analysis is completed.

From (1.14) and (1.16) a particular choice of similarity solution (out of a three parameter family) is necessary in order to have the proper asymptotic behavior. This solution occurs so long as the coefficient of the exponential decay in (1.14) is below a crucial constant. Berezin and Karpman point out that this coefficient [in our case $3^{1/4}|b_0(0)|/\sqrt{\pi}$ must be less than 0.8 in order for a numerical integration of the $f(-x/t^{1/3})$ equation to lead to oscillatory behavior for $x/t^{1/3} \ll -1$. In our case, the maximum allowable value of $3^{1/4}|b_0(0)|/\sqrt{\pi}$ is $3^{1/4}/\sqrt{\pi} \approx 0.74$.

Recent work by Zakharov and Shabat[8] and Ablowitz, Kaup, Newell, and Segur[9] has shown that the inverse scattering method is applicable for other nonlinear evolution equations. In particular, the equation

$$iu_t + u_{xx} + 2\alpha u^2 u^* = 0, \quad \alpha = \mp 1, \tag{1.17}$$

which describes the evolution of an almost monochromatic wave train in a dissipation free system[10,11] has been solved. When $\alpha = -1$, the scattering problem is self-adjoint and thus only the continuous spectrum arises. In this case, the asymptotic solution is given, for all x, by

$$u(x,t) \sim \frac{1}{2\sqrt{\pi t}} \frac{b_0(-x/4t)}{1 - \frac{1}{4}|b_0(-x/4t)|^2} e^{i(x^2/4t) - i(\pi/4)} + \cdots. \tag{1.18}$$

The reflection coefficient $b_0(k)$ is determined from the initial conditions. When $\alpha = +1$, the final state is one of individual and paired permanent waves (corresponding to the discrete spectrum of the scattering problem) superimposed on an algebraically decaying field of self-similar form, much like (1.18), which arises from the continuous part of the spectrum. Note for t large, $x = O(t^{1/2})$, (1.18) reduces to the similarity solution found by Benney and Newell.[10]

2. SIMILARITY SOLUTION FOR THE KdV EQUATION

We seek solutions to the KdV equation (1.2) in the form suggested in the previous section

$$u(x,t) = (1/t^{2/3})f(\eta), \quad \eta = -x/t^{1/3}, \tag{2.1}$$

and obtain the following ordinary differential equation for f,

$$3f''' + 3ff' + 2f + \eta f' = 0. \tag{2.2}$$

From the Gel'fand–Levitan equation (1.4), we expect the relevant solution as $\eta \to -\infty$ is the one which decays exponentially

$$f \sim a(-\eta)^{1/4} e^{(2/3\sqrt{3})(-\eta)^{3/2}}. \tag{2.3}$$

In Ref. 5 it is shown that if we begin at $\eta = -\infty$ with (2.3) for $|a| < 0.8$ and integrate through to $\eta = +\infty$ with (2.2), the asymptotic state on the right is

$$f \sim d\eta^{1/4} e^{(2i/3)\sqrt{3}\eta^{3/2}} + (*) + \cdots. \tag{2.4}$$

As previously mentioned, the fact that the reflection coefficient is less than unity ensures that this solution branch is achieved. In order to obtain the criterion for matching, we derive (2.4) as well as its next correction. For η large and positive, the dominant terms in (2.4) are the first and fourth provided that the order of f is less than the order of η for large f. For convenience we use the small parameter ϵ $(0 < \epsilon \ll 1)$ to scale η

$$\eta = \xi/\epsilon \tag{2.5}$$

and anticipating the result, scale

$$f = (1/\epsilon^{1/4})g. \tag{2.6}$$

Introducing the fast scale

$$\theta = h(\xi)/\epsilon^{3/2},$$

to describe the fine oscillatory structure in addition to the order one scale ξ. By usual two-timing considerations

$$\frac{d}{d\xi} = \frac{h}{\epsilon^{3/2}}\frac{\partial}{\partial\theta} + \frac{\partial}{\partial\xi}. \tag{2.7}$$

Substitution into (2.2) gives us the equation for $g(\theta, \xi) = g_0 + \epsilon^{3/4}g_1 + \epsilon^{3/2}g_2 + \cdots$,

$$\begin{aligned}&\left[\left(3h'^3\frac{\partial^3}{\partial\theta^3} + \xi h'\frac{\partial}{\partial\theta}\right) + \epsilon^{3/2}\left(9h'h''\frac{\partial^2}{\partial\theta^2}\right.\right.\\ &\left.\left. + 9h'^2\frac{\partial^3}{\partial\theta^2\partial\xi} + \xi\frac{\partial}{\partial\xi} + 2\right)\right]\\ &\times(g_0 + \epsilon^{3/4}g_1 + \epsilon^{3/2}g_2 + \epsilon^{9/4}g_3)\\ &= \frac{-3}{2}\epsilon^{3/4}\left(h'\frac{\partial}{\partial\theta} + \epsilon^{3/2}\frac{\partial}{\partial\xi}\right)\\ &\times[g_0^2 + 2\epsilon^{3/4}g_0g_1 + \xi^{3/2}(g_0g_2 + g_1^2)].\end{aligned} \tag{2.8}$$

Solving iteratively, we find

$$g_0 = B(\xi)e^{i\theta} + B^*(\xi)e^{-i\theta}, \tag{2.9}$$

$$g_1 = A(\xi) + \frac{1}{6(h')^2}B^2e^{2i\theta} + \frac{1}{6(h')^2}B^{*2}e^{-2i\theta}, \tag{2.10}$$

with the following side conditions which are necessary to remove θ^2 and θ terms at the orders $O(\epsilon^{3/4})$, $O(\epsilon^{3/2})$ and $O(\epsilon^{9/4})$:

$$(h')^2 = \tfrac{1}{3}\xi, \tag{2.11}$$

$$2\xi\frac{dB}{d\xi} - \frac{1}{2}B = 3ih'\left(\frac{1}{6(h')^2}BB^* + A\right)B, \tag{2.12}$$

$$\xi\frac{dA}{d\xi} + 2A = -3\frac{d}{d\xi}(BB^*). \tag{2.13}$$

Solving, and writing f in terms of η, it is found that

$$f = d\eta^{1/4}e^{i[(2/3\sqrt{3})\eta^{3/2} - (3/4\sqrt{3})d^2\log\eta + \theta_0]} + (*) - d^2\eta^{-1/2} + \cdots. \tag{2.14}$$

Note that while the structure of the solution is found, the amplitude d remains indeterminate, and via (2.3) and (2.14) it is not clear how a and d are related.

3. ASYMPTOTIC EVALUATION OF THE GEL'FAND-LEVITAN EQUATION

In this section we outline a procedure by which consistent asymptotic solutions to the Gel'fand–Levitan equation (1.4) may be obtained. We require that the initial data $u(x, 0)$ gives no discrete eigenvalues in the corresponding Schrodinger equation; namely, no solitons are present. As previously discussed the solution has different representations in the three regions (i) $x \gg t^{1/3}$, (ii) $x = O(t^{1/3})$, (iii) $x \ll -t^{1/3}$. In region (i) the asymptotic solution may be found by using the asymptotic expression for $B(\xi, t)$ and iterating successively on the Gel'fand–Levitan equation. In region (ii), one may obtain an expression for the long time behavior of $B(\xi, t)$ in terms of an integral closely related to the Airy function. In this regime, successive iteration on the Gel'fand–Levitan equation does not lead to a uniform asymptotic expansion for $K(x, y, t)$ and $u(x, t)$. Nevertheless, as discussed in the introduction, the structure of the solution in this region is suggested by the long time behavior of $B(\xi, t)$ to be $(1/t^{2/3})f(-x/t^{1/3})$. In region (iii), one can again successfully solve the Gel'fand–Levitan equation from a knowledge of the asymptotic behavior of $B(\xi, t)$. However, the analysis is more subtle for several reasons and involves the successive interaction between a mean-like term and a rapidly oscillating term which is in phase with $B(\xi, t)$. These interactions arise from the integral term

$$\int_x^\infty K(x, z)B(y + z)dz \tag{3.1}$$

in the Gel'fand–Levitan equation.

The analysis in region (i) is the most straightforward. Given

$$B(\xi, t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} b_0(k)e^{ik\xi + 8ik^3t}dk,$$

a steepest descents analysis yields

$$B(\xi, t) \sim \frac{1}{\sqrt{96\pi t}}\frac{b_0(i\sqrt{\xi/24t})}{[(1/24)(\xi/t)]^{1/4}}e^{-(2/3\sqrt{24})(\xi/t)^{3/2}t} + \cdots \tag{3.2}$$

for t large and positive and ξ/t positive and nonzero. [Note that $b_0(i\sqrt{\xi/24t})$ is real]. The steepest descents analysis requires that the integration contour be moved from the real k axis so that it passes through one of the critical points $k = +i\sqrt{\xi/24t}$. Therefore, in order for (3.2) to hold, $b_0(k)$ can have no poles between the original and the new contour. Note that there cannot be any poles in the intervening region as the poles of $b_0(k)$ in the upper half plane are related to the solitons, which we have assumed not to be present. However, if solitons did exist, the present analysis would, with modified reflection coefficients, describe the state in the region $V_st \gg x$ where V_s is the speed of the slowest solitary wave. Since the integral (3.1) will have terms which are not only algebraically but exponentially smaller than $B(\xi, t)$ it is consistent to leading order to set

$$K(x,y,t) \sim -B(x+y,t)$$
$$= \frac{-1}{\sqrt{96\pi t}} \frac{b_0(i\sqrt{(1/24)[(x+y)/t]})}{\{(1/24)[(x+y)/t]\}^{1/4}}$$
$$\times e^{(-2/3\sqrt{24})[(x+y)/t]^{3/2}t} + \cdots. \quad (3.3)$$

From (1.3), we readily deduce the asymptotic state of $u(x,t)$ for $x/t^{1/3} \gg 1$,

$$u(x,t) \sim \frac{3^{1/4}}{\sqrt{\pi t}} \left(\frac{x}{t}\right)^{1/4} b_0\left(i\sqrt{\frac{x}{12t}}\right) e^{(-2/3\sqrt{3})(x/t)^{3/2}t} \quad (3.4)$$

which for small x/t agrees with the Karpman similarity solution if we take

$$a = (3^{1/4}/\sqrt{\pi}) b_0(0). \quad (3.5)$$

Note in particular the requirement that $|b_0(0)| \leq 1$ implies that $|a| \leq 3^{1/4}/\sqrt{\pi}$.

The analysis in region (iii) $(x/t^{1/3}) \ll -1$ is not straightforward since the integral (3.1) requires that we must use the long time behavior of $B(\xi,t)$ in the three regions. However, we will circumvent this difficulty by using some reflection properties of the KdV equation (1.2). We can readily see that if $u = f(x,t)$ is a solution of (1.2) with $u(x,0) = g(x)$, then $u = +f(-x,-t) = f_R(x,t)$ is a solution with $u(x,0) = g(-x) = g_R(x)$. Let us look, therefore, for the solution $f_R(x,t)$ in the limit $t \to -\infty$ $x > 0$ and then after completing the analysis relate this to $f(x,t)$ as $t \to +\infty$. Suppose the reflection coefficient corresponding to $g(-x)$ is $b_1(k)$. Then,

$$B_R(\xi,t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} b_1(k)\, e^{-i|t|[k(\xi/t)+8k^3]} dk \quad (3.6)$$

and using a stationary phase analysis

$$B_R(\xi,t) \underset{t\to-\infty}{\sim} \frac{\beta}{(|t|)^{1/2}} b\left(\frac{-\xi}{t}\right) e^{i\theta(\xi,t)} + (*), \quad (3.7)$$

where β, θ, and b are given by

$$\beta = 1/\sqrt{96\pi},$$
$$b\left(\frac{-\xi}{t}\right) = \frac{b_1(\sqrt{(1/24)(-\xi/t)})}{[(1/24)(-\xi/t)]^{1/4}}, \quad (3.8)$$
$$\theta(\xi,t) = (2/3\sqrt{24})(-\xi/t)^{3/2}|t| - \pi/4.$$

Given that the initial data satisfies the condition

$$\int |g(-x)|(1+|x|)dx < \infty$$

it has been shown by Faddeev[12] that the Gel'fand–Levitan equation can be solved by a Neumann series. At first glance, it appears that the series generated in this way is not asymptotic as $|t| \to \infty$, but on closer examination it is possible to isolate the leading asymptotic contributions of each of the terms of the Neumann series and hence of $K(x,x;t)$. The dominant terms consist of two essential parts; one part is an oscillatory component in phase with B_R with amplitude proportional to $1/|t|^{1/2}$ and the other part is an order one mean component generated by the D.C. interaction of $K(x,z)$ with $B_R(y+z)$. In what follows we show how these terms can be obtained by appropriately analyzing and then summing the Neumann series. After obtaining these results we remark on how this solution may be obtained by directly applying stationary phase analysis to (1.4). It is convenient to rescale variables,

$$x = \bar{x}|t|, \quad y = \bar{y}|t|, \quad z = \bar{z}|t|, \quad (3.9)$$

whereupon the Gel'fand–Levitan equation (1.4) reads

$$H(\bar{x},\bar{y},|t|) + B_R(\bar{x}+\bar{y}) + |t| \int_{\bar{x}}^{\infty} H(\bar{x},\bar{z},|t|) B_R(\bar{y}+\bar{z}) d\bar{z} = 0, \quad (3.10)$$

where $K(x,y,t) = H(\bar{x},\bar{y},|t|)$.

We write

$$H(\bar{x},\bar{y},|t|) = \sum_{n=0}^{\infty} H_n(\bar{x},\bar{y},|t|) \quad (3.11)$$

and solve (3.10) recursively. The first term in (3.11) is given by

$$H_0(\bar{x},\bar{y},|t|) \sim -B_R(\bar{x}+\bar{y})$$
$$= -[\beta b(\bar{x}+\bar{y})/|t|^{1/2}]\, e^{i\theta(\bar{x}+\bar{y})} + (*) + \cdots \quad (3.12)$$

where $\theta(\bar{x}+\bar{y}) = (2/3\sqrt{24})(\bar{x}+\bar{y})^{3/2}|t| - \pi/4$ and β and b are as given in (3.8). The second term $H_1(\bar{x},\bar{y},|t|)$ contains two components, one of which $(\bar{H}_1)$ is an order one mean term when $\bar{y} = \bar{x}$ and the second (H_1) is a rapidly oscillating term, always of order $1/|t|$, which is essentially the second harmonic of the oscillation $e^{i\theta(\bar{x}+\bar{y})}$. The latter term plays no central role in the analysis; it does not return at subsequent order to reproduce first harmonic terms of order $1/|t|^{1/2}$ and to this order can be omitted. In what follows we will only write down the terms which ultimately provide the principal contributions (as $|t| \to \infty$) to H_1, H_2, etc. The asymptotic analysis on these leading order contributions is postponed until some subtle points connected with the analysis are discussed. We find,

$$H_1(\bar{x},\bar{y},|t|) \sim \bar{H}_1(\bar{x},\bar{y},|t|)$$
$$= \beta^2 \int_{\bar{x}}^{\infty} b(\bar{x}+\bar{z})\, b^*(\bar{y}+\bar{z})\, e^{i[\theta(\bar{x}+\bar{z})-\theta(\bar{y}+\bar{z})]} d\bar{z} + (*). \quad (3.13)$$

The next term $H_2(\bar{x},\bar{y},|t|)$ has a leading order contribution denoted $H_2^{(1)}(\bar{x},\bar{y},|t|)$ which is generated by the interaction between the mean $\bar{H}_1(\bar{x},\bar{z},|t|)$ and the oscillatory $B_R(\bar{y}+\bar{z})$. A stationary phase analysis shows this term to be in phase with $B_R(\bar{x}+\bar{y})$ and of order $1/|t|^{1/2}$:

$$H_2(\bar{x},\bar{y},|t|) \sim H_2^{(1)}(\bar{x},\bar{y},|t|)$$
$$= -\beta^3|t|^{1/2} \int_{\bar{x}}^{\infty}\int_{\bar{x}}^{\infty} b(\bar{y}+\bar{z})\, b(\bar{x}+\bar{k})\, b^*(\bar{z}+\bar{k})$$
$$\times e^{i[\theta(\bar{y}+\bar{z})+\theta(\bar{x}+\bar{k})-\theta(\bar{z}+\bar{k})]} d\bar{z}d\bar{k} + (*). \quad (3.14)$$

Thereafter, the pattern repeats. The principal contribution to $H_3(\bar{x},\bar{y},|t|)$ is a mean like term which we call $\bar{H}_3(\bar{x},\bar{y},|t|)$ (order one and, for large $|t|$, independent of $|t|$ when $\bar{y} = \bar{x}$). This term is generated by the DC component of the product of $H_2^{(1)}(\bar{x},\bar{z},|t|)$ and $B_R(\bar{x}+\bar{z})$ in the integral $\int_{\bar{x}}^{\infty} H_2^{(1)}(\bar{x},\bar{z},|t|) B_R(\bar{x}+\bar{z}) d\bar{z}$ and is given by

$$H_3(\bar{x},\bar{y},|t|) \sim \bar{H}_3(\bar{x},\bar{y},|t|)$$
$$= \beta^4|t| \int_{\bar{x}}^{\infty}\int_{\bar{x}}^{\infty}\int_{\bar{x}}^{\infty} b^*(\bar{y}+\bar{z})\, b(\bar{z}+\bar{l})\, b(\bar{x}+\bar{k})\, b^*(\bar{k}+\bar{l})$$
$$\times e^{i[\theta(\bar{z}+\bar{l})+\theta(\bar{x}+\bar{k})-\theta(\bar{l}+\bar{k})-\theta(\bar{y}+\bar{z})]} d\bar{z}d\bar{k}d\bar{l} + (*). \quad (3.15)$$

The principal contribution of $H_4(\bar{x},\bar{y},|t|)$ is a term of order $1/|t|^{1/2}$ for large $|t|$ and in phase with both $B_R(\bar{x}+\bar{y})$ and $H_2^{(1)}(\bar{x},\bar{y},|t|)$:

$$H_4(\bar{x},\bar{y},|t|) \sim H_4^{(1)}(\bar{x},\bar{y},|t|)$$
$$= -\beta^5|t|^{3/2} \int_{\bar{x}}^{\infty}\int_{\bar{x}}^{\infty}\int_{\bar{x}}^{\infty}\int_{\bar{x}}^{\infty} b(\bar{y}+\bar{z})\, b^*(\bar{z}+\overline{m})$$

$$\times b(\bar{m}+\bar{l})\,b(\bar{x}+\bar{k})\,b^*(\bar{k}+\bar{l})$$
$$\times e^{i[\theta(\bar{y}+\bar{z})+\theta(\bar{m}+\bar{l})+\theta(\bar{x}+\bar{k})-\theta(\bar{k}+\bar{l})-\theta(\bar{z}+\bar{m})]}$$
$$\times d\bar{z}d\bar{k}d\bar{l}d\bar{m} + (*). \tag{3.16}$$

The principal contribution in $H_5(\bar{x},\bar{y},|t|)$ will be a mean-like $H_5(\bar{x},\bar{y},|t|)$ given by

$$H_5(\bar{x},\bar{y},|t|) \sim \bar{H}_5(\bar{x},\bar{y},|t|)$$
$$= \beta^6 |t|^2 \int_{\bar{x}}^{\infty}\int_{\bar{x}}^{\infty}\int_{\bar{x}}^{\infty}\int_{\bar{x}}^{\infty}\int_{\bar{x}}^{\infty} b^*(\bar{y}+\bar{z})\,b(\bar{z}+\bar{n})$$
$$\times b^*(\bar{n}+\bar{m})\,b(\bar{m}+\bar{l})\,b(\bar{x}+\bar{k})\,b^*(\bar{k}+\bar{l})$$
$$\times e^{i[\theta(\bar{z}+\bar{n})+\theta(\bar{m}+\bar{l})+\theta(\bar{x}+\bar{k})-\theta(\bar{l}+\bar{k})-\theta(\bar{n}+\bar{m})-\theta(\bar{y}+\bar{z})]}$$
$$\times d\bar{z}d\bar{k}d\bar{l}d\bar{m}d\bar{n} + (*). \tag{3.17}$$

The asymptotic ($|t| \to \infty$) behavior of the principal contributions can be determined by means of stationary phase methods. Specifically, we shall repeatedly use the result given by Lewis,[13] that

$$\int_D g(\mathbf{p})\,e^{i\lambda\phi(\mathbf{p})}\,d\mathbf{p} \sim \left(\frac{2\pi}{\lambda}\right)^{n/2} [\det(\phi_{jk})]^{-1/2}$$
$$g(\mathbf{q})\,e^{i\lambda\phi(\mathbf{q})+(i\pi/4)\,\mathrm{sig}(\phi_{jk})}, \quad \lambda \to \infty, \tag{3.18}$$

where the domain of integration D is in E^n (n-dimensional Euclidean space) and contains one stationary point at $\mathbf{p}=\mathbf{q}$, i.e., $\mathrm{grad}_{\mathbf{p}}\,\phi = 0$. ϕ_{jk} is the matrix of second partial derivatives evaluated at the stationary point. Finally, $\mathrm{sig}\phi_{jk}$ is the sum of the signs of the eigenvalues of (ϕ_{jk}). There are some slight modifications which we shall need. Consider $n=2$ and, in particular, the integral given in (3.14) for $H_2^{(1)}(\bar{x},\bar{y},|t|)$. The stationary point is at $\bar{z}=\bar{x}$, $\bar{k}=\bar{y}$ and lies on the boundary of D. In this case the contribution from the stationary point will be $\frac{1}{2}$ that given above. However, we are primarily interested in $u(x,t) = 12(d/dx)K(x,x,t) = (12/|t|)(d/d\bar{x})\,H(\bar{x},\bar{x},|t|)$. But when $y=x$ (implies $\bar{y}=\bar{x}$), the stationary point lies at the corner of D. The principal contribution is then $\frac{1}{4}$ that given above. It has been shown[14] that the transition from the value $\frac{1}{2}$ to $\frac{1}{4}$ as the stationary point $(\bar{x},\bar{y})$ tends to the corner $(\bar{x},\bar{x})$ along the boundary of D involves Fresnel-like integrals. The change takes place fairly abruptly in the region $y-x = O(1/|t|^{1/2})$, i.e. the transition is smooth but involves derivatives which are of order $|t|^{1/2}$. A similar situation occurs in the higher order integrals.

Applying the stationary phase formula (3.18) (suitably modified at all corners of D) to the sequences $H_0, H_2^{(1)}$, $H_4^{(1)}$, and $\bar{H}_1, \bar{H}_3$ and $\bar{H}_5$ a pattern is seen to emerge and the result to all orders (in bb^*) can be generated. Using the definitions (3.8), we find

$$K(x,x;t) \sim \frac{-\beta b(2\bar{x})}{|t|^{1/2}}\,\frac{1}{1-\frac{1}{4}|b_1(\sqrt{\bar{x}/12})|^2}\,e^{i\theta(2\bar{x})} + (*)$$
$$+ 2\beta^2 \int_{\bar{x}}^{\infty} b(\bar{x}+\bar{z})\,b^*(\bar{x}+\bar{z})\,\frac{d\bar{z}}{1-\sigma|b_1(\sqrt{(\bar{x}+\bar{z})/24}|^2} \tag{3.19}$$

where $\sigma = \frac{1}{2}$ when $\bar{z}-\bar{x} = 0(1)$ and $\sigma = \frac{1}{4}$ when $\bar{z}-\bar{x}=0$.

Before discussing the derivation of $u(x,t)$ from $K(x,x;t)$, it must be mentioned that the result for $K(x,y;t)$ and hence $K(x,x;t)$ can be obtained in a surprisingly direct manner. Looking for a solution of (1.4) as $|t| \to \infty$ of the form

$$K(x,y,t) \sim M(\bar{x},\bar{y};t)$$
$$+ (1/|t|^{1/2})\,h(\bar{x},\bar{y})\,e^{i\theta(\bar{x}+\bar{y})} + (*) + \cdots, \tag{3.20}$$

yields the relation

$$M(\bar{x},\bar{y};t) = -\beta\int_{\bar{x}}^{\infty} h(\bar{x},\bar{z})\,b^*(\bar{y}+\bar{z})\,e^{i[\theta(\bar{x}+\bar{z})-\theta(\bar{y}+\bar{z})]}$$
$$\times dz + (*) \tag{3.21}$$

and

$$e^{i\theta(\bar{x}+\bar{y})}[h(\bar{x},\bar{y}) + b(\bar{x}+\bar{y})]$$
$$- |t|\int_{\bar{x}}^{\infty}\int_{\bar{x}}^{\infty} h(\bar{x},\bar{k})\,b^*(\bar{z}+\bar{k})\,b(\bar{y}+\bar{z})$$
$$\times e^{i[\theta(\bar{x}+\bar{k})-\theta(\bar{z}+\bar{k})+\theta(\bar{y}+\bar{z})]}\,d\bar{z}d\bar{k} = 0, \tag{3.22}$$

after neglecting all terms of order $1/|t|$. Equation (3.22) is readily solved by use of the stationary phase result (3.18) to find

$$h(\bar{x},\bar{y}) = \frac{-\beta b_1(\sqrt{(\bar{x}+\bar{y})/24})}{[(\bar{x}+\bar{y})/24]^{1/4}}\,\frac{1}{[1-\sigma|b_1(\sqrt{(\bar{x}+\bar{y})/24})|^2]} \tag{3.23}$$

with σ as previously defined. Similarly, the mean agrees with (3.19) when $\bar{y}=\bar{x}$. Since the jump in σ remains localized about $y-x = O(1/|t|^{1/2})$ the basic assumptions in deriving (3.20–3.23) remain satisfied and agreement with the Neumann series is established.

Finally, in order to compute $u(x,t) = (12/|t|)(d/d\bar{x})$ $H(\bar{x},\bar{x};|t|)$ from (3.19) some care must be exercised. Naively, it might be suspected that, in the second term of (3.19), σ should take the value $\frac{1}{2}$ which is its value for most of the range of integration. However, proceeding in this way, the mean term does not quite match to the Berezin–Karpman similarity solution given in (2.14). The difficulty is the existence of the rapid transition (which we loosely call a hiccough) in $h(\bar{x},\bar{z})$ as $\bar{z}\to\bar{x}$. Therefore, even though the expression (3.19) is a valid asymptotic representation for $H(\bar{x},\bar{x},|t|)$ it is not readily differentiable. It is "allowable" to differentiate the rapidly oscillating term since the principal contribution to $u(x,t)$ comes from the differentiation of the rapid phase $\theta(2\bar{x})$ and therefore to leading order the hiccough may be ignored. To determine the mean of $u(x,t)$, however, it is necessary to differentiate the expression for $H(\bar{x},\bar{x},|t|)$ before the asymptotic analysis is attempted.

The differentiation of $\bar{H}_1(\bar{x},\bar{x})$ is straightforward since σ does not enter in (3.13). The differentiation of $\bar{H}_3(\bar{x},\bar{x},|t|)$ can be carried through after first making the transformations

$$\bar{z}+\bar{x}=\bar{z}', \quad \bar{k}+\bar{x}=\bar{k}', \quad \bar{l}-\bar{x}=\bar{l}' \tag{3.24}$$

to remove the explicit $\bar{x}$ dependence from the integrand. The domain of integration in the variables $\bar{z}', \bar{k}'$, and $\bar{l}'$ is $(2\bar{x},\infty)$, $(2\bar{x},\infty)$ and $(0,\infty)$ and differentiation with respect to $\bar{x}$ yields two integrals which may be evaluated by stationary phase methods (here again the stationary point $\bar{k}'=2\bar{x}$, $\bar{l}'=0$ is on the corner of the domain) and which yield equal contributions. The differentiation of $\bar{H}_5(\bar{x},\bar{x},|t|)$ may be carried through after first making the transformations

$$\bar{z}+\bar{x}=\bar{z}', \quad \bar{k}+\bar{x}=\bar{k}', \quad \bar{m}+\bar{x}=\bar{m}', \quad \bar{l}-\bar{x}=\bar{l}',$$
$$\bar{n}-\bar{x}=\bar{n}'. \tag{3.25}$$

The domain of integration in the variables $\bar{z}', \bar{k}', \bar{m}', \bar{l}'$, and $\bar{n}'$ is $(2\bar{x},\infty)$, $(2\bar{x},\infty)$, $(2\bar{x},\infty)$, $(0,\infty)$, and $(0,\infty)$ and differentiation with respect to $\bar{x}$ yields three integrals (each of four dimensions) which may be evaluated by stationary phase methods (stationary point $\bar{k}'=2\bar{x}$, $\bar{m}'=2\bar{x}$, $\bar{l}'=0$, $\bar{n}'=0$ is on the corner of the four dimensional domain). Each of the three integrals yield equal contributions. In four dimensions, the contribution from

a stationary point which lies at a corner is $\frac{1}{16}$ of the contribution of a stationary point which lies in the interior. Again, after calculating the first few terms a pattern emerges. Writing the mean of $u(x,t)$ as $\bar{u}(x,t)$, we find

$$\bar{u}(x,t) \sim -\frac{3^{1/2}}{\pi}\left|\, b_1(\sqrt{\bar{x}/12})\,\right|^2 \times\left[1+\frac{2}{2^2}\left|b_1(\sqrt{\bar{x}/12})\right|^2+\frac{3}{2^4}\left|b_1(\sqrt{\bar{x}/12})\right|^4+\cdots\right]. \quad (3.26)$$

The next term in the mean of $H(\bar{x},\bar{x},|t|), \bar{H}_7(\bar{x},\bar{x},|t|)$ will be a seven dimensional integral. A transformation similar to (3.24) and (3.25) yields four integrations from $(2\bar{x},\infty)$ and three from $(0,\infty)$. Differentiation produces four integrals whose asymptotic behavior can be evaluated by stationary phase methods where the stationary points are corners of the domain of integration. The four integrals produce equal contributions and add a term $(4/2^6)|b_1(\sqrt{\bar{x}/12})|^6$ inside the bracket in (3.26). In general, we find $\bar{H}_{2n+1}(\bar{x},\bar{x},|t|)$ yields a contribution of $[(n+1)/2^{2n}]\,|\,b_1(\sqrt{\bar{x}/12})\,|^{2n}$ inside the bracket in (3.26). The coefficients of this series can be summed to

$$1+\sum_{n=1}^{\infty}\frac{n+1}{2^{2n}}\left|\, b_1\left(\sqrt{\frac{\bar{x}}{12}}\right)\right|^{2n}=\frac{1}{[1-\frac{1}{4}\,|\,b_1(\bar{x}/12)\,|^2]^2} \quad (3.27)$$

and thus

$$u(x,t) \sim \frac{-3^{1/4}\bar{x}^{1/4}\,b_1(\sqrt{\bar{x}/12})}{\sqrt{\pi}\,|t|^{1/2}[1-\frac{1}{4}\,|\,b_1(\sqrt{\bar{x}/12})\,|^2]} \times e^{(2i/3\sqrt{3})\bar{x}^{3/2}|t|+i\pi/4}+(*) -\left(\frac{3^{1/4}}{\sqrt{\pi}}\,\frac{b_1(\sqrt{\bar{x}/12})}{[1-\frac{1}{4}\,|\,b_1(\sqrt{\bar{x}/12})\,|^2]}\right)^2\frac{1}{\bar{x}^{1/2}\,|t|},\quad \bar{x}=\frac{x}{|t|}. \quad (3.28)$$

In order to find the long time behavior of $u(x,t)$ when $u(x,0)=g(x)$, we make the sign reversals

$$x\to -x, \qquad t\to -t$$

and find the asymptotic behavior of $u(x,t)$ which is given by (1.16). For small values of $-x/t$, (3.28) to this order agrees precisely with the Berezin–Karpman similarity solution (2.14) for large $-x/t^{1/3}$. The undetermined amplitude d in both the Berezin–Karpman solution and the solution arising in the WKB formulation [see Eq. (3.33)] is therefore given by

$$d=(-3^{1/4}/\sqrt{\pi})[b_1(0)/(1-\tfrac{1}{4}\,|\,b_1(0)\,|^2)]. \quad (3.29)$$

For large $-x/t$, if

$$b_1(k)=O(1/k) \quad (3.30)$$

we find

$$u \sim -\frac{C}{(-x/t)^{1/4}\,t^{1/2}}\,e^{(2i/3\sqrt{3})(-x/t)^{3/2}t+i\pi/4}+(*) -\frac{C^2}{(-x/t)^{3/2}\,t}, \quad C \text{ a constant}, \quad (3.31)$$

which agrees with the similarity solution (4.34) found from a WKB formulation.

4. WKB THEORY

In this section, we describe the solution in the range $x \ll -t^{1/3}$ using a WKB approach. While the results in this case merely corroborate our previous analysis, in general it may provide the only feasible means for examining the region of rapid oscillation. The idea of this method is to average out the fine oscillations and describe the dynamical evolution of the parameters related to the coarse structure. We introduce the scales

$$\theta=\frac{\Theta(X,T)}{\mu}, \qquad X=\mu x, \qquad T=\mu t, \quad (4.1)$$

where μ is a measure of the ratio of the fine to the coarse structure. Defining

$$\theta_x=\Theta_X=k, \qquad \theta_t=\Theta_T=-\omega, \quad (4.2)$$

we find

$$\frac{\partial}{\partial x}=k\frac{\partial}{\partial\theta}+\mu\frac{\partial}{\partial X}, \qquad \frac{\partial}{\partial t}=-\omega\frac{\partial}{\partial\theta}+\frac{\partial}{\partial T}. \quad (4.3)$$

Substituting (4.3) into (1.2) and looking for solutions

$$u=f(\theta,\mu)+\mu u^{(1)}+\mu^2u^{(2)}+\cdots, \quad (4.4)$$

f satisfies the ordinary differential equation

$$-\omega f'+kff'+k^3f'''=0 \quad (4.5)$$

which, when integrated twice, gives

$$\tfrac{1}{2}k^2f'^2+\tfrac{1}{6}f^3-\tfrac{1}{2}cf^2=E_1f+E_2, \quad (4.6)$$

where E_1 and E_2 are integration constants and $c=\omega/k$. Solutions which are 2π-periodic may be written

$$f=\beta+(\alpha-\beta)\,cn^2\left(\frac{K(m)}{\pi}\,\theta\,|\,m\right), \qquad m=\frac{\alpha-\beta}{\alpha-\gamma}, \quad (4.7)$$

where

$$\alpha+\beta+\gamma=3c, \quad (4.8)$$

$$\alpha\beta+\alpha\gamma+\beta\gamma=E_1, \quad (4.9)$$

$$\alpha\beta\gamma=E_2, \quad (4.10)$$

and

$$\sqrt{(\alpha-\gamma)/12}\,(1/k)=K(m)/\pi. \quad (4.11)$$

In the above formulas $K(m)$ is the complete elliptic integral of the first kind. All the quantities, α,β,γ,E_1, E_2,k, and ω are functions of the slow variables X and T. In order that u be periodic, certain compatibility relations must be satisfied. These are (4.11), conservation of waves

$$k_T+\omega_X=0 \quad (4.12)$$

together with two relations on E_1 and E_2. It is convenient to eliminate the four variables ω,k,E_1, and E_2 in terms of the three roots α,β, and γ. The following relations were obtained by Whitham[6]:

$$\left(\frac{\partial}{\partial T}+a_1\frac{\partial}{\partial X}\right)(\beta+\gamma)=0, \quad (4.13)$$

$$\left(\frac{\partial}{\partial T}+a_2\frac{\partial}{\partial X}\right)(\gamma+\alpha)=0, \quad (4.14)$$

$$\left(\frac{\partial}{\partial T}+a_3\frac{\partial}{\partial X}\right)(\alpha+\beta)=0. \quad (4.15)$$

The propagation speeds a_1,a_2, and a_3 are given by

$$a_1=c-\frac{\alpha-\beta}{3}\,\frac{K}{K-E},$$

$$a_2 = c - \frac{\alpha-\beta}{3}\,\frac{(1-m)K}{E-(1-m)K},$$

$$a_3 = c + \frac{\alpha-\beta}{3}\,\frac{(1-m)K}{mE}, \tag{4.16}$$

where $c = \frac{1}{3}(\alpha+\beta+\gamma)$ and $K(m)$ and $E(m)$ are the complete elliptic integrals of the first and second kind, respectively. For small m, f tends to a cosine whereas for $m \to 1$, f tends to a solitary wave. We examine the small m limit and show that these solutions take on the form of the similarity solution. In powers of $(\alpha-\beta)/4$ (which we call B), the wave speeds are

$$a_1 = \gamma - B + [B^2/2(\alpha-\gamma)] + \cdots, \tag{4.17}$$

$$a_2 = \gamma + B + [B^2/2(\alpha-\gamma)] + \cdots, \tag{4.18}$$

$$a_3 = A - [B^2/(\alpha-\gamma)] + \cdots, \quad A = [(\alpha+\beta)/2]. \tag{4.19}$$

It may be verified that to this order $a_3 \sim \bar{f} = \gamma + (\alpha-\beta)\times E(m)/K(m)$ = mean of f. Substituting (4.17)–(4.19) into Eqs. (4.13)–(4.15) and appropriately combining the first two equations, we obtain

$$\left[\frac{\partial}{\partial T} + \left(\gamma + \frac{B^2}{2(\alpha-\gamma)}\right)\frac{\partial}{\partial X}\right](A+\gamma) + 2B\,\frac{\partial B}{\partial X} = 0, \tag{4.20}$$

$$\left[\frac{\partial}{\partial T} + \left(\gamma + \frac{B^2}{2(\alpha-\gamma)}\right)\frac{\partial}{\partial X}\right]B + \tfrac{1}{2}B\,\frac{\partial}{\partial X}(A+\gamma) = 0, \tag{4.21}$$

$$\left[\frac{\partial}{\partial T} + \left(A - \frac{B^2}{\alpha-\gamma}\right)\frac{\partial}{\partial X}\right]A = 0. \tag{4.22}$$

We will show that the similarity solutions of these equations agree with those found in Sec. 2. Since the equations were derived on the basis of small B, to first order we have

$$\left(\frac{\partial}{\partial T} + \gamma\,\frac{\partial}{\partial X}\right)(A+\gamma) = 0, \tag{4.23}$$

$$\left(\frac{\partial}{\partial T} + A\,\frac{\partial}{\partial X}\right)A = 0. \tag{4.24}$$

The only similarity solution which is consistent with $\alpha-\gamma$ positive and order one is

$$\gamma = +X/T, \quad A = 0. \tag{4.25}$$

Looking for solutions B of the form $(-X)^p T^q$ in (4.21) then $q + p + \frac{1}{2} = 0$. If, in addition, we demand that the solutions match to the similarity solution then $p = \frac{1}{4}$. Formally, therefore, set

$$A = \frac{1}{T^{2/3}}\,\bar{A}(\eta), \quad B = \frac{1}{T^{2/3}}\,\bar{B}(\eta),$$

$$\gamma = \frac{1}{T^{2/3}}\,\bar{\gamma}(\eta), \quad \eta = -\frac{X}{T^{1/3}} \tag{4.26}$$

in (4.20)–(4.22). The resulting equations are satisfied to $O(\eta^{-2})$ by

$$\bar{A} = 0, \quad \bar{\gamma} = -\eta, \quad \bar{B} = d\eta^{1/4} - (d^3/32)\eta^{-5/4}. \tag{4.27}$$

To leading order, the mean is given by,

$$\bar{f} = B^2/\gamma = -d^2\eta^{-1/2}. \tag{4.28}$$

The phase function Θ satisfies the equation

$$\Theta_T + c\Theta_X = 0, \quad c = \gamma/3 = X/3T, \tag{4.29}$$

which integrates to give

$$\Theta = F(-X/T^{1/3}). \tag{4.30}$$

To leading order, this solution may be identified with the similarity solution by the choice

$$F = (2/3\sqrt{3})\,(-X/T^{1/3})^{3/2} \tag{4.31}$$

and then the phase

$$\theta = \Theta/\mu = (2/3\sqrt{3})\,(-x/t^{1/3}) + \theta_0. \tag{4.32}$$

For large t, therefore, the WKB approach leads to solutions which have the form

$$u \sim \frac{d}{t^{2/3}}\,\eta^{1/4}\,e^{i(2/3\sqrt{3})\eta^{3/2}} + (*) - \frac{d^2}{t^{2/3}\eta^{1/2}} + \cdots. \tag{4.33}$$

These solutions correspond to the similarity solution found in Sec. 2 and as previously discussed, in the small amplitude limit, the oscillation amplitude d is indeterminate. Note also that (4.33) suggests the form in which to seek asymptotic solutions of the Gel'fand–Levitan equation (1.4). In Sec. III the amplitudes have been determined and are related to the reflection coefficient of the inverse scattering problem associated with (1.1).

As a final remark, we point out that small amplitude similarity solutions to (4.20)–(4.21) also exist in the range $X/T \ll -1$. It may readily be verified that

$$\gamma \sim +\frac{x}{t}, \quad A \sim 0, \quad B \sim \frac{1}{t^{1/2}}\left(\frac{-x}{t}\right)^{-m} \quad \text{all } m > 0, \tag{4.34}$$

satisfies (4.20)–(4.22) to leading order. These solutions agree with the asymptotic solution of the Gel'fand–Levitan equation for $x/t \ll -1$ given by (1.16) since $b_1(\sqrt{(1/12)(-x/t)})$ must decay as $(-x/t)^{-n}$.

CONCLUDING REMARKS

Specifically we have shown when no solitons are present the long-time solution of the KdV equation consists of oscillations and exponential decay matched together by a solution of self-similar form. The basic ideas and results of our study allow us to form a relatively complete qualitative picture of the asymptotic solution even when solitons are present. Indeed the asymptotic picture would essentially be the same as above after a time sufficiently long for the solitons to form, interact, and completely emerge from the initial state after which time they would propagate to the right in the exponentially small regime. Formulas [(1.14)–(1.16)] will have the same structure.

It is remarkable that the asymptotic behavior of the solution takes on a form comprised of the simpler ingredients of evolution equations. Specifically, permanent waves, similarity solutions, rapid oscillation, and exponential (linear) decay. While it has been already seen how the permanent solitary waves fit into the KdV theory, we now can also see the central role the similarity solution plays in the overall picture. At this time a pragmatic (and optimistic) approach to solving nonlinear equations of evolution might be based upon a strategy of finding, understanding, and piecing together these special components.

ACKNOWLEDGMENTS

The authors wish to thank their colleagues Martin Kruskal and Harvey Segur for many lively discussions and constructive comments. The work was supported by N.S.F. Grants GP32839X and GA27727A1. One of us (M.J.A.) is grateful to the Applied Mathematics Summer Institute (Dartmouth 1972) and ONR for support during the initial phases of this work.

---

[1]D. S. Korteweg and G. deVries, Phil. Mag. R. Soc. Lond. **39**, 422 (1895).
[2]C. S. Gardner, J. M. Green, M. D. Kruskal, and R. M. Miura, Phys. Rev. Lett. **19**, 1095 (1967).
[3]I. M. Gel'fand and B. M. Levitan, Am. Math. Soc. Transl. **1**, 253 (1955).
[4]I. Kay and H. E. Moses, J. Appl. Phys. **27**, 1503 (1956).
[5]I. A. Berezin and V. I. Karpman, Zh. Eksp. Teor. Fiz. **46**, 1880 (1964) [Sov. Phys.-JETP **19** (5), 1265 (1964)].
[6]G. B. Whitham, Proc. R. Soc. A **283**, 238 (1965).
[7]M. D. Kruskal and N. J. Zabusky, "Progress on the Fermi-Pasta-Ulam nonlinear string problem," Princeton Plasma Physics Laboratory Annual Report MA77-Q-21, p. 301 (unpublished).
[8]V. E. Zakharov and A. B. Shabat, Zh. Eksp. Teor. Fiz. **61**, 118 (1971) [Sov. Phys.-JETP **34**, 62 (1972)].
[9]M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur (submitted to Phys. Rev. Lett.).
[10]D. J. Benney and A. C. Newell, J. Math. & Phys. **46**, 133 (1967).
[11]V. I. Bespalov, A. G. Litvak, and V. I. Talanov, *Second All-Union Symposium on Nonlinear Optics, 1966. Collection of Papers* (Nauka, Moscow, 1968).
[12]L. K. Fadeev, Dokl. Akad. Nauk SSSR **121**, 63 (1958).
[13]R. M. Lewis, *ORSI Symposium Proceedings*, Delft, Holland, 1966.
[14]N. Bleistein and R. A. Handelsman, J. Math. Anal. Appl. 27, 434 (1969).

A geometric generalization of Hooke's law

Edward N. Glass*

Department of Physics, University of Cincinnati, Cincinnati, Ohio 45221

Jeffrey Winicour†

Department of Physics, University of Pittsburgh, Pittsburgh, Pennsylvania 15213

Aerospace Research Laboratories, Wright-Patterson AFB, Ohio 45433

(Received 22 February 1973)

A geometrical version of Hooke's law for prestressed materials is established using techniques suggested by general relativistic elasticity theory. The concept of dragged (D) tensor fields is used to present a simple coordinate independent form of Hook's law for prestressed materials: $\bar{P}^{ij} = D P^{ij} + C^{ijkl}\eta_{kl}$, where P^{ij} is the initial stress, C^{ijkl} are the second order elastic coefficients, and η_{kl} is the strain. The effects of initial stresses on Hooke's law have recently been given empirical significance by Wallace and the work here agrees with and simplifies his formalism. The elastic behavior of neutron stars, which requires a general relativistic treatment, is a case in which large prestresses of the type considered here are as important as the elastic modulii themselves. The version of Hooke's law developed here is the classical limit of the general relativistic theory of stellar elasticity. A simple derivation of the seismic response of a Newtonian self-gravitating body to elastic perturbations is presented.

1. INTRODUCTION

The importance of the methods of tensor calculus to elasticity theory has not been a matter of great dispute since the publication of the classic books by Voigt and Brillouin.[1] The purpose of this paper is to point out some additional methods of differential geometry which are essential in the treatment of general relativistic elastic systems. These methods are the use of dragged tensor fields[2,3] and their infinitesimal counterparts, the Lie derivative of tensor fields. From the historical view of differential geometry, they are not modern methods, although they do play a central role in the modern theory.

The use of these methods leads to an attractive conceptual simplification of the formulation of Hooke's law for prestressed materials in Newtonian physics. The consequences of initial stresses become empirically important when the stresses are comparable to the elastic moduli, as in the interior of the earth. In addition, initial stresses are empirically important for understanding the stress dependence of various ultrasonic phenomena.[4]

Astrophysics has provided the consumate example of prestressed elastic material. Both the cores and crusts of neutron stars are thought to be in elastic states. The self-gravitation of a neutron star provides an enormous initial stress comparable to the elastic modulii. The relativistic elasticity theory[5-9] of such systems has as its classical limit the Hooke's law presented here.

The language and notation used in this paper is that of field theorists and relativists rather than classical elasticity theorists and, in particular, is common to papers[5-9] on general relativistic elasticity theory and general relativistic hydrodynamics.

We base our presentation on a knowledge of tensor calculus at the level of Brillouin's book[1,10]. A summary of the properties of dragging and of Lie differentiation is given in Sec. 2. Apart from a choice of symbols, our treatment is based upon those of Schouten[2] and of Yano.[3] We apply these techniques to give a geometrical formula- of the generalized Hooke's law in Sec. 3. In Sec. 4, we give an application to perturbations from elastic equilibrium such as small seismic disturbances

2. GEOMETRICAL PRELIMINARIES

We wish to give a geometrical description of the three-dimensional Euclidean space E^3 of classical physics.

The space E^3 consists of a set of points $\{p\}$ endowed with a three-dimensional Euclidean geometry. As is customary, we introduce a coordinate system X which associates with each point p a triple of numbers x^i_p $(i = 1, 2, 3)$. When clarity is not sacrificed, we will delete the indices p and i.

Tensor fields are represented in the coordinate system X by expressions

$$T^{ij\cdots}{}_{kl\ldots}(x).$$

In accordance with the kernel-index notation,[2,3] the kernel symbol T specifies the tensor field and the indices $i, j, k, l, \cdots$ refer to its representation in the coordinate system X. In particular, the Euclidean geometry of E^3 is described by the metric tensor $g_{ij}(x)$. In a Cartesian coordinate system, the Euclidean metric assumes the standard values δ_{ij}. However, the use of tensor calculus makes it equally easy to work in an arbitrary coordinate system. We need only replace partial derivatives by covariant derivatives when the two are not equivalent. We denote partial differentiation by a comma and covariant differentiation by a semicolon.

Coordinate transformations reflect the various ways we may associate points with triples of numbers. Under a transformation from a system X to a system X', the coordinates of the point p transform according to

$$x'^i_p = x'^i(x^j_p),$$

where the functions $x'^i(x)$ have nonvanishing Jacobian. Coordinate transformations are *passive* in the sense that they do not involve transformations of the spatial points p but only a formal relabeling of the coordinates of p. Under coordinate transformations tensor fields transform according to the example

$$T^{i'}{}_{k'}(x'_p) = x^{i'}{}_{,a}\, x^b{}_{,k'}\, T^a_b(x_p).$$

In accordance with the kernel-index notation, the kernel symbol T remains unchanged since it refers to the same tensor field and only the indices are primed to indicate that the representation of the tensor field is being given in the X' system. A slight awkwardness of notation occurs for scalar fields. We have

$$\phi'(x'_p) = \phi(x_p) = \phi|_p,$$

Copyright © 1973 by the American Institute of Physics

all terms being equal to the value of the scalar field ϕ at the point p. Here the prime following ϕ does not imply a change in the kernel symbol ϕ but indicates a different functional representation of ϕ in the X' system.

More important to the description of elastic deformations is the group of *active* transformations $\{D\}$ which map spatial points p into new spatial points Dp. There is a one-to-one correspondence between *active* point transformations and *passive* coordinate transformations. This correspondence has the potential for causing a great deal of confusion. To avoid such confusion, *we will henceforth consider only active point transformations.*

The transformation D drags the point p with coordinates x^i_p into the point Dp with coordinates x^i_{Dp}. There is a natural way in which such a map drags a tensor field

$$T^{ij\cdots}{}_{kl\cdots}(x)$$

into a new tensor field[11]

$$DT^{ij\cdots}{}_{kl\cdots}(x).$$

To formulate this, let us first consider scalar fields.

We define the dragged scalar field $D\phi$ by the conditions

$$D\phi(x^i_{Dp}) = \phi(x^i_p)$$

or equivalently

$$D\phi|_{Dp} = \phi|_p.$$

In accordance with the kernel-index notation, the scalar field $D\phi$ differs from the original scalar field ϕ (except in special cases such as when ϕ is a constant). In terms of the inverse map D^{-1} we have

$$D\phi(x_p) = \phi(x_{D^{-1}p}). \tag{2.1}$$

To define the dragging of a covariant vector field, we use the fact that it may be written in the form

$$T_i = f\phi_{,i} + g\psi_{,i} + h\tau_{,i},$$

where f, g, h, ϕ, ψ, and τ are scalar fields. We then define

$$DT_i = Df(D\phi)_{,i} + Dg(D\psi)_{,i} + Dh(D\tau)_{,i}. \tag{2.2}$$

The rule for dragging contravariant vector fields follows from demanding the scalar relations

$$DV^i(x_{Dp})DT_i(x_{Dp}) = V^i(x_p)T_i(x_p) \tag{2.3}$$

for arbitrary T_i. The rule for dragging a general tensor field then follows by building it up as sums and products of vector fields. For example, if

$$T^i_j = S^i T_j + U^i V_j,$$

then

$$DT^i_j = DS^i DT_j + DU^i DV_j. \tag{2.4}$$

To formulate the concept of Lie differentiation, consider the one parameter family of transformations D_ϵ which map the point p with coordinates x^i_p into the point $D_\epsilon p$ with coordinates

$$x^i_{D_\epsilon p} = x^i_p + \epsilon\xi^i(x_p). \tag{2.5}$$

This maps each tensor field $T^{ij\cdots}{}_{kl\cdots}$ into a one parameter family of tensor fields $D_\epsilon T^{ij\cdots}{}_{kl\cdots}$. For each ϵ the difference

$$\Delta_\epsilon T^{ij\cdots}{}_{kl\cdots} = T^{ij\cdots}{}_{kl\cdots} - D_\epsilon T^{ij\cdots}{}_{kl\cdots} \tag{2.6}$$

is again a tensor field. The Lie derivative is defined by

$$\underset{\xi}{\mathcal{L}} T^{ij\cdots}{}_{kl\cdots} = \lim_{\epsilon\to 0}\epsilon^{-1}\Delta_\epsilon T^{ij\cdots}{}_{kl\cdots}. \tag{2.7}$$

For scalar fields, we have from Eq. (2.1)

$$\underset{\xi}{\mathcal{L}}\phi = \lim_{\epsilon\to 0}\epsilon^{-1}[\phi(x)] - D_\epsilon\phi(x)] = \lim_{\epsilon\to 0}\epsilon^{-1}[\phi(x) - \phi(x-\epsilon\xi)],$$

so that

$$\underset{\xi}{\mathcal{L}}\phi = \phi_{,i}\xi^i = \phi_{;i}\xi^i. \tag{2.8}$$

Similarly, for covariant vector fields we have

$$\underset{\xi}{\mathcal{L}}T_i = T_{i;j}\xi^j + T_j\xi^j{}_{;i} \tag{2.9}$$

and for contravariant vector fields,

$$\underset{\xi}{\mathcal{L}}T^i = T^i{}_{;j}\xi^j - T^j\xi^i{}_{;j}. \tag{2.10}$$

For tensor fields, the following rules apply:

$$\underset{\xi}{\mathcal{L}}T^{ij} = T^{ij}{}_{;k}\xi^k - T^{ik}\xi^j{}_{;k} - T^{kj}\xi^i{}_{;k} \tag{2.11}$$

and

$$\underset{\xi}{\mathcal{L}}T_{ij} = T_{ij;k}\xi^k + T_{ik}\xi^k{}_{;j} + T_{kj}\xi^k{}_{;i}. \tag{2.12}$$

Nowhere in the above discussion of dragging and Lie differentiation have any metrical properties been used, and it should be emphasized that these two concepts are metric independent. However, when a symmetric metric is introduced such as the Euclidean metric of E^3, it is easy to check that covariant differentiation and partial differentiation are equivalent in Eqs. (2.8)-(2.12). For further details see Chap. 1 of Ref. 3.

3. HOOKE'S LAW

We consider finite deformations of an elastic body for which the third and higher order elastic coefficients vanish. In the absence of initial stresses, Hooke's law states that the stress tensor $\overline{P}^{ij}$ and the strain tensor η_{ij} are connected by a linear relationship

$$\overline{P}^{ij} = C^{ijkl}\eta_{kl}. \tag{3.1}$$

Here the tensor C^{ijkl} describes the second order elastic coefficients appropriate to the deformation, i.e., adiabatic or isothermal. They possess the Voigt symmetry

$$C^{ijkl} = C^{klij} = C^{(kl)ij},$$

where the parentheses denote symmetrization:

$$T^{(ij)} = \tfrac{1}{2}(T^{ij} + T^{ji}).$$

Our goal is to extend Eq. (3.1) to include effects of initial stresses. Proceeding by means of geometrical arguments, we begin by discussing the geometrical significance of the strain tensor.

A. Strain

The Euclidean metric g_{ij} determines the distance between material points in an elastic body. For neigh-

boring points p_1 and p_2 with infinitesimal separation vector

$$dx^i(p_1) = x^i_{p_2} - x^i_{p_1},$$

we have the standard formula

$$ds^2(p_1, p_2) = g_{ij}(x_{p_1})dx^i(p_1)dx^j(p_1).$$

Since we are dealing with infinitesimals, it does not matter whether we evaluate g_{ij} at p_1 or at p_2. Under a deformation D, the material point at the spatial point p moves to the spatial point Dp. The infinitesimal separation vector $dx^i(p_1)$ is dragged along into the separation vector at Dp_1 given by

$$Ddx^i(Dp_1) = x^i_{Dp_2} - x^i_{Dp_1}.$$

The resulting distance between neighboring material points originally at p_1 and p_2 is then given by

$$ds^2(Dp_1, Dp_2)|_{p_1} = D^{-1}g_{ij}(x_{p_1})dx^i(p_1)dx^j(p_1). \quad (3.2)$$

To evaluate Eq. (3.2) at the point p_1, we apply scalar relations analogous to Eq. (2.3). This becomes transparent upon rewriting Eq. (3.2) as

$$ds^2(Dp_1, Dp_2) = DD^{-1}g_{ij}(x_{Dp_1})Ddx^i(Dp_1)Ddx^j(Dp_1).$$

Thus

$$ds^2(Dp_1, Dp_2)|_{p_1} = D^{-1}g_{ij}(x_{p_1})dx^i(p_1)dx^j(p_1).$$

In elasticity theory, the strain tensor corresponding to the finite deformation D is conventionally defined[12] by

$$ds^2(Dp_1, Dp_2)|_{p_1} - ds^2(p_1, p_2) = 2\eta_{ij}(x_{p_1})dx^i(p_1)dx^j(p_1). \quad (3.3)$$

Consequently, we find from Eq. (3.3) that the strain matrix is given by

$$2\eta_{ij}(x) = D^{-1}g_{ij}(x) - g_{ij}(x). \quad (3.4)$$

The subgroup of point transformations $\{G\} \subset \{D\}$ for which

$$ds^2(Gp_1, Gp_2) = ds^2(p_1, p_2)$$

plays a specially important role in classical physics. It is called the Euclidean group. If we do not consider reflections, it is well known that the Euclidean group consists of combinations of rigid rotations and translations. These are precisely the transformations under which the physical properties of a system isolated from external forces are invariant. Consistent with this invariance principle, we see from Eqs. (3.3) and (3.4) that there is no strain associated with Euclidean transformations,

$$2\eta_{ij} = G^{-1}g_{ij} - g_{ij} = 0. \quad (3.5)$$

In defining the strain tensor, Eq. (3.3) we might equally well have proceeded on *a priori* grounds to define a tensor τ_{ij} by the prescription

$$ds^2(p_1, p_2) - ds^2(D^{-1}p_1, D^{-1}p_2)|_{p_1} = 2\tau_{ij}(x_{p_1})dx^i(p_1)dx^j(p_1). \quad (3.6)$$

We will refer to τ_{ij} as the *stretch* tensor associated with D. Analogous to Eq. (3.4), we have

$$2\tau_{ij}(x) = g_{ij}(x) - Dg_{ij}(x). \quad (3.7)$$

In some respects, the stretch tensor is more natural to the formulation of Hooke's law than the strain tensor. This will be discussed below.

We now consider the case of small deformations, describing these by

$$x^i_{Dp} = x^i_p + \xi^i(x_p). \quad (3.8)$$

We consider only terms of first order in the vector field ξ^i. Equation (3.8) is equivalent to Eq. (2.5) for the case $\epsilon = 1$. From the discussion below Eq. (2.5) it follows that for an arbitrary tensor field we can write

$$DT^{ij\cdots}{}_{kl\cdots} = T^{ij\cdots}{}_{kl\cdots} - \underset{\xi}{\mathcal{L}}T^{ij\cdots}{}_{kl\cdots} \quad (3.9)$$

and

$$D^{-1}T^{ij\cdots}{}_{kl\cdots} = T^{ij\cdots}{}_{kl\cdots} + \underset{\xi}{\mathcal{L}}T^{ij\cdots}{}_{kl\cdots}. \quad (3.10)$$

Applying this to Eq. (3.4), we find for the strain tensor

$$2\eta_{ij} = \underset{\xi}{\mathcal{L}}g_{ij}. \quad (3.11)$$

Similarly, Eq. (3.7) gives for the stretch tensor

$$2\tau_{ij} = \underset{\xi}{\mathcal{L}}g_{ij}.$$

Hence in the limit of small deformations the strain tensor and stretch tensor are equivalent. Using the fact that the covariant derivative of the metric tensor vanishes, we see that Eq. (2.12) implies

$$\underset{\xi}{\mathcal{L}}g_{ij} = 2\xi_{(i;j)} \quad (3.12)$$

so that

$$\eta_{ij} = \xi_{(i;j)}. \quad (3.13)$$

The conditions for infinitesimal Euclidean transformations become

$$\xi_{(i;j)} = 0. \quad (3.14)$$

B. Stress

The equation of motion for an element of a material body is

$$\rho a^i = F^i. \quad (3.15)$$

Here ρ is the mass density, a^i is the local acceleration of a material point in the body, and F^i is the force per unit volume. All internal interactions may be described by a stress tensor P^{ij} so that

$$F^i = P^{ij}{}_{;j} + B^i, \quad (3.16)$$

where the B^i are the external body forces. The equilibrium conditions for a nonrotating body become

$$\rho a^i = F^i = P^{ij}{}_{;j} + B^i = 0. \quad (3.17)$$

Here ρ is a scalar field and P^{ij} is a tensor field. Note that it is also common in elasticity theory to work with the scalar density $g^{1/2}\rho$ and the tensor density

$$T^{ij} = g^{1/2}P^{ij}, \quad (3.18)$$

where g is the determinant of the metric.[10]

Now consider an isolated body in equilibrium which satisfies Eq. (3.17). (We will consider only the case $B^i = 0$ here. See Sec. 4 below for an example of nonzero B^i.) After the body undergoes a deformation D, Eq. (3.17) will in general no longer hold for the resulting force field $\bar{F}^i$ and the resulting stress field $\bar{P}^{ij}$. However, if D is restricted to the subgroup of Euclidean transformations $\{G\}$, then the principle of Euclidean invariance demands that the stress field be dragged along, i.e.,

$$\bar{P}^{ij}(x) = GP^{ij}(x). \tag{3.19}$$

This simply states that the stress field is translated and rotated along with the body. Under Euclidean transformations the metric satisfies Eq. (3.5) so that covariant differentiation commutes with the dragging operation. Hence the resulting force field still vanishes, i.e.,

$$\bar{F}^i = \bar{P}^{ij}{}_{;j} = (GP^{ij})_{;j} = G(P^{ij}{}_{;j}) = GF^i = 0.$$

C. Stress and strain

We are now ready to formulate the generalized Hooke's law relating the stresses and strains in a prestressed material. Two criteria must be satisfied:

(i) When there are no initial stresses, the law must reduce to the usual form Eq. (3.1).

(ii) When the deformation corresponds to a Euclidean transformation (no strain) in some small region of the body, the law must reduce to Eq. (3.19) in that region.

For an arbitrary finite deformation, the most direct generalization is

$$\bar{P}^{ij} = DP^{ij} + C^{ijkl}\eta_{kl}. \tag{3.20}$$

For a realistic body, there will of course be additional terms in the stress-strain relation associated with third and higher order elastic coefficients. By using Eq. (3.4) to describe the strain matrix, Eq. (3.20) becomes

$$\bar{P}^{ij} = DP^{ij} + \tfrac{1}{2}C^{ijkl}(D^{-1}g_{kl} - g_{kl}). \tag{3.21}$$

This form suggests the alternative generalization of Hooke's law

$$\bar{P}^{ij} = DP^{ij} + \tfrac{1}{2}C^{ijkl}(g_{kl} - Dg_{kl}), \tag{3.22}$$

which corresponds to substituting the stretch tensor for the strain tensor in Eq. (3.20). The virtue of Eq. (3.22) is that, to determine the stress at the point p in the deformed state, knowledge of the background stress and metric is needed only at $D^{-1}p$. On the other hand, Eq. (3.21) depends upon the background stress at $D^{-1}p$ and the metric at Dp. The choice between the two possibilities Eq. (3.21) or (3.22), should be decided by which leads to the simpler properties for the higher order elastic coefficients.[4] We will not pursue this matter further here. In the following, we restrict our attention to small deformations for which the strain and stretch tensors are equal so that Eqs. (3.21) and (3.22) are equivalent.

We return to the description of small deformations given in Eq. (3.8). From Eq. (3.9) we have

$$DP^{ij} = P^{ij} - \underset{\xi}{\mathcal{L}}P^{ij},$$

and from Eq. (3.11)

$$\eta_{ij} = \tfrac{1}{2}\underset{\xi}{\mathcal{L}}g_{ij}.$$

Hooke's law, Eq. (3.20), then takes the form

$$\bar{P}^{ij} = P^{ij} - \underset{\xi}{\mathcal{L}}P^{ij} + \tfrac{1}{2}C^{ijkl}\underset{\xi}{\mathcal{L}}g_{kl}. \tag{3.23}$$

Note that careful distinction between the contravariant and covariant stress tensors is important here. We have[3]

$$g_{ik}g_{jl}\underset{\xi}{\mathcal{L}}P^{kl} = \underset{\xi}{\mathcal{L}}P_{ij} - g_{ik}P^{kl}\underset{\xi}{\mathcal{L}}g_{jl} - g_{jl}P^{kl}\underset{\xi}{\mathcal{L}}g_{ik}.$$

Using this to lower the indices in Eq. (3.23), we find

$$\bar{P}_{ij} = P_{ij} - \underset{\xi}{\mathcal{L}}P_{ij} + [\tfrac{1}{2}C_{ij}{}^{kl} + 2\delta_{(i}{}^{k}P_{j)}{}^{l}]\underset{\xi}{\mathcal{L}}g_{kl}.\}$$

We can write out the Lie derivatives in Eq. (3.23) by means of Eqs. (2.11) and (3.12). Hooke's law for small deformations then becomes

$$\bar{P}^{ij} = P^{ij} - P^{ij}{}_{;k}\xi^k + P^{ik}\xi^j{}_{;k} + P^{jk}\xi^i{}_{;k} + C^{ijkl}\xi_{(k;l)}. \tag{3.24}$$

Wallace[4] works with the stress tensor density defined in Eq. (3.18). He finds

$$\bar{T}^{ij}(\bar{x}) = T^{ij}(x) + T^{ik}\xi^j{}_{,k} + T^{jk}\xi^i{}_{,k} - T^{ij}\xi^k{}_{,k} + C^{ijkl}\eta_{kl}, \tag{3.25}$$

where $\bar{x}^i = x^i + \xi^i$. This implies

$$\bar{T}^{ij}(x) = T^{ij}(x) - T^{ij}{}_{,k}\xi^k + T^{ik}\xi^i{}_{,k} + T^{jk}\xi^i{}_{,k} - T^{ij}\xi^k{}_{,k} + C^{ijkl}\eta_{kl}.$$

We now use the relations

$$T^{ij} = g^{1/2}P^{ij}, \qquad \bar{T}^{ij} = \bar{g}^{1/2}\bar{P}^{ij},$$

and the density property of the metric determinant

$$\bar{g}^{1/2} = \left|\frac{\partial x^i}{\partial \bar{x}^j}\right| g^{1/2} = (1 - \xi^i{}_{,i})g^{1/2},$$

to obtain

$$\bar{P}^{ij} = P^{ij} - P^{ij}{}_{,k}\xi^k + P^{ik}\xi^j{}_{,k} + P^{jk}\xi^i{}_{,k} + C^{ijkl}\eta_{kl}.$$

Here the partial derivatives may be replaced by covariant derivatives [see the discussion following Eq. (2.12)]. Hence we arrive back at Eq. (3.24) and thus establish agreement with the work of Wallace.

4. ELASTIC PERTURBATIONS

A. The general perturbation equation

We now use the preceding analysis to treat small perturbations of an elastic system about its equilibrium state. According to Eq. (3.17), the equilibrium configuration satisfies

$$\rho a^i = P^{ij}{}_{;j} + B^i = 0, \tag{4.1}$$

provided we neglect effects of rotation. The perturba-

tion is described by the time dependent deformation D_t,

$$x_{D_t p}{}^i = x^i{}_p + \xi^i(x_p, t), \tag{4.2}$$

and only terms of first order in the deformation are kept. The coordinates of the deformation are in a frame attached to the center of mass of the system and the motion of the center of mass is considered to be of order ξ^2 (thus excluding large motions of the whole system). We have

$$\bar{a}^i = \ddot{\bar{\xi}}^i = \frac{\partial^2 \xi^i}{\partial t^2}, \tag{4.3}$$

where a "bar" is used to denote quantities in the deformed state. The equation of motion becomes

$$\bar{\rho}\bar{a}^i = \bar{P}^{ij}{}_{;j} + \bar{B}^i. \tag{4.4}$$

Here

$$\bar{\rho}(x_p) = \rho(x_p) - \rho_{;i}\xi^i - \rho\xi^i{}_{;i}, \tag{4.5}$$

where the first correction term is due to displacement of the material and the second is due to expansion of the material. The elastic stresses $\bar{P}^{ij} - P^{ij}$ are given by Eq. (3.24). More specific information concerning the nature of the body forces is necessary to determine $\bar{B}^i$ (an explicit case is treated below). Collecting these results, we expand the equation of motion (4.4) about the equilibrium condition (4.1) to obtain the perturbation equation.

$$\rho\ddot{\xi}^i = [C^{ijkl}\xi_{(k;l)}]_{;j} + 2P^{k(i}\xi^{j)}{}_{;kj} + B^i{}_{;k}\xi^k - \xi^i{}_{;k}B^k + \bar{B}^i - B^i. \tag{4.6}$$

In the absence of body forces, the perturbation equation can be rewritten as

$$\rho\ddot{\xi}^i = [C^{ijkl}\xi_{(k;l)}]_{;j} + P^{ik}\xi^j{}_{;jk} + 2P_{jk}\xi^{(i;k)j} - P_{jk}\xi^{(k;j)i}. \tag{4.7}$$

In this form it is clear that if the strain vanishes then the acceleration vanishes! If we additionally assume that the background stress is simply hydrostatic,

$$P^{ij} = -Pg^{ij}, \tag{4.8}$$

then Eq. (4.7) simplifies to the form

$$\rho\ddot{\xi}^i = [C^{ijkl}\xi_{(k;l)} - 2P\xi^{(i;j)}]_{;j}. \tag{4.9}$$

B. Self-gravitating body forces

Now, as an explicit example, we consider the case of a self-gravitating body such as the earth. The gravitational forces can be described as body forces by means of the gravitational potention ϕ,

$$B^i = -\rho\phi^{;i},$$

where ϕ satisfies the Poisson equation and

$$\bar{B}^i = -\bar{\rho}\bar{\phi}^{;i}.$$

If the perturbations are restricted to those whose characteristic wavelength is small compared to the radius of the body, then the approximation

$$\bar{\phi} = \phi$$

is quite good. By using Eq. (4.5), the perturbation equation (4.6) becomes

$$\rho\ddot{\xi}^i = [C^{ijkl}\xi_{k;l}]_{;j} + 2P^{k(i}\xi^{j)}{}_{;kj} + \rho\phi^{;j}(\delta^i{}_j\xi^k{}_{;k} + \xi^i{}_{;j}) - \rho\phi^{;i}{}_j\xi^j. \tag{4.10}$$

Note, for regions in the body undergoing no strain Eq. (4.10) becomes

$$\ddot{\xi}^i = \underset{\xi}{\mathcal{L}}\phi^{;i},$$

which corresponds to a small Euclidean displacement of the region with respect to the net mass distribution.]

To further simplify our considerations, we take our self-gravitating body to be homogeneous, isotropic, and hydrostatically supported. Thus

$$P^{;i} = -\rho\phi^{;i}, \tag{4.11}$$

where P is defined in Eq. (4.8). For an isotropic body, there are only two independent elastic modulii, the Lamé parameters, which will be constant for our homogeneous model:

$$C^{ijkl} = \lambda g^{ij}g^{kl} + 2\mu g^{i(k}g^{l)j}. \tag{4.12}$$

Combining the physical conditions (4.11) and (4.12) with the perturbation equation (4.10), we find

$$\rho\ddot{\xi}^i = (\lambda + \mu)\Theta^{;i} + \mu\nabla^2\xi^i - P(\Theta^{;i} + \nabla^2\xi^i) - P^{;i}\Theta - P^{;j}\xi^i{}_{;j} - \rho\phi^{;i}{}_j\xi^i, \tag{4.13}$$

where $\Theta = \xi^i{}_{;i}$ is the volume expansion. The compressional and shear waves in the elastic body are obtained by taking the divergence and curl of Eq. (4.13). The compression wave equation is

$$\rho\ddot{\Theta} = (\lambda + 2\mu - 2P)\nabla^2\Theta + 4\pi G\rho^2\Theta - P_{;i}(3\Theta^{;i} + \nabla^2\xi^i) \tag{4.14}$$

and the shear wave equation is

$$\rho(\ddot{\xi}^{[i;k]}) = (\mu - P)\nabla^2(\xi^{[i;k]}) - \nabla^2\xi^{[i}P^{;k]} - P^{;j}\xi^{[i}{}_{;j}{}^{k]} - (\rho\xi^j)_{;[k}\phi^{;i]}{}_j, \tag{4.15}$$

where $\xi^{[i;k]} = \frac{1}{2}(\xi^{i;k} - \xi^{k;i})$, and ϵ_{ikl} the Levi-Civita density should be contracted into the entire equation.

For the lower mantle of the earth (1000-3000 km. below the surface) the elastic modulii λ and μ and the hydrostatic pressure P all have the same order of magnitude[13] ($\sim 10^{12}$ dynes/cm^2). It is clear from the wave equations above that in this region the velocity of the waves and their dispersion are direct functions of the gravitational stresses.

In addition to geophysical applications there are astrophysical consequences of gravitationally stressed elastic media. For instance, if neutron star cores are in elastic states, they would manifest general relativistic elasticity in which the velocity of sound waves is comparable to the velocity of light. A general relativistic treatment of elastic bodies with a classical limit giving Hooke's law as presented here appears elsewhere.[9]

---

*Present Address: Department of Physics, Trenton State College, Trenton, New Jersey 08625.

†Supported in part by the National Science Foundation.

[1]W. Voight, *Lehrbuch der Kristallphysik* (Tuebner, Berlin, 1910). L. Brillouin, *Les tenseurs en mécanique et en elasticite* (Masson, Paris, 1938).

[2]J. A. Schouten, *Ricci-Calculus* (Springer-Verlag, Berlin, 1954), 2nd ed.

[3]K. Yano, *The Theory of Lie Derivatives and Its Applications* (North-Holland, Amsterdam, 1957).

[4]D. C. Wallace, in *Solid State Physics*, edited by F. Seitz *et. al.* (Academic, New York, 1970), Vol. 25.

[5]F. J. Dyson, Ap. J. **156,** 529 (1969).

[6]W. C. Hernandez, Phys. Rev. A **1,** 1013 (1970).

[7]A. Papepetrov, Ann. Inst. Henri Poincare **16,** 63 (1972).

[8]B. Carter and H. Quintana, Proc. Phys. Soc. Lond. A **331,** 57 (1972).

[9]E. Glass and J. Winicour, J. Math. Phys. **13,** 1934 (1972).

[10]E. Brillouin, *Tensors in Mechanics and Elasticity* (Academic, New York, 1964).

[11]This differs slightly from the modern notation of differential geometry which symbolizes a diffeomorphism by ϕ and its induced linear transformations in the tangent and cotangent spaces by ϕ_* and ϕ^*, respectively. Here we use the symbol D to emphasize the active mapping of one point into another by an elastic deformation and the accompanying notion of dragging.

[12]F. Murnaghan, *Finite Deformation of an Elastic Solid* (Wiley, New York, 1951).

[13]R. Haddon and K. Bullen, Phys. Earth Planet. Inter. **2,** 35 (1969).

Large coupling expansions for eigenenergies and Regge trajectories of the Yukawa potential

H. J. W. Müller-Kirsten

Fachbereich Physik der Universität Trier-Kaiserslautern, Kaiserslautern, Germany

N. Vahedi-Faridi

Department of Physics, Pars College, Teheran, Iran

(Received 7 November 1972)

Simple and straightforward procedures are developed for deriving complete large coupling expansions for the wavefunctions, eigenenergies, and Regge trajectories of the generalized Yukawa potential. The procedures may easily be applied to similar problems.

1. INTRODUCTION

Large coupling solutions to various problems arising in particle physics have attracted considerable interest recently, particularly in view of the now almost classical desire to understand strong interactions in terms of purely field-theoretic formulations. It is well known that attempts in this direction are severely handicapped by the fact that the relevant coupling parameter is a factor of about 1000 larger than the fine structure constant of quantum electrodynamics. Since many expansions of physically relevant quantities studied presently are asymptotic in nature (e.g., high energy expansions), there seems to be no plausible reason for ignoring asymptotic large coupling solutions. In fact, some time ago it has been emphasized by Dingle[1] that it is precisely the physicist's approach to a problem which will commonly lead to asymptotic rather than convergent expansions. Field-theoretical models for large coupling constants are scarce (one such example was studied by Schiff[2]) except in the form of the Bethe–Salpeter equation. Large coupling solutions of the Bethe–Salpeter equation of scalar ϕ^3 theory have recently been studied by Cheng and Wu[3]; in the frame of a quark model they have been studied by Böhm, Joos, and Krammer.[4] Large coupling solutions of the Bethe–Salpeter equation of the Wick–Cutkosky model have been derived by one of us.[5]

Although the ultimate aim of physicists is a fully relativistic particle theory, considerable insight has always been drawn from simpler, nonrelativistic potential models, for which complete—or almost complete—solutions can be obtained. Moreover, these nonrelativistic models are directly applicable in nuclear or atomic physics and so justify separate investigation in their own right, apart from purely mathematical interest. Nonrelativistic wave equations for large coupling constants of the potential have been studied by several authors. Cheng and Wu[6] calculated the approximate behavior of Regge trajectories for the Yukawa potential. Extensive investigations of the energy eigenvalues for the Yukawa potential were carried out by Iafrate and Mendelsohn[7] (see also Zauderer[8]). The basic large coupling expansion obtained by these authors for the energy eigenvalues [Eq. (61) of Ref. 7(a)] is identical with an expansion obtained previously by one of us[9,10] [Eq. (25) of Ref. 9 or (131) of Ref. 10] by inverting a high energy asymptotic expansion of the Regge trajectories.[11] Other cases for which large coupling solutions have been derived are the Gauss potential[12] and its generalized form.[13]

In the present note we extend previous investigations of the Yukawa potential[9–11,15] to the case of large coupling constants. In particular we use a perturbation procedure which we have been using in our previous and numerous other investigations.[16] In Sec. 2 we derive the large coupling expansion of the energy eigenvalues for a generalized Yukawa potential. The approach is simpler and more general than that of Iafrate and Mendelsohn.[7] In Sec. 3 we derive large coupling expansions for the Regge trajectories. In either case the solutions are obtained in the form of complete asymptotic expansions (as distinguished from asymptotic approximations). Finally, in Sec. 4, we give a brief discussion of our results.

2. EIGENENERGIES

We consider the radial Schrödinger equation in the form used by Iafrate and Mendelsohn,[7]

$$-\frac{1}{2}\frac{d^2\psi(r)}{dr^2}-\frac{1}{r}\frac{d\psi(r)}{dr}+V(r)\psi(r)=\frac{l(l+1)}{2r^2}\psi(r)=E\psi(r)$$

with a generalized central field potential of the form

$$V(r)=-\frac{Z}{r}B(\lambda r),\ B(\lambda r)=\sum_{j=0}^{\infty}B_j(\lambda r)^j,\qquad B_0=1.$$

Here $\hbar=c=1$; the reduced mass is also taken equal to 1, whereas in all our previous investigations[9–12,15] it was taken equal to $\frac{1}{2}$. In the limit $\lambda\to 0$, V reduces to the Coulomb potential, i.e., for $E<0$ we have

$$E/Z^2=-1/2n^2+\epsilon\Delta,\qquad \epsilon=\lambda/Z, \tag{1}$$

where the principal quantum number $n=N+l+1$, $N=0,1,2,\cdots$. Changing the independent variable of the radial wave equation to y, where

$$r=(n/2Z)y,$$

and setting $\psi(y)=e^{-y/2}y^l u(y)$, we obtain the equation

$$D_a u(y)=\left[\epsilon\alpha y+n\sum_{t=2}^{\infty}\epsilon^t B_t\left(\frac{ny}{2}\right)^t\right]u(y), \tag{2}$$

where

$$D_a\equiv y\frac{d^2}{dy^2}+(b-y)\frac{d}{dy}-a,$$

$$\alpha=\tfrac{1}{2}n^2(B_1-\Delta),\qquad a=-N,\qquad b=2l+2. \tag{3}$$

We observe that to order zero in ϵ we have

$$u=u^{(0)}=\Phi(a,b;y)\equiv\Phi(a), \tag{4}$$

where Φ is a confluent hypergeometric function which becomes a normalizable polynomial if its expansion is broken off after a finite number of terms, i.e., when $a=-N$ (the case of the Coulomb potential). In solving

Copyright © 1973 by the American Institute of Physics

(2) by our perturbation procedure, we use the recurrence relation of the confluent hypergeometric functions. This may be written

$$y\Phi(a) = (a, a+1)\Phi(a+1) + (a,a)\Phi(a) + (a, a-1)\Phi(a-1), \tag{5}$$

where

$$\begin{aligned}(a, a+1) &= a = l + 1 - n,\\ (a,a) &= b - 2a = 2n,\\ (a, a-1) &= a - b = -l - 1 - n.\end{aligned} \tag{6}$$

By repeated application of (5) we obtain

$$y^m \Phi(a) = \sum_{j=m}^{-m} S_m(a,j)\Phi(a+j). \tag{7}$$

The coefficients S_m satisfy the following recurrence relation:

$$\begin{aligned}S_m(a,r) = {} & S_{m-1}(a, r-1)\cdot(a+r-1, a+r)\\ & + S_{m-1}(a,r)\cdot(a+r, a+r)\\ & + S_{m-1}(a, r+1)\cdot(a+r+1, a+r)\end{aligned} \tag{8}$$

with $S_0(a,0) = 1$, all $S_0(a, i \neq 0) = 0$, and $S_m(a,r) = 0$ for $|r| > m$. Thus the first approximation of u, i.e., (4), leaves uncompensated

$$\begin{aligned}R^{(0)}(a) &\equiv \left[\epsilon\alpha y + n\sum_{t=2}^{\infty}\epsilon^t B_t\left(\frac{ny}{2}\right)^t\right]\Phi(a)\\ &= \sum_{i=1}^{\infty}\epsilon^i \sum_{j=-i}^{i}[a, a+j]_i\Phi(a+j),\end{aligned} \tag{9}$$

where

$$\begin{aligned}[a, a+1]_1 &= \alpha a,\\ [a,a]_1 &= \alpha(b-2a),\\ [a, a-1]_1 &= \alpha(a-b),\end{aligned} \tag{10}$$

and for $i > 1$

$$[a, a+j]_i = n(n/2)^i B_i S_i(a,j).$$

We now observe that

$$D_{a+j} = D_a - j \quad \text{and so} \quad D_a\Phi(a+j) = j\Phi(a+j).$$

Thus a term $\mu\Phi(a+j)$ in $R^{(0)}(a)$ may be removed by adding to $u^{(0)}$ the contribution $\mu\Phi(a+j)/j$ except, of course, when $j = 0$. Hence to $O(1)$ in ϵ we have

$$u = u^{(0)} + u^{(1)}, \qquad u^{(1)} = \sum_{i=1}^{\infty}\epsilon^i \sum_{\substack{j=-i\\ j\neq 0}}^{i}\frac{[a, a+j]_i}{j}\Phi(a+j) \tag{11}$$

together with

$$0 = \epsilon[a,a]_1\Phi(a). \tag{12}$$

Since $u^{(0)} = \Phi(a)$ leaves uncompensated $R^{(0)}(a)$, we find that $u^{(1)}$ leaves uncompensated

$$R^{(1)}(a) = \sum_{i=1}^{\infty}\epsilon^i \sum_{\substack{j=-i\\ j\neq 0}}^{i}\frac{[a, a+j]_i}{j}R^{(0)}(a+j).$$

Proceeding as before we now have to $O(2)$ in ϵ

$$\begin{aligned}u &= u^{(0)} + u^{(1)} + u^{(2)}\\ u^{(2)} &= \sum_{i=1}^{\infty}\epsilon^i \sum_{\substack{j=-i\\ j\neq 0}}^{i}\frac{[a, a+j]_i}{j}\sum_{i'=1}^{\infty}\epsilon^{i'}\sum_{\substack{j'=-i'\\ j+j'\neq 0}}^{i'}\frac{[a+j, a+j+j']_{i'}}{j+j'}\\ &\quad\times\Phi(a+j+j')\end{aligned} \tag{13}$$

together with

$$0 = \left(\epsilon[a,a]_1 + \sum_{i=1}^{\infty}\epsilon^i\sum_{\substack{j=-i\\ j\neq 0}}^{i}\frac{[a, a+j]_i}{j}\sum_{i'=1}^{\infty}\epsilon^{i'}[a+j, a]_{i'}\right)\Phi(a). \tag{14}$$

Still higher contributions are readily obtained in a similar way. At first sight our expressions may look complicated, but once it is realized that the coefficients have been constructed so that they can be obtained by simple considerations in terms of "steps" from a to $a+j$ and from there back to a, they may be obtained almost directly from (9) and (10). The solution u may therefore be written

$$u(y) = \Phi(a) + \sum_{i=1}^{\infty}\epsilon^i\sum_{\substack{j=-i\\ j\neq 0}}^{i}P_i(a,j)\Phi(a+j), \tag{15}$$

where

$$P_1(a, \pm 1) = \frac{[a, a\pm 1]_1}{\pm 1},$$

$$P_2(a, \pm 2) = \frac{[a, a\pm 2]_2}{\pm 2} + \frac{[a, a\pm 1]_1}{\pm 1}\cdot\frac{[a\pm 1, a\pm 2]_1}{\pm 2},$$

and so on. In fact a recursion relation for the P's can be written down.[15] The equation from which Δ and hence the eigenvalues may be determined—i.e., (12) to $O(1)$, (14) to $O(2)$—is readily seen to be

$$\begin{aligned}0 = {} & \epsilon[a,a]_1 + \epsilon^2\left([a,a]_2 + \frac{[a, a+1]_1}{1}[a+1, a]_1\right.\\ & \left. + \frac{[a, a-1]_1}{-1}[a-1, a]_1\right)\\ & + \epsilon^3\left([a,a]_3 + \frac{[a, a+1]_2}{1}[a+1, a]_1\right.\\ & + \frac{[a, a+1]_1}{1}[a+1, a]_2 + \frac{[a, a-1]_2}{-1}[a-1, a]_1\\ & + \frac{[a, a-1]_1}{-1}[a-1, a]_2\\ & + \frac{[a, a+1]_1}{1}\frac{[a+1, a+1]_1}{1}[a+1, a]_1\\ & \left. + \frac{[a, a-1]_1}{-1}\frac{[a-1, a-1]_1}{-1}[a-1, a]_1\right) + \cdots.\end{aligned} \tag{16}$$

In view of the widespread applicability of our procedure we describe briefly a simple method for writing down the various terms contributing to the coefficient of any power of ϵ in the expansion (16). As an example we consider the coefficient of ϵ^4. We call $[a, a+j]_i$ a "step" of order i from a to $a+j$. In going from a back to a in various steps so that the sum of the orders of the steps in one cycle from a back to a is 4, we have—clearly—only one cycle consisting of one step, i.e., $[a,a]_4$, which we write $(4)^1$, the superscript indicating the number of cycles. A cycle consisting of a step of order 1 followed by a step of order 3 is written (1, 3). Thus in order to ensure that no terms are missed out, it is best to write down first the various combinations of steps contributing; thereafter it is easy to count the number of cycles arising from each combination. In the example under discussion we have altogether the following contributions:

$$\begin{aligned}&(4)^1\\ &+ (3,1)^2 + (2,2)^4 + (1,3)^2\\ &+ (2,1,1)^4 + (1,2,1)^4 + (1,1,2)^4\\ &+ (1,1,1,1)^4.\end{aligned}$$

Hence we have altogether 25 terms. The following example illustrates our procedure:

$$(1,2,1)^4 = \frac{[a,a+1]_1}{1}\frac{[a+1,a+1]_2}{1}[a+1,a]_1 + \frac{[a,a-1]_1}{-1}\frac{[a-1,a-1]_2}{-1}[a-1,a]_1 + \frac{[a,a+1]_1}{1}\frac{[a+1,a-1]_2}{-1}[a-1,a]_1 + \frac{[a,a-1]_1}{-1}\frac{[a-1,a+1]_2}{1}[a+1,a]_1.$$

Evaluating in (16) terms up to and including the power ϵ^4, calculating Δ and substituting into (1), we obtain

$$\frac{E}{Z^2} = -\frac{1}{2n^2} + B_1\epsilon + \epsilon^2\frac{B_2}{2}(-x+3n^2) + \epsilon^3\frac{n^2}{2}B_3(-3x+5n^2+1) + \frac{\epsilon^4 n^2}{8}\left[B_4\cdot\{3x^2-6x(5n^2+1)+5n^2(7n^2+5)\} + B_2^2\cdot\{3x^2-n^2(7n^2-5)\}\right] + O(\epsilon^5), \tag{17}$$

where $x = l(l+1)$ and $n = N+l+1$, $N = 0,1,2,\cdots$. If $B(\lambda r) = \exp(-\lambda r)$, the expansion (17) reduces to the result obtained by Iafrate and Mendelsohn.[7] We remark finally that the solution (15) is valid in the domain $r < O(1/Z)$.

3. REGGE TRAJECTORIES

For our present purposes it is convenient to use the radial wave equation in the following form ($\hbar = c =$ reduced mass $= 1$):

$$-\frac{1}{2}\frac{d^2(r\psi)}{dr^2} + \left[\frac{l(l+1)}{2r^2} + V(r) - E\right](r\psi) = 0,$$

where for bound states $E < 0$. It is easiest to consider a pure Yukawa potential instead of the generalized potential used above:

$$V(r) = -(Z/r)e^{-\lambda r}.$$

Setting

$$r = e^z, \quad r\psi = \phi e^{z/2},$$

we obtain

$$\left(\frac{d^2}{dz^2} + 2Ee^{2z} - (l+\tfrac{1}{2})^2 + 2Ze^z\exp(-\lambda e^z)\right)\phi = 0. \tag{18}$$

Thus the solution for the Coulomb potential ($\lambda = 0$) at energy zero is related to that for an S-wave exponential potential at energy $-\frac{1}{2}(l+\frac{1}{2})^2$.

We first derive the approximate asymptotic behavior of l—or rather $(l+\frac{1}{2})^2$—for large values of Z. We do this as in the case of the Mathieu equation[16] by locating the minimum of the function v defined by

$$v(z) = -2Ee^{2z} - 2Ze^z\exp(-\lambda e^z). \tag{19}$$

Differentiating v, we have at the position $z = z_0$ of the minimum

$$0 = v^{(1)}(z_0) = -2e^z[2Ee^z + Z(1-\lambda e^z)\exp(-\lambda e^z)]$$

where the superscript (i) of v indicates the ith derivative with respect to z. Solving for z_0 we have

$$z_0 = -\ln\lambda + \ln\left[1 + \frac{2Ee^{z_0}}{Z\exp(-\lambda e^{z_0})}\right].$$

For large Z and sufficiently small values of E we may expand the second logarithm and obtain

$$z_0 = -\ln\lambda + \frac{2Ee}{\lambda Z}\left[1 + \frac{4Ee}{\lambda Z} + O\left\{\left(\frac{E}{Z}\right)^2\right\}\right] \tag{20}$$

or

$$r_0 = e^{z_0} \simeq 1/\lambda.$$

We now expand $v(z)$ in the neighborhood of z_0, i.e., we write

$$v(z) = v(z_0) + \sum_{i=2}^{\infty}\frac{(z-z_0)^i}{i!}v^{(i)}(z_0). \tag{21}$$

Here

$$v(z_0) = -\frac{2Z}{\lambda e}\left(1+\frac{eE}{\lambda Z}\right) - \frac{e^2}{\lambda Z^2}\left(\frac{2E}{\lambda}\right)^3 + O\left(\frac{1}{Z^3}\right)$$

and

$$v^{(2)}(z_0) = \frac{2Z}{\lambda e}\left[1 - \frac{4eE}{\lambda Z}\right] + O\left(\frac{1}{Z^2}\right). \tag{22}$$

Thus in the neighborhood of z_0 we may approximate the radial wave equation by

$$\left(\frac{d^2}{dz^2} - (l+\tfrac{1}{2})^2 - v(z_0) - \frac{(z-z_0)^2}{2!}v^{(2)}(z_0)\right)\phi = 0. \tag{23}$$

The behavior of the "eigenvalues" $(l+\frac{1}{2})^2$ may now be determined by comparing (23) with the equation of parabolic cylinder functions. We set

$$z - z_0 = x/h, \quad h = [2v^{(2)}(z_0)]^{1/4} \tag{24}$$

and obtain

$$\left(\frac{d^2}{dx^2} - \frac{(l+\frac{1}{2})^2 + v(z_0)}{h^2} - \tfrac{1}{4}x^2\right)\phi = 0.$$

It follows that

$$(l+\tfrac{1}{2})^2 \simeq -v(z_0) - \tfrac{1}{2}qh^2,$$

where q is exactly an odd integer if the wavefunction is required to vanish exponentially at infinity; otherwise q is only approximately an odd integer, and its deviation from an odd integer can (in principle) be calculated in the form of an asymptotic expansion. By setting

$$(l+\tfrac{1}{2})^2 = -v(z_0) - \tfrac{1}{2}qh^2 - \Delta/h^4, \tag{25}$$

where Δ defines the remainder, the entire original wave equation becomes

$$D_q\phi = \frac{2}{h^2}\left[\frac{\Delta}{h^4} - \sum_{i=3}^{\infty}\frac{x^i v^{(i)}(z_0)}{i!h^i}\right]\phi \tag{26}$$

with

$$D_q \equiv -2\frac{d^2}{dx^2} - q + \tfrac{1}{2}x^2.$$

Equation (26) is now in a form suitable for application of our perturbation method. To a first approximation

$\phi = \phi^{(0)}$ is simply a parabolic cylinder function

$$\phi^{(0)} = \phi_q = D_{(q-1)/2}(x), \qquad D_q\phi_q = 0.$$

This function is well known to possess the recurrence formula

$$x\phi_q = (q, q+2)\,\phi_{q+2} + (q, q-2)\phi_{q-2},$$

where

$$\left.\begin{aligned}(q, q+2) &= 1\\ (q, q-2) &= \tfrac{1}{2}(q-1)\end{aligned}\right\}. \tag{27}$$

For higher powers we have

$$x^i\phi_q = \sum_{j=2i}^{-2i} S_i(q,j)\phi_{q+j}$$

and a recurrence relation may easily be written down for the coefficients S. The first approximation then leaves uncompensated terms amounting to

$$R_q^{(0)} = \frac{2}{h^2}\left(\frac{\Delta}{h^4}\phi_q - \sum_{i=3}^{\infty}\frac{v^{(i)}(z_0)}{i!h^i}\sum_{j=2i}^{-2i} S_i(q,j)\phi_{q+j}\right),$$

which we write as

$$R_q^{(0)} = \frac{2}{h^2}\sum_{i=3}^{\infty}\frac{1}{h^i}\sum_{j=2i}^{-2i}[q, q+j]_i\,\phi_{q+j},$$

where

$$\begin{aligned}[q,q]_4 &= \Delta - \frac{v^{(i)}(z_0)}{i!}S_4(q,0),\\ [q, q+j]_i &= -\frac{v^{(i)}(z_0)}{i!}S_i(q,j), \qquad 2i \geq j,\end{aligned} \tag{28}$$

for $j \neq 0$ when $i = 4$.

Since $D_{q+j} = D_q - j$, $D_q\phi_{q+j} = j\phi_{q+j}$, a term $\mu\phi_{q+j}$ in $R_q^{(0)}$ may be removed by adding to $\phi^{(0)}$ the contribution $\mu\phi_{q+j}/j$ except, of course, when $j = 0$. Thus the next order contribution of ϕ becomes

$$\phi^{(1)} = \frac{2}{h^2}\sum_{i=3}^{\infty}\frac{1}{h^i}\sum_{\substack{j=2i\\ j\neq 0}}^{-2i}\frac{[q, q+j]_i}{j}\,\phi_{q+j}.$$

In its turn this contribution leaves uncompensated

$$\begin{aligned}R_q^{(1)} &= \frac{2}{h^2}\sum_{i=3}^{\infty}\frac{1}{h^i}\sum_{\substack{j=2i\\ j\neq 0}}^{-2i}\frac{[q, q+j]_i}{j}R_{q+j}^{(0)}\\ &= \left(\frac{2}{h^2}\right)^2\sum_{i=3}^{2\infty}\frac{1}{h^i}\sum_{\substack{j=2i\\ j\neq 0}}^{-2i}\frac{[q, q+j]_i}{j}\sum_{i'=3}^{\infty}\frac{1}{h^{i'}}\sum_{j'=2i'}^{-2i'}\\ &\quad\times[q+j, q+j+j']_{i'}\,\phi_{q+j+j'}.\end{aligned}$$

Proceeding in this way we obtain the solution $\phi = \phi^{(0)} + \phi^{(1)} + \cdots$ valid for $|\ln r - \ln r_0| < 0(1/h)$ together with an eigenvalue equation from which Δ may be determined. The latter is obtained by setting equal to zero the sum of the terms in ϕ_q in $R_q^{(0)}, R_q^{(1)}, \ldots$ which have been unaccounted for so far. Thus

$$\begin{aligned}0 &= \frac{2}{h^2}\sum_{i=3}^{\infty}\frac{1}{h^i}[q,q]_i\\ &\quad+\left(\frac{2}{h^2}\right)^2\sum_{i=3}^{\infty}\frac{1}{h^i}\sum_{\substack{j=2i\\ j\neq 0}}^{-2i}\frac{[q, q+j]_i}{j}\sum_{i'=3}^{\infty}\frac{1}{h^{i'}}[q+j, q]_{i'}.\end{aligned} \tag{29}$$

In evaluating this expansion we recall that some of the coefficients $[q,q]_i$ are zero, as follows from the recurrence relation of the functions ϕ_q; e.g., $[q,q]_1 = [q,q]_3 = [q,q]_5 = \cdots = 0$. Calculating Δ and substituting into (25), we find

$$\begin{aligned}(l+\tfrac{1}{2})^2 &= -v(z_0) - \tfrac{1}{2}qh^2 - \tfrac{3}{2}(q^2+1)v^{(4)}(z_0)/4!h^4\\ &\quad - \tfrac{5}{2}q(q^2+5)v^{(6)}(z_0)/6!h^6 + O(v^{(8)}/h^8).\end{aligned} \tag{30}$$

This is an expansion in descending powers of Z as may be seen by observing that

$$v^{(i)}(z_0) = O(Z) = O(h^4).$$

Taking the square root, we obtain

$$\begin{aligned}l &= -\tfrac{1}{2} + [-v(z_0)]^{1/2}\left[1 + \frac{q}{4}\frac{h^2}{v(z_0)}\right.\\ &\quad\left. + \frac{1}{32v(z_0)}\left(\frac{(q^2+1)v^{(4)}(z_0)}{h^4} - \frac{q^2h^4}{v(z_0)}\right) + \cdots\right]\end{aligned}$$

or

$$\begin{aligned}l &= \left(\frac{2Z}{\lambda e}\right)^{1/2}\left(1 + \frac{eE}{2\lambda Z}\right) - \frac{1}{2}\\ &\quad - \frac{q}{4}2^{1/2}\left[1 - \frac{5}{2}\frac{eE}{\lambda Z}\right] + O\left(\frac{1}{Z}\right).\end{aligned} \tag{31}$$

More terms can readily be calculated when required.

We mentioned earlier that q is an odd integer. This follows from the requirement of normalizability of the solutions $r\psi$ for $E < 0$, i.e.,

$$\int_0^{\infty}|r\psi(r)|^2dr < \infty.$$

The solution we derived above—i.e., $r\psi = r^{1/2}\phi$, $\phi = \phi_q + O(1/h)$—is valid only in a small region of the range of integration, i.e. near

$$r_0 = (1/\lambda)[1 + O(E/Z)].$$

At the margin of this domain the solution has to be matched (i.e., continued analytically) to branches valid in the outside domains—e.g., to the Jost solution behaving like $\exp(iE^{1/2}r)$ for $r \to \infty$. As in customary quantum mechanics considerations it may be seen that our solution possesses zeros in its domain of validity for $Z \to \infty$ (i.e., is periodic) provided q is an odd integer.

Although we derived our solution for $E < 0$ it is clear that the Regge trajectories—obtained effectively as poles of the S matrix in the region of negative energies—must be valid for positive energies as well provided these energies are (as assumed) small in magnitude compared with Z. The imaginary part of the trajectories (which vanishes for $Z \to \infty$) is too small to show up in the terms we have calculated. However, its existence for $E > 0$ may be seen by looking at h. Inserting the first approximation of $v^{(2)}$ we have

$$h = \left[\frac{4Z}{\lambda e}\left(1 - \frac{4eE}{\lambda Z}\right)\right]^{1/4}.$$

Thus h^2 is complex, and l—which is a function of h^2—develops an imaginary part for $E > 0$.

4. CONCLUDING REMARKS

Our main objective here is to demonstrate that simple and straightforward perturbation procedures may

be developed to obtain complete large coupling expansions of energy eigenvalues and Regge trajectories for the Yukawa potential. In either case we derived one branch (valid in a restricted domain of r) of the solution; in Sec. 2, for instance, the so-called regular solution which is valid near $r = 0$. A complete investigation of the high energy solutions[15] and the example of the Gauss potential[12] allow us to assume that it is not difficult to derive other branches such as the Jost solutions (and thence the S matrix and the large coupling expansion of the phase shift). There are also some related problems of general interest. We mention in particular: the calculation of Regge trajectories of the Yukawa potential when E and Z are of comparable order of magnitude, and the calculation of Regge trajectories of the Gauss and other related potentials. The large coupling expansion (31) of the Regge trajectories for the Yukawa potential shows that for sufficiently large coupling constants the zero energy intercept may rise above 1. In a physical theory this may be taken to imply an upper limit for the possible value of the coupling constant provided $\alpha(0)$ for the Pomeranchuk trajectory is less than or equal to 1. However, this restriction for the Pomeranchuk trajectory is unnecessary as has been pointed out by Cheng and Wu.[3] These authors conjecture that the correct way to extend Regge theory to relativistic processes may be to define the Pomeranchuk trajectory as the J-plane singularity of the potential, rather than the scattering amplitude itself. From this point of view the Pomeranchuk trajectory is located at $J > 1$. It is interesting, therefore, to study large coupling solutions of more realistic models.

[1]R. B. Dingle, "The role of difference equations in perturbation theory," University of St. Andrews preprint 1964 (unpublished).
[2]L. I. Schiff, Phys. Rev. **92**, 766 (1953).
[3]H. Cheng and T. T. Wu, Phys. Rev. D **5**, 3192 (1972).
[4]M. Böhm, H. Joos, and M. Krammer, Nuovo Cimento A **7**, 21 (1972).
[5]H. J. W. Müller, Nuovo Cimento A **10**, 76 (1972).
[6]H. Cheng and T. T. Wu, Phys. Rev. D **5**, 3189 (1972).
[7]G. J. Iafrate and L. B. Mendelsohn, (a) Phys. Rev. **182**, 244 (1969); (b) Phys. Rev. A **2**, 561 (1970).
[8]E. Zauderer, J. Math. Phys. **13**, 42 (1972).
[9]H. J. W. Müller, Ann. Phys. (Leipz.) **16**, 255 (1965).
[10]H. J. W. Müller, in *Lectures in Theoretical High Energy Physics*, edited by H. H. Aly (Interscience, New York, 1968), p. 371.
[11]H. J. W. Müller, Ann. Phys. (Leipz.) **15**, 395 (1965); Physica **31**, 688 (1965).
[12]H. J. W. Müller, J. Math. Phys. **11**, 355 (1970).
[13]L. K. Sharma, Chap. 8, thesis, A.P.S. University of Rewa, India (1971).
[14]I. V. Simenog, Ukr. Phys. J. **13**, 184 (1969); Kiev Inst. for Theor. Physics, Preprint 68/80, 1968 (unpublished).
[15]H. J. W. Müller and K. Schilcher, J. Math. Phys. **9**, 255 (1968).
[16]References to applications to Mathieu, spheroidal and ellipsoidal wave equations may be found in H. J. W. Müller, Z. Physik **205**, 149 (1967). Applications to the Bethe-Salpeter equation may be found in references quoted in Ref. 5 above.

Erratum and Addendum: On the inverse problem for a hyperbolic dispersive partial differential equation. II [J. Math. Phys. 14, 406 (1973)]

V. H. Weston and R. J. Krueger

Division of Mathematical Sciences, Purdue University, West Lafayette, Indiana 47907
(Received 14 March 1973)

The following hypothesis was omitted from the final theorem: "Assume $N_1, N_2, N_1^{\perp}, N_2^{\perp}$ are invariant subspaces under the operators in (3)."

However, the results of that paper may be made stronger if we require that the solution of system (1) also satisfy, in a neighborhood of the end points, Eqs. (24b) and (25b) of Paper I, namely,

$$0 = \int_{-\mu}^{\mu} S_+(y,t)\psi(y)dy, \qquad 2l-\mu \le t < 2l-\mu+\delta,$$
$$0 = \int_{\mu}^{2l-\mu} S_-(y,t)\phi(y)dy, \qquad -\mu-\delta < t \le -\mu. \tag{7}$$

Taking the adjoint of (7) in conjunction with (3), we may again relate ϕ, ψ to the solution of an initial-boundary value problem as was done in Lemma 2. In this case we do not require these functions to satisfy the additional restrictions given in that lemma. This corresponds to having an incident wave $u_+^i(s)$ which is discontinuous at $s = 2\mu - 2l$, $s = 0$, and a wave $u_-^i(s)$ discontinuous at $s = 0$, $s = 2\mu$. As in Lemma 2, we can show that the solution of this initial-boundary value problem is zero for $x = 0$, $2l - 2\mu < t < 2l$ and $x = l$, $l < t < 3l - 2\mu$. Now, from the theory of weak solutions, the discontinuities in the incident waves propagate according to an exponential law, and thus prevent us from concluding that $u \equiv 0$ in the rectangle R described in Lemma 1. However, the adjoint of (7) will show that $u_+^r(s) = u_-^r(s) = 0$ for $s = 2l - 2\mu$ which implies that $u_+^i(s), u_-^i(s)$ are continuous at $s = 0$. We therefore observe that, in R, u is a Riemann function for the adjoint of the equation given in (2) and is required to vanish at $x = t = l - \mu$. If the Riemann function does not vanish, there is no discontinuity in the incident waves. Applying the alternative theorem for compact operators, the following is obtained.

Theorem: The only solution of systems (1) and (7) is the trivial solution if the Riemann function for the adjoint to the partial differential equation, given in system (2), associated with the Riemann boundary conditions on the characteristics $t = x + 2l - 2\mu$, $t = 2l - x$, does not vanish at $x = t = l - \mu$.

The condition that the Riemann function vanish is a strong restriction. It can be shown using successive approximations that if the coefficients of the equation satisfy the relations $\pm A = B$, $B \le 0$, $C \ge 0$, then the Riemann function will not vanish at $x = t = l - \mu$. For more general conditions on the coefficients there may exist values of μ for which the Riemann function will vanish at that point, and hence the solution is not unique. An example of this has been developed and will be published in future work.

Erratum: Gauge Invariant Decompositon of Yang-Mills Potentials [J. Math. Phys. 13, 1704 (1972)]

Richard P. Treat

Department of Physics, West Virginia University, Morgantown, West Virginia 26506
(Received 15 February 1972)

Equation (3.2) should read:

$$\partial_\alpha b^{*\alpha} - b_\alpha \times b^{*\alpha} = \partial_\alpha b^\alpha.$$

Line 1, column 1, p. 1705 should read:

10 which are required....

Line 39, column 1, p. 1706 should read:

where it is required[10] that....

Line 40, column 2, p. 1711 should read:

a functional of b_κ....

Equation (A10), column 2, p. 1713, should read:

$$\mathfrak{D}_{xx'}{}^{ij'} = \delta_i^{j'}/|x - x'|.$$

Copyright © 1973 by the American Institute of Physics