High-energy behavior of renormalized Feynman amplitudes

Edward B. Manoukian

Centre de Recherches Mathématiques, Université de Montréal, Montréal, Province de Québec H3C 3J4, Canada
(Received 28 November 1977)

The high-energy polynomial *and* logarithmic behavior of renormalized Feynman amplitudes, involving *subtractions*, is derived with total generality when *some* or all of the external momentum components of the graphs in question become large in Euclidean space nonexceptionally. This is achieved by explicitly carrying out the subtractions of renormalization as dictated, for example, by the improved Dyson–Salam renormalization scheme directly in momentum-space.

I. INTRODUCTION

The high-energy polynomial *and* logarithmic behavior of renormalized Feynman amplitudes A, involving subtractions, is derived with total generality when all or, more generally, *some* of the external momentum components of the graphs in question become large in Euclidean space nonexceptionally. This is obtained by explicitly carrying out the subtractions of renormalization as dictated by the improved Dyson–Salam renormalization scheme[1] directly in momentum space in obtaining the final expression for the renormalized Feynman integrand R or by the scheme of Bogoliubov–Parasiuk–Hepp–Zimmermann as given by Zimmermann.[2] The latter form is more suitable for this particular problem but is equivalent to the former one for the final expression for R.

In our investigations, the earlier work of Weinberg[3] and Fink,[4] which is also formulated in momentum space, and the ingenious method of determining the asymptotic behavior of special types of functions, of the type of Feynman integrals, is used as a basis. The analysis follows by a straightforward dimensional analysis of the subtracted out Feynman integrand followed by a detailed analysis of the so-called maximizing subspaces in determining the behavior of A in the Weinberg–Fink sense.

Section II is restricted to the simpler problem of the polynomial behavior of A and is then extended to the polynomial and logarithmic behavior of A in Sec. III where the maximizing subspaces are analyzed. An example is then worked out in detail. A general account of dimensional analysis of subtracted out Feynman integrands will be given in a subsequent work. The degree of divergence will be chosen to coincide with the dimensionality of a graph throughout the analysis.

II. POLYNOMIAL BEHAVIOR

Let the $4n$ integration variables of the renormalized Feynman integrand R and the $4n'$ components of the external independent momenta be combined to denote the components of a $4(n+n')$-vector in a Euclidean space $R^{4(n+n')}$. Let

$$\mathbf{P}\in R^{4(n+n')}, \tag{1}$$

such that any of the integration variables and any of the external independent momentum components may be written as a linear combination of the components of $\mathbf{P}$.

Let G, corresponding to the unrenormalized Feynman integrand I_G, denote the proper and connected graph associated with the renormalized amplitude A. For each line l in G carrying a momentum Q_l, we introduce a vector $\mathbf{V}_l$ in $R^{4(n+n')}$ such that $\mathbf{V}_l\cdot\mathbf{P}=Q_l$ for each component Q_l of the corresponding 4-vector. We denote the collection of all the $\mathbf{V}_l$ vectors collectively by $\mathbf{V},\cdots$. Let I be any chosen subspace of $R^{4(n+n')}$ associated with the $4n$ integration variables and let E be chosen to be any space in $R^{4(n+n')}$ such that I and E are disjoint and $R^{4(n+n')}=I+E$. [For definite and precise definition of $\Lambda(I)$, which we adopt, we refer the reader to Ref. 3.]

Let U be the set of all subspaces $\subseteq R^{4(n+n')}$ such that for any $S\in U$, $\Lambda(I)S=S_r$ $(\subseteq E)$ in reference to the renormalized, subtracted out[1,2] Feynman amplitude

$$A(\mathbf{L}_1\eta_1\cdots\eta_m+\cdots+\mathbf{L}_r\eta_r\cdots\eta_m+\cdots\mathbf{L}_m\eta_m+\mathbf{C}), \tag{2}$$

$$S_r=\{\mathbf{L}_1,\ldots,\mathbf{L}_r\},\quad 0<r\leq m,\quad 0<m\leq 4n', \tag{3}$$

where $\mathbf{L}_1,\ldots,\mathbf{L}_r$, are independent vectors in E spanning the space S_r and $\mathbf{C}$ is a vector confined to a finite region W in E. The parameters $\eta_1,\eta_2,\cdots$ are real and positive and taken to be independently large.

In reference to a space $S'\subseteq R^{4(n+n')}$, denote the asymptotic coefficient of the bound of R, in general, by $\alpha(S')$. All the maximizing subspaces $(\in U)$ for the bound of A, with respect to the parameter η_r in (2), are, by definition, included in the following subclass C of subspaces in U.

Definition of class C: Let $\{S',S'',\cdots\}$ be the subclass $C\subseteq U$ of all, arbitrarily labeled, subspaces $\subseteq R^{4(n+n')}$ such that for any $S',S''\in C$, $\alpha(S')+\dim S'=\alpha(S'')+\dim S''$, and for any $S\in U$, $\alpha(S)+\dim S\leq\alpha(S')+\dim S'$, and if the equality holds, then $S\in C$.

We now introduce the subclass C_0 of C from which the polynomial bound of A is readily determined (Theorem 1).

Definition of class C_0: Let $\{S',S'',\cdots\}$ be the given subclass $T\subseteq U$ of all, arbitrarily labeled, subspaces and $G',G'',\cdots$ diagrams $(\subseteq G)$ associated with them, i.e., with all their $\mathbf{V}$'s not orthogonal to $S',S'',\cdots$, respectively, such that the following are true. Let G'_i be *any* one of the connected components of G'. Let G'_{i0} be the subdiagram of G'_i corresponding to all those lines in G'_i forming closed loops, which means G'_i/G'_{i0} is either singly connecting (i.e., involving no closed

 0022-2488/78/1905-0917$1.00

loops) or empty (if G'_{i0} is not empty in the latter case), and $\dim S' - \dim S_r = 4\sum_i L(G'_{i0})$, where the sum corresponds to the connected parts of G'. $L(G)$ denotes the number of independent loops of G. [At least some of the **V**'s, associated with the external variables, in each proper and connected components of G'_{i0} must be not orthogonal to S'.] The subdiagrams G' and G'_i were arbitrary. Now let U_M denote the set of all the above diagrams $G', G'', \cdots$ respecting the just-mentioned properties, $U_M = \{G', G'', \cdots\}$. Let $U^0_M \subseteq U_M$ consist of all those diagrams in U_M maximizing the expression $\{d(G'), d(G''), \cdots\}$. I.e., if $\bar{S} \in T$, associated with it a diagram $\bar{G}$, then $d(\bar{G}) \leq d(\underline{G}')$ with $\bar{G} \in U_M$, $G' \in U^0_M$ and if the equality holds then $\bar{G} \in U^0_M$. Denote the class $\{S', \cdots\}$ of all the spaces with which are associated the diagrams in U^0_M by C_0.

A straightforward dimensional analysis of R with the subtractions given in the Zimmermann form establishes the following result.

Theorem 1: The asymptotic coefficient of $A(\mathbf{L}_1\eta_1\cdots\eta_m + \cdots + \mathbf{L}_r\eta_r\cdots\eta_m + \cdots + \mathbf{C})$ associated with η_r is given by

$$\alpha_I(S_r) = d(S')\big|_{S' \in C_0}, \tag{4}$$

where S' is any space in C_0 and $d(S')$ denotes the dimensionality $d(G')$ of diagram G' associated with S', and $C_0 \subseteq C$.

The simplicity of Theorem 1 is that although the theory involves subtractions, one may restrict oneself to a subclass C_0, in determining the polynomial bound of A, in such a way "as if" the theory involves no subtractions.

In the next section the class C will be determined and we finally obtain not only the polynomial but also the logarithmic behavior of the renormalized amplitudes A. When the theory involves subtractions the set $C - C_0$ is, in general, not empty and the knowledge of the larger class C is essential. It is thus important to note that the class C cannot be obtained without carrying out the subtractions of renormalization.

III. POLYNOMIAL AND LOGARITHMIC BEHAVIOR

In this section we complete our study in determining both the polynomial and logarithmic behavior of the renormalized, subtracted out, Feynman amplitudes. To do this we must first determine the larger class C of the maximizing subspaces for the bound of A relative to S_r ($\subseteq E$) in Eqs. (1) and (2).

As a preparation for Theorem 2 we need the following construction.

Let S be a given space in U. Suppose G' is a subdiagram of G such that $I_{G/G'}$ is independent of λ, the parameter associated with S in R. Here all the momenta, internal and/or external, in those lines in G with their **V**'s not orthogonal to S have been scaled by λ. Let $G'_1, \ldots, G'_n$ be the connected components of G'. Let $G'_{i1}, \ldots, G'_{iN_i}$ denote the proper and connected parts of G'_i, $i = 1, \ldots, n$ and $G'_{i0} \equiv \cup_{j=1}^{N_i} G'_{ij}$. By definition, with $G'_i \supset G'_{i0}$, G'_i is constructed out of $G'_{i1}, \ldots, G'_{iN_i}$ with the latter connected with one another by singly connecting lines, i.e., G'_i/G'_{i0} involves no closed loops. Let all the **V**'s associated with the external variables, in G'_i/G'_{i0} be not orthogonal to S for all $i = 1, \ldots, n$.

In general, let $G'_{ij1}, G'_{ij2}, \cdots$ be any proper and connected (if $\neq \emptyset$) subdiagrams contained in a G'_{ij} such that $G'_{ij1} \cap G'_{ij2} = \emptyset, \cdots$, pairwise and

$$d(G'_{ij1}) \geq 0, \quad d(G'_{ij2}) \geq 0, \quad \cdots, \tag{5}$$

with $j \in [1, \ldots, N_i]$, $i \in [1, \ldots, n]$, and *such that* all the **V**'s associated with the internal variables in the lines of $G'_{ij1}, G'_{ij2}, \cdots$ are orthogonal to S.

Let

$$L(S) \equiv \sum_{\substack{1 \leq j \leq N_i \\ 1 \leq i \leq n}}{}' L(G'_{ij}/G'_{ij1} \cup G'_{ij2} \cup \cdots), \tag{6}$$

where the sum in (6) is restricted only to those terms with all the **V**'s, associated with the internal variables in $G'_{ij}/G'_{ij1} \cup G'_{ij2} \cup \cdots$ not orthogonal to S.

Finally the space S and diagram G' are so chosen that the following are true:

(a) Let $\tilde{G}'$ be any other diagram, similarly constructed as G', with corresponding subdiagrams $\tilde{G}'_{ij}, \ldots, \tilde{G}'_{ijk}, \cdots$ as defined above for G', with $\tilde{G}'$ formally obtained from G' by adding or deleting a set of lines and vertices to G' with

$$L(S) = \sum{}' L(\tilde{G}'_{ij}/\tilde{G}'_{ij1} \cup \tilde{G}'_{ij2} \cup \cdots), \tag{7}$$

in reference to S similarly defined as in (6). Then G' is such that

$$d(\tilde{G}') \leq d(G'). \tag{8}$$

(b) For any proper and connected diagrams

$$\tilde{G}_B \supseteq \bigcup_{s=1}^{k} G'_{i_s}, \ldots, \tilde{G}_{B'} \supseteq \bigcup_{s'=1}^{k'} G'_{j_{s'}}, \cdots \tag{9}$$

(if any), where

$$\{G'_{i_1}, \ldots, G'_{i_k}\}, \ldots, \{G'_{j_1}, \ldots, G'_{j_{k'}}\}, \cdots \tag{10}$$

are *any* disjoint subsets of the set

$$\{G'_1, \ldots, G'_n\},$$

with

$$\tilde{G}_B \cap \tilde{G}_{B'} = \emptyset, \quad \text{pairwise} \tag{11}$$

and

$$I_{(\cup_B \tilde{G}_B) \cup ((\cup_{i=1}^{n}{}' G'_i)\,/\,\cup_{i=1}^{n} G'_i)}, \text{ is independent of } \eta, \tag{12}$$

where $\cup_B \tilde{G}_B$ is the union of the proper and connected diagrams as in (9) corresponding to the subsets appearing in (10) and $\cup_{i=1}^{n}{}' G'_i$ corresponds to the subdiagrams not appearing in the subsets in (10), then

$$\sum_B d(\tilde{G}_B) + \sum_{i=1}^{n}{}' d(G'_i) \leq \sum_{i=1}^{n} d(G'_i) = d(G'), \tag{13}$$

for *any* possible subsets as in (10) with any possible proper and connected diagrams as in (9), (11), and (13). Equation (13) is also true for all diagrams $\tilde{G}', \cdots$, as G', as given in (a) with the extreme left-hand side of (13) replaced by corresponding expressions for $\tilde{G}_B$'s defined in reference to $\tilde{G}'$ and the sum $\sum'_i d(G'_i)$ replaced by a corresponding expression $\sum'_i d(\tilde{G}'_i)$ with the right-hand

side of (13) unchanged. The conditions (a) and (b) are not difficult to follow (see below Theorem 2 and the construction of class C to be given below).

Then a straightforward dimensional analysis by using the subtractions in the Zimmermann form yields the following.

Theorem 2:

$$\deg_\chi R \leq d(G') - 4L(S), \tag{14}$$

with $L(S)$ as defined in (6). The equality in (14) may hold if one of the following is true for each $G'_{ij} \subset G'_i$, with $G' = \cup_{i=1}^n G'_i$, in each of the following cases:

Case 1: All the **V**'s, associated with the internal variables, in a $G'_{ij}/G'_{ij1} \cup G'_{ij2} \cup \cdots$ are not orthogonal to S.

(i) $d(G'_{ij}) < 0$, $G'_{ij1}, G'_{ij2}, \cdots = \emptyset$.

(ii) $d(G'_{ij}) \geq 0$, $G'_{ij1}, G'_{ij2}, \cdots = \emptyset$, and some of the **V**'s associated with the external variables in G'_{ij} are not orthogonal to S.

(iii) $d(G'_{ij}) < 0$, $G'_{ij1}, G'_{ij2}, \cdots \neq \emptyset$, and at least some of the **V**'s associated with external variables in each $G'_{ij1}, G'_{ij2}, \cdots$ are not orthogonal to S [the latter may be formally relaxed for those G'_{ijk} with $d(G'_{ijk}) \equiv 0$ from dimensional analysis alone].

(iv) $d(G'_{ij}) \geq 0$, $G'_{ij1}, G'_{ij2}, \cdots \neq \emptyset$, at least some of the **V**'s associated with the external variables in each $G'_{ij1}, G'_{ij2}, \cdots$ are not orthogonal to S [unless formally for $d(G'_{ijk}) \equiv 0$] and some of the **V**'s associated with the external variables to $G'_{ij}/G'_{ij1} \cup G'_{ij2} \cup \cdots$ are not orthogonal to S.

Case 2: All the **V**'s, associated with the internal variables, in a G'_{ij} are orthogonal to S.

(v) $d(G'_{ij}) \geq 0$ and some of the **V**'s associated with the external variables in G'_{ij} are not orthogonal to S.

If all the **V**'s, associated with the internal variables, in G' are orthogonal to S, then the above (i)—(v) criteria collapse to the last one [(v)] for all G'_{ij}'s in G' with $L(S) = 0$.

Condition (8) just guarantees the fact that the degree of R, with respect to λ, cannot be increased further beyond the right-hand side value of Eq. (14) by a rearrangement of subdiagrams. Condition (13) guarantees the fact that we cannot find Taylor operations corresponding to the subdiagrams $\tilde{G}_B, \ldots, \tilde{G}_{B'}, \cdots$ in (9) and (11) which may further increase $\deg_\lambda R$ beyond the bound given in Eq. (14). The latter derives from the following. Let

$$Y_{\tilde{G}_B} = I_{\tilde{G}_B/G'_{ij} \cup G'_{kl} \cup G'_{mn} \cup \cdots} (1 - T_{G'_{ij}}) \times Y_{G'_{ij}} (1 - T_{G'_{kl}}) Y_{G'_{kl}} (- T_{G'_{mn}}) Y_{G'_{mn}} \cdots,$$

be an expression corresponding to a subset in (10) such that

$$\deg_\eta (1 - T_{G'_{ij}}) Y_{G'_{ij}} \leq d(G'_{ij}) - 4L_{ij}(S^0),$$

where $L_{ij}(S^0) \neq 0$ is the number of independent loops in $G'_{ij}/G'_{ij1} \cup G'_{ij2} \cup \cdots$, for example. Then we have explicitly for $\tilde{G}_B \supset \cup_s G'_{i_s}$ with $d(\tilde{G}_B) \geq 0$,

$$\deg_\eta (- T_{\tilde{G}_B}) Y_{\tilde{G}_B} < d(G_B) - 4 \sum'_{[(i,j),(k,l),\cdots]} L_{ij}(S^0).$$

Summing over all such $\tilde{G}_B$'s in (9) leads to the condition stated in Eq. (13). On the other hand, (13) and (8) are equivalent if, for example, the diagram $(\cup_B \tilde{G}_B) \cup (\cup_{i=1}^{n'} G'_i)$ coincides with $\tilde{G}'$. [Condition (13) also guarantees the fact that we cannot find a space $\tilde{S} \in U$, as S given above, which may increase the expression $\deg_{\tilde{\chi}} R + 4L(\tilde{S})$, beyond the value $d(G')$, where $\tilde{\lambda}$ is the parameter associated with $\tilde{S}$ and $L(\tilde{S})$ is defined similarly to $L(S)$.]

Theorem 2, together with the definition of class C_0 and the formal definition of class C gives us the information needed to readily *construct* the class C.

A. Construction of class C

Let $\{S'_1, S''_1, \cdots\}$ be the collection of all, arbitrary labeled, spaces belonging to class C_0 with $G'_1, G''_1, \cdots$ diagrams associated with them with their **V**'s not orthogonal to $S'_1, S''_1, \cdots$, respectively. By definition each connected part of G'_1, for example, may be constructed out of proper and connected parts connected with one another by singly connecting lines. Let

$$G'_{10}, G'_{120}, \ldots, G'_{1n0}, \quad \text{be the collection of all the proper and connected parts of } G'_1. \tag{15}$$

For any $i \in [1, \ldots, n]$, let

$$\{g^i_1, g^i_2, \ldots, g^i_{t'_i}\} \tag{16}$$

be the set of the empty and *all* proper subdiagrams $\subseteq G'_{1i0}$, respectively, such that

$$d(g^i_{jk}) \geq 0, \quad j = 2, \ldots, t'_i, \tag{17}$$

for each kth proper and connected part of g^i_j. Consider the set

$$K_{i_1 i_2 \cdots i_n} = \{g^1_{i_1}, g^2_{i_2}, \ldots, g^n_{i_n}\}, \tag{18}$$

with

$$1 \leq i_1 \leq t'_1, \ldots, 1 \leq i_n \leq t'_n. \tag{19}$$

For each fixed set $\{i_1, \ldots, i_n\}$, we define a

$$\text{generalized diagram: } (G'_1 | g^1_{i_1} \cup \cdots \cup g^n_{i_n}), \tag{20}$$

by shrinking $g^1_{i_1}, \ldots, g^n_{i_n}$ in G'_1 to points and by replacing their corresponding integrands $I_{g^1_{i_1}}, \ldots, I_{g^n_{i_n}}$, by

$$\prod_{k(1)} (- T_{g^1_{i_1 k(1)}}) I_{g^1_{i_1}}, \ldots, \prod_{k(n)} (- T_{g^n_{i_n k(n)}}) I_{g^n_{i_n}}, \tag{21}$$

respectively in $I_{G'_1}$, where the $k(j)$'s correspond to the proper and connected components of g^j_{ij}. Note that since $G'_{1j0} \cap G'_{1k0} = \emptyset$, $j \neq k$ and $g^j_{i_j} \subseteq G'_{1j0}$, all the elements in $K_{i_1 i_2 \cdots i_n}$ in (18) are disjoint.

As a function of the *external variables* to g^j_{ij}, we introduce a generalized vertex function defined by

$$f_{g^j_{ij}} \equiv \prod_{k(j)} (- T_{g^j_{i_j k(j)}}) I_{g^j_{i_j k(j)}}. \tag{22}$$

[In general, we may also allow any other permissible Taylor operations between $(- T_{g^j_{i_j k(j)}})$ and $I_{g^j_{i_j k(j)}}$ in (22) [and (21)] in defining $f_{g^j_{ij}}$, but it is no loss of generality as given in (22) for the foregoing analysis.] The generalized diagram $(G'_1 | g^1_{i_1} \cup \cdots \cup g^n_{i_n})$ is then determined from G'_1 by replacing the subdiagrams $g^1_{i_1}, \ldots, g^n_{i_n}$ in the latter, as a function of their external variables, by the corresponding $f_{g^1_{i_1}}, \ldots, f_{g^n_{i_n}}$, respectively, in the analytic ex-

pression for $I_{G_1'}$. In this notation, $g_{i_1}^1,\dots,g_{i_n}^n$ will be treated as vertices in $(G_1'|g_{i_1}^1\cup\cdots\cup g_{i_n}^n)$ and vertex functions $f_{g_{i_1}^1},\dots,f_{g_{i_n}^n}$, in their *external variables*. In particular it is worth noting that, by definition, $(G_1'|\phi)\equiv G_1'$ and $(G_1'|G_1')$ if admissible, means just a vertex function as a function $f_{G_1'}$ in the external variables of G_1'.

Let $S'_{i_1\cdots i_n}$ be any space in U such that all the **V**'s in $G_1'/g_{i_1}^1\cup\cdots\cup g_{i_n}^n$ are not orthogonal to $S'_{i_1\cdots i_n}$ and at least some of the **V**'s in each of the proper and connected parts of $g_{i_1}^1,\dots,g_{i_n}^n$, associated with their external variables, are not orthogonal to $S'_{i_1\cdots i_n}$ [the latter may be formally relaxed for those proper and connected parts $g_{i_{jk}}^j$ and $d(g_{i_{jk}}^j)\equiv 0$, from dimensional analysis alone] with

$$\dim S'_{i_1\cdots i_n}-\dim S_r=4L(S'_{i_1\cdots i_n})\equiv 4L(G_1'/g_{i_1}^1\cup\cdots\cup g_{i_n}^n).$$

By definition, we, arbitrarily, label the subspaces in correspondence with the labeling of the diagrams associated with them. Accordingly the space $S'_{i_1\cdots i_n}$ is not uniquely labeled because, "in general," one may formally generate differently labeled spaces $\tilde S_{i_1\cdots i_n},\cdots$ by varying the number of lines in $g_{i_1}^1,\dots,g_{i_n}^n$ with the corresponding **V**'s, associated with the external variables, not orthogonal to $\tilde S_{i_1\cdots i_n},\cdots$ with $4L(S'_{i_1\cdots i_n})=4L(\tilde S_{i_1\cdots i_n})=\cdots$. Accordingly we will let $S'_{i_1\cdots i_n}$ denote, "in general" for convenience, a class of all such subspaces with which are "associated" the generalized diagram $(G_1'|g_{i_1}^1\cup\cdots\cup g_{i_n}^n)$ without introducing further notation.

By definition, $S'_{11\cdots1}\equiv S_1'$, in our previous notation. In particular it is worth noting that if G_1' were (totally) singly connecting (which may be defined as a tree) then $t_1'=1$.

Let $t'\equiv t_1'x\cdots xt_n'$ and relabel the set $\{S_1',\dots,S'_{i_1\cdots i_n},\cdots\}$ for all $1\le i_1\le t_1',\dots,1\le i_n\le t_n'$ by $\{S_1',S_2',\dots,S_{t'}'\}$. Repeat the above construction corresponding to the spaces S_1'', S_1''', $\cdots$ in C_0.

From definition of the classes C_0, C and Theorem 2 we readily see that the classes C_0 and C are given by

$$C_0=\{S_1',S_1'',\cdots\},\tag{23}$$

$$C=\{S_1',\dots,S_{t'}';S_1'',\dots,S_{t''}'';\cdots\}.\tag{24}$$

An explicit example of the above construction will be given later.

We will learn from Theorem 3 below, in determining the logarithmic asymptotic coefficients, that its conclusion does not change if an element in C is counted twice (or more) in it. [On the other hand, in general, an element in C may stand for a class of subspaces as mentioned above. The subspaces in C (and C_0) are arbitrarily labeled in correspondence with the labeling of the subdiagrams associated with them.]

B. Logarithmic asymptotic coefficients $\beta_I(S_r)$

Write I as a decomposition to, arbitrarily chosen and arbitrarily labeled, one-dimensional $4n$ disjoint subspaces, $I=I_1+I_2+\cdots+I_{4n}$, associated with the $4n$ integration variables. Introduce $4n$ parameters, $p_1,\dots,p_{4n}$ which may take on only the values 0 or 1. If all the maximizing subspaces for the I_1 integration relative to S_r, in reference to A [Eqs. (2) and (3)], after performing the $I_2+\cdots+I_{4n}$ integration, have the same dimension, then set $p_1=0$, and 1 otherwise. If all the maximizing subspaces for a I_i integration, after performing the $I_{i+1}+\cdots+I_{4n}$ integration, relative to *any* of the maximizing subspaces for the I_{i-1} integration after performing the $I_i+\cdots+I_{4n}$ integration have the same dimension, then $p_i=0$, and 1 otherwise for $i=2,\dots,4n$. Then according to Fink,[4] the logarithmic asymptotic coefficient associated with A is given by

$$\beta_I(S_r)=\sum_{i=1}^{4n}p_i.\tag{25}$$

By repeating the elementary Lemma 1 of Ref. 4 we arrive at the following lemma.

Lemma 1:

(i) Let $S'\in C$, then $\Lambda(I_2+\cdots+I_{4n})S'$ is a maximizing subspace for the I_1 integration relative to S_r after performing the $I_2+\cdots+I_{4n}$ integration.

(ii) Any maximizing subspace for the I_1 integration relative to S_r after performing the $I_2+\cdots+I_{4n}$ integration belongs to the class

$$\{\Lambda(I_2+\cdots)S',\Lambda(I_2+\cdots)S'',\cdots\}\equiv\Lambda(I_2+\cdots+I_{4n})C,$$

where $C=\{S',S'',\cdots\}$ and $\Lambda(I_2+\cdots+I_{4n})C$ consists of all such maximizing subspaces for the I_1 integration.

Again by repeating Lemma 1 for the I_2 integration, we obtain the following lemma.

Lemma 2:

(i) $\Lambda(I_3+\cdots I_{4n})S''$ (with $S''\in C$) is a maximizing subspace for the I_2 integration, after performing the $I_3+\cdots+I_{4n}$ integration, relative to $\Lambda(I_2+\cdots+I_{4n})S'$, for any $S'\in C$, if and only if, $\Lambda(I_2+\cdots+I_{4n})S''=\Lambda(I_2+\cdots+I_{4n})S'$.

(ii) All the maximizing subspaces for the I_2 integration, after performing the $I_3+\cdots+I_{4n}$ integration, relative to $\Lambda(I_2+\cdots+I_{4n})S'$, for any $S'\in C$, are given through $\{\Lambda(I_3+\cdots)S'',\Lambda(I_3+\cdots)S''',\dots,\Lambda(I_3+\cdots)S'\}$, with $S'',S''',\cdots$, those elements in C such that $\Lambda(I_2+\cdots)S''=\Lambda(I_2+\cdots)S'''=\cdots=\Lambda(I_2+\cdots)S'$.

Generalizing Lemma 2 to an arbitrary I_i integration, inductively, we readily arrive at the following in which we give the elementary details for the convenience of the reader. Suppose all those spaces $\tilde S,\tilde S'',\cdots$ in C with

$$\Lambda(I_{i-1}+\cdots)\tilde S'=\Lambda(I_{i-1}+\cdots)\tilde S''=\cdots=\Lambda(I_{i-1}+\cdots)S'_{i-1},$$

are maximizing subspaces for the $I_{i-1}+\cdots I_{4n}$ integration relative to $\Lambda(I_{i-1}+\cdots)S'_{i-1}$ with S'_{i-1} some space in C. Then

$$\alpha_{I_{i-1}+\cdots}(\Lambda(I_{i-1}+\cdots)S'_{i-1})=\max_{\Lambda(I_{i-1})\bar S=\Lambda(I_{i-1}+\cdots)S'_{i-1}}[\alpha_{I_i+\cdots}(\bar S)+\dim\bar S$$

$$
\begin{aligned}
&- \dim\Lambda(I_{i-1}+\cdots)S'_{i-1}] \geq \alpha_{I_i+\cdots}(\Lambda(I_i+\cdots)S'_i) \\
&+ \dim\Lambda(I_i+\cdots)S'_i - \dim\Lambda(I_{i-1}+\cdots)S'_{i-1} \\
&= \max_{\Lambda(I_i+\cdots)\tilde{S}=\Lambda(I_i+\cdots)S'_i} [\alpha(\tilde{S}) + \dim\tilde{S} \\
&- \dim\Lambda(I_{i-1}+\cdots)S'_{i-1}] \\
&\geq \alpha(S'') + \dim S'' - \dim\Lambda(I_{i-1}+\cdots)S'_{i-1},
\end{aligned}
$$

where S'_i and $S'' \in C$ are such that

$$
\begin{aligned}
&\Lambda(I_i+\cdots)S'' = \Lambda(I_i+\cdots)S'_i, \\
&\Lambda(I_{i-1}+\cdots)S'' = \Lambda(I_{i-1}+\cdots)S'_i \\
&\qquad = \Lambda(I_{i-1}+\cdots)S'_{i-1},
\end{aligned}
$$

i.e., S'' is a maximizing subspace for the $I_{i-1}+\cdots+I_{4n}$ integration relative to $\Lambda(I_{i-1}+\cdots)S'_{i-1}$, by hypothesis. Hence the chain of the above inequalities becomes one of the equalities leading to the conclusion that, in particular, $\Lambda(I_i+\cdots)S'_i$ is a maximizing subspace for the I_i integration, after performing the $I_{i+1}+\cdots+I_{4n}$ integration, relative to $\Lambda(I_{i-1}+\cdots)S'_{i-1}$. We also readily, and similarly, learn that all the maximizing subspaces for the $I_i+\cdots+I_{4n}$ integration relative to $\Lambda(I_i+\cdots)S'_i$ are all those $S', S'', \cdots$ in C with

$$
\begin{aligned}
&\Lambda(I_i+\cdots)S' = \Lambda(I_i+\cdots)S'' = \cdots \\
&\qquad = \Lambda(I_i+\cdots)S'_i,
\end{aligned}
$$

and all the maximizing subspaces for the I_{i+1} integration, after performing the $I_{i+2}+\cdots I_{4n}$ integration, relative to $\Lambda(I_i+\cdots)S'_i$ are all the elements in

$$
\{\Lambda(I_{i+1}+\cdots)S', \ \Lambda(I_{i+1}+\cdots)S'', \ldots, \Lambda(I_{i+1}+\cdots)S'_i\}
$$

with $S', S'', \cdots$, those elements in C satisfying the above relation. This leads to the following lemma.

Lemma 3: There is a space S'_i in C, such that $\Lambda(I_i+\cdots)S'_i$ is a maximizing subspace for the I_{i-1} integration, after performing the $I_i+\cdots+I_{4n}$ integration, relative to any one of the maximizing subspaces for the I_{i-2} integration after performing the $I_{i-1}+\cdots+I_{4n}$ integration. All the maximizing subspaces for the I_i integration, after performing the $I_{i+1}+\cdots+I_{4n}$ integration, relative to $\Lambda(I_i+\cdots)S'_i$ are given by

$$
\{\Lambda(I_{i+1}+\cdots)S'_{i+1}, \ \Lambda(I_{i+1}+\cdots)S''_{i+1}, \ldots, \Lambda(I_{i+1}+\cdots)S'_i\},
$$

with $S'_{i+1}, S''_{i+1}, \cdots$, all those elements in C such that

$$
\begin{aligned}
&\Lambda(I_i+\cdots)S'_{i+1} = \Lambda(I_i+\cdots)S''_{i+1} = \cdots \\
&\qquad = \Lambda(I_i+\cdots)S'_i.
\end{aligned}
$$

The logarithmic asymptotic coefficient $\beta_I(S_r)$ associated with A is then determined from the class C as given in the following theorem.

Theorem 3:

$$
\beta_I(S_r) = \sum_{i=1}^{4n} p_i, \tag{26}
$$

where:

(1) $p_1 = 0$, if all the elements in

$$
\{\dim\Lambda(I_2+\cdots)S'_1 - \dim S_r, \ \dim\Lambda(I_2+\cdots)S'' - \dim S_r, \ \cdots\} \tag{27}
$$

are equal, with $\{S', S'', \cdots\} = C$, and $p_1 = 1$ otherwise;

(2) $p_i = 0$, $i \in [2, 3, \ldots, 4n]$, if all the elements in

$$
\{\dim\Lambda(I_{i+1}+\cdots)S'_{i+1} - \dim S_r, \ \dim\Lambda(I_{i+1}+\cdots)S''_{i+1} - \dim S_r, \ \cdots\} \tag{28}
$$

are equal, and $p_i = 1$ otherwise, for all those S'_{i+1}, $S''_{i+1}, \cdots \in C$, such that

$$
\Lambda(I_i+\cdots)S'_{i+1} = \Lambda(I_i+\cdots)S''_{i+1} = \cdots = \Lambda(I_i+\cdots)S'_i, \tag{29}
$$

where $\Lambda(I_i+\cdots)S'_i$ is a maximizing subspace for the I_{i-1} integration, after performing the $I_i+\cdots+I_{4n}$ integration, relative to any one of the maximizing subspaces for the I_{i-2} integration after performing the $I_{i-1}+\cdots+I_{4n}$ integration, with S'_i some element in C.

For convenience, the dimension of the maximizing subspaces in (27) and (28) were measured relative to the dimension of S_r.

Theorem 3 is general and fully determined from the class C. It will be simplified further in Theorem 4. However, it is instructive to apply Theorem 3 to the example given in the Appendix. In the example depicted in Fig. 1 and given in the Appendix, let $I_1+\cdots+I_4$ be arbitrarily associated with the variables $k_{24}^{g_1}$ and $I_5+\cdots+I_8$ with $k_{13}^{g_2}$, with nonvanishing Jacobian,

$$
J\left(\frac{k_1, k_2}{k_{24}^{g_1}, k_{13}^{g_2}}\right).
$$

Then

$$
\begin{aligned}
p_1 &\to \{\dim\Lambda(I_2+\cdots)S'_1 - \dim S_r, \ldots, \dim\Lambda(I_2+\cdots)S'_4 - \dim S_r\} \\
&= \{1, 0, 1, 0\}, \quad p_1 = 1, \\
p_2 &\to \{\dim\Lambda(I_3+\cdots)S'_4 - \dim S_r\} = \{0\}, \quad p_2 = 0, \\
p_3 &\to \{\dim\Lambda(I_4+\cdots)S'_4 - \dim S_r\} = \{0\}, \quad p_3 = 0, \\
p_4 &\to \{\dim\Lambda(I_5+\cdots)S'_4 - \dim S_r\} = \{0\}, \quad p_4 = 0, \\
p_5 &\to \{\dim\Lambda(I_6+\cdots)S'_2 - \dim S_r, \ \dim\Lambda(I_6+\cdots)S'_4 - \dim S_r\} \\
&= \{1, 0\}, \quad p_5 = 1
\end{aligned}
$$

(see definition of S'_2 in the Appendix),

$$
\begin{aligned}
p_6 &\to \{\dim\Lambda(I_7+\cdots)S'_4 - \dim S_r\} = \{0\}, \quad p_6 = 0, \\
p_7 &\to \{\dim\Lambda(I_8+\cdots)S'_4 - \dim S_r\} = \{0\}, \quad p_7 = 0, \\
p_8 &\to \{\dim S'_4 - \dim S_r\} = \{0\}, \quad p_8 = 0.
\end{aligned}
$$

Hence we finally have for the example given in the Appendix,

$$
\alpha_I(S_r) = 1, \tag{30}
$$

$$
\beta_I(S_r) = 2. \tag{31}
$$

For any $S' \in C$, S' is a maximizing subspace for the $(I_2+\cdots+I_{4n})$ integration relative to S_r and also by the same reasoning as that leading to Lemma 3 we learn that S' is a maximizing subspace for the $I_i+\cdots+I_{4n}$ integration relative to $\Lambda(I_i+\cdots I_{4n})S'$ for any $1 \leq i \leq 4n$. By the application of Theorem 3 and by conveniently choosing S' to be an element *in* C_0 we arrive at the following simpler version of Theorem 3.

Theorem 4: Let $S', S'', \cdots$ be all those spaces in C such that

$$\Lambda(I_i+\cdots)S' = \Lambda(I_i+\cdots)S'' = \cdots = \Lambda(I_i+\cdots)S_1', \tag{32}$$

with $S_1' \in C_0$, then

$$\beta_I(S_r) = \sum_{i=1}^{4n} p_i,$$

with $p_i = 0$, if all the elements in

$$\{\dim\Lambda(I_{i+1}+\cdots)S_1' - \dim S_r, \ldots, \dim\Lambda(I_{i+1}+\cdots)S'' - \dim S_r, \dim\Lambda(I_{i+1}+\cdots)S' - \dim S_r\} \tag{33}$$

are equal, and $p_i = 1$, otherwise, with $i = 1, \ldots, 4n$.

In determining $\beta_I(S_r)$ from Theorem 4, the subspace $S_1'(\in C_0)$ will be called a reference space. In many applications, Theorem 4 simplifies the analysis, since a specific reference space has been chosen, and this is especially the case for examples with $C_0 = \{S_1'\}$ and by the very construction of class C.

For the example given in the Appendix with the variables chosen as before, an immediate application of Theorem 4 (and Theorem 1) leads to:

$$\begin{aligned}
p_1 &\to \{\dim\Lambda(I_2+\cdots)S_1' - \dim S_r, \ldots, \dim\Lambda(I_2+\cdots)S_4' - \dim S_r\} \\
&= \{1,0,1,0\}, \quad p_1 = 1, \\
p_i &\to \{\dim\Lambda(I_{i+1}+\cdots)S_1' - \dim S_r\} = \{0\}, \quad p_i = 0, \\
&\quad i = 2,3,4,6,7,8, \\
p_5 &\to \{\dim\Lambda(I_6+\cdots)S_1' - \dim S_r, \dim\Lambda(I_6+\cdots)S_3' - \dim S_r\} \\
&= \{5,4\}, \quad p_5 = 1,
\end{aligned} \tag{34}$$

and hence

$$\alpha_I(S_r) = 1, \tag{35}$$

$$\beta_I(S_r) = 2, \tag{36}$$

as before.

IV. SUMMARY

Definition of the class C_0, the construction of the class C, and the very general Theorem 4 (or Theorem 3) and Theorem 1, essentially contain our main results which we now summarize. Other examples, including special cases, then follow from the application of these very general results in a standard manner.

We give the following brief, but formal, summary of our results, which is helpful for practical applications in obtaining the power and logarithmic behavior of the renormalized, subtracted out, amplitudes A.

(i) In determining the power of η_r in A, i.e., $\alpha_I(S_r)$, one may, formally, read off $\alpha_I(S_r)$ by inspecting the *maximum* dimensionality $d(G')$ (i.e., of a maximizing G') of a subdiagram of $G' \subseteq G$ with G' associated with a space S' (i.e., all its V's are not orthogonal to S') with $\Lambda(I)S' = S_r$ and the number of independent loops of G' equal to $\dim S' - \dim S_r$. G', of course, may in general contain a subdiagram (or subdiagrams) with no closed loops. The association of a subidagram G' with a space S' and the introduction of the vectors V, here, is as *already* given, precisely as in Ref. 3. The analysis of carrying out the subtractions in determining the power law of A, i.e., the power of η, shows that this may be done "as if" the theory involves no subtractions.

(ii) In determining the logarithmic asymptotic coefficients $\beta_I(S_r)$, all the maximizing subspaces for the bound of A need to be analyzed. In the previous section we have shown how to construct all these maximizing subspaces. This may be, formally, summarized through the following. Let $\{G_1', G_1'', \cdots\}$ be the set of all subdiagrams $\subseteq G$ maximizing the expression for the bound of A as given *in* (i), i.e., in such a way as if the theory involves no subtractions. Let $\{S_1', S_1'', \cdots\}$ be the set of subspaces with which are associated the subdiagrams $G_1', G_1'', \cdots$ [with $\Lambda(I)S_1' = S_r$, etc.]. Let $\{\emptyset, g_1, \cdots\}$ be the set of all proper, but not necessarily connected, subdiagrams $\subseteq G_1'$, such that each connected part of $g_1, \cdots$ is divergent (i.e., with nonnegative dimensionality), including the case with $g_0 \equiv \emptyset$. [By definition, a proper but not necessarily connected diagram means that the number of its connected parts does not increase upon cutting any one of its internal lines.] For each G_1', introduce the set of all subdiagrams $\{G_1', \ldots, (G_1'/g_i), \cdots\}$ with $g_i \in \{\emptyset, g_1, \cdots\}$ as just defined. Let $\{S_1', S_2', \cdots\}$ be the set of all spaces such that all the V's in $\cdots, (G_1'/g_i), \cdots$ are not orthogonal to $\cdots, S_i', \cdots$, respectively, with all the V's in (G/G_1') orthogonal to $S_1', S_2', \cdots$, respectively, and the number of independent loops of (G_1'/g_i) equal to $\dim S_i' - \dim S_r$. By repeating the above construction from the subdiagrams $G_1'', \cdots$, as well, we obtain, formally, the class of all the maximizing subspaces for the bound of A, $\{S_1', S_2', \ldots, S_1'', S_2'', \cdots\}$. The logarithmic asymptotic coefficients are then readily obtained by the application of Theorems 3 or 4. An explicit example of the above construction has been given in the text.

The above two points then briefly, but formally, summarize our results for the determination of the behavior of A. The keen reader will, however, appreciate going through the analysis and the more precise definitions as given in the text.

ACKNOWLEDGMENTS

I would like to thank my colleages Jiri Patera and Pavel Winternitz of the Centre de Recherches Mathématiques of the University of Montreal and Robert T. Sharp of the department of physics of McGill University for their very kind hospitality which made this work possible. I would also like to take this opportunity to give my sincere thanks to Professor Steven Weinberg for the keen interest he has shown in the authors's work on renormalization.

APPENDIX

Consider the self-energy graph in Fig. 1 with S_r corresponding to the vector q by the fermion.

Let Q_{ij} denote the momentum flowing from the ith to

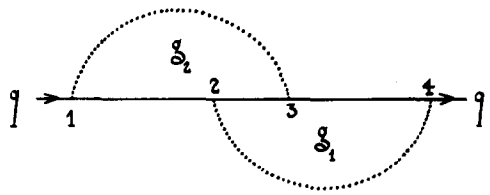

FIG. 1. A self-energy graph G of the fermion with a $\bar{\psi}\psi\phi$ coupling contribution to it with the dotted line denoting a scalar boson and the fermion is of spin $\frac{1}{2}$.

the jth vertex in Fig. 1 with Q_{12} and Q_{34} denoting the momenta carried by the virtual scalar boson. The former may be written as

$$Q_{ij}=k_{ij}+q_{ij}, \quad i,j\in[1,2,3,4], \quad Q_{ij}=-Q_{ji}, \tag{A1}$$

where k_{ij} is a linear combination of the integration variables and q_{ij} is a function of the external variable q of G. A canonical choice of variables[2,1] is given as the solution of

$$q_{ij}=u_i-u_j, \tag{A2}$$

with

$$\begin{aligned} &q_{12}+q_{13}=q_{24}+q_{34}=q,\\ &q_{12}+q_{32}+q_{42}=0,\\ &q_{13}+q_{43}+q_{23}=0. \end{aligned} \tag{A3}$$

Similarly, in reference to the subdiagram g_1, we write

$$Q_{ij}=k_{ij}^{g_1}+q_{ij}^{g_1}, \quad i,j\in[2,3,4], \tag{A4}$$

with $q_{ij}^{g_1}$ as the solution of

$$q_{ij}^{g_1}=v_i-v_j, \tag{A5}$$

$$\begin{aligned} &Q_{12}+q_{32}^{g_1}+q_{42}^{g_1}=0,\\ &Q_{13}+q_{43}^{g_1}+q_{23}^{g_1}=0, \end{aligned} \tag{A6}$$

$$q_{24}^{g_1}+q_{34}^{g_1}=q.$$

In reference to the subdiagram g_2, we write

$$Q_{ij}=k_{ij}^{g_2}+q_{ij}^{g_2}, \quad i,j\in[1,2,3], \tag{A7}$$

with $q_{ij}^{g_2}$ as the solution of

$$q_{ij}^{g_2}=\omega_i-\omega_j, \quad Q_{42}+q_{12}^{g_2}+q_{32}^{g_2}=0, \tag{A8}$$

$$q+q_{31}^{g_2}+q_{21}^{g_2}=0, \quad q_{13}^{g_2}+q_{23}^{g_2}+Q_{43}=0. \tag{A9}$$

The above system is readily solved to give a solution of the form:

$$\begin{aligned} Q_{12}&=Ak_1+Ck_2+q/2,\\ Q_{13}&=-Ak_1-Ck_2+q/2,\\ Q_{32}&=-(A-B)k_1-(C-D)k_2,\\ Q_{34}&=-Bk_1-Dk_2+q/2,\\ Q_{24}&=Bk_1+Dk_2+q/2, \end{aligned} \tag{A10}$$

$$\begin{aligned} k_{24}^{g_1}&=\left(B-\frac{A}{3}\right)k_1+\left(D-\frac{C}{3}\right)k_2, \quad q_{24}^{g_1}=\frac{A}{3}k_1+\frac{C}{3}k_2+\frac{q}{2},\\ k_{34}^{g_1}&=-\left(B-\frac{A}{3}\right)k_1-\left(D-\frac{C}{3}\right)k_2,\\ q_{34}^{g_1}&=-\frac{A}{3}k_1-\frac{C}{3}k_2+\frac{q}{2}, \end{aligned} \tag{A11}$$

$$\begin{aligned} k_{32}^{g_1}&=\left(B-\frac{A}{3}\right)k_1+\left(D-\frac{C}{3}\right)k_2, \quad q_{32}^{g_1}=-\frac{2A}{3}k_1-\frac{2C}{3}k_2,\\ k_{12}^{g_2}&=-\left(A-\frac{B}{3}\right)k_1-\left(C-\frac{D}{3}\right)k_2, \quad q_{13}^{g_2}=-\frac{B}{3}k_1-\frac{D}{3}k_2+\frac{q}{2},\\ k_{12}^{g_2}&=\left(A-\frac{B}{3}\right)k_1+\left(C-\frac{D}{3}\right)k_2, \quad q_{12}^{g_2}=\frac{B}{3}k_1+\frac{D}{3}k_2+\frac{q}{2},\\ k_{32}^{g_2}&=-\left(A-\frac{B}{3}\right)k_1-\left(C-\frac{D}{3}\right)k_2, \quad q_{32}^{g_2}=\frac{2B}{3}k_1+\frac{2D}{3}k_2, \end{aligned} \tag{A12}$$

where the constants A,B,C,D are arbitrary to the extent that $k_{lm}^{g_i}\neq 0$, $k_{lm}^{G}\neq 0$ for arbitrary k_1 and k_2, $i=1,2$, and $l,m\in[1,2,3,4]$.

Construction of the class C: The class C_0 is given by $C_0=\{S_1'\}$ with the whole graph G associated with S_1': $G_1'\equiv G$ and $G_{10}'\equiv G$, i.e., G_1'/G_{10}' is empty. From the construction of C we readily obtain

$$C=\{S_1', S_2', S_3', S_4'\}, \tag{A13}$$

with which are, respectively, associated the diagrams,

$$(G|\emptyset)\equiv G, \quad (G|g_1), \quad (G|g_2), \quad (G|G). \tag{A14}$$

[Note that since $d(g_1)=0=d(g_2)$, it follows that $(G|g_i)=(G/g_i)$.] The spaces $S_1',\ldots,S_4'$ are precisely defined as follows. Label the coordinate axes by (k_1,k_2,q). Then

$$S_1'=\{\mathbf{L}_1,\mathbf{L}_2,\mathbf{L}_3\}$$

with

$$\mathbf{L}_1=(1,0,0), \quad \mathbf{L}_2=(0,1,0), \quad \mathbf{L}_3(0,0,1);$$

$$S_2'=\{\mathbf{L}_1',\mathbf{L}_2'\}$$

with

$$\mathbf{L}_1'=(\alpha_2,\beta_2,0), \quad \mathbf{L}_2'=(0,0,1),$$

respectively, where α_2 and β_2 are any nontrivial consistent solutions of

$$\alpha_2\left(B-\frac{A}{3}\right)+\beta_2\left(D-\frac{C}{3}\right)=0; \tag{A15}$$

$S_3'=\{\mathbf{L}_1'',\mathbf{L}_2''\}$ with $\mathbf{L}_1''=(\alpha_3,\beta_3,0)$, $\mathbf{L}_2''=(0,0,1)$, respectively, where α_3 and β_3 are any nontrivial consistent solutions of

$$\alpha_3\left(\frac{B}{3}-A\right)+\beta_3\left(\frac{D}{3}-C\right)=0, \tag{A16}$$

and $S_4'=\{\mathbf{L}\}$ with $\mathbf{L}=(0,0,1)$. [It should be noted that k_1,k_2,q are formally 4-vectors.]

[1] E. B. Manoukian, Phys. Rev. D 14, 966, 2202(E) (1976).
[2] W. Zimmermann, Commun. Math. Phys. 15, 208 (1969).
[3] S. Weinberg, Phys. Rev. 118, 838 (1960).
[4] J. P. Fink, J. Math. Phys. 9, 1389 (1968).

Spectral theory for the acoustic wave equations with generalized Neumann boundary conditions in exterior domain

K. -H. Chen[a)]

SIAM Institute for Mathematics and Society, Visiting Department of Genetics, Medical School of Stanford University, Stanford, California 94305

C. C. Yang

Research Mathematician, Space Systems Division, Naval Research Laboratory, Washington, D. C. 20375
(Received 14 June 1976)

In this paper we studied the spectral theory of the operator (either in the exterior domain or in the whole space) that is induced from problems in the n-dimensional Euclidean space for the hyperbolic linear partial differential equations with the generalized Neumann boundary condition. The resulting theory provides a foundation for studying the wave operator and scattering operator involved in scattering theory and possibly also for studying the respective inverse problem.

I. INTRODUCTION

The following is a preliminary report on some recent theoretical investigation pertaining to scattering theory. Further study and the physical implications of these results will be discussed elsewhere. The scheme for direct scattering theory is outlined first, and an important topic is indicated for the inverse problem.

If $u(x)$ is the difference between the instantaneous pressure and the equilibrium pressure, $\rho(x)$ is the equilibrium density of the medium, $c(x)$ is the local speed of sound, Ω is an open connected subset of R^n with bounded complement, D_t denotes $\partial/\partial t$, and

$$A \equiv -c^2(x)\rho(x)\nabla\cdot\frac{1}{\rho(x)}\nabla, \quad t\in R,\ x\in\Omega,$$

then the acoustic wave equation is

$$D_t^2 u = -Au, \tag{1}$$

with the initial conditions

$$u(0,x)=f(x) \text{ and } D_t u(0,x)=g(x), \quad x\in\Omega. \tag{2}$$

and the *generalized Neumann boundary condition*

$$\int c^{-2}(x)\rho^{-1}(x)\mathrm{Au}(x)v(x)dx = \int \rho^{-1}(x)\nabla u(x)\cdot\nabla v(x)dx, \quad x\in\Omega \text{ and } v,\nabla v\in L^2(\Omega). \tag{3}$$

If the domain Ω considered here has a smooth boundary $\partial\Omega$, say C^2, then boundary condition (3) is equivalent to the *classical Neumann boundary condition*

$$\nu\cdot\nabla u(t,x)=0, \quad x\in\partial\Omega, \tag{4}$$

with ν denoting the outward unit normal to $\partial\Omega$ at x.

This acoustic wave propagation problem is considered here as a perturbed system in contrast to the following, named "the unperturbed system":

$$D_t^2 v(t,x)=A_0 v(t,x)=-\nabla^2 v(t,x), \quad x\in\Omega, \tag{5}$$

with the same initial and boundary conditions (2) and (3) of the perturbed problem.

We impose here the general assumptions that apply throughout this paper:

Assumption 1: The exterior domain Ω has the *finite tiling property*: There exists an open set O in R^n, compact sets $K_1,\ldots,K_N$ in R^n, and nonzero vectors $x^{(1)},\ldots,x^{(N)}$ such that

$$\partial\Omega\subset 0, \tag{6}$$

$$O\cap\Omega\subset\cup K_j, \quad 1\le j\le N, \tag{7}$$

and

$$\{x=x_0+tx^{(j)}: 0<t<1\}\subset\Omega, \quad x_0\in\Omega\cap K_j. \tag{8}$$

This property of an exterior domain is due to Wilcox.[1] Here we would not exlude the case $\Omega=R^n$.

Assumption 2: The density $\rho(x)$ is $C^2(\Omega)$ and real valued, and for some constant $J>1$

$$J\ge\rho(x)\ge J^{-1}, \quad x\in\Omega. \tag{9}$$

Also,

$$\rho(x)\to 1 \text{ when } |x|\to\infty, \tag{10}$$

and

$$D^\alpha\frac{1}{\rho(x)} \text{ behaves like } o(|x|^{-1}) \text{ when } |x|\to\infty \tag{11}$$

for all α, $1\le|\alpha|=\alpha_1+\cdots+\alpha_n\le 2$.

Assumption 3: The local speed $c(x)$ is $C^1(\Omega)$ and real valued, and for some constant $K>1$

$$K\ge c(x)\ge K^{-1}, \quad x\in\Omega. \tag{12}$$

Also,

$$c(x)\to 1 \text{ when } |x|\to\infty \tag{13}$$

and

$$\nabla\ln[c(x)]=o(|x|^{-1}) \text{ when } |x|\to\infty. \tag{14}$$

Assumption 4: The "Stummel condition" is satisfied by $q(x)\equiv c^{-2}(x)\rho^{-1}(x)$, that is, for some $a>0$,

$$\sup\int|q(y)|^2|x-y|^{-n+4-a}dy<+\infty, \text{ if } n\ge 4, \tag{15}$$

[a)] Most portion of the research work was carried on while K. H. Chen was at the Department of Mathematics, The University of New Orleans and at the Department of Biomathematics, University of Alabama in Birmingham.

 0022-2488/78/1905-0924$1.00

where the integration variable y runs in the disk $\{|x-y|<1\}\cap\Omega$ and the supremum is taken on $x\in\Omega$.

The differential operator A_0 defined by (5) on Ω, with Assumption 1, subject to the generalized Neumann boundary condition, is a self-adjoint, nonnegative operator in the Hilbert space $L_2(\Omega)$. Its spectrum is the closed interval $[0, \infty)$ and is absolutely (spectral) continuous and without eigenvalues. These interesting results are proved by Wilcox.[1] With these four general assumptions we will show that the operator A defined by (1) subject to the generalized Neumann boundary condition is also a self-adjoint, nonnegative operator in the same Hilbert space $L_2(\Omega)$ and that its spectrum contains the interval $(0, \infty)$, is contained in $[0, \infty)$, and is absolutely continuous. The only uncertainty occurs when the origin is to be an eigenvalue. These results are discussed in the next section.

The scattering operator S is unitary if the Møller wave operators $W_\pm$ are orthogonal,

$$W_\pm^* W_\pm = I, \tag{16}$$

and are complete,

$$W_\pm W_\pm^* = I - E(0+), \tag{17}$$

where $E(\lambda)$ is the resolution of the identity for A. Therefore, in the third section, we will discuss the existence of the Møller wave operators $W_\pm$ and properties (16) and (17).

Both relations (16) and (17) could be proved by employing the expansion principle of the generalized eigenfunctions, which requires the principle of the limiting absorption. This latter principle also plays a key role in the proof of the spectrum to be absolutely continuous. In Sec. 4, we will study the principle in detail.

II. SOLUTIONS OF THE ACOUSTIC WAVE EQUATION EQUATION WITH THE NEUMANN BOUNDARY CONDITION AND ITS SPECTRAL THEORY

A solution of the mixed initial-boundary-value problem of Eqs. (1)–(3) will be constructed by using a famous spectral theorem. Then the absolute continuity of the spectrum of A will be studied. The resulting theory provides a preparation for constructing the wave operator and scattering operator.

We recall here the initial-boundary-value problem:

$$D_t^2 = -Au \equiv c^2(x)\rho(x)\nabla\cdot\frac{1}{\rho(x)}\nabla u, \quad t>0,\ x\in\Omega, \tag{18}$$

$$u(0,x)=f(x),$$

and

$$D_t u(0,x)=g(x), \quad x\in\Omega, \tag{19}$$

and

$$\int c^{-2}(x)\rho^{-1}(x)Au(x)v(x)dx = \int\rho^{-1}(x)\nabla u(x)\cdot\nabla v(x)dx, \quad x\in\Omega \text{ and } v,\nabla v\in L^2(\Omega). \tag{20}$$

The formulation of the problem will be based on the following function spaces:

$$u\in L_2(\Omega) \Longleftrightarrow \int |u(x)|^2dx<\infty,\ x\in\Omega, \tag{21}$$

in the Lebesque measure,

$$u\in H_m(\Omega) \Longleftrightarrow D^\alpha u\in L_2(\Omega), \quad |\alpha|\le m, \tag{22}$$

$$u\in L_2(\Omega;c^2\rho) \Longleftrightarrow (c^2\rho)^{-1/2}u\in L_2(\Omega), \tag{23}$$

$$u\in H_m(\Omega;c^2\rho) \Longleftrightarrow (c^2\rho)^{-1/2}D^\alpha u\in L_2(\Omega), \quad |\alpha|\le m, \tag{24}$$

and

$$u\in X_i \Longleftrightarrow \rho^{-1/2}\nabla^j u\in L^2(\Omega) \quad (1\le j\le i). \tag{25}$$

These spaces are Hilbert spaces with respect to the following inner products respectively:

$$(u,v)=\int_\Omega u(x)\overline{v(x)}dx, \quad x\in\Omega, \tag{26}$$

$$(u,v)_m=\sum(D^\alpha u, D^\alpha v), \quad |\alpha|\le m, \tag{27}$$

$$\langle u,v\rangle=\int c^{-2}(x)\rho^{-1}(x)u(x)\overline{v(x)}dx, \quad x\in\Omega, \tag{28}$$

$$\langle u,v\rangle_m=\sum\langle D^\alpha u, D^\alpha v\rangle, \quad |\alpha|\le m, \tag{29}$$

and

$$[u,v]_i=\sum\frac{1}{\rho(x)}\nabla^j u(x)\overline{\nabla^j v(x)}dx, \quad x\in\Omega \text{ and } 1\le j\le i\le 2. \tag{30}$$

It is clear from the characteristics (9) of ρ in Assumption 2 and (12) of $c(x)$ in Assumption 3 that $H_0(\Omega)=L_2(\Omega)$, $L_2(\Omega;c^2\rho)=H_0(\Omega;c^2\rho)$, and $H_m(\Omega;c^2\rho)$ is equivalent to $H_m(\Omega)$ in the view of their norms.

Suppose that the boundary $\partial\Omega$ of Ω is sufficiently smooth, say C^2. Then Green's theorem implies that

$$\begin{aligned}\langle Au,v\rangle &= -\int\left[\nabla\cdot\frac{1}{\rho(x)}\nabla u(x)\right]v(x)dx\\ &=\int\frac{1}{\rho(x)}\nabla u(x)\cdot\overline{\nabla v(x)}dx\\ &=\int\left[\nu\cdot\frac{1}{\rho(x)}\nabla\rho u(x)\right]\overline{v(x)}dS(x), \quad x\in\partial\Omega,\end{aligned} \tag{31}$$

where ν is the unit outward normal of $\partial\Omega$ at x. This means that

$$-\langle Au,v\rangle+[u,v]_1=\int\frac{1}{\rho(x)}[\nu\cdot\nabla u(x)]\cdot v(x)dS(x), \quad x\in\partial\Omega. \tag{32}$$

Therefore, for $u\in X_2$, u satisfies the classical Neumann boundary condition $\nu\cdot\nabla u(t,x)=0$ if and only if u satisfies the relation (20), which in the new notation is

$$\langle Au,v\rangle=[u,v]_1, \quad v\in X_1\cap L_2(\Omega;c^2\rho). \tag{33}$$

Definition 1: A function $u\in X_2$ is said to satisfy the *generalized Neumann condition* if and only if (33) holds.

This definition does not require the assumption of the smoothness of the boundary $\partial\Omega$ of Ω. However, it defines the classical Neumann condition if $\partial\Omega$ is smooth.

Furthermore, a definition is introduced for a closed subspace in X_2:

$u \in \mathcal{H} \Leftrightarrow u \in X_2$ and satisfies (33). (34)

The construction of a solution of the initial-boundary-value problem is based on the linear operator A in H_2 given by (18) with domain

$$D(A) = \mathcal{H} \tag{35}$$

Theorem 1: A is a self-adjoint operator on the Hilbert space $L_2(\Omega; c^2\rho)$. Moreover, $A \geq 0$.

The verification of the assertions is based on the following result. *Lemma* 1: Let $\mathcal{H}$ be a Hilbert space and let $L: \mathcal{H} \to \mathcal{H}$ be a linear operator densely defined in $\mathcal{H}$. Assume that $L \subset L^*$, the adjoint of L, $L \geq 0$, and that the range $R(I+L)$ of $I+L$ is $\mathcal{H}$. Then L is self-adjoint.

Proof of Lemma 1: $L \geq 0$ indicates that the deficiency indices of the symmetric operator L and of L^*, an extension of L, are equal.[2] Condition $R(I+L) = \mathcal{H}$ implies the deficiency index is zero and hence L is self-adjoint.

Proof of Theorem 1: Verify the conditions of Lemma 1 for $\mathcal{H}$ defined by (34) and $L = A$. The space $C_0^\infty(\Omega)$ of infinitely-many-times continuously differentiable functions with compact support in Ω is a subset of $D(A) = \mathcal{H}$; hence $D(A)$ is dense in $L_2(\Omega; c^2\rho)$. Let u and v be any two elements of $D(A)$. Then (33) and (30) yield

$$\langle Au, v\rangle = [u, v]_1 = \int \frac{1}{\rho(x)} \nabla u(x) \cdot \overline{\nabla v(x)} dx = \langle u, Av\rangle. \tag{36}$$

Consequently, $D(A) \subset D(A^*)$ and $A^*u = Au$ for all $u \in D(A)$; that is, $A \subset A^*$.

The assertion $A \geq 0$, also one of the conditions to be checked, follows from the result yielded by (33) and (30) that for all $u \in D(A)$

$$\langle Au, u\rangle = [u, u]_1 = \int |\rho^{-1/2}(x) \nabla u(x)|^2 dx, \quad x \in \Omega. \tag{37}$$

The only condition left to be verified, $R(I+A) = L_2(\Omega; c^2\rho)$, means that for each f in $L_2(\Omega; c^2\rho)$ there exists an element u of $D(A)$ such that

$$\langle u, v\rangle + \langle Au, v\rangle = \langle f, v\rangle \quad \text{for all } v \text{ in } H_1(\Omega; c^2\rho). \tag{38}$$

This, together with (33), is equivalent to

$$\langle u, v\rangle + [u, v]_1 = \langle f, v\rangle \quad \text{for all } v \text{ in } H_1(\Omega; c^2\rho). \tag{39}$$

Sufficiently, if the equivalent inner product for $H_1(\Omega; c^2\rho)$ that we use is

$$\{u, v\} = \langle u, v\rangle + [u, v]_1, \quad u, v \in H_i\ (\Omega; c^2\rho), \tag{40}$$

we need to verify the existence of u such that

$$\{u, v\} = \langle f, v\rangle, \quad v \in H_1(\Omega; c^2\rho). \tag{41}$$

However,

$$|\langle f, v\rangle| \leq \langle f, f\rangle^{1/2} \langle v, v\rangle^{1/2} \leq \text{const}\{f, f\}^{1/2}\{v, v\}^{1/2}.$$

The Riesz representation theorem in the Hilbert space $(H_1(\Omega; c^2\rho), \{\cdot, \cdot\})$ yields the existence of an element u in $H_1(\Omega; c^2\rho)$ satisfying (41) and then (39). On the other hand, (39) implies (38) for all v in $C_0^\infty(\Omega)$. Thus $Au = f - u$ with the member on the right side in $L_2(\Omega, c^2\rho)$; hence Au is also in $L_2(\Omega, c^2\rho)$. Moreover, because the validity of (38) itself is implied when $c_0^\infty(\Omega)$ is dense in $H_1(\Omega, c^2\rho)$, the combination of (38) and (39) ensures that u satisfies the generalized Neumann condition (33). Thus $R(I+A) = \mathcal{H}$ is verified, and the proof of theorem 1 is complete.

Therefore, the Kato (Ref. 2, p. 331) second representation theorem ensures the following corollary:

Corollary: A has a nonnegative square root $A^{1/2}$ whose domain $D(A^{1/2}) = H_1(\Omega; c\rho^{1/2})$ has the inner product

$$\{u, v\} = \sum \int c^{-1}(x)\rho^{-1/2}(x) D^\alpha u(x) \overline{D^\alpha v(x)} dx, \quad x \in \Omega, \quad |\alpha| \leq 1. \tag{42}$$

Furthermore, $A^{1/2}$ satisfies the relation

$$\{A^{1/2}u, A^{1/2}u\} = \sum (\rho^{-1/2}(x) D_j u, \rho^{-1/2} D_j u), \quad 1 \leq j \leq n \tag{43}$$

From the results in Theorem 1 and Corollary 1, the argument of Wilcox,[1] with slight adjustment, gives the following theorem.

Theorem 2: For each f in $D(A)$ and g in $D(A^{1/2})$ there exists a uniquely defined strict solution u with finite energy of the initial-boundary-value problem (18), (19), and (33) with $t \in R$ such that

$$u \in C^2[R, L_2(\Omega, c^2\rho)] \cap C^1[R, H_1(\Omega, c\rho^{1/2})] \cap C(R, \mathcal{H}),$$

and u has the energy integral in the two equivalent forms

$$\begin{aligned} E(u, \Omega, t) &= \{D_t u(t), D_t u(t)\} + \sum \{D_j u(t), D_j u(t)\}, \\ &\qquad j = 1, \ldots, n, \\ &= \{D_t u(t), D_t u(t)\} + \{A^{1/2}u(t), A^{1/2}u(t)\}, \end{aligned} \tag{44}$$

and has the constancy of energy

$$E(u, \Omega, t) = \sum \{D_j f, D_j f\} + \{g, g\}. \tag{45}$$

From the spectral theorem for A and the associated operator calculus, we have the following theorem.

Theorem 3: For real-valued functions f in $L_2(\Omega)$ and g in $D(A^{-1/2})$, define

$$h = f + iA^{-1/2}g \in L_2(\Omega). \tag{46}$$

Then the solution in $L_2(\Omega)$ defined by

$$u(t) = (\cos tA^{1/2})f + (A^{-1/2} \sin tA^{1/2})g \tag{47}$$

with bounded coefficient operators, satisfies

$$U(t, x) = \text{Re} v(t, x), \tag{48}$$

where $v(t, x)$ is the complex valued solution in $L_2(\Omega)$ defined by

$$v(t, 0) = \exp(-itA^{1/2})h. \tag{49}$$

Because the operator A is nonnegative, its spectrum $\sigma(A)$ is contained in the interval $[0, \infty)$. The nonexistence of the positive eigenvalue is ensured by the following theorem.

Theorem 4 (*Mochizuki*[3]): Assume Assumptions 2–4. Then, $u = 0$ in Ω is the only $L_2(\Omega)$ solution of the equation

$$-\nabla \cdot \frac{1}{\rho(x)} \nabla u - \lambda c^{-2}(x)\rho^{-1}(x) u = 0, \quad x \in \Omega, \ \lambda > 0. \tag{50}$$

Proof: It suffices to check the conditions imposed by Mochizuki. His first three conditions are presented by Assumptions 2—4 by setting $a_{jk}(x)=\delta_{jk}\rho^{-1}(x)$, $b_j=0$, $p=0$, and $q(x)=\lambda c^{-2}(x)\rho^{-1}(x)$. Particularly, if $n\leq 3$, the "Stummel condition"

$$\sup\int |q(y)|^2 dy<\infty,\ y\in\Omega:\ |x-y|<1,\ x\in\Omega, \tag{51}$$

is a consequence of the boundedness and the smoothness of $c(x)$ and $\rho(x)$ on Ω. His fourth condition is given by his Remark 1.2 and by (9), (11), and (14). The smoothness assumption on $c(x)$ and $\rho(x)$ implies the Hölder condition (cf. Friedman Ref. 4, p. 23). This condition gives the unique continuation property in his last condition according to his Remark 1.1, which is referred to in Refs. 5 and 6.

Under the conditions just checked, stronger result on the growth of solutions has been concluded by Mochizuki.[3] It will be recalled here because of the need in the principle of limiting absorption in Sec. 4. Denote by $S(t)$ the sphere with radius t: $S(t)=\{x|\ |x|=t\}$ $(t>0)$.

Theorem 5 (Mochizuki[3]): Assume Assumptions 2—4. If u is a not identically vanishing solution of Eq. (49), then for any $a>0$,

$$\liminf_{t\to\infty} t^a\int |\lambda c^{-2}(x)\rho^{-1}(x)|\ |u(x)|^2 + |x\circ\rho^{-1}(x)\nabla u(x)|^2 dS=\infty,\quad x\in S(t). \tag{52}$$

The employment of the standard principle of limiting absorption yields the absolute continuity of the continuous spectrum of A. The principle will be discussed in Sec. 4 and there also will be provided a proof for the following statement.

Theorem 6: The resolution $E(s)$ of the identity for the operator A is absolutely continuous on any closed interval in $(0,\infty)$.

Remark: Whether $\lambda=0$ is an eigenvalue or belongs to the continuous spectrum is not yet clear.

To be an important preparation for the discussion of the principle of limiting absorption, the argument used in the proof of Theorem 1 will be employed again to prove the local coercivity of the operator A. More precisely, this argument is utilized to the operator $A_t=A+t$ on Ω_t with the same kind of boundary condition on the boundary of Ω_t. Here $\Omega_t=\{x\Omega|\ |x|<t\}$. Denote by $\langle u,v\rangle_{m,t}$ the inner product $\langle u,v\rangle_m$ with integral considered on Ω_t on $(u,v)_t$ for (u,v). Then, the local coercivity is the following.

Theorem 7: For a local $H_2(\Omega_t;\ c^2\rho)$ function satisfying the boundary condition (33) with sufficiently large $t>0$, there exist constants $d(t)$ such that

$$\langle u,u\rangle_{1,t}^{1/2}\leq d(t)\{(u,u)_{t+1}^{1/2}+(Au,Au)_{t+1}^{1/2}\}. \tag{53}$$

III. DISCUSSION OF THE MØLLER WAVE OPERATORS OPERATORS

As indicated in the Introduction, all properties of A studied in the last section hold for A_0. Moreover, the spectrum of A_0 is $[0,\ \infty]$ and is absolutely continuous. These are results of Wilcox.[1] These results and those of the last section ensure the existence of the unitary groups $\exp(-iL_0t)$ and $\exp(-iLt)$, $-\infty<t<\infty$, associated with the self-adjoint operators $L_0=A_0^{1/2}$ and $L=A^{1/2}$.

The strong limits

$$W_\pm=\text{s-lim}\ \exp(iLt)\exp(-iL_0t),\quad t\to\pm\infty \tag{54}$$

are called the Møller wave operators. This yields

$$\lim\|\exp(-iLt)h-\exp(-iL_0t)h_\pm\|=0,\quad t\to\pm\infty, \tag{55}$$

and

$$h=W_\pm h_\pm. \tag{56}$$

The following map is the scattering operator:

$$S:f_-\to f_+=Sf_-. \tag{57}$$

It is required to be unitary on $L_2(\Omega)$, which is an easy consequence of the orthogonality (16) and the completeness (17) of the Møller wave operators.

The existence of the strong limit (54) is proved by employing the vanishing, when $t\to\pm\infty$, of the local energy of the solution to the unperturbed system and employing the decaying in rate $-1-\delta$ of $|t|$ for the derivatives of the first two orders of the solution to the unperturbed system.

The coincidence of the Møller wave operator $W_\pm$ with the stationary wave operators $U_\pm$ and the properties of the orthogonality and the completeness for $U_\pm$ guarantee the corresponding properties for $W_\pm$. The proof of the coincidence and of these two properties for $U_\pm$ is based on the expansion principle of the generalized eigenfunctions of A, which has been employed by a dozen different authors, including Ikebe,[7] Mochizuki,[8,9] and Wilcox.[1] This principle will be given in the follow-up article. A portion of the article involves the principle of limiting absorption, which will be studied in the next section.

IV. PRINCIPLE OF LIMITING ABSORPTION

There is no doubt about the role played by the principle of limiting absorption in the eigenfunction expansion, which is a keystone for scattering theory. However, in the three-dimensional inverse scattering theory for the Schrödinger equation, both Faddeev[10] and Newton[11] studied the "Volterra" integral equation, where the principle of limiting absorption again played an essential role. We will apply the principle of limiting absorption to the acoustic wave equation. This should bring some light to the study of the inverse problem of the scattering theory for the acoustic wave in three dimensions, which will extend results of Ref. 12. At the end of this section, there is given the proof of Theorem 4 on the absolute continuity of spectrum. The argument of this section follows that of Mochizuki.[13]

For a real number a, let $L_{2,a}(\Omega)$, $L_{2,a}(\Omega;\ c^2\rho)$, $H_{m,a}(\Omega)$, $H_{m,a}(\Omega;\ c^2\rho)$, and $X_{t,a}$, be the spaces consisting the functions u that $(1+|x|)^a u(x)$ belong to the respective spaces $L_2(\Omega)$, $L_2(\Omega;\ c^2\rho)$, $H_m(\Omega)$, $H_m(\Omega;\ c^2\rho)$, and X_t. In the rest, a is a positive number. By the principle of limiting absorption here, we mean that for a positive (negative) number μ and a set of complex numbers K,

with positive ImK and nonnegative (nonpositive) real part ReK, imaginary parts having μ as the limit, and for the set of functions $(A-K^2)^{-1}f$ for each f in $L_{2,1+a}(\Omega;c^2\rho)$, their limits exist and are in the space $L_{2,-1-a}(\Omega;c^2\rho)$, subject to the following uniqueness condition—the radiation condition.

Definition: A solution $u(x)$ in $L_{2,-1-a}(\Omega;c^2\rho)$ of the equation

$$Au(x)=K^2u(x)+f(x) \quad \text{in } \Omega,\ f\in L_{2,1+a}(\Omega;c^2\rho), \tag{58}$$

[satisfying the boundary condition (33)] is said to satisfy the incoming (outgoing) radiation condition if and only if

$$(\nabla - iK\tilde{x})u(x)\in (H_{-1+a})^n \quad (\tilde{x}=x/|x|). \tag{59}$$

Then, upon a very delicate inequality analysis Mochizuki[3] proved the powerful inequality

$$((\nabla - iK\tilde{x})u, \overline{(\nabla - iK\tilde{x})u})_{-1+a}$$

$$\leq C\{(1+|K|^2)(u,u)_{-1-a}+(f,f)_{1+a}\}, \tag{60}$$

where $C>0$ is a constant independent of K, f, and u.

The principle of limiting absorption consists of two parts—uniqueness and existence. We take care of uniqueness part first and then the existence part.

Theorem 8: Equation (58) has at most one outgoing (incoming) solution.

Proof: The following holds:

$$2\mathrm{Im}K|\mathrm{Re}K|(u,u)\leq((1+|x|)^{1+a}f,f)^{1/2}((1+|x|)^{-1-a}u,u)^{1/2}; \tag{61}$$

$$\liminf_{t\to\infty} t^a\int\{\rho^{-1}(x)\nabla u(x)\cdot\overline{\nabla u(x)}$$

$$+\mu^2|u(x)|^2\}\,dS=0 \quad (x\in S_t). \tag{62}$$

They are inequality (3.7) and the last inequality in the proof of Theorem 3.1 in Ref. 13, p. 42. There the boundary condition is different from (33). But both arguments fit our problem because the imaginary part of the right side of (33) with $v=u$ vanishes, which is the only adjustment necessary.

Now Eq. (61) implies the assertion of this theorem for the case Im$K>0$. We suppose $K=\mu$ is real. Then, (62) and (52) contradict each other. The proof is then complete.

Next is the statement of the existence part of the principle; its proof is outlined afterward.

Theorem 9: Let μ be any positive (negative) number and let $K_\pm$ be a set of complex numbers K, with Im$K>0$ and Re$K\geq 0$ (Re$K\leq 0$), having μ as its limit. Let $u(x,K)=(A-K^2)^{-1}f$ for a given f in $L_{2,1+a}(\Omega,c^2\rho)$. Then, the $L_{2,-1-a}(\Omega;c^2\rho)$ limit of $u(x,K)$ exists as $u(x,\mu)$ and belongs to $L_{2,-1-a}(\Omega;c^2\rho)$. The function $u(x,\mu)$ is a unique outgoing (incoming) solution of (58) with $K=\mu$.

Outline of Proof: Let K_n be a sequence in $K_\pm$ convergent to μ. Let f_n be a bounded sequence in $L_{2,1+a}(\Omega;c^2\rho)$. Then, by A being self-adjoint, $u_n=(A-K_n^2)^{-1}f_n$ exists and is a solution of (58) with $K=K_n$ and $f=f_n$ in $L_2(\Omega;c^2\rho)$. Moreover, u_n satisfies the outgoing (incoming) radiation condition. Next, by applying the inequality for (61), Theorem 7 and the Rellich compactness criterion, we have that $\{u_n\}$ is precompact in $L_{2,-1-a}(\Omega;c^2\rho)$ if and only if $\{u_n\}$ is bounded in the same space.

This and inequality (60) give the claim that if $\{u_n\}$ converges in $L_{2,-1-a}(\Omega;c^2\rho)$ to u, then u satisfies the radiation condition with $K=\mu$. Moreover, upon this claim and Theorem 8 by building up a contradiction, there is a constant C depending only on $K_\pm$ that

$$\sup(u(x,K),u(x,K))^{1/2}_{-1-a}\leq C(f,f)^{1/2}_{1+a} \quad (K\in K_\pm), \tag{63}$$

where $u(x,K)=(A-K^2)^{-1}f$.

Thus far, $\{u(x,K_n)\}$ is compact in $L_{2,-1-\epsilon}(\Omega;c^2\rho)$. And then, there exists a subsequence $\{u(x,K_n)\}$ which converges to an outgoing (incoming) solution $u(x,\mu)$ of (58) with $K=\mu$. Finally, the previous uniqueness theorem yields the assertion and the theorem is proved.

This proof indicates that for a given K as above and f in $L_{2,1+a}(\Omega;c^2\rho)$, there exists a unique outgoing (incoming) solution of (58). Then, treating $R(K^2)=(A-K^2)^{-1}$, with Im$K>0$, as the operator from $L_{2,1+a}(\Omega;c^2\rho)$ to $L_{2,-1-a}(\Omega;c^2\rho)$, we have as a consequence of the proof the following.

Theorem 10: The operator $R(K^2)$ is bounded and depends continuously on K.

Finally, we find that the absolute continuity of the continuous spectrum is a corollary of this last theorem, which is spelled out more precisely as follows and has the same proof as Mochizuki (Ref. 13, Theorem 3.4).

Theorem 11: The continuous spectrum of the self-adjoint operator A is absolutely continuous with respect to the Lebesque measure.

ACKNOWLEDGMENT

For his kindness in giving K. H. Chen his preprints and his lecture notes, we are indebted to Professor K. Mochizuki.

[1] C. H. Wilcox, *Scattering Theory for the d'Alembert Equation in Exterior Domains*, Lecture Notes, Math 442 (Springer, New York, 1975).
[2] T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1966), p. 268.
[3] K. Mochizuki, "Growth Properties of Solutions of Second Order Elliptic Differential Equations," to appear in J. Math., Kyoto Univ.
[4] A. Friedman, *Partial Differential Equations* (Holt, Rinehart, and Winston, New York, 1969).
[5] E. M. Landis, "On Some Properties of Solutions of Elliptic Equations," Dokl. Akad. Nauk SSSR 107, 640—43 (1956) (Russian).
[6] M. H. Protter, "Unique Continuation for Elliptic Equations," Trans. Amer. Math. Soc. 95, 81—91 (1960).
[7] T. Ikebe, "Scattering for the Schrodinger Operator in an Exterior Domain," J. Math. Kyoto Univ. 7, 93—112 (1967).

[8]K. Mochizuki, "Spectral and Scattering Theory for Symmetric Hyperbolic Systems in an Exterior Domain," Publ. RIMS, Kyoto Univ., Ser. A 5, 219–58 (1969).
[9]K. Mochizuki, "Scattering Theory for Wave Equations with Dissipative Terms," to appear in Proc. Jpn. Acad.
[10]L. D. Faddeev, "Three-Dimensional Inverse Problem in the Quantum Theory of Scattering," Preprint ITP-7k-106E (Kiev, 1971).
[11]R. G. Newton, "The Gel'fand–Levitan Method in the Inverse Scattering Problem," in *Scattering Theory in Mathematical Physics*, Proc. NATO Adv. Study Inst. (Reidel, Boston, 1974), pp. 193–235.
[12]H. H. Szu, C. E. Carroll, C. C. Yang, and S. Ahn, "A new functional equation in the plasma inverse problem and its analytic properties," J. Math. Phys. 17, 1236 (1976).
[13]K. Mochizuki, lecture notes on Spectral and Scattering Theory for Second Order Elliptic Differential Operators in an Exterior Domain, Seminar in Differential Operators in an Exterior Domain, Seminar in Differential Equations, University of Utah, 1972.

The use of the symmetric group in the construction of multispinor Lagrangians[a)]

M. L. Larsen[b)] and W. W. Repko

Department of Physics, Michigan State University, East Lansing, Michigan 48824
(Received 18 July 1977)

The construction of Lagrange functions for nth rank multispinor (Bargmann–Wigner) fields is developed using the symmetric group S_n. Restrictions on the number of fields present in the Lagrangian and their couplings to one another are obtained by constructing differential operators which transform irreducibly under S_n. The technique is illustrated by a brief discussion of the second and third rank cases and a more detailed fourth rank example. The Lagrangians thus obtained are precisely those specified by Lorentz invariance, but the method used greatly facilitates their construction.

1. INTRODUCTION

Since the introduction of the Dirac formulation for the spin-$\frac{1}{2}$ field,[1] a variety of higher spin formulations has been developed.[2–5] Most of these have been applied in situations involving relatively low spin where problems associated with the construction of Lagrange functions are few, if any. In this paper, we consider the problem of constructing Lagrange functions for Bargmann–Wigner fields[4] of arbitrary spin. As originally introduced, a Bargmann–Wigner field of spin s is represented by a totally symmetrical multispinor $\psi_{\alpha_1\alpha_2\cdots\alpha_{2s}}$, $1 \leq \alpha_i \leq 4$, which satisfies the equations[6]

$$\begin{aligned}(\gamma\cdot\partial)_{\alpha_1\alpha_1'}\psi_{\alpha_1'\alpha_2\cdots\alpha_{2s}} &= -m\psi_{\alpha_1\alpha_2\cdots\alpha_{2s}}\\ &\vdots \qquad\qquad \vdots\\ (\gamma\cdot\partial)_{\alpha_{2s}\alpha_{2s}'}\psi_{\alpha_1\alpha_2\cdots\alpha_{2s}'} &= -m\psi_{\alpha_1\alpha_2\cdots\alpha_{2s}}.\end{aligned} \tag{1}$$

These equations can also be applied to multispinors whose symmetry under the operation of permuting indices is more complicated, and the result is again a field with a definite spin.[7]

The problem of determining a Lagrange function which depends on the fields and their first derivatives and leads to Eqs. (1) is easily solved for the cases $s=1,2$. The former is nothing but the well-known Lagrange function for the Dirac field, while the latter can be constructed by analogy as

$$L = -\bar{\psi}\left[\frac{(\gamma\cdot\partial)_1 + (\gamma\cdot\partial)_2}{2} + m\right]\psi, \tag{2}$$

where the subscripts denote the operations

$$\begin{aligned}[(\gamma\cdot\partial)_1\psi]_{\alpha\beta} &= (\gamma\cdot\partial)_{\alpha\alpha'}\psi_{\alpha'\beta},\\ [(\gamma\cdot\partial)_2\psi]_{\alpha\beta} &= (\gamma\cdot\partial)_{\beta\beta'}\psi_{\alpha\beta'}, \quad \text{etc.}\end{aligned} \tag{3}$$

In order to obtain the Bargmann–Wigner equations for fields with $s>2$, it is necessary to introduce auxiliary fields into the Lagrange function.[8,9] These fields are introduced in such a way that the Bargmann–Wigner equations result for the primary fields while, at the same time, the field equations for the auxiliary fields imply that they vanish. The Lagrange function for spin $\frac{3}{2}$ can be written in the form[8,9]

$$\begin{aligned}L = &-\bar{\psi}[(\gamma\cdot\partial)_1 + m]\psi + \tfrac{2}{3}\bar{\chi}[(\gamma\cdot\partial)_1 - (\gamma\cdot\partial)_3 + 3m]\chi\\ &+\bar{\Omega}[(\gamma\cdot\partial)_3 - m]\Omega - (1/\sqrt{3})[\bar{\chi}(\gamma\cdot\partial)_1\psi + \bar{\psi}(\gamma\cdot\partial)_1\chi]\\ &+\tfrac{1}{2}[\bar{\chi}(\gamma\cdot\partial)_3\Omega + \bar{\Omega}(\gamma\cdot\partial)_3\chi].\end{aligned} \tag{4}$$

Here, $\psi_{(\alpha\beta\gamma)}$ is totally symmetrical in its three indices, $\Omega_{[\alpha\beta\gamma]}$ is totally antisymmetrical, and $\chi_{[\alpha\beta]\gamma}$ is a field of mixed symmetry such that[10]

$$\chi_{[\alpha\beta]\gamma} = -\chi_{[\beta\alpha]\gamma}, \tag{5a}$$

$$\chi_{[\alpha\beta]\gamma} + \chi_{[\beta\gamma]\alpha} + \chi_{[\gamma\alpha]\beta} = 0. \tag{5b}$$

The choice of auxiliary fields in the third rank case is not exceedingly difficult because the options for various symmetries are limited.[11] For higher spin cases, the symmetry type, number, and couplings of the auxiliary fields are no longer so obvious. Thus, we seek a systematic way to determine the properties of the auxiliary fields in order to produce a Lagrange function such that the Bargmann-Wigner equations hold for the desired field and the other fields vanish.

The requirement that the Lagrangian be invariant under the symmetric group fixes the nature of the auxiliary fields and their couplings. Only multispinors which transform irreducibly under the symmetric group correspond to fields of definite spin when the Bargmann–Wigner equations are applied to them. Given this, it is natural to ask whether the kinetic operators $(\gamma\cdot\partial)_i$, $i=1,2,\ldots,n$, can be arranged to have definite symmetric group transformation properties. Once it is established that this is possible, the Clebsch–Gordan series for S_n provides a simple way of determining the independent couplings. The number of such couplings is the same as that obtained by any "trial and error" method, the advantage being that a simple character table calculation replaces the more usual tedious manipulation of spinor indices.

In the next section we develop the relationship between the symmetric group and multispinor Lagrange functions. Section 3 contains an application of this development to the fields of spin 1, $\frac{3}{2}$, and the spin-0 part of a fourth-rank multispinor. We conclude with some discussion about the general case.

2. THE SYMMETRIC GROUP AND MULTISPINOR LAGRANGIANS

The symmetric group S_n is the group of $n!$ permu-

[a)]Work supported in part by the National Science Foundation.
[b)]Present address: Towson State University, Baltimore, MD 21204.

 0022-2488/78/1905-0930$1.00

tations on n objects. Methods exist for finding the number of irreducible representations and their characters and are described in many standard texts.[12]

We assert that the multispinor Lagrangian is invariant under symmetric group transformations. The kinetic terms are of the bilinear form $\bar{F}\hat{O}G$, where F and G are fields which, without loss of generality, may be chosen to transform as irreducible representations of the symmetric group appropriate for the spinor rank. The operator $\hat{O}$ is a matrix of the dimension required to connect the representations F and G and has elements that are linear combinations of $(\gamma\cdot\partial)_i$.

We first seek a convenient representation for the elements of $\hat{O}$. Derivative operators which transform under the symmetric group may be constructed from the projection operator

$$P^{(m)}_{ik}=\frac{d_m}{g}\sum_s [\Gamma^{(m)}(s)]_{ik}P(s). \tag{6}$$

Here d_m is the order of the representation, g is the order of the group, $\Gamma(s)$ if the matrix representation of the permutation s, and P is the particular transformation corresponding to this permutation. If this operator acts on an arbitrary function F, the result is $F^{(m)}$, a function belonging to the irreducible representation m, provided the result is not zero.

For multispinors of rank n there are clearly n linearly independent matrices $(\gamma\circ\partial)_i$. We apply the operator in Eq. (6) to $(\gamma\circ\partial)_1$ and generate a single operator from the identity representation

$$\alpha=\frac{1}{n}\sum_{i=1}^{n}(\gamma\circ\partial)_i, \tag{7}$$

and $(n-1)$ other operators β_k from the $(n-1,1)$ representation[12]

$$\beta_k=[1/\sqrt{n(n-1)}]\sum_{i=1}^{k}(\gamma\cdot\partial)_i-k(\gamma\circ\partial)_{k+1}/\sqrt{k(k+1)},$$
$$k=1,2,\ldots,(n-1). \tag{8}$$

The operators β_k then form a one column matrix that transforms as the $(n-1,1)$ representation of S_n. Since α and the β_k exhaust all the linear combinations of $(\gamma\circ\partial)_i$, the elements of the operators $\hat{O}$ may all be expressed as linear combinations of α and β_k. Thus, the kinetic operators as well as the fields themselves may be chosen to have definite transformation properties under the symmetric group.

Application of symmetric group invariance then provides a convenient way of enumerating the various couplings. This follows from the fact that the coupling $\bar{F}\hat{O}G$ represents a Kronecker product of three representations and must itself transform as the identity. If (λ) is the representation to which F belongs and (μ) is the representation to which G belongs, then from a general result of group theory the Clebsch–Gordan series

$$(\lambda)\times(\mu)=\sum_j a_j(j) \tag{9}$$

must contain either the representation (n) or $(n-1,1)$ for a nonvanishing coupling $\bar{F}\hat{O}G$ to occur. If the series contains (n), then an α-type coupling is possible:

$$\bar{F}\alpha G;$$

if $(n-1,1)$, then the coupling contains a matrix $\hat{O}$ with elements that are linear combinations of the β_k.

The components of $\hat{O}$ may be constructed as follows. For a coupling $\bar{F}\hat{O}G$ with F belonging to the (λ) representation and G belonging to the (μ) representation, the js-component of $\hat{O}$ may be computed from

$$[\hat{O}(\lambda,\mu)]_{js}\equiv\sum_k \beta_k\langle\lambda j,(n-1)k\,|\,\mu s\rangle, \tag{10}$$

where for brevity $(n-1)$ denotes the $(n-1,1)$ representation. $\langle\lambda j,(n-1)k\,|\,\mu s\rangle$ is the Clebsch–Gordan coefficient connecting the (λ) and the $(n-1,1)$ representations to the (μ) representation.

The requirement that all fields in the Lagrangian transform as irreducible representations of the symmetric group assures a nonredundant and exhaustive choice of fields. Moreover, the symmetry of $[\hat{O}G]$ under a permutation of its multispinor indices is the same as that of $\bar{F}$ under the same permutation. As a consequence, $\bar{F}$ may be varied as though all of its components were independent. Thus, use of symmetric group invariance not only facilitates the writing of the Lagrangian, but also simplifies the process of obtaining the field equations.

3. APPLICATIONS

A. Second rank multispinor

The multispinor field with two indices has two symmetries corresponding to the irreducible representations of S_2: (2) and (1^2). The symmetric field is denoted $\psi_{(\alpha\beta)}$ and must satisfy the Bargmann–Wigner equations

$$(\gamma\circ\partial)_{\alpha_1\alpha_1'}\psi_{\alpha_1'\alpha_2}=-m\psi_{\alpha_1\alpha_2},$$
$$(\gamma\circ\partial)_{\alpha_2\alpha_2'}\psi_{\alpha_1\alpha_2'}=-m\psi_{\alpha_1\alpha_2}, \tag{11}$$

if ψ is to represent a spin-1 particle. The totally antisymmetric field is designated $\Omega_{[\alpha\beta]}$.

The totally symmetric operator α is

$$\alpha=\tfrac{1}{2}[(\gamma\cdot\partial)_1+(\gamma\cdot\partial)_2], \tag{12}$$

and β is

$$\beta=\tfrac{1}{2}[(\gamma\circ\partial)_1-(\gamma\circ\partial)_2]. \tag{13}$$

The most general Lagrangian takes the form

$$L=-\bar{\psi}(\alpha+m)\psi+a[\bar{\psi}\beta\Omega+\bar{\Omega}\beta\psi]+b\bar{\Omega}\alpha\Omega+cm\bar{\Omega}\Omega. \tag{14}$$

Variation of the Lagrangian gives the field equations

$$-(\alpha+m)\psi+a\beta\Omega=0,$$
$$b\alpha\Omega+cm\Omega+a\beta\psi=0. \tag{15}$$

It is easy to check that ψ satisfies the Bargmann–Wigner equations and Ω vanishes if and only if

$$a=b=c=0. \tag{16}$$

In this simple case, the formalism shows that the auxiliary field Ω is unnecessary, in agreement with Eq. (2).

B. Third rank multispinor

There are three irreducible representations in S_3 corresponding to the partitions of 3: (3), (2,1), (1^3), or $(\bar{3})$, the representation (2,1) being two dimensional. We thus require three multispinor fields in the Lagrangian with symmetries as indicated below.

$$\psi_{(\alpha\beta\gamma)},\ \Xi_{\alpha\beta\gamma}=\begin{pmatrix}\chi_{(\alpha\beta)\gamma}\\ \chi_{[\alpha\beta]\gamma}\end{pmatrix},\ \Omega_{[\alpha\beta\gamma]}.$$

The derivative operators α and β_k are obtained from the projection operator, Eq. (6), namely

$$\alpha=\tfrac{1}{3}[(\gamma\cdot\partial)_1+(\gamma\cdot\partial)_2+(\gamma\cdot\partial)_3], \tag{17}$$

$$\beta_1=(1/2\sqrt{3})[(\gamma\cdot\partial)_1-(\gamma\cdot\partial)_2], \tag{18}$$

$$\beta_2=\tfrac{1}{6}[(\gamma\cdot\partial)_1+(\gamma\cdot\partial)_2-2(\gamma\cdot\partial)_3]. \tag{19}$$

By inspection of the Clebsch–Gordan series connecting the respective representations, the following couplings are seen to be symmetric group invariant:

$$\begin{array}{lll}\bar{\psi}\alpha\psi & \Xi\hat{O}(2,3)\psi & \Xi\hat{O}(2,\bar{3})\Omega\\ \bar{\psi}\hat{O}(3,2)\Xi & \bar{\Omega}\alpha\Omega & \Xi\alpha\Xi\ .\end{array}$$

The notation for the $\hat{O}$ matrices indicates the representations connected by them. The (2,1) representation is designated by 2 for brevity. Once the Clebsch–Gordan coefficients have been calculated, formula (10) may be used to obtain the following expressions:

$$\hat{O}(2,2)=\begin{pmatrix}-\beta_1 & \beta_2\\ \beta_2 & \beta_1\end{pmatrix},$$

$$\hat{O}(3,2)=(\beta_2,\beta_1), \tag{20}$$

$$\hat{O}(\bar{3},2)=(\beta_1,-\beta_2).$$

An appropriate choice of constants weighting each coupling in the Lagrangian leads to field equations which imply the Bargmann–Wigner equations on $\psi_{(\alpha\beta\gamma)}$ and the vanishing of $\Xi_{\alpha\beta\gamma}$ and $\Omega_{[\alpha\beta\gamma]}$. The Lagrangian thus constructed is[8,9]

$$\begin{aligned}\mathcal{L}=&-\bar{\psi}(\alpha+m)\psi+\tfrac{1}{2}\bar{\Xi}\hat{O}(2,2)\Xi-\bar{\Omega}(\alpha+m)\Omega+m\bar{\Xi}\Xi\\&+\tfrac{1}{2}[\bar{\psi}\hat{O}(3,2)\Xi+\bar{\Xi}\hat{O}(2,3)\psi]+\tfrac{1}{2}[\bar{\Xi}\hat{O}(2,\bar{3})\Omega\\&+\bar{\Omega}\hat{O}(\bar{3},2)\Xi].\end{aligned} \tag{21}$$

C. Fourth rank multispinor

The partitions of 4: (4), (3,1), (2^2), $(2,1^2)$, and (1^4) correspond to the irreducible representations of S_4. It is therefore possible to construct fields with the following symmetries:

(4) $\psi_{(\alpha\beta\gamma\delta)}$,

$$(3,1)\quad \Xi_{\alpha\beta\gamma\delta}=\begin{bmatrix}\chi'_{(\alpha\beta\gamma)\delta}\\ \chi''_{(\alpha\beta)\gamma\delta}\\ \chi'''_{[\alpha\beta](\gamma\delta)}\end{bmatrix},$$

$$(2^2)\quad \phi_{\alpha\beta\gamma\delta}=\begin{bmatrix}\phi'_{(\alpha\beta)(\gamma\delta)}\\ \phi''_{[\alpha\beta][\gamma\delta]}\end{bmatrix},$$

$$(2,1^2)\quad \zeta_{\alpha\beta\gamma\delta}=\begin{bmatrix}\zeta'_{(\alpha\beta)[\gamma\delta]}\\ \zeta''_{[\alpha\beta](\gamma\delta)}\\ \zeta'''_{[\alpha\beta\gamma]\delta}\end{bmatrix},$$

(1^4) $\Omega_{[\alpha\beta\gamma\delta]}$.

These functions fulfill the following cyclic relations over four indices:

$$\begin{aligned}&\chi_{\alpha\beta\gamma\delta}+\chi_{\delta\alpha\beta\gamma}+\chi_{\gamma\delta\alpha\beta}+\chi_{\beta\gamma\delta\alpha}=0,\\&\zeta_{\alpha\beta\gamma\delta}-\zeta_{\delta\alpha\beta\gamma}+\zeta_{\gamma\delta\alpha\beta}-\zeta_{\beta\gamma\delta\alpha}=0,\\&\phi_{\alpha\beta\gamma\delta}-\phi_{\delta\alpha\beta\gamma}-\phi_{\gamma\delta\alpha\beta}+\phi_{\beta\gamma\delta\alpha}=0.\end{aligned} \tag{22}$$

ψ transforms under the identity representation and Ω under the alternating representation. The derivative operators obtained as before are

$$\begin{aligned}&\alpha=\tfrac{1}{4}[(\gamma\cdot\partial)_1+(\gamma\cdot\partial)_2+(\gamma\cdot\partial)_3+(\gamma\cdot\partial)_4],\\&\beta_1=(1/2\sqrt{6})[(\gamma\cdot\partial)_1-(\gamma\cdot\partial)_2],\\&\beta_2=(1/6\sqrt{2})[(\gamma\cdot\partial)_1+(\gamma\cdot\partial)_2-2(\gamma\cdot\partial)_3],\\&\beta_3=\tfrac{1}{12}[(\gamma\cdot\partial)_1+(\gamma\cdot\partial)_2+(\gamma\cdot\partial)_3-3(\gamma\cdot\partial)_4].\end{aligned} \tag{23}$$

For illustrative purposes we choose the simplest nontrivial case, that of spin 0, and establish that the field ϕ defined above has only one positive energy component if it satisfies the Klein–Gordon equation and

$$(\alpha+m)\phi=0. \tag{24}$$

Consider the element $\phi_{[\alpha\beta][\gamma\delta]}$. In the rest frame Eq. (24) becomes

$$-m[\gamma_4^{(1)}+\gamma_4^{(2)}+\gamma_4^{(3)}+\gamma_4^{(4)}]_{\alpha\beta\gamma\delta,\alpha'\beta'\gamma'\delta'}\phi_{\alpha'\beta'\gamma'\delta'}=-m\phi_{\alpha\beta\gamma\delta}. \tag{25}$$

In standard representation the elements of the diagonal matrix γ_4 are $(1,1,-1,-1)$ so that ϕ vanishes unless its indices correspond to the first or second rows of γ_4. But the antisymmetry of the indices leaves only one independent solution: $\phi_{[12][12]}$. For negative energies only $\phi_{[34][34]}$ survives. Thus the field is a spin-0 object. For auxiliary fields only the fields with lower symmetry are selected: ζ and Ω.

The couplings among these fields as determined for the previous cases are

$$\begin{array}{ll}\bar{\phi}\alpha\phi & \bar{\zeta}\hat{O}(\bar{3},\bar{4})\Omega\\ \bar{\phi}\hat{O}(2,\bar{3})\zeta & \bar{\Omega}\hat{O}(\bar{4},\bar{3})\zeta\\ \bar{\zeta}\hat{O}(\bar{3},2)\phi & \bar{\Omega}\alpha\Omega\\ \bar{\zeta}\alpha\zeta & \bar{\zeta}\hat{O}(\bar{3},\bar{3})\zeta\end{array}$$

As before, the argument of $\hat{O}$ is an abbreviation for the two representations connected by $\hat{O}$. Upon closer examination, the coupling $\bar{\Omega}\alpha\Omega$ is found to vanish as a consequence of the total antisymmetry of Ω. The $\hat{O}$ operators are found to be[13]

$$\hat{O}(\bar{3},2)=\frac{1}{\sqrt{3}}\begin{bmatrix}-\sqrt{2}\beta_3+\beta_2 & -\beta_1\\ -\beta_1 & -\sqrt{2}\beta_3-\beta_2\\ -\sqrt{2}\beta_1 & \sqrt{2}\beta_2\end{bmatrix},$$

$$\hat{O}(\bar{3},\bar{3})=\frac{1}{\sqrt{3}}\begin{bmatrix}\beta_3+\sqrt{2}\beta_2 & -\sqrt{2}\beta_1 & \beta_1\\ -\sqrt{2}\beta_1 & \beta_3-\sqrt{2}\beta_2 & -\beta_2\\ \beta_1 & -\beta_2 & -2\beta_3\end{bmatrix}, \tag{26}$$

$$\hat{O}(\bar{4},\bar{3})=\sqrt{2/3}\,(-\beta_1,\beta_2,-\beta_3),$$

and, in terms of these operators, the Lagrangian is

$$\mathcal{L} = -\bar{\phi}(\alpha+m)\phi + (1/\sqrt{3})[\bar{\phi}\hat{O}(2,\bar{3})\zeta + \bar{\zeta}\hat{O}(\bar{3},2)\phi] - (2/\sqrt{6})[\bar{\Omega}\hat{O}(\bar{4},\bar{3})\zeta + \bar{\zeta}\hat{O}(\bar{3},\bar{4})\Omega] - \tfrac{2}{9}\bar{\zeta}(\alpha-m)\zeta + m\bar{\Omega}\Omega. \quad (27)$$

We now show that this leads to the desired field equations.

Variation of Eq. (27) leads to the field equations

$$-(\alpha+m)\phi + (1/\sqrt{3})\hat{O}(2,\bar{3})\zeta = 0, \quad (28a)$$

$$(1/\sqrt{3})\hat{O}(\bar{3},2)\phi - \tfrac{2}{9}(\alpha-m)\zeta - (2/\sqrt{6})\hat{O}(\bar{3},\bar{4})\Omega = 0, \quad (28b)$$

$$-(2/\sqrt{6})\hat{O}(\bar{4},\bar{3})\zeta + m\Omega = 0. \quad (28c)$$

Eliminating Ω between Eq. (28b) and (28c), one obtains

$$(m/\sqrt{3})\hat{O}(\bar{3},2)\phi - \tfrac{2}{3}[\tfrac{1}{3}(\alpha-m)m + \hat{O}(\bar{3},\bar{4})\hat{O}(\bar{4},\bar{3})]\zeta = 0. \quad (29)$$

To combine this with Eq. (28a), we must first operate on it with $\hat{O}(2,\bar{3})$ and use the identities

$$\hat{O}(2,\bar{3})\hat{O}(\bar{3},\bar{4})\hat{O}(\bar{4},\bar{3}) = -\tfrac{2}{9}(4\alpha^2 - \Box^2)\hat{O}(2,\bar{3}), \quad (30)$$

$$\hat{O}(2,\bar{3})\hat{O}(\bar{3},2) = -\tfrac{2}{9}(\alpha^2 - \Box^2) - \hat{Q}(2,2), \quad (31)$$

where $\hat{Q}(2,2)$ is a 2×2 matrix quadratic in the derivative operators which does not transform as any particular representation of S_4. Thus we have

$$(m/\sqrt{3})[-\tfrac{2}{9}(\alpha^2 - \Box^2) - \hat{Q}(2,2)]\phi - \tfrac{2}{3}[\tfrac{1}{3}(\alpha-m)m - \tfrac{2}{9}(4\alpha^2 - \Box^2)]\hat{O}(2,\bar{3})\zeta = 0. \quad (32)$$

Eliminating $\hat{O}(2,\bar{3})\zeta$ between this and Eq. (28a) gives

$$[\tfrac{2}{9}(4\alpha^2 - \Box^2)\alpha + \tfrac{1}{9}(4\alpha^2 - \Box^2)m - \tfrac{1}{2}m\hat{Q}(2,2) + \tfrac{1}{3}m^3]\phi = 0. \quad (33)$$

If we now multiply Eq. (33) by $\hat{O}(\bar{3},2)$, and exploit the identities

$$\hat{O}(\bar{3},2)\hat{Q}(2,2) = \tfrac{2}{9}(4\alpha^2 - \Box^2)\hat{O}(\bar{3},2), \quad (34)$$

$$(4\alpha^2 - \Box^2)\alpha\hat{O}(\bar{3},2) = 0, \quad (35)$$

we find

$$\hat{O}(\bar{3},2)\phi = 0. \quad (36)$$

Now, the identity $(4\alpha^2 - \Box^2)\alpha\beta_k = 0$ implies

$$(4\alpha^2 - \Box^2)\alpha\hat{O} = 0, \quad (37)$$

for all $\hat{O}$. From this and Eq. (28a) we have

$$(4\alpha^2 - \Box^2)\alpha(\alpha+m)\phi = 0. \quad (38)$$

With this result and the identity $\alpha\hat{Q}(2,2) = 0$, operating on Eq. (33) with α leads to

$$[(4\alpha^2 - \Box^2)\alpha + 3m^3]\alpha\phi = 0, \quad (39)$$

which, together with Eq. (38), gives

$$(\alpha+m)\alpha\phi = 0. \quad (40)$$

From Eqs. (36) and (28a), we then find

$$\hat{O}(\bar{3},2)\hat{O}(2,\bar{3})\zeta = 0. \quad (41)$$

Thus operating on Eq. (28b) with $\hat{O}(\bar{3},2)\hat{O}(2,\bar{3})$ produces

$$(4\alpha^2 - \Box^2)\hat{O}(\bar{3},\bar{4})\Omega = 0, \quad (42)$$

where use has been made of the relation

$$\hat{O}(\bar{3},2)\hat{O}(2,\bar{3})\hat{O}(\bar{3},\bar{4}) = -\tfrac{4}{9}(4\alpha^2 - \Box^2)\hat{O}(\bar{3},\bar{4}). \quad (43)$$

This relation also implies

$$(4\alpha^2 - \Box^2)(\alpha-m)\zeta = 0, \quad (44)$$

which, with Eq. (37), leads to

$$(4\alpha^2 - \Box^2)\hat{O}(\bar{4},\bar{3})\zeta = 0. \quad (45)$$

Thus taking $(4\alpha^2 - \Box^2)$ through Eq. (28c) gives

$$(4\alpha^2 - \Box^2)\Omega = 0. \quad (46)$$

In similar fashion one obtains

$$(4\alpha^2 - \Box^2)\hat{O}(2,\bar{3})\zeta = 0 \quad (47)$$

from Eq. (44), and hence Eq. (28a) can be written

$$(4\alpha^2 - \Box^2)(\alpha+m)\phi = 0. \quad (48)$$

Returning now to Eq. (33) and using this last result together with Eqs. (31) and (36), one finds

$$(\alpha^2 - m^2)\phi = 0. \quad (49)$$

This relation and Eq. (40) imply

$$(\alpha+m)\phi = 0, \quad (50)$$

which, together with Eq. (36), insures that ϕ satisfies the Bargmann—Wigner equations.

We now show that the fields ζ and Ω vanish. The remaining field equations are

$$\tfrac{1}{9}(\alpha-m)\zeta + (1/\sqrt{6})\hat{O}(\bar{3},\bar{4})\Omega = 0, \quad (51a)$$

$$-(1/\sqrt{6})\hat{O}(\bar{4},\bar{3})\zeta + \tfrac{1}{2}m\Omega = 0. \quad (51b)$$

Operating on Eq. (51a) with $(1/\sqrt{6})\hat{O}(\bar{4},\bar{3})$ and on Eq. (51b) with $\frac{1}{9}(\alpha-m)$ and adding the results, ζ is eliminated. Using the identity

$$\hat{O}(\bar{4},\bar{3})\hat{O}(\bar{3},\bar{4}) = -\tfrac{2}{9}(\alpha^2 - \Box^2), \quad (52)$$

and Eq. (46), we arrive at

$$(2\alpha^2 + \alpha m - m^2)\Omega = 0. \quad (53)$$

Now we consider $\alpha\Omega$ in detail. The totally antisymmetric nature of $\Omega_{\alpha\beta\gamma\delta} = \epsilon_{\alpha\beta\gamma\delta}\omega$ assures that $[\alpha\Omega]_{\alpha\beta\gamma\delta}$ is also totally antisymmetric. Using this fact, we have

$$[(\gamma\circ\partial)_1 + (\gamma\cdot\partial)_2 + (\gamma\cdot\partial)_3 + (\gamma\circ\partial)_4]\Omega \sim \mathrm{Tr}[\gamma\circ\partial]\omega = 0.$$

Thus, Eq. (53) implies

$$\Omega = 0. \quad (54)$$

Equations (51) now reduce to

$$(\alpha-m)\zeta = 0, \quad (55a)$$

$$\hat{O}(\bar{4},\bar{3})\zeta = 0, \quad (55b)$$

and from Eqs. (50) and (28a)

$$\hat{O}(2,\bar{3})\zeta = 0. \quad (56)$$

Use of these equations and the appropriate identities leads to

$$\hat{O}(\bar{3},\bar{3})\zeta = \hat{O}(3,\bar{3})\zeta = 0, \quad (57)$$

where[13]

$$\hat{O}(3,\bar{3}) = \frac{1}{\sqrt{2}}\begin{bmatrix} \beta_2 & \beta_1 & 0 \\ -\beta_3 & 0 & \beta_1 \\ 0 & -\beta_3 & -\beta_2 \end{bmatrix}. \quad (58)$$

Equations (55)—(57) imply that ζ satisfies the Bargmann—Wigner equations. This fact and the symmetry of ζ are sufficient to insure its vanishing.

4. CONCLUSION

It has been shown that the symmetric group provides a useful tool in the construction of multispinor Lagrangians of second, third, and fourth rank. The irreducible representations of S_n for these dimensions correspond to the multispinor fields which must be introduced, while the direct product series provides information about the nature and number of the couplings which survive. Moreover, the operators in the kinetic terms may be constructed directly using general group theoretical methods and the transformation matrices of S_n.

In third and fourth rank examples of Sec. 3, we have omitted all details of the arguments leading to the numerical coefficients in the Lagrange functions, Eqs. (21) and (27). They are, of course, determined by starting with unspecified coefficients and choosing them such that the desired equations emerge. This process requires a strategy for reducing the field equations by systematically eliminating auxiliary fields in favor of the primary field until the Bargmann—Wigner equations have been obtained. In the examples we have considered it was always possible to obtain an operator K such that $K\psi = \psi$ where ψ is the primary field. K has the property that it annihilates the kinetic operator $\hat{O}$ which couples the primary field to the auxiliary field of next lower symmetry. These conditions are sufficient to obtain the Bargmann—Wigner equations on the primary field, and, with a suitable adjustment of the constants, the auxiliary fields can be made to vanish. We believe this to be a general procedure for obtaining the desired field equations.

[1] P.A.M. Dirac, Proc. R. Soc. (London) A **117**, 610 (1928).
[2] M. Fierz and W. Pauli, Proc. R. Soc. (London) A **173**, 211 (1939).
[3] W. Rarita and J. Schwinger, Phys. Rev. **60**, 61 (1941).
[4] V. Bargmann and E.P. Wigner, Proc. Natl. Acad. Sci. USA **34**, 211 (1946).
[5] A partial list of more recent developments includes: S.N. Gupta, Phys. Rev. **95**, 1334 (1954); L.L. Foldy, Phys. Rev. **102**, 568 (1956); H. Umezawa and A. Visconti, Nucl. Phys. **1**, 348 (1956); S. Weinberg, Phys. Rev. **133**, B 1318 (1964); D.L. Weaver, C.L. Hammer, and R.H. Good, Phys. Rev. **135**, B241 (1964); W.K. Tung, Phys. Rev. **156**, 1385 (1967); S.-J. Chang, Phys. Rev. **161**, 1308 (1967); W.J. Hurley, Phys. Rev. D **3**, 2339 (1971); R.A. Krajcik and M.M. Nieto, Phys. Rev. D **13**, 924 (1976); The last reference contains an extensive bibliography.
[6] We take the matrices γ_μ to be Hermitian matrices satisfying $\gamma_\mu\gamma_\nu + \gamma_\nu\gamma_\mu = 2\delta_{\mu\nu}$.
[7] A. Salam, R. Delbourgo, and J. Strathdee, Proc. R. Soc. (London) A **284**, 146 (1965).
[8] G.S. Guralnik and T.W.B. Kibble, Phys. Rev. **139**, B712 (1965); **150**, 1406(E) (1966).
[9] S.N. Gupta and W.W. Repko, Phys. Rev. **165**, 1415 (1968).
[10] Henceforth, we denote symmetrical indices by enclosing them in parentheses and antisymmetrical indices by enclosing them in square brackets.
[11] For an alternative approach to the construction of Lagrange functions for the Bargmann—Wigner fields, see S.-J. Chang, Phys. Rev. **161**, 1316 (1967).
[12] See, for example, M. Hamermesh, *Group Theory* (Addison-Wesley, Reading, Mass., 1962), which has also served as the source of the notation used throughout this paper.
[13] M.L. Larsen, Ph.D. thesis (Michigan State University, 1976).

Definition of polarization of a spin 1/2 particle in an external electromagnetic field

S. S. Sidhu and R. H. Good, Jr.

Physics Department, The Pennsylvania State University, University Park, Pennsylvania 16802
(Received 8 June 1977)

The 3-vector polarization of a Dirac particle with Pauli anomalous moment, in external electric and magnetic fields, is shown to be given by a certain ratio of components of the wavefunction. The result is valid in the same approximation as leads to the Bargmann–Michel–Telegdi classical equations for the polarization.

I. INTRODUCTION

The polarization of a free Dirac particle may be discussed either in terms of a 4-vector operator

$$\mathbf{T} = \gamma_5(i\boldsymbol{\gamma} - \mathbf{p}) = \beta\boldsymbol{\sigma} - \gamma_5\mathbf{p}, \tag{1a}$$

$$T_4 = \gamma_5(i\gamma_4 - iH) = i\boldsymbol{\sigma}\cdot\mathbf{p} \tag{1b}$$

or a 3-vector operator

$$\mathbf{O} = p^{-2}[(H/|H|)\,\boldsymbol{\sigma}\cdot\mathbf{p}\mathbf{p} + \mathbf{p}\times(\beta\boldsymbol{\sigma}\times\mathbf{p})]. \tag{2}$$

Here T_μ is closely related to the Bargmann–Wigner operator[1] and $\mathbf{O}$ was introduced by Stech.[2] The properties and interrelations of these operators were reviewed by Fradkin and Good[3]; their conventions and notation are used in this present paper. The two operators are related by an operator analog of the rest-to-lab Lorentz transformation

$$\mathbf{T} = \mathbf{O} + (|H|+1)^{-1}\mathbf{O}\cdot\mathbf{p}\mathbf{p}, \tag{3a}$$

$$T_4 = i(H/|H|)\,\mathbf{O}\cdot\mathbf{p}. \tag{3b}$$

Every pure plane-wave state is polarized in the sense that, given a four-component wavefunction Ψ, there is a unique unit vector $\mathbf{s}$ such that

$$\mathbf{O}\cdot\mathbf{s}\Psi = \Psi. \tag{4}$$

The actual determination of the polarization of a particle state (having $H/|H| = +1$) may be made easily by using the relation

$$\Psi_2/\Psi_1 = \exp(i\phi)\tan\tfrac{1}{2}\theta \tag{5}$$

where θ, ϕ are the spherical polar angles of $\mathbf{s}$.

There is a limited carry-over of these ideas to the case of a Dirac particle with Pauli anomalous moment in external electric and magnetic fields. Fradkin and Good[3] defined the polarization 4-vector as

$$\mathbf{T} = \beta\boldsymbol{\sigma} - \gamma_5\boldsymbol{\pi}, \tag{6a}$$

$$T_4 = i\boldsymbol{\sigma}\cdot\boldsymbol{\pi}. \tag{6b}$$

An important justification for this definition is that, when averages over a narrow wavepacket are taken so as to get a classical treatment of polarization, $\langle T_\mu\rangle$ satisfies the Bargmann–Michel–Telegdi equations.[4] The 3-vector polarization, as an aspect of the Dirac wavefunction for an interacting particle, has not been discussed previously.

The purpose of this paper is to show that, in the same approximation as leads to the BMT equations, the 3-vector polarization is well defined and given again by Eq. (5).

It is expected that this result will have application in following the change of polarization through a barrier penetration. The BMT[4] or Thomas[5] classical equations[6] ordinarily apply in following the polarization throughout the classically allowed regions on opposite sides of a barrier. If the connection between the wavefunctions on the two sides of the barrier can be found, perhaps by using a WKB type of approximation, then Eq. (5) gives the connection between the polarizations and hence between the classical solutions in the two allowed regions.

II. BASIC EQUATIONS

The Dirac equation with Pauli anomalous moment term included is

$$(\gamma_\mu\pi_\mu + \tfrac{1}{2}e\hbar\mu F_{\mu\nu}\gamma_\mu\gamma_\nu - i)\,\Psi = 0. \tag{7}$$

As shown in detail in Ref. 3 the classical equations for the orbit and the BMT equations for the polarization apply when there is a narrow wavepacket such that one can write

$$\pi_\mu\Psi(\mathbf{x}, t) = \langle\pi_\mu\rangle\,\Psi(\mathbf{x}, t) \tag{8}$$

and when all components $e\hbar F_{\mu\nu}$ are small compared to unity. In Eq. (8) $\langle\pi_\mu\rangle$ denotes the classical value, a function of t only. Consider a classically allowed region for a particle solution so that $-i\langle\pi_4\rangle$ is real and positive. As in Ref. 3 the notation $-i\langle\pi_4\rangle = \bar{\gamma}$ is used; it is the classical rest–plus–kinetic energy of the particle.

III. POLARIZATION DISCUSSION

Let the four-component wavefunction be written as two two-component functions

$$\Psi = \begin{bmatrix}\Psi_a\\ \Psi_b\end{bmatrix} \tag{9}$$

(if the system is nonrelativistic Ψ_a and Ψ_b are the large and small components). The upper half of Eq. (7) leads directly to

$$\Psi_b = \frac{1}{1+\bar{\gamma}}\,\boldsymbol{\sigma}\cdot\langle\boldsymbol{\pi}\rangle\Psi_a \tag{10}$$

in first approximation, the terms in $e\hbar F_{\mu\nu}$ disregarded. With this result everything can be expressed in terms of the two-component function Ψ_a.

 0022-2488/78/1905-0935$1.00

Consider two functions Ψ_1 and Ψ_2, for two packets that follow the same classical orbit but that may otherwise be different. Let the polarization 4-vector be defined by

$$T_\mu = \gamma_5 (i\gamma_\mu - \pi_\mu) \tag{11}$$

[this is a different definition of the fourth component than used in Ref. 3, Eq. (6) above, but the difference involves only terms proportional to $e\hbar F_{\mu\nu}$ which are negligible in the discussion of the BMT limit]. A straightforward calculation leads to

$$\Psi_1^\dagger \Psi_2 = \frac{2\bar{\gamma}}{1+\bar{\gamma}} \Psi_{1a}^\dagger \Psi_{2a}, \tag{12}$$

$$\Psi_1^\dagger \mathbf{T} \Psi_2 = \frac{2\bar{\gamma}}{1+\bar{\gamma}} \Psi_{1a}^\dagger \left(\boldsymbol{\sigma} + \frac{\langle \boldsymbol{\pi} \rangle \boldsymbol{\sigma} \cdot \langle \boldsymbol{\pi} \rangle}{\bar{\gamma}+1} \right) \Psi_{2a}, \tag{13a}$$

$$\Psi_1^\dagger T_4 \Psi_2 = i \frac{2\bar{\gamma}}{1+\bar{\gamma}} \Psi_{1a}^\dagger \boldsymbol{\sigma} \cdot \langle \boldsymbol{\pi} \rangle \Psi_{2a}. \tag{13b}$$

This shows that, within the present approximation, one can replace a discussion of the polarization 4-vector T_μ in the space of the four-component wavefunction Ψ by a discussion of the Pauli spin matrices $\boldsymbol{\sigma}$ in the space of the two components Ψ_a. One must make due allowance for a different normalization and for the instantaneous rest-to-lab Lorentz transformation.

As is well known, any two component wavefunction Ψ_a is polarized in the sense that

$$\boldsymbol{\sigma} \cdot \mathbf{s} \Psi_a = \Psi_a, \tag{14}$$

where the spherical polar angles of the polarization direction $\mathbf{s}$ are given by

$$\Psi_{a2}/\Psi_{a1} = \exp(i\phi) \tan\tfrac{1}{2}\theta. \tag{15}$$

The conclusion is that the particle has a definite 3-vector polarization with direction given by the ratio of the top two components of the four-component wavefunction. Although the ratio is a function of $\mathbf{x}$ and t, the polarization in this approximation has meaning only at the wavepacket so one can evaluate the ratio at $\mathbf{x}(t)$ as given by the classical orbit.

[1] V. Bargmann and E. P. Wigner, Proc. Natl. Acad. Sci. USA 34, 211 (1948).
[2] B. Stech, Z. Physik 144, 214 (1956).
[3] D. M. Fradkin and R. H. Good, Jr., Rev. Mod. Phys. 33, 343 (1961).
[4] V. Bargmann, L. Michel, and V. L. Telegdi, Phys. Rev. Lett. 2, 435 (1959).
[5] L. H. Thomas, Phil. Mag. 3, 1 (1927).
[6] For a review of the classical equations see, for example, R. H. Good, Jr. and T. J. Nelson, *Classical Theory of Electric and Magnetic Fields* (Academic, New York, 1971), Sec. 30.

A second order calculation of the adiabatic invariant of a charged particle spiraling in a longitudinal magnetic field[a),b)]

R. Chehab

Laboratoire de l'Accélérateur Linéaire, Université de Paris-Sud, Orsay, France 91405
(Received 27 June 1977)

The change of the action integral of a charged particle spiraling in a slowly varying longitudinal magnetic field is investigated. The method used to solve this two-dimensional problem is a generalization of Vandervoort's analysis of a one-dimensional system. It is based on a canonical transformation of the usual conjugate coordinates and momenta into a set of four variables of the action-angle type. The new canonical equations are solved by a method of iteration, and the solution is used to calculate the change of the adiabatic invariant. Our results are analogous to those obtained with the approximations derived by Hertweck and Schlüter and by Chandrasekhar. We have compared our analytic results to those obtained by numerical integration of the equation of motion in the case of a magnetic solenoid with a slowly decreasing field.

1. INTRODUCTION

It is well known that an adiabatic invariant is a quantity which remains constant during an infinitely slow variation of the external parameters of the physical system under consideration. Generally, the action integrals are adiabatic invariants for particles undergoing periodic motions.

The adiabatic invariance of a quantity J to the nth order is clearly stated[1] if a positive constant M exists such that the change ΔJ of J satisfies

$$|\Delta J| < \epsilon^n \circ M,$$

for all "sufficiently small" ϵ, where the parameter of smallness ϵ represents the rate of variation of the external parameters. The behavior of adiabatic invariants when external parameters vary slowly between two constant values has been investigated by several authors.[2-7] Many of these investigations are based on a perturbation theory in which the independent variable is expressed in terms of a unit that increases in inverse proportion to the parameter of smallness ϵ. These studies provide results which are asymptotically valid but are not sufficient to give the rate of change of the adiabatic invariants for finite values of ϵ.

Hertweck and Schlüter,[3] Chandrasekhar,[5] and Parker[7] have developed alternative methods to calculate the change of the adiabatic invariants corresponding to slowly time-varying magnetic fields in the case of finite values of the parameter of smallness.

We present here an analysis based on the Hamiltonian formalism which enables us to evaluate the change of the adiabatic invariant of an ultrarelativistic particle spiraling in a magnetic field which is time constant but which varies slowly longitudinally between two different values of the field.

[a)] Work supported by the "Institut National de Physique Nucléaire et de Physique des Particules du C. N. R. S."
[b)] This study is a part of a thesis submitted to the University Paris XI for the degree of "Docteur-Ingénieur," June 2, 1975 and registered by the C. N. R. S. under the number A. O. 11504.

2. VARIATION OF THE ACTION INTEGRAL

We consider a longitudinal magnetic field (Fig. 1) of axial symmetry and ultrarelativistic particles whose transverse momenta are much smaller than their scalar momentum. The curvature of the lines of forces is neglected.

The action integral for the transverse motion of a particle in a constant magnetic field is given by

$$A = \oint \sum_i p_i \, dq_i = \pi p_\perp^2 / eB, \tag{1}$$

where (q_i, p_i) are the conjugate variables (x, p_x, y, p_y), $p_\perp$ the transverse momentum, and B the field strength.

If the magnetic field B changes slowly, so does the cyclotron frequency eBc/P, where P, the scalar momentum, is constant. It follows that the motion is no longer strictly periodic and the contour of integration of the integral in Eq. (1) cannot be accurately defined. In order to avoid this difficulty, we follow Vandervoort's method of solving his one-dimensional harmonic oscillator problem and make a canonical transformation of the set of conjugate variables (x, p_x, y, p_y) into a new set of variables (Q_1, P_1, Q_2, P_2) of the "action-angle" type.

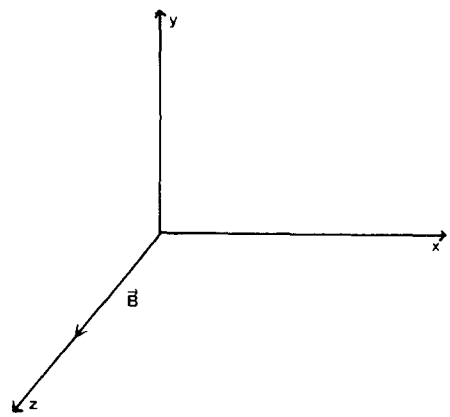

FIG. 1. System of coordinates.

 0022-2488/78/1905-0937$1.00

The Hamiltonian associated with the independent variable z is

$$G = -p_z = -eA_z - \left[\left(\frac{H-eV}{c}\right)^2 - m_0^2c^2 - (p_x - eA_x)^2 - (p_y - eA_y)^2\right]^{1/2}, \tag{2}$$

where p_x, p_y, p_z are the conjugate momenta, A_x, A_y, A_z, the potential vector components and where H is the Hamiltonian associated with the independent variable t (time) and V the scalar potential.

We identify H with energy to obtain

$$H^2/c^2 - m_0^2c^2 = (mv)^2 = P^2,$$

where m is the relativistic mass and P, the scalar momentum.

Taking into account the restrictive hypothesis on the relative values of the momenta, we can expand the Hamiltonian G around $p_x = 0$ and $p_y = 0$. We find

$$G(z) = -P\left(1 - \frac{1}{2P^2}(p_x^2 + p_y^2) - \frac{1}{2P^2}[(eA_x)^2 + (eA_y)^2] + \frac{1}{P^2}(p_x eA_x + p_y eA_y)\right). \tag{3}$$

Since the magnetic field has an axial symmetry, we can write the transverse components of the potential vector A in terms of the longitudinal component $\bar{B}(z)$ of the magnetic field B:

$$A_x = -\frac{y}{2}\bar{B}(z) + \frac{y}{16}(x^2 + y^2)\bar{B}''(z) + \cdots,$$
$$A_y = \frac{x}{2}\bar{B}(z) - \frac{x}{16}(x^2 + y^2)\bar{B}''(z) + \cdots, \tag{4}$$

where the prime denotes differentiation with respect to z. (This notation will be used throughout the article.)

The Hamiltonian G can be written as

$$G(z) = -P + \frac{1}{2P}(p_x^2 + p_y^2) + \frac{P}{2}\left(\frac{e\bar{B}}{2P}\right)^2(x^2 + y^2) - \frac{e\bar{B}}{2P}(xp_y - yp_x) - \frac{1}{16}\left(\frac{e\bar{B}}{2P}\right)\bar{B}''(z)(x^2 + y^2)^2 + \frac{1}{16}\left(\frac{e\bar{B}''}{P}\right)(x^2 + y^2)(xp_y - yp_x), \tag{5}$$

if we limit the expansion to second order in the derivatives of $\bar{B}$.

We shall choose a generating function of the form

$$F = F(q_i, Q_i, k), \tag{6}$$

where $k = e\bar{B}/P$. In this way we have

$$\frac{\partial F}{\partial q_i} = p_i, \qquad \frac{\partial F}{\partial Q_i} = -P_i, \qquad \frac{\partial F}{\partial z} = K - G, \tag{7}$$

and the new Hamiltonian K is given by

$$K = G + k'\frac{\partial}{\partial k}F(q_i, Q_i, k). \tag{8}$$

The canonical equations associated with the new variables are

$$Q_i' = \frac{\partial K}{\partial P_i} = \frac{\partial G}{\partial P_i} + k'\frac{\partial}{\partial P_i}\left(\frac{\partial F}{\partial k}\right), \qquad P_i' = -\frac{\partial K}{\partial Q_i} = -\frac{\partial G}{\partial Q_i} - k'\frac{\partial}{\partial Q_i}\left(\frac{\partial F}{\partial k}\right). \tag{9}$$

The canonical transformation[8]

$$x = \sqrt{1/kP}\,(\sqrt{2P_1}\sin Q_1 + P_2),$$
$$p_x = \tfrac{1}{2}\sqrt{kP}\,(\sqrt{2P_1}\cos Q_1 - Q_2),$$
$$y = \sqrt{1/kP}\,(\sqrt{2P_1}\cos Q_1 + Q_2),$$
$$p_y = \tfrac{1}{2}\sqrt{kP}\,(-\sqrt{2P_1}\sin Q_1 + P_2) \tag{10}$$

leads to the new set of variables (Q_1, P_1, Q_2, P_2). In our case, we use the following generating function:

$$F(q_i, Q_i, k) = \frac{\sqrt{kP}}{2}(\sqrt{kP}\cdot y - 2Q_2)x - \frac{\tan Q_1}{2}(\sqrt{kP}\cdot y - Q_2)^2. \tag{11}$$

The Hamiltonian K can therefore be written in a simpler form:

$$K = -P + kP_1 + \frac{k'}{2k}\sqrt{2P_1}\,(P_2\cos Q_1 - Q_2\sin Q_1) - \frac{k''}{16kP}\cdot 2P_1[2P_1 + P_2^2 + Q_2^2 + 2\sqrt{2P_1}(P_2\sin Q_1 + Q_2\cos Q_1)] \times \left[1 + \frac{1}{\sqrt{2P_1}}(P_2\sin Q_1 + Q_2\cos Q_1)\right], \tag{12}$$

where the fourth term contains the second order contributions from the Hamiltonian G. These contributions may be neglected in the applications made below, and so they will be.

We may observe that P_1, Q_2, and P_2 are constant when k is constant. Using Eqs. (1) and (10), we can verify that the product of P_1 by 2π gives the action integral. The squares of Q_2 or P_2 have the same dimensions as action integrals.

The canonical equations are

$$Q_1' = k + \frac{k'}{2k}\frac{1}{\sqrt{2P_1}}(P_2\cos Q_1 - Q_2\sin Q_1),$$
$$P_1' = \frac{k'}{2k}\sqrt{2P_1}(P_2\sin Q_1 + Q_2\cos Q_1),$$
$$Q_2' = \frac{k'}{2k}\sqrt{2P_1}\cos Q_1,$$
$$P_2' = \frac{k'}{2k}\sqrt{2P_1}\sin Q_1. \tag{13}$$

From Eq. (13) it is then possible to derive a first approximation by ignoring terms of the order of k'/k.

$$Q_1^{(0)} = \int^z k(\xi)\,d\xi + \psi, \quad P_1^{(0)} = \text{const},$$
$$Q_2^{(0)} = \text{const}, \quad P_2^{(0)} = \text{const}. \tag{14}$$

A second approximation may be obtained by using an iteration method:

$$\begin{aligned} Q_1^{(1)} &= Q_1^{(0)} + \int_{z_0}^z \frac{k'}{2k}\cdot (2P_1^{(0)})^{-1/2}(P_2^{(0)}\cos Q_1^{(0)} \\ &\quad - Q_2^{(0)}\sin Q_1^{(0)})\,d\xi, \\ Q_2^{(1)} &= Q_2^{(0)} + \int_{z_0}^z \frac{k'}{2k}(2P_1^{(0)})^{1/2}\cos Q_1^{(0)}\,d\xi, \\ P_2^{(1)} &= P_2^{(0)} + \int_{z_0}^z \frac{k'}{2k}(2P_1^{(0)})^{1/2}\sin Q_1^{(0)}\,d\xi. \end{aligned} \tag{15}$$

Using Eqs. (13) and (15), we can evaluate the quantity P_1'. We thus obtain

$$\frac{P_1'}{P_1} = \frac{k'}{k}\cdot\frac{1}{\sqrt{2P_1}}(P_2^{(1)}\sin Q_1^{(1)} + Q_2^{(1)}\cos Q_1^{(1)}). \tag{16}$$

Let us define a quantity Δ by the expression

$$P_1 = P_1^{(0)}(1+\Delta),$$

where Δ is of the first order in k'/k.

The quantity P_1, conjugate of Q_1, coincides with $P_1^{(0)}$ when the movement is strictly periodic. Using Eq. (15) and Taylor expansions of $\sin Q_1^{(1)}$ and $\cos Q_1^{(1)}$, we can develop the right-hand side of Eq. (16). Retaining terms of order no higher than the second one in k'/k, we obtain

$$\begin{aligned} \frac{P_1'}{P_1} &= \frac{k'}{k}\cdot(2P_1^{(0)})^{-1/2}\cdot\{[P_2^{(0)}\sin Q_1^{(0)} + Q_2^{(0)}\cos Q_1^{(0)}]\left(1-\frac{\Delta}{2}\right) \\ &\quad + \sin Q_1^{(0)}\int_{z_0}^z \frac{k'}{2k}(2P_1^{(0)})^{1/2}\sin Q_1^{(0)}\,d\xi + \cos Q_1^{(0)} \\ &\quad \times \int_{z_0}^z \frac{k'}{2k}(2P_1^{(0)})^{1/2}\cos Q_1^{(0)}\,d\xi \\ &\quad + (P_2^{(0)}\cos Q_1^{(0)} - Q_2^{(0)}\sin Q_1^{(0)})\cdot\int_{z_0}^z \frac{k'}{2k} \\ &\quad \cdot (2P_1^{(0)})^{-1/2}[P_2^{(0)}\cos Q_1^{(0)} - Q_2^{(0)}\sin Q_1^{(0)}]\,d\xi\}. \end{aligned} \tag{17}$$

From Eq. (17) we deduce

$$\frac{d\Delta}{dz} = \frac{k'}{k}\cdot(2P_1^{(0)})^{-1/2}[P_2^{(0)}\sin Q_1^{(0)} + Q_2^{(0)}\cos Q_1^{(0)}].$$

We may define an angle σ according to

$$\sigma = \sin^{-1}\left\{\frac{Q_2^{(0)}}{[(Q_2^{(0)})^2 + (P_2^{(0)})^2]^{1/2}}\right\}. \tag{18}$$

Integrating Eq. (17) and replacing $2P_1$ by J and $2P_1^{(0)}$ by J_0, we obtain

$$\begin{aligned} \frac{J}{J_0} &= \exp\left(\left(\frac{(Q_2^{(0)})^2 + (P_2^{(0)})^2}{J_0}\right)^{1/2}\int_{z_0}^z \frac{k'}{k}\sin(Q_1^{(0)}+\sigma)\,d\xi\right. \\ &\quad + \left[\int_{z_0}^z \frac{k'}{2k}\sin Q_1^{(0)}\,d\xi\right]^2 + \left[\int_{z_0}^z \frac{k'}{2k}\cos Q_1^{(0)}\,d\xi\right]^2 \\ &\quad + \frac{1}{4}\left(\frac{(Q_2^{(0)})^2 + (P_2^{(0)})^2}{J_0}\right)\left\{\left[\int_{z_0}^z \frac{k'}{k}\cos(Q_1^{(0)}+\sigma)\,d\xi\right]^2\right. \\ &\quad \left.\left. - \left[\int_{z_0}^z \frac{k'}{k}\sin(Q_1^{(0)}+\sigma)\,d\xi\right]^2\right\}\right). \end{aligned} \tag{19}$$

Let us define the quantities

$$U(\sigma, z) = \int_{z_0}^z \frac{k'}{k^2}\sin(Q_1^{(0)}+\sigma)\,dQ_1^{(0)}$$

and

$$V(\sigma, z) = \int_{z_0}^z \frac{k'}{k^2}\cos(Q_1^{(0)}+\sigma)\,dQ_1^{(0)}.$$

By expanding the exponential term of Eq. (19) and retaining terms up to $(k'/k^2)^2$, we obtain

$$\begin{aligned} \frac{J}{J_0} &= 1 + \left(\frac{(Q_2^{(0)})^2 + (P_2^{(0)})^2}{J_0}\right)^{1/2}\times U(\sigma, z) + \frac{1}{4}\left(\frac{(Q_2^{(0)})^2 + (P_2^{(0)})^2}{J_0}\right) \\ &\quad \times\{[U(\sigma, z)]^2 + [V(\sigma, z)]^2\} + \tfrac{1}{4}\{[U(0, z)]^2 + [V(0, z)]^2\}. \end{aligned} \tag{20}$$

Expression (20) is an expansion of J, the action integral, to second order in the parameter of smallness k'/k^2. Two remarks can be made in order to simplify Eq. (20).

First, we can write the canonical angular momentum as [see Eq. (10)]

$$\begin{aligned} p_\phi &= xp_y - yp_x \\ &= \tfrac{1}{2}[(P_2^{(0)})^2 + (Q_2^{(0)})^2 - 2P_1^{(0)}]. \end{aligned}$$

This leads to

$$\frac{(Q_2^{(0)})^2 + (P_2^{(0)})^2}{2P_1^{(0)}} = 1 + \frac{2p_\phi}{J_0}. \tag{21}$$

Secondly, we can verify that if k'/k^2 is constant or is a weakly monotonic function of z, the following relation holds:

$$[U(\sigma, z)]^2 + [V(\sigma, z)]^2 = [U(0, z)]^2 + [V(0, z)]^2.$$

Equation (20) may therefore be rewritten:

$$\begin{aligned} \frac{J}{J_0} &= 1 + \left(1 + \frac{2p_\phi}{J_0}\right)^{1/2}\cdot U(\sigma, z) + \frac{1}{2}\left(1 + \frac{p_\phi}{J_0}\right) \\ &\quad \times\{[U(\sigma, z)]^2 + [V(\sigma, z)]^2\}. \end{aligned} \tag{22}$$

If we average Eq. (22) over equally weighted values of ψ, we obtain

$$\langle J/J_0\rangle_\psi = 1 + \tfrac{1}{2}(1 + p_\phi/J_0)\times\{[U(\sigma, z)]^2 + [V(\sigma, z)]^2\}. \tag{23}$$

Expression (22) and (23) are similar to those derived by Vandervoort in the case of the one-dimensional harmonic oscillator. Here we take coupling into account by including terms which contain p_ϕ.

3. COMPARISON OF OUR RESULTS WITH THE APPROXIMATIONS OF HERTWECK-SCHLÜTER AND CHANDRASEKHAR

The transverse motion of an ultrarelativistic particle in a time constant but longitudinally varying magnetic field B is described by the following equation:

$$P\frac{d^2X}{dz^2} + ie\bar{B}\frac{dX}{dz} + \tfrac{1}{2}ie\frac{d\bar{B}}{dz}X = 0, \tag{24}$$

The scalar momentum P is constant, and X is a quantity defined by $X = x + iy$, where x and y are the Cartesian coordinates of the particle. Following the method presented by Chandrasekhar[5] for the evaluation of the change of the adiabatic invariant, we are led to the expression

$$J/J_0 = 1 - 2S\rho\cos(\chi + \theta) + (1 + S^2)\rho^2. \tag{25}$$

In this equation, ρ represents the modulus and χ the phase of the slowly varying parameter,

$$R(z) = \tfrac{1}{2}\int_{z_0}^{z} (k'/k)_\xi \cdot \exp[-i\int^{\xi} k(\eta)\,d\eta]\,d\xi. \tag{26}$$

The quantity S is the ratio of the modula of the two exponential type solutions of Eq. (24).

In order to compare these results to ours, we use the relations

$$p_\phi = \tfrac{1}{2}e\bar{B}_0|X_0|^2 + (P/2i)[X_0^* X_0' - X_0 X_0'^*]$$

and

$$J_0 = P^2|X_0'|^2/e\bar{B}_0,$$

where * indicates the conjugate value. We can easily verify that the ratio S is related to the invariants p_ϕ and J_0 according to

$$S^2 = 1 + 2p_\phi/J_0. \tag{27}$$

Using Eq. (26), we are led to the relation

$$\rho^2 = |R(z)|^2 = \tfrac{1}{4}\{[U(\sigma, z)]^2 + [V(\sigma, z)]^2\}. \tag{28}$$

Equation (22), which we obtain above, can therefore be rewritten:

$$J/J_0 = 1 + S\cdot U(\sigma, z) + (1 + S^2)\rho^2,$$

which is quite similar to Eq. (25) derived by using Chandrasekhar's method.

4. APPLICATION TO A MAGNETIC SOLENOID WITH A SLOWLY DECREASING FIELD

Magnetic solenoids with a slowly decreasing field are used to improve the energy acceptance of positron accelerators.[9] Usually such systems—called "adiabatic"—have a field B decreasing according to

$$\bar{B} = \frac{\bar{B}_0}{1 + \alpha z},$$

where $\bar{B}_0$ is the initial value of $\bar{B}$ and α is a constant.

For such a field, the parameter of smallness is given by

$$\epsilon = \frac{k'}{k^2} = \frac{P}{e\bar{B}^2}\cdot\frac{d\bar{B}}{dz}.$$

This parameter remains constant for any particle of fixed energy spiraling in the magnetic field.

In order to study the behavior of the adiabatic invariant, we shall compare the results obtained by direct application of Eq. (22) to those obtained by numerical integration of the equations of motion.

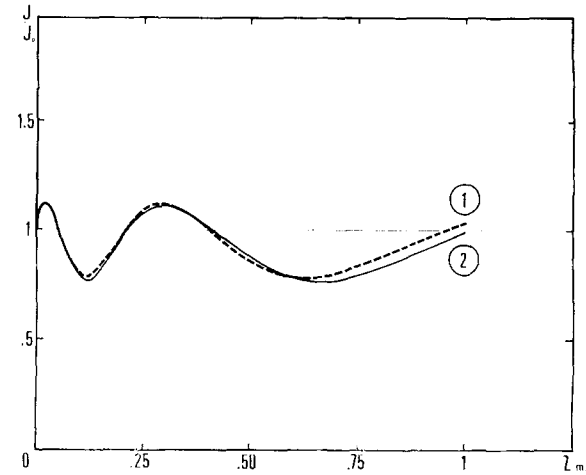

FIG. 3. Behavior of the adiabatic invariant for the magnetic field $B = B_0/(1+\alpha z)$. Curve 1 is obtained via Eq. (22); curve 2, via numerical integration. $B_0 = 2$ Tesla, $\alpha = 12$ m^{-1}, $x_0 = 0.41$ mm, $y_0 = 1.13$ mm, $x_0' = 101.5$ mrd, $y_0' = 110.5$ mrd, $P = 10$ MeV/c, and $\epsilon = 0.2$.

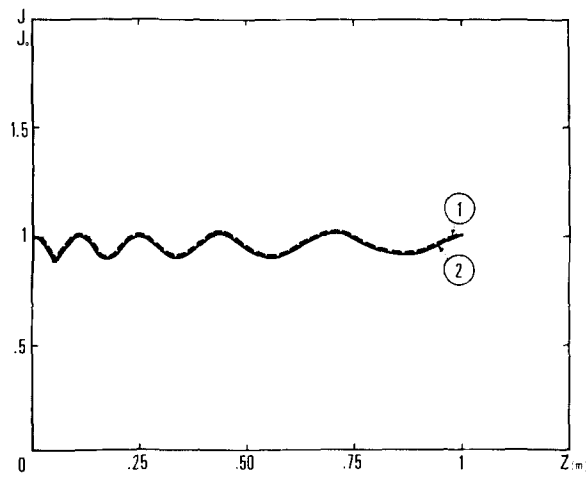

FIG. 2. Behavior of the adiabatic invariant for the magnetic field $B = B_0/(1+\alpha z)$. Curve 1 is obtained via Eq. (22); curve 2, via numerical integration. $B_0 = 10$ Tesla, $\alpha = 3.75$ m^{-1}, $x_0 = 0.41$ mm, $y_0 = 1.13$ mm, $x_0' = 61$ mrd, $y_0' = 66.3$ mrd, $P = 40$ MeV/c, and $\epsilon = 0.05$.

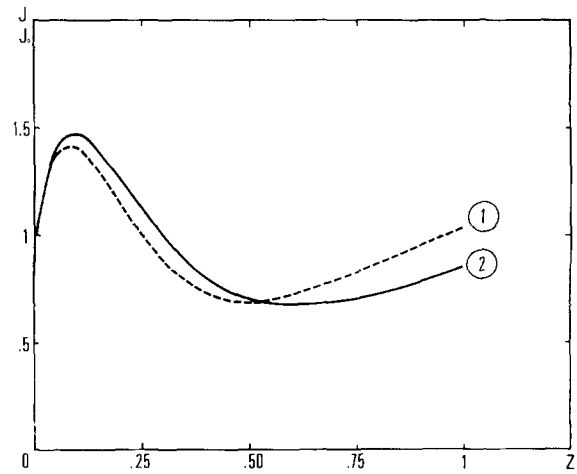

FIG. 4. Behavior of the adiabatic invariant for the magnetic field $B = B_0/(1+\alpha z)$. Curve 1 is obtained via Eq. (22); curve 2, via numerical integration. $B_0 = 2$ Tesla, $\alpha = 12$ m^{-1}, $x_0 = 0.41$ mm, $y_0 = 1.13$ mm, $x_0' = 101.5$ mrd, $y_0' = 110.5$ mrd, $P = 20$ MeV/c, and $\epsilon = 0.4$.

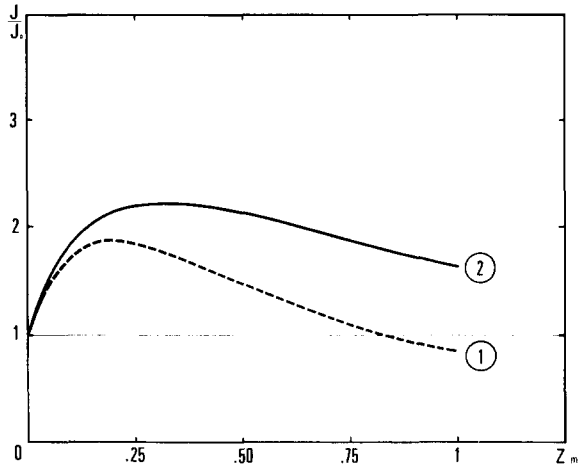

FIG. 5. Behavior of the adiabatic invariant for the magnetic field $B=B_0/(1+\alpha z)$. Curve 1 is obtained via Eq. (22); curve 2, via numerical integration. $B_0=2$ Tesla, $\alpha=12$ m^{-1}, $x_0=0.41$ mm, $y_0=1.13$ mm, $x_0'=101.5$ mrd, $y_0'=110.5$ mrd, $P=30$ MeV/c, and $\epsilon=0.6$.

A. Comparison of our analytic approach with direct numerical calculation

Some preliminary remarks may be made concerning the zeroth order expression of the invariants using the initial conditions of the motion of the particle $[x_0, y_0, x_0' = dx_0/dz,\ y_0' = dy_0/dz,\ P]$. First we can write

$$\begin{aligned} p_\phi &= P(x_0 y_0' - y_0 x_0') + \tfrac{1}{2} e\bar{B}_0 (x_0^2 + y_0^2), \\ 2P_1^{(0)} &= P^2 (x_0'^2 + y_0'^2)/e\bar{B}_0, \\ Q_2^{(0)} &= \sqrt{k_0 P}\,[y_0 - x_0'/k_0], \\ P_2^{(0)} &= \sqrt{k_0 P}\,[x_0 + y_0'/k_0]. \end{aligned} \tag{29}$$

Secondly, we observe that the angles ψ and σ can be exactly calculated using Eqs. (10) and (18) and the initial conditions.

Numerical integration of the equation of motion given above (Eq. 24) has been made by the Runge–Kutta method of the fourth order.

The adiabatic invariant J expressed as:

$$J = P^2 (x'^2 + y'^2)/e\bar{B} \tag{30}$$

has been calculated. The results regarding the quotient J/J_0 evaluated for various initial conditions and magnetic field shapes have been compared to the determination given by Eq. (22) in the same conditions. Figures 2–5 illustrate these comparisons. As expected, we observe that the smaller ϵ is, the better the agreement is between our analytic second order approximation and the (quasi) exact numerical results.

B. Energy acceptance of a solenoid in a transport system

If we study the behavior of a many-particle beam (rather than a single-particle trajectory), we have to average the expression of the adiabatic invariant over equally weighted values of ψ. Since the parameter ϵ is constant, expression (23) becomes

$$\langle J/J_0 \rangle_\psi = 1 + \tfrac{1}{2}\epsilon^2 [1 + p_\phi/J_0] \cdot \times \{ [\int_{z_0}^{z} \sin(Q_1^{(0)} + \sigma)\, dQ_1^{(0)}]^2 + [\int_{z_0}^{z} \cos(Q_1^{(0)} + \sigma)\, dQ_1^{(0)}]^2 \}. \tag{31}$$

For given initial coordinates x_0, x_0', y_0, y_0' and a given magnetic field law, we may deduce from Eq. (31) an upper limit on the particle energies, provided that an upper bound is put on the growth of the adiabatic invariant. Such a limit provides valuable information on the band of energies accepted by the solenoid.

Using the Hamiltonian formalism, we have derived an expression of the change of the adiabatic invariant for an ultrarelativistic particle spiraling in a longitudinally varying magnetic field. Our expression is of the second order in the parameter of smallness. The constant coefficients of this expansion are functions of the canonical angular momentum (constant to the order of our expansion) and of the initial value of the action integral.

Our results are in agreement with those of Hertweck and Schlüter and of Chandrasekhar. We have applied our theoretical analysis to the particular case of charged particles spiraling along a nonuniform magnetic field, and compared our predictions to a precise numerical integration of the exact equation of motion. Both results are in close agreement.

ACKNOWLEDGMENTS

The research reported in this paper is a part of a thesis prepared at Orsay under the supervision of Professor J. Haïssinski. I am deeply indebted to Professor J. Haïssinski for much guidance and encouragement. I wish to express my gratitude to Dr. P. Lapostolle for several valuable remarks.

[1] A. Lenard, Ann. Phys. (N.Y.) 6, 261–76 (1959).
[2] R. Kulsrud, Phys. Rev. 106, 205–07 (1957).
[3] F. Hertweck and A. Schlüter, Z. Naturforsch. **12a**, 844–49 (1957).
[4] C. Gardner, Phys. Rev. 115, 791–94 (1959).
[5] S. Chandrasekhar, *Plasma Physics* (Univ. of Chicago Press, Chicago, 1960), p. 48..
[6] P. Vandervoort, Ann. Phys. (N.Y.) 12, 436–43 (1961).
[7] L. Parker, Nuovo Cimento B **XL**, 99–108 (1965).
[8] P. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), p. 297.
[9] R. Helm, "The Positron Source," in *The Stanford Two-mile Accelerator*, edited by R. Neal (Benjamin, New York, 1968).

The graded Lie groups SU(2,2/1) and OSp(1/4)[a)]

Feza Gürsey and Louis Marchildon

Physics Department, Yale University, New Haven, Connecticut 06520
(Received 31 May 1977)

We study the graded Lie groups corresponding to the graded Lie algebras SU(2,2/1) and OSp(1/4). General finite group transformations are parametrized, and nonlinear representations are obtained on coset spaces. Jordan and traceless algebras are constructed which admit these groups as automorphism groups.

I. INTRODUCTION

Supersymmetry (Fermi—Bose symmetry) has in recent years more and more attracted the attention of physicists. Ever since Wess and Zumino proposed[1] Lagrangian densities giving rise to action functionals invariant under the interchange of Bose with Fermi coordinates, and noticed[2] that field theories constructed from such Lagrangians seem to behave better with respect to ultraviolet infinities, many authors have attempted to explore further the properties and consequences of such theories. The most striking of these developments has been supergravity,[3] which combines the spin-2 gravitational field with a spin-$\frac{3}{2}$ massless Majorana field in a supersymmetric way.

From the mathematical point of view, supersymmetric theories were soon recognized to be intimately related to the structures called graded Lie group (GLG's) and graded Lie algebras (GLA's).[4] Indeed the algebraic structure of the original Wess-Zumino transformations is the one of the GLA SU(2, 2/1), which has the conformal algebra times U(1) as its Lie subalgebra and a four-component complex spinor as its odd part. More recently, MacDowell and Mansouri[5] have put into light the OSp(1/4) structure of supergravity.

In most of these developments it has been enough to focus one's attention to graded Lie algebras, that is, to those elements of the graded Lie groups which are infinitesimally close to the unit element. Graded Lie groups themselves have been considered sometime ago by Berezin and Kac,[6] and Salam and Strathdee[7] have used them to some extent in introducing the notion of superfields.

In this paper we will attempt to study more closely two important graded Lie groups, namely SU(2, 2/1) and OSp(1/4). General finite SU(2, 2/1) transformations will be parametrized in Sec. II, and the subgroup OSp(1/4) will subsequently be identified. Having done so, we will turn our attention to two important classes of representations of these groups, which may turn out to be quite relevant for physical applications.

In Sec. III we will consider nonlinear representations on coset spaces. Nonlinear representations of Lie groups are closely related to the theory of spontaneous symmetry breakdown in gauge field theories. Indeed, if $V(\phi)$ is the Higgs potential to be minimized, with the scalar fields ϕ belonging to a definite representation R of the gauge group G, then $V(\phi)$ has its minimum when $\phi = \phi_0(x)$, where $\phi_0(x)$ obeys a certain algebraic relation like the Michel—Radicati relations.[8] These equations are solved by putting $\phi_0(x)$ in the canonical form $W_R(x)\phi_0$, where W_R is an element of the representation R of G with parameters depending on x. Let H be the subgroup of G that leaves ϕ_0 invariant. Then $\phi_0(x)$ takes the form

$$\phi_0(x) = W_R^{(G/H)}(x)\phi_0.$$

Now we can introduce new scalar fields $\eta_0(x)$ by putting

$$\phi(x) = W_R^{(G/H)}(x)[\eta_0(x) + \phi_0], \tag{1.1a}$$

while a field $\psi(x)$, belonging to the representation S, can be written as

$$\psi(x) = W_S^{(G/H)}(x)\psi_0(x). \tag{1.1b}$$

Here W_S is the S transformation which corresponds to W_R. Since gauge theories are invariant under arbitrary local gauge transformations, the physics is the same whether described by means of the fields ϕ and ψ, or η_0 and ψ_0. Moreover, if we make an arbitrary local gauge transformation of the fields ϕ, ψ into new fields ϕ', ψ', then the latter can still be written in a form similar to Eq. (1.1). The parameters of the new matrices $W_R'^{(G/H)}$ and $W_S'^{(G/H)}$ are then fields which form a coset representation of the gauge group G. They transform linearly under an element of the subgroup H of G, but nonlinearly under an element of the coset G/H. Should the Lie group G be replaced by a graded Lie group, the parameters of the transformation W will also transform nonlinearly under G/H.

It may be remarked, as was originally pointed out, that the superspace of Salam and Strathdee can itself be regarded as a coset space.[7] Take for instance OSp(1/4)/Sp(4), which is the coset of a graded Lie group with respect to its Lie subgroup. It has four elements, which can be related to the Grassmann elements θ_α that transform like a Majorana spinor under Sp(4). Now since the position space x^μ in a de Sitter universe is associated with Sp(4)/Sl(2, c), the superspace (x^μ, θ_α) is just the coset OSp(1/4)/Sl(2, c). Generalizations of superspace (extended superspace) obtained by taking cosets of extended supergravity groups with respect to $\mathrm{Sl}(2,c) \otimes G_{\mathrm{int}}$ (G_{int} being the internal symmetry group) have been considered by Gell-Mann and co-workers.[9]

Another possibility for superspace is obtained by taking the coset SU(2, 2/1)/WS$\otimes$ SO(1, 1)$\otimes$ U(1), where WS is the familiar Wess—Zumino superalgebra of dimension 14, SO(1, 1) is the dilatation group, and U(1) is the one-parameter chiral group. This coset has

[a)]Research (Yale Report COO-3075-177) supported in part by the U.S. Energy Research and Development Administration under Contract No. EY-76-C-02-3075.

 0022-2488/78/1905-0942$1.00

dimension 8 and includes the coset $SU(2,2)/P \otimes SO(1,1)$ of the conformal group with respect to the Poincaré group extended by dilatations, which has dimension 4 and can be identified with the position space x^μ. Extended superspace can then be defined by starting from a graded Lie group G which admits $H = WS \otimes SO(1,1) \otimes U(1) \otimes G_{int}$ as a subgroup, and identifying the extended superspace with G/H. In this way one is again led to the study of nonlinear representations of G on the coset space G/H.

Among the linear representations of the classical Lie groups, very important ones are those defined on Jordan and traceless algebras. In fact the classical Lie groups can be regarded as automorphism groups of the Jordan algebras of matrices that are Hermitian with respect to the conjugation operation (denoted by a bar) in the division algebras of real ($\mathbb{R}$), complex ($\mathbb{C}$), and quaternionic ($\mathbb{Q}$) numbers, the conjugation being trivial for $\mathbb{R}$. The Jordan product is the symmetrized product

$$A \cdot B = \tfrac{1}{2}(AB + BA) \quad (A = \bar{A}^T, \quad B = \bar{B}^T),$$

under which Hermitian matrices are closed. The corresponding automorphism groups for $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{Q}$ are just the orthogonal, unitary, and symplectic groups. If we consider 3×3 octonionic matrices over $\mathbb{R}$ or $\mathbb{C}$ that are Hermitian with respect to octonionic conjugation, then we get the automorphism groups F_4 and E_6 respectively, while the exceptional group G_2 appears as the automorphism group of the octonion algebra.[10]

In this formulation the Jordan algebras of Hermitian matrices form the starting point. Lie groups arise as the automorphism groups of the Jordan algebras. Now in axiomatic quantum mechanics[11] the Jordan algebra has a direct physical interpretation as the algebra of observables that are represented by Hermitian matrices. Local observables are fields which are associated with infinite dimensional Jordan algebras. Suppose that we enlarge this traditional definition of observables by also adding fermion fields as local observables. Integrating over space we obtain nonlocal observables that can be represented by finite matrices. A fermionic charge in supersymmetry will be just such an operator. It will be associated with a matrix with entries that are Grassmann numbers and which is Hermitian with respect to the antisymmetrical metric C. Such observables will be closed in a generalized Jordan algebra. Then, supergroups will arise as automorphism groups of these Jordan superalgebras[12] of bosonic and fermionic observables. In Secs. IV and V we develop the algebras of generalized observables that admit the supergroups SU(2, 2/1) and OSp(1/4) as automorphism groups.

II. THE GLG SU(2,2/1)

We introduce a four-component object $\psi \equiv (\psi_1, \psi_2, \psi_3, \psi_4)^T$ and a one-component object ϕ, which we put together in a five-component column $X = \binom{\psi}{\phi}$. The components of ψ are complex anticommuting numbers (i.e., complex numbers whose real and imaginary parts are odd elements of a Grassmann algebra), whereas ϕ is a complex commuting number (i.e., one whose real and imaginary parts are even elements of the same algebra). Let $Y = \binom{\chi}{\omega}$ be another such five-component column, and consider the following bilinear form,

$$(X, Y) \equiv \bar{\psi}\gamma_5\chi + \phi^*\omega \equiv \psi^\dagger C\chi + \phi^*\omega. \tag{2.1}$$

Here $\psi^\dagger$ means $(\psi^T)^*$, the star acting on the imaginary unit. We have introduced a matrix $C = \gamma^0\gamma_5$, which is real and antisymmetrical.[13] Due to the anticommuting character of ψ and χ it is clear that

$$(X, Y) = (Y, X)^*. \tag{2.2}$$

We want to investigate the most general linear transformation acting on X, which transforms $X = \binom{\psi}{\phi}$ into $X' = \binom{\psi'}{\phi'}$ and which leaves the bilinear form (2.1) invariant. Let us first consider transformations which do not mix ψ and ϕ. We have

$$X' = \begin{pmatrix}\psi'\\ \phi'\end{pmatrix} = U(U, u)X = \begin{bmatrix} U & 0 \\ 0 & u \end{bmatrix}\begin{pmatrix}\psi\\ \phi\end{pmatrix}, \tag{2.3}$$

where U is a 4×4 matrix with complex commuting entries, and u is a complex commuting number. Clearly then

$$(X', Y') = \psi'^\dagger C\chi' + \phi'^*\omega'$$
$$= \psi^\dagger U^\dagger C U\chi + \phi^* u^* u\omega.$$

Since ψ, χ, ϕ, and ω are arbitrary, this will be equal to (X, Y) as given in (2.1) if and only if[14]

$$U^\dagger C U = C, \tag{2.4a}$$

$$u^* u = 1. \tag{2.5}$$

Equation (2.5) means that u is an arbitrary complex commuting number of unit "modulus." Defining a unitary matrix $M \equiv (1/\sqrt{2})(I + i\gamma_5)$ one can rewrite (2.4a) as follows,

$$(MUM^\dagger)^\dagger(MCM^\dagger)(MUM^\dagger) = MCM^\dagger. \tag{2.4b}$$

Since $MCM^\dagger = -i\rho_3$, Eq. (2.4b) shows that U is unitarily equivalent to a generalized U(2, 2) matrix. By "generalized" we mean a matrix whose entries are general complex commuting numbers instead of ordinary complex numbers. (From now on the word "generalized" will be implicit when dealing with such matrices.) If we consider a U which is infinitesimally close to the unit matrix, i.e., $U = I + \delta U$, then (2.4a) and (2.4b) imply that

$$(\delta U)^\dagger C + C(\delta U) = 0, \tag{2.6a}$$

$$(M\delta UM^\dagger)^\dagger\rho_3 + \rho_3(M\delta UM^\dagger) = 0. \tag{2.6b}$$

Hence $M\delta UM^\dagger$ belongs to the Lie algebra of U(2, 2). It is interesting that any matrix U satisfying Eq. (2.4a) can be obtained from an element of the Lie algebra [i.e., a matrix satisfying (2.6a)] through exponentiations. This is shown in the Appendix. Also, the set of all U's of the form $\exp(\alpha I)$, where α is a purely imaginary commuting number, forms an invariant subgroup of (generalized) U(2, 2). The quotient group is SU(2, 2), the conformal group. Summarizing, Eqs. (2.4a) and (2.5) imply that the group of the transformations that leave (2.1) invariant and do not mix ψ and ϕ is $U(2,2) \otimes U(1)$, an invariant subgroup of which is $SU(2,2) \otimes U(1)$.

Let us now look at those linear transformations of X into X' which leave (2.1) invariant but mix ψ and ϕ. It is convenient to consider infinitesimal transformations first. We write

$$\begin{pmatrix}\psi'\\ \phi'\end{pmatrix}=\begin{bmatrix}I & \mathcal{E}_1\\ \mathcal{E}_2^\dagger & 1\end{bmatrix}\begin{pmatrix}\psi\\ \phi\end{pmatrix}, \tag{2.7}$$

where $\mathcal{E}_1$ and $\mathcal{E}_2$ are four-columns made out of infinitesimal complex anticommuting numbers. From (2.7) we get (to first order)

$$\begin{aligned}(X', Y') &= (\psi^\dagger+\mathcal{E}_1^\dagger\phi^*)C(\chi+\mathcal{E}_1\omega)+(-\psi^\dagger\mathcal{E}_2+\phi^*)(\mathcal{E}_2^\dagger\chi+\omega)\\ &= \psi^\dagger C\chi+\phi^*\omega+\psi^\dagger(C\mathcal{E}_1-\mathcal{E}_2)\omega+\phi^*(\mathcal{E}_1^\dagger C+\mathcal{E}_2^\dagger)\chi.\end{aligned}$$

Again, since ψ, χ, ϕ, and ω are arbitrary, this is equal to (X, Y) if and only if $-\mathcal{E}_2^\dagger=\mathcal{E}_1^\dagger C$. Thus the infinitesimal generator is simply

$$\begin{bmatrix}0 & \mathcal{E}_1\\ -\mathcal{E}_1^\dagger C & 0\end{bmatrix}. \tag{2.8}$$

We are now in a position to evaluate the finite group elements corresponding to generators of the form (2.8). We write

$$\Theta=\begin{bmatrix}0 & \theta\\ -\theta^\dagger C & 0\end{bmatrix}, \tag{2.9}$$

with θ finite. The group element generated by Θ is then simply obtained through exponentiation, it is $\exp(\Theta)$. In evaluating the exponential, one takes advantage of the fact that the components of θ are anticommuting numbers. Since there are four components, each one involving two anticommuting numbers (the real and imaginary parts), no term in the exponential contributes which is of degree higher than eight in the components of θ. Thus we can write

$$\exp(\Theta)=\sum_{n=0}^{8}\frac{1}{n!}\Theta^n.$$

An explicit computation shows that

$$\exp(\Theta)=\begin{bmatrix}I-\theta\theta^\dagger C\left(\frac{1}{a}(1-\cos\sqrt{a})\right) & \left(\frac{1}{\sqrt{a}}\sin\sqrt{a}\right)\theta\\ \left(\frac{-1}{\sqrt{a}}\sin\sqrt{a}\right)\theta^\dagger C & \cos\sqrt{a}\end{bmatrix}, \tag{2.10}$$

where $a=a(\theta)\equiv\theta^\dagger C\theta$. By $(1/\sqrt{a})\sin\sqrt{a}$, for instance, we mean the series expansion of this function up to the eighth power in the components of θ, or, equivalently, up to the fourth power in a.

Equation (2.10) gives the group element which corresponds to the generator (2.9). On can check directly that a transformation of X into X' by means of (2.10) leaves the bilinear form (2.1) invariant.

As is shown in the Appendix, a general element of the group of all linear transformations $X\to X'$ leaving (2.1) invariant can now be obtained as the product of a transformation of the form (2.3) (an "even" transformation) with one of the form (2.10) (an "odd" transformation). From this general group one can extract a "one-parameter" invariant subgroup by taking all the elements of the form $\exp(\alpha I^{(5)})$, where α is a purely imaginary commuting number and $I^{(5)}$ is the identity matrix in five dimensions. Let λ be the 5×5 matrix given by

$$\lambda=\begin{bmatrix}-I^{(4)} & 0\\ 0 & 1\end{bmatrix}. \tag{2.11}$$

The generators of the quotient group are then those generators $\{\mathcal{O}\}$ of the general group for which $\mathrm{Tr}(\mathcal{O}\lambda)=0$. The quotient group has $SU(2,2)\otimes U(1)$ for its even part, and the same odd part as the general group. It is in fact the graded Lie group $SU(2,2/1)$.[15] The odd transformations then coincide with the coset space $SU(2,2/1)/SU(2,2)\otimes U(1)$. One can define a representation of the subgroup $SU(2,2)\otimes U(1)$ on the coset space as follows:

$$\exp(\Theta)\to\begin{bmatrix}U & 0\\ 0 & u\end{bmatrix}\exp(\Theta)\begin{bmatrix}U^{-1} & 0\\ 0 & u^*\end{bmatrix}=\exp(\Theta'), \tag{2.12}$$

where, as is easily shown

$$\Theta'\equiv\begin{bmatrix}0 & \theta'\\ -\theta'^\dagger C & 0\end{bmatrix},\quad \theta'=u^*U\theta. \tag{2.13}$$

Hence θ transforms by a phase under U(1), and under SU(2,2) it transforms like the four-dimensional spinor representation.

For later use we give another, more convenient parametrization of Eq. (2.10). We introduce two objects ζ and σ as

$$\zeta=\zeta(\theta)\equiv\left(\frac{1}{\sqrt{a}}\tan\sqrt{a}\right)\theta, \tag{2.14a}$$

$$\sigma^2=\sigma(\zeta)^2\equiv(1+\zeta^\dagger C\zeta)^{-1}. \tag{2.14b}$$

Here σ is the "positive" square root of σ^2, i.e., the one whose series expansion starts with $+1$. A simple calculation now shows that (2.10) can be rewritten as

$$\exp(\Theta)\equiv K(\zeta)=\begin{bmatrix}I-\frac{\sigma^2}{1+\sigma}\zeta\zeta^\dagger C & \sigma\zeta\\ -\sigma\zeta^\dagger C & \sigma\end{bmatrix}. \tag{2.15}$$

Furthermore, one easily checks that

$$[K(\zeta)]^{-1}=K(-\zeta). \tag{2.16}$$

We also introduce a 4×4 matrix $k(\zeta)$ defined as

$$\sigma(\zeta)k(\zeta)=I-\frac{\sigma^2}{1+\sigma}\zeta\zeta^\dagger C. \tag{2.17}$$

It satisfies the relation

$$[\sigma(\zeta)k(\zeta)]^{-1}=I+\frac{\sigma}{1+\sigma}\zeta\zeta^\dagger C. \tag{2.18}$$

We close this section by extracting from SU(2,2/1) the important subgroup OSp(1/4). We take the odd generator Θ appearing in Eq. (2.9) and restrict the four-component complex spinor θ in it by means of a Majorana type constraint, i.e.,

$$\theta^c\equiv i\gamma^2\theta^*=\theta. \tag{2.19}$$

Then θ depends on four, rather than eight, anticommuting parameters. In the finite group element (2.15), ζ also satisfies the Majorana condition. As regards the even elements of SU(2,2/1), we must restrict them to that subgroup of $SU(2,2)\otimes U(1)$ which preserves the Majorana character of θ under (2.12) and (2.13). For this to be the case one must have

$$u=1,\quad (i\gamma^2)U^*(i\gamma^2)=U. \tag{2.20}$$

With the help of the matrix $M=(1/\sqrt{2})(I+i\gamma_5)$ introduced earlier, one can rewrite the second part of (2.20) as follows,

$$MUM^\dagger=(Mi\gamma^2M)(M^\dagger U^*M)(M^\dagger i\gamma^2M^\dagger)$$
$$=(i\gamma^2)(M^\dagger U^*M)(i\gamma^2). \tag{2.21}$$

For a (finite or infinitesimal) generator δU of U, Eq. (2.21) implies that

$$M\delta UM^\dagger=(i\gamma^2)(M\delta UM^\dagger)^*(i\gamma^2). \tag{2.22}$$

Thus the even generators of the subgroup we are considering must satisfy Eq. (2.22) together with (2.6b). The latter means that the generators are arbitrary linear combinations, with real commuting coefficients, of the following sixteen matrices:

$$iI,\quad i\gamma^\mu,\quad i\sigma^{\mu\nu},\quad i\gamma^\mu\gamma_5,\quad \gamma_5.$$

The former restricts these to $i\gamma^\mu$ and $i\sigma^{\mu\nu}$, i.e., to ten matrices. But it is well known that $i\gamma^\mu$ and $i\sigma^{\mu\nu}$ generate the de Sitter group SO(3,2), which has Sp(4) as its covering group. Hence we have a subset of SU(2,2/1) whose odd part essentially consists of Majorana spinors and whose even part is the group Sp(4). This is in fact the graded Lie group OSp(1/4).

It is clear from the foregoing that, with our conventions for Dirac matrices, the fifteen matrices $i\gamma^\mu$, $i\sigma^{\mu\nu}$, $i\gamma^\mu\gamma_5$, and γ_5 generate the conformal group. The translation generators correspond to $\frac{1}{2}(I+\gamma_5)\gamma^\mu$. In this connection it may be worthwhile to point out that γ_5 has in fact nothing to do with the pseudoscalar operator. The four-component spinor on which the four-dimensional representation of SU(2,2) acts is called a twistor and when γ_5 is diagonal it decomposes under Sl(2,c) into two two-component spinors. One of these is a genuine spinor that is invariant under translations, whereas the other one changes under translations. The matrix γ_5 leaves the latter invariant, and changes the sign of the former.

III. NONLINEAR REALIZATION OF SU(2,2/1) AND OSp(1/4)

We mentioned in Sec. II, and show in the Appendix, that a general SU(2,2/1) element Ω can be written as the product of an odd times an even transformation, namely

$$\Omega=K(\zeta)/\!/(U,u) \tag{3.1}$$

with $K(\zeta)$ as in Eq. (2.15) and $/\!/(U,u)$ as in (2.3) [U and u being restricted to SU(2,2)$\otimes$ U(1)]. The group elements $K(\zeta)$ parametrize the (left) cosets of the decomposition of SU(2,2/1) with respect to its subgroup SU(2,2)$\otimes$ U(1). In Eqs. (2.12) and (2.13) we have shown that one can define a linear representation of SU(2,2) $\otimes$ U(1) on this coset space. In this section we will construct a representation of the full group SU(2,2/1) on the coset space. Such a representation will no longer be linear.

We start by considering the following problem. Given two elements Ω and Ω_1 of SU(2,2/1) what is the decomposition, analogous to (3.1), of their product $\Omega_1'=\Omega\Omega_1$? In other words, if Ω is given by (3.1) and Ω_1 is equal to $K(\xi)/\!/(V,v)$, find the objects ξ', V', and v' such that

$$\Omega_1'=K(\xi')/\!/(V',v')=\Omega\Omega_1=K(\zeta)/\!/(U,u)K(\xi)/\!/(V,v). \tag{3.2}$$

Let us first insert a conveniently written unit matrix in the right-hand side of (3.2). We have

$$\Omega\Omega_1=K(\zeta)/\!/(U,u)K(\xi)/\!/^{-1}(U,u)/\!/(U,u)/\!/(V,v).$$

Clearly, $/\!/(U,u)/\!/(V,v)$ is equal to $/\!/(UV,uv)$. Making use of Eqs. (2.12) and (2.13), together with the fact that ζ and θ [as given in (2.15) and (2.13)] both transform in the same way under SU(2,2)$\otimes$ U(1), we get

$$\Omega\Omega_1=K(\zeta)K(u^*U\xi)/\!/(UV,uv). \tag{3.3}$$

Define $\tilde{\xi}$ as $\tilde{\xi}\equiv u^*U\xi$. If we suceed in writing the product $K(\zeta)K(\tilde{\xi})$ as

$$K(\zeta)K(\tilde{\xi})=K(\xi_0)/\!/(W,w), \tag{3.4}$$

then the decomposition (3.2) of Ω_1' will result with

$$\xi'=\xi_0,\quad V'=WUV,\quad v'=wuv. \tag{3.5}$$

So we turn our attention to the decomposition (3.4). We write both the left- and right-hand sides of Eq. (3.4) explicitly. From (2.15), (2.17), and (2.3) we have:

$$K(\zeta)K(\tilde{\xi})=\sigma(\zeta)\sigma(\tilde{\xi})\begin{bmatrix} k(\zeta) & \zeta \\ -\zeta^\dagger C & 1\end{bmatrix}\begin{bmatrix} k(\tilde{\xi}) & \tilde{\xi} \\ -\tilde{\xi}^\dagger C & 1\end{bmatrix}, \tag{3.6}$$

$$K(\xi_0)/\!/(W,w)=\sigma(\xi_0)\begin{bmatrix} k(\xi_0) & \xi_0 \\ -\xi_0^\dagger C & 1\end{bmatrix}\begin{bmatrix} W & 0 \\ 0 & w\end{bmatrix}. \tag{3.7}$$

Equating corresponding matrix elements in (3.6) and (3.7) yields the following relations:

$$\sigma(\zeta)\sigma(\tilde{\xi})\{k(\zeta)k(\tilde{\xi})-\zeta\tilde{\xi}^\dagger C\}=\sigma(\xi_0)k(\xi_0)W, \tag{3.8}$$

$$\sigma(\zeta)\sigma(\tilde{\xi})\{k(\zeta)\tilde{\xi}+\zeta\}=\sigma(\xi_0)\xi_0 w, \tag{3.9}$$

$$\sigma(\zeta)\sigma(\tilde{\xi})\{-\zeta^\dagger Ck(\tilde{\xi})-\tilde{\xi}^\dagger C\}=\sigma(\xi_0)(-\xi_0^\dagger C)W, \tag{3.10}$$

$$\sigma(\zeta)\sigma(\tilde{\xi})\{-\zeta^\dagger C\tilde{\xi}+1\}=\sigma(\xi_0)w. \tag{3.11}$$

From (3.8)–(3.11) it is now possible to determine ξ_0, W, and w explicitly. To get ξ_0, we substitute $\sigma(\xi_0)w$ as given by (3.11) into (3.9), and obtain in a straightforward way

$$\xi_0=[1-\zeta^\dagger C\tilde{\xi}]^{-1}\{k(\zeta)\tilde{\xi}+\zeta\}. \tag{3.12}$$

As regards w we expect that it is a complex commuting number of modulus unity. Its phase is then determined by Eq. (3.11). As a consistency check, let us verify indeed $ww^*=1$. From (3.12) we have

$$\xi_0^\dagger C\xi_0=\{(1-\zeta^\dagger C\tilde{\xi})^{*-1}(1-\zeta^\dagger C\tilde{\xi})^{-1}\}$$
$$\times\{[\tilde{\xi}^\dagger k(\zeta)^\dagger+\zeta^\dagger]C[k(\zeta)\tilde{\xi}+\zeta]\}. \tag{3.13}$$

With the help of the following identities:

$$k(\zeta)^\dagger Ck(\zeta)=\frac{1}{\sigma(\zeta)^2}[C-\sigma(\zeta)^2C\zeta\zeta^\dagger C], \tag{3.14}$$

$$\zeta^\dagger Ck(\zeta)=\zeta^\dagger C, \tag{3.15}$$

$$k(\zeta)^\dagger C\zeta=C\zeta, \tag{3.16}$$

one obtains, after some algebraic manipulations,

$$1+\xi_0^\dagger C\xi_0\equiv\frac{1}{\sigma(\xi_0)^2}$$
$$=\frac{1}{[1-\zeta^\dagger C\tilde{\xi}][1-\zeta^\dagger C\tilde{\xi}]^*}\frac{1}{\sigma(\zeta)^2}\frac{1}{\sigma(\tilde{\xi})^2}. \tag{3.17}$$

A comparison of Eq. (3.17) with (3.11) then shows that indeed $ww^*=1$.

There remains to determine W. This is done most easily by means of Eq. (3.8), together with identity (2.18). We obtain

$$W=\sigma(\zeta)\sigma(\tilde{\xi})\left(I+\frac{\sigma(\xi_0)}{I+\sigma(\xi_0)}\,\xi_0\xi_0^\dagger C\right)[k(\zeta)k(\tilde{\xi})-\zeta\tilde{\xi}^\dagger C], \tag{3.18}$$

with ξ_0 and $\sigma(\xi_0)$ given in terms of ζ and $\tilde{\xi}$ by Eqs. (3.12) and (3.17).

It is not difficult to check that W as given in (3.18) is consistent with Eq. (3.10). By means of rather tedious algebraic manipulations one can also verify explicitly that W indeed satisfies (2.4a).

So we have obtained the form of the objects ξ_0, W, and w appearing in the decomposition (3.4). Thus, through Eq. (3.5), the more general decomposition (3.2) also results. In the course of doing this calculation, perhaps interesting in itself, we have also arrived at what we were looking for at the beginning of this section, namely, a representation of the full group on the coset space. Indeed Eq. (3.12), written in full, reads

$$\xi'=[1-\zeta^\dagger C(u^*U\xi)]^{-1}\{k(\zeta)(u^*U\xi)+\zeta\}. \tag{3.19}$$

Equation (3.19) means that under the group action $\Omega=K(\zeta)U(U,u)$, the coset parametrized by ξ goes over into the one parametrized by ξ'. From the way (3.19) has been constructed, it is clear that it satisfies the group product law. This can also be checked explicitly. Under a transformation belonging to $SU(2,2)\otimes U(1)$, ξ transforms linearly as in (2.13), whereas it transforms nonlinearly under an element of $SU(2,2/1)/SU(2,2)\otimes U(1)$.

Equation (3.19) can also be arrived at in a somewhat different way. Consider the five-component column $X=\binom{\psi}{\phi}$ which was introduced in Sec. II. Under a general group transformation $\Omega=K(\zeta)U(U,u)$, X is transformed into $X'=\binom{\psi'}{\phi'}$. By means of Eqs. (2.15), (2.17), and (2.3) we get

$$\psi'=\sigma(\zeta)[k(\zeta)U\psi+\zeta u\phi], \tag{3.20a}$$

$$\phi'=\sigma(\zeta)[-\zeta^\dagger CU\psi+u\phi]. \tag{3.20b}$$

Now if we define new variables ξ and ξ' as

$$\xi=\psi/\phi,\quad \xi'=\psi'/\phi', \tag{3.21}$$

we obtain, upon taking the ratio of (3.20a) over (3.20b),

$$\xi'=\frac{k(\zeta)U\psi+\zeta u\phi}{-\zeta^\dagger CU\psi+u\phi}=\frac{k(\zeta)(u^*U\xi)+\zeta}{1-\zeta^\dagger C(u^*U\xi)}.$$

This is the same as Eq. (3.19). That is, the nonlinear transformation properties of the coset space $SU(2,2/1)/SU(2,2)\otimes U(1)$ coincide with those of ψ/ϕ, where $\binom{\psi}{\phi}$ transforms as the fundamental linear representation of $SU(2,2/1)$.

Clearly, all the calculations we have done in this section with the group $SU(2,2/1)$ can be carried to its subgroup $OSp(1/4)$. The spinors ζ, ξ, and ξ' are then Majorana spinors; u, v, v', and w are equal to 1; and U, V, V', and W satisfy (2.20). Equation (3.19) then provides a nonlinear representation of $OSp(1/4)$ on the coset space $OSp(1/4)/SO(3,2)$.

IV. JORDAN AND TRACELESS (WEDGE) ALGEBRAS FOR SU(2,2/1)

We have seen that the action of a general element Ω of $SU(2,2/1)$ on the five-component column X is such that

$$X\to X'=\Omega X=KUX. \tag{4.1}$$

The matrices Ω form a five-dimensional[16] irreducible representation of $SU(2,2/1)$. Another, inequivalent five-dimensional irreducible representation can be obtained by considering the transformation properties of $X^\dagger$ instead of X. One has

$$X^\dagger\to(X')^\dagger=(KUX)^\dagger=X^\dagger U^\dagger\bar{K}=X^\dagger\bar{\Omega}, \tag{4.2}$$

where, as one easily finds

$$\bar{K}\equiv\bar{K}(\zeta)=\begin{bmatrix} I-\dfrac{\sigma^2}{1+\sigma}\,C\zeta\zeta^\dagger & -\sigma C\zeta \\ \sigma\zeta^\dagger & \sigma \end{bmatrix}. \tag{4.3}$$

$\bar{K}(\zeta)$ differs from $K(\zeta)^\dagger$ in the sign of its upper right-hand entry. This comes from the anticommutation properties of ζ and ψ. We note the following identity,

$$\bar{K}(\zeta)\equiv K(-C\zeta). \tag{4.4}$$

We now introduce a 5×5 unitary matrix $\mathcal{C}$ defined as

$$\mathcal{C}\equiv\begin{bmatrix} C & 0 \\ 0 & 1 \end{bmatrix}. \tag{4.5}$$

Since Ω is such that it leaves the bilinear form (2.1) invariant, It is clear that

$$\bar{\Omega}\mathcal{C}\Omega=\mathcal{C}. \tag{4.6}$$

From (4.2) we obtain for the transformation properties of $X^\dagger\mathcal{C}$

$$X^\dagger\mathcal{C}\to(X')^\dagger\mathcal{C}=X^\dagger\bar{\Omega}\mathcal{C}=(X^\dagger\mathcal{C})\Omega^{-1}, \tag{4.7}$$

the last step following from Eq. (4.6). Clearly the transformation properties of $X^\dagger\mathcal{C}$ are unitarily equivalent to those of $X^\dagger$.

We can consider the direct product of the 5 and $\bar{5}$ representations, which we write as follows,

$$\Xi=XX^\dagger\mathcal{C}=\begin{bmatrix} \psi\psi^\dagger C & \psi\phi^* \\ \phi\psi^\dagger C & \phi\phi^* \end{bmatrix}. \tag{4.8}$$

Under the group action (4.1) we have

$$\Xi\to\Xi'=X'(X^\dagger\mathcal{C})'=\Omega XX^\dagger\mathcal{C}\Omega^{-1}=\Omega\Xi\Omega^{-1}. \tag{4.9}$$

Thus Ξ transforms as a 25-dimensional representation of $SU(2,2/1)$. Such a representation however is not irreducible. Indeed, recalling the 5×5 matrix λ defined in Eq. (2.11), one easily finds that

$$\mathrm{Tr}(\Xi\lambda)\equiv\mathrm{Tr}(XX^\dagger\mathcal{C}\lambda)=X^\dagger\mathcal{C}X\equiv(X,X). \tag{4.10}$$

Hence $\mathrm{Tr}(\Xi\lambda)$ is invariant under the group action (4.9), which means that it transforms as an irreducible part of Ξ. The 5×5 matrix Ξ_0, defined as

$$\Xi_0 \equiv \Xi + \tfrac{1}{3} I \,\mathrm{Tr}(\Xi\lambda) \tag{4.11}$$

then satisfies

$$\mathrm{Tr}(\Xi_0\lambda) = 0, \tag{4.12}$$

and it transforms irreducibly like the 24-dimensional adjoint representation of SU(2, 2/1).[17]

Let us rewrite the components of Ξ given in Eq. (4.8) as follows,

$$\Xi = \begin{bmatrix} \psi\psi^\dagger C & \psi\varphi^* \\ \varphi\psi^\dagger C & \varphi\varphi^* \end{bmatrix} = \begin{bmatrix} AC & B \\ B^\dagger C & D \end{bmatrix}. \tag{4.13a}$$

Here we see that A is a 4×4 matrix made out of complex commuting numbers (each entry is the product of two anticommuting numbers), D is a commuting number, and B is a four-column made out of anticommuting numbers. Furthermore we have

$$A^\dagger = -A, \tag{4.14}$$

$$D^\dagger = D. \tag{4.15}$$

We note however that the matrix (4.13a) has been constructed in a very special way, namely, as the product of X times $X^\dagger C$. We now consider a general 5×5 matrix Ξ, defined as

$$\Xi \equiv \begin{bmatrix} AC & B \\ B^\dagger C & D \end{bmatrix}, \tag{4.13b}$$

the only requirements on A, B, and D being that (i) Eqs. (4.14) and (4.15) be satisfied, (ii) A and D be made out of complex commuting numbers, and (iii) B be made out of complex anticommuting numbers. Such a matrix depends in general on 25 (commuting or anticommuting) parameters. We denote by $\mathcal{H}$ the space of all such matrices.

Let Ξ_1 and Ξ_2 be two elements of $\mathcal{H}$. We define the following symmetrical product of Ξ_1 and Ξ_2, which we denote by a dot,

$$\Xi_1 \circ \Xi_2 = \tfrac{1}{2}(\Xi_1\Xi_2 + \Xi_2\Xi_1). \tag{4.16}$$

Let us check that this product closes in $\mathcal{H}$. We have

$$\Xi_1 \circ \Xi_2 = \frac{1}{2}\left\{\begin{pmatrix} A_1C & B_1 \\ B_1^\dagger C & D_1 \end{pmatrix}\begin{pmatrix} A_2C & B_2 \\ B_2^\dagger C & D_2 \end{pmatrix} + \begin{pmatrix} A_2C & B_2 \\ B_2^\dagger C & D_2 \end{pmatrix}\begin{pmatrix} A_1C & B_1 \\ B_1^\dagger C & D_1 \end{pmatrix}\right\}$$

$$= \frac{1}{2}\begin{pmatrix} (A_1CA_2 + B_1B_2^\dagger + A_2CA_1 + B_2B_1^\dagger)C & (A_1CB_2 + B_1D_2 + A_2CB_1 + B_2D_1) \\ (B_1^\dagger CA_2 + D_1B_2^\dagger + B_2^\dagger CA_1 + D_2B_1^\dagger)C & (B_1^\dagger CB_2 + D_1D_2 + B_2^\dagger CB_1 + D_2D_1) \end{pmatrix}$$

$$= \begin{pmatrix} AC & B \\ B_0^\dagger C & D \end{pmatrix}.$$

From the fact that Ξ_1 and Ξ_2 are of the form (4.13b) with entries satisfying the requirements (i), (ii), and (iii), it is not difficult to show that $\Xi_1 \circ \Xi_2$ is also of the form (4.13b) (i.e., $B_0 = B$) and that A, B, and D satisfy the three requirements.

Having defined such a symmetrical product we can easily check that it satisfies the Jordan identity, namely

$$\Xi_1 \circ \{\Xi_2 \circ (\Xi_1 \circ \Xi_1)\} = (\Xi_1 \circ \Xi_2) \circ (\Xi_1 \circ \Xi_1). \tag{4.17}$$

That is, we have endowed the space $\mathcal{H}$ with the structure of a Jordan algebra.

Let Ω be an element of SU(2, 2/1). In analogy with Eq. (4.9) we define a mapping of $\mathcal{H}$ onto $\mathcal{H}$ as follows,

$$\Xi \to \Xi' = \Omega\Xi\Omega^{-1}. \tag{4.18}$$

That such a mapping sends an element of $\mathcal{H}$ into another element of $\mathcal{H}$ can be seen by decomposing Ω as $K\mathcal{U}$ and verifying explicitly that both $\mathcal{U}\Xi\mathcal{U}^{-1}$ and $K\Xi K^{-1}$ belong to $\mathcal{H}$. The fact that the mapping (4.18) is "onto" follows immediately from the observation that an element Ξ_1 in $\mathcal{H}$ is obtained through (4.18) from the element $\Omega^{-1}\Xi_1\Omega$.

Clearly, the mapping (4.18) is an automorphism of the Jordan product (4.16). Indeed we have, for any Ξ_1 and Ξ_2 in $\mathcal{H}$

$$\begin{aligned} \Xi_1' \circ \Xi_2' &= \tfrac{1}{2}(\Xi_1'\Xi_2' + \Xi_2'\Xi_1') \\ &= \tfrac{1}{2}(\Omega\Xi_1\Omega^{-1}\Omega\Xi_2\Omega^{-1} + \Omega\Xi_2\Omega^{-1}\Omega\Xi_1\Omega^{-1}) \\ &= \tfrac{1}{2}\Omega(\Xi_1\Xi_2 + \Xi_2\Xi_1)\Omega^{-1} \\ &= (\Xi_1 \circ \Xi_2)'. \end{aligned} \tag{4.19}$$

Hence any Ω belonging to SU(2, 2/1) induces an automorphism of the Jordan algebra defined by (4.18), and SU(2, 2/1) is a group of these automorphisms.

One can define another algebra, which is reminiscent of the wedge algebra of Gell-Mann, Michel, and Radicati,[8] that also admits SU(2, 2/1) as an automorphism group. We first consider the subspace $\mathcal{H}_0$ of the space $\mathcal{H}$ which is spanned by all the elements Ξ_0 of $\mathcal{H}$ for which [recall Eq. (4.12)]

$$\mathrm{Tr}(\Xi_0\lambda) = 0.$$

Now for any Ω in SU(2, 2/1) and any Ξ in $\mathcal{H}$ the following identity holds,

$$\mathrm{Tr}\{\Omega\Xi\Omega^{-1}\lambda\} = \mathrm{Tr}\{\Xi\lambda\}. \tag{4.20}$$

Thus the mapping (4.18) sends an arbitrary element of $\mathcal{H}_0$ into some other element of $\mathcal{H}_0$. Let Ξ_{01} and Ξ_{02} both belong to $\mathcal{H}_0$. We define the following product of these two elements, which is manifestly closed in $\mathcal{H}_0$,

$$\Xi_{01} \vee \Xi_{02} \equiv \Xi_{01} \circ \Xi_{02} + \tfrac{1}{3} I \,\mathrm{Tr}\{\Xi_{01} \circ \Xi_{02}\lambda\}. \tag{4.21}$$

From Eqs. (4.19) and (4.20) it is clear that the mapping (4.18) is an automorphism of this product. Hence SU(2, 2/1) is a group of automorphisms of the wedge algebra.

We have seen that SU(2, 2/1) is a group of automorphisms of both the Jordan and the wedge algebras. Strictly speaking however it is not the most general group of such automorphisms. Indeed, as one can convince onself of, any automorphism g of the Grassmann algebra will also induce an automorphism of the algebra defined by (4.16) and (4.21). Moreover, g is different from any mapping of the type (4.18). Let G denote the group of all the automorphisms of the Grassmann algebra, and consider the product of two automorphisms g and Ω of the Jordan and wedge algebras. Clearly this product is also an automorphism, and one can show that the set of all such products forms a group $\mathcal{G}$. It may be interesting to point out that $\mathcal{G}$ is in fact the semidirect product $G \circledS \mathrm{SU}(2, 2/1)$. That is, both G and SU(2, 2/1) are subgroups of $\mathcal{G}$. The latter is an invariant subgroup with the property that $\mathcal{G}/\mathrm{SU}(2, 2/1) \sim G$, whereas the former is not an invariant subgroup.

The set of all the elements Ξ_0 of the wedge algebra spans a space on which the group SU(2, 2/1) is rep-

resented as in (4.18). That representation however is equivalent to the adjoint representation. Indeed, recalling Eq. (4.13b) and the defining properties of Ξ_0 one can see that any object of the form $i\Xi_0$ is nothing but a generator of SU(2, 2/1) on which the adjoint representation is defined as $i\Xi_0 \to i\Xi_0' = \Omega(i\Xi_0)\Omega^{-1}$. This situation is very reminiscent of the one encountered in dealing with SU(n), when the wedge algebra is defined on those same λ-matrices that generate the group. In the case of the symplectic or orthogonal groups however, symmetric and antisymmetric algebras which admit Sp($2n$) and SO(n) as automorphism groups are most simply constructed on different spaces. Analogously, we will consider in the next section the example of the GLG OSp(1/4) and construct, on spaces different from the space of the generators, Jordan and wedge algebras which will indeed admit OSp(1/4) as a group of automorphisms.

V. JORDAN AND WEDGE ALGEBRAS FOR OSp(1/4)

We recall the way the subgroup OSp(1/4) of SU(2, 2/1) was obtained in Sec. II. For the even part we take all the elements of the form (2.3) which satisfy Eq. (2.20) in addition to (2.4a) and (2.5). Explicitly

$$U = \begin{bmatrix} U & 0 \\ 0 & 1 \end{bmatrix}, \tag{5.1}$$

$$U^\dagger C U = C, \quad (i\gamma^2)U^*(i\gamma^2) = U. \tag{5.2}$$

On the other hand, the odd part of OSp(1/4) is spanned by elements like (2.15), with ζ satisfying the Majorana constraint. In this case (2.15) can be re-expressed as

$$K(\zeta) = \begin{bmatrix} I - \dfrac{\sigma^2}{1+\sigma}\zeta\zeta^T(i\gamma^2 C) & \sigma\zeta \\ -\sigma\zeta^T(i\gamma^2 C) & \sigma \end{bmatrix} \tag{5.3}$$

with

$$\zeta = i\gamma^2\zeta^*, \quad \sigma^2 = [1 + \zeta^T(i\gamma^2 C)\zeta]^{-1}. \tag{5.4}$$

A general element of OSp(1/4) can then be written as $\Omega = K(\zeta)U(U)$, and it is not difficult to show that it leaves the following bilinear form invariant,

$$(X, Y) = \psi^T(i\gamma^2 C)\chi + \phi\omega. \tag{5.5}$$

Here ψ and χ are four-component Majorana spinors, whereas ϕ and ω are one-component real commuting scalars.

The adjoint representation of OSp(1/4) can be realized on the elements of the corresponding GLA. Indeed the odd and even elements of the GLA are respectively

$$\Theta = \begin{bmatrix} 0 & \theta \\ -\theta^T(i\gamma^2 C) & 0 \end{bmatrix}, \quad L = \begin{bmatrix} L & 0 \\ 0 & 0 \end{bmatrix}, \tag{5.6}$$

with θ and L satisfying

$$\theta = i\gamma^2\theta^*, \quad L^\dagger C + CL = 0, \quad (i\gamma^2)L^*(i\gamma^2) = L. \tag{5.7}$$

If $\mathcal{O}$ stands for an arbitrary element of the GLA one obtains a realization of the adjoint representation of the group as

$$\mathcal{O} \to \mathcal{O}' = \Omega\mathcal{O}\Omega^{-1}. \tag{5.8}$$

Each such mapping of the GLA onto itself is in fact an automorphism.

The Jordan and wedge algebras for OSp(1/4) can be constructed in a way very similar to the corresponding construction for SU(2, 2/1). We start with the five-component column $X = \begin{pmatrix} \psi \\ \phi \end{pmatrix}$, where ψ is a Majorana spinor and ϕ a real commuting scalar. The group OSp(1/4) can then be represented on X as $X \to X' = \Omega X$. Introducing a 5×5 matrix Γ as

$$\Gamma \equiv \begin{bmatrix} i\gamma^2 & 0 \\ 0 & 1 \end{bmatrix}, \tag{5.9}$$

one then finds for the transformation properties of $X^T\Gamma C$,

$$X^T\Gamma C = X^\dagger C \to X^\dagger\bar{\Omega}C = X^\dagger C\Omega^{-1} = (X^T\Gamma C)\Omega^{-1}. \tag{5.10}$$

We now consider the set of all 5×5 matrices Ξ of the form

$$\Xi = \begin{bmatrix} A(i\gamma^2 C) & B \\ B^T(i\gamma^2 C) & D \end{bmatrix}, \tag{5.11}$$

where A and D are made out of commuting numbers, B is made out of anticommuting numbers, and A, B, and D satisfy

$$A^T = -A, \quad (i\gamma^2)A^*(i\gamma^2) = A, \tag{5.12}$$

$$B = i\gamma^2 B^*, \tag{5.13}$$

$$D = D^*. \tag{5.14}$$

We see that a general matrix Ξ depends on eleven parameters. Clearly, the matrix $XX^T\Gamma C$ is of the form (5.11). Given two matrices Ξ_1 and Ξ_2 of the form (5.11), their symmetrized product

$$\Xi_1 \circ \Xi_2 = \tfrac{1}{2}(\Xi_1\Xi_2 + \Xi_2\Xi_1) \tag{5.15}$$

is also of that form. The Jordan identity is easily checked, which means that we have a Jordan algebra of the Ξ-matrices. One can then define a mapping of the set of these matrices onto itself as follows,

$$\Xi \to \Xi' = \Omega\Xi\Omega^{-1}, \tag{5.16}$$

Ω being an arbitrary element of OSp(1/4). Clearly, such a mapping is an automorphism of the Jordan algebra.

A wedge algebra can also be defined in this case, again by considering the set of all matrices Ξ_0 which are of the form (5.11) and in addition satisfy the constraint $\mathrm{Tr}(\Xi_0\lambda) = 0$. This constraint is preserved under the mapping (5.16), as well as under the wedge product

$$\Xi_{01} \vee \Xi_{02} \equiv \Xi_{01} \cdot \Xi_{02} + \tfrac{1}{3} I\,\mathrm{Tr}\{\Xi_{01} \circ \Xi_{02}\lambda\}. \tag{5.17}$$

Again, OSp(1/4) is a group of automorphisms of the wedge algebra. The space of all matrices Ξ_0 corresponds to the ten-dimensional representation of OSp(1/4) which occurs in the direct product decomposition

$$5 \otimes 5 = 1 \oplus 10 \oplus 14.$$

Since the wedge and Jordan algebras have dimensionalities 10 and 11 respectively, it is clear that they differ from the adjoint representation, of dimension 14. It is interesting to look a little bit more closely into the representation defined on the wedge algebra. To do this let us first consider the set of all those generators of SU(2, 2/1) which are orthogonal to the generators of OSp(1/4), i.e., the generators of the coset space

SU(2, 2/1)/OSp(1/4). For the odd part we want elements like (2.9) such that θ does not have any Majorana part. Clearly then $\theta = -(i\gamma^2\theta^*)$, and we have

$$\Theta_\perp = \begin{bmatrix} 0 & \theta \\ \theta^T(i\gamma^2 C) & 0 \end{bmatrix}, \quad \theta = -i\gamma^2\theta^*. \tag{5.18}$$

For the even part we want the elements of the Lie algebra of SU(2, 2)⊗U(1), other than the ones included in (5.6). They are of the form

$$\mathcal{L}_\perp = \begin{bmatrix} L & 0 \\ 0 & l \end{bmatrix}, \quad \mathrm{Tr}(\mathcal{L}_\perp \lambda) = 0, \tag{5.19}$$

with L and l satisfying

$$L^\dagger C + CL = 0, \quad (i\gamma^2)L^*(i\gamma^2) = -L, \quad l = -l^*. \tag{5.20}$$

Now take the direct sum of the spaces spanned by elements of the form $\Theta_\perp$ and $\mathcal{L}_\perp$. It is not difficult to see that this precisely corresponds to the set of all elements of the form $i\Xi_0$. This means that the elements of the wedge algebra in fact parametrize the cosets of the decomposition of SU(2, 2/1) with respect to OSp(1/4), and the representation of OSp(1/4) on the elements of the wedge algebra is in fact a representation on the coset space SU(2, 2/1)/OSp(1/4).

VI. CONCLUSION

We have shown that groups like SU(2, 2/1) and OSp(1/4), which are relevant to supergravity and supersymmetric theories in general can be expressed as finite transformations in a manifold which one may interpret as a coset space. Such supergroups can also be related to automorphisms of algebras of extended observables comprising bosons and fermions. It is apparent that much of these properties are in fact more general than the context in which we have introduced them, and can be derived for a large class of graded Lie groups. Our results may be used, for instance, to obtain nonlinear transformation laws of fields in locally supersymmetric gauge theories. We hope to pursue these lines in a future report.

APPENDIX

Let $X = \begin{pmatrix} \psi \\ \phi \end{pmatrix}$. We consider a general linear transformation of the type $X \to X' = \begin{pmatrix} \psi' \\ \phi' \end{pmatrix}$, which we write as

$$X' = \Omega X = \begin{bmatrix} A & B \\ E^\dagger & D \end{bmatrix} \begin{pmatrix} \psi \\ \phi \end{pmatrix}. \tag{A1}$$

Here A and D are made out of complex commuting numbers, whereas B and $E^\dagger$ are made out of complex anticommuting numbers. Requiring invariance of the bilinear form (2.1) under the transformation (A1) we get

$$\begin{aligned} &\psi^\dagger C\chi + \phi^*\omega \\ &\quad = \psi'^\dagger C\chi' + \phi'^*\omega' \\ &\quad = (\psi^\dagger A^\dagger + \phi^* B^\dagger)C(A\chi + B\omega) + (-\psi^\dagger E + \phi^* D^*)(E^\dagger\chi + D\omega) \\ &\quad = \psi^\dagger(A^\dagger CA - EE^\dagger)\chi + \psi^\dagger(A^\dagger CB - ED)\omega \\ &\qquad + \phi^*(B^\dagger CA + D^*E^\dagger)\chi + \phi^*(B^\dagger CB + D^*D)\omega. \end{aligned}$$

Since ψ, χ, ϕ, and ω are arbitrary it is not difficult to show that one must have[18]

$$A^\dagger CA - EE^\dagger = C, \tag{A2}$$

$$B^\dagger CB + D^*D = 1, \tag{A3}$$

$$A^\dagger CB - ED = 0 = B^\dagger CA + D^*E^\dagger. \tag{A4}$$

We now try to find the most general A, B, E, and D which satisfy these equations. In order to do so we will write down several Ansätze which are suggested by the final form we want to reach.

We know that E is a four-component anticommuting object. It is not difficult to see that there always exists another four-column ζ such that

$$E = \sigma C\zeta, \quad \sigma^2 = (1 + \zeta^\dagger C\zeta)^{-1}. \tag{A5}$$

On the other hand, A is a 4×4 matrix with complex commuting entries. Again one can show that A can always be written as

$$A = \left(I - \frac{\sigma^2}{1+\sigma}(H\zeta)(H\zeta)^\dagger C\right)H, \tag{A6}$$

where ζ and σ are the specific ones appearing in Eq. (A5). H is a general 4×4 matrix with complex commuting entries. If we substitute (A5) and (A6) into (A2) we now have

$$\begin{aligned} &C + (\sigma C\zeta)(\sigma C\zeta)^\dagger \\ &\quad = C + EE^\dagger = A^\dagger CA \\ &\quad = H^\dagger\left(I - \frac{\sigma^2}{1+\sigma}C(H\zeta)(H\zeta)^\dagger\right)C\left(I - \frac{\sigma^2}{1+\sigma}(H\zeta)(H\zeta)^\dagger C\right)H \\ &\quad = H^\dagger CH - \frac{2\sigma^2}{1+\sigma}(H^\dagger CH)\zeta\zeta^\dagger(H^\dagger CH) \\ &\qquad + \left(\frac{\sigma^2}{1+\sigma}\right)^2[(H^\dagger CH)\zeta(\zeta^\dagger H^\dagger CH\zeta)\zeta^\dagger(H^\dagger CH)]. \end{aligned} \tag{A7}$$

It is easy to see that a sufficient condition for (A7) to hold is that $H^\dagger CH = C$. We will now show that this is also a necessary condition.

Suppose that the Grassmann algebra we are working with has $2n$ generators. That is, there are $2n$ linearly independent real anticommuting numbers which are not odd degree polynomials in other anticommuting numbers. [In the SU(2, 2/1) case for instance, $n = 4$.] Then the following holds true: An arbitrary $m \times m$ commuting matrix H, provided that its complex part H_0 is not singular (by "complex part" we mean the part which is not the coefficient of any product of anticommuting numbers), can always be written as follows:

$$H = H_0H_2H_4\cdots H_{2n}, \tag{A8}$$

$$H_{2k} = I + h_{2k}, \tag{A9}$$

where h_{2k} does not contain any product of anticommuting numbers whose degree is less than $2k$. It is not difficult to convince oneself that such a decomposition holds true. Indeed if H_0, $H_2, \cdots,$ and H_{2k-2} are known, and since H_0 has an inverse, one can always find an h_{2k} such that the product $H_0H_2\cdots H_{2k}$ matches H in $(2k)$th degree of products of anticommuting numbers. But then, no further contribution in $(2k)$th degree is produced, so (A8) and (A9) follow. We note however that this decomposition is far from being unique.

The left- and right-hand sides of Eq. (A7) must agree in all degrees of products of anticommuting numbers.

Independently of the validity of (A8) and (A9), the zeroth degree resulting equation is simply

$$H_0^\dagger C H_0 = C. \tag{A10}$$

This shows, among other things, that H_0 is not singular, which means that (A8) and (A9) now hold. We now proceed by induction to show that there exists a decomposition such that $H_{2k}^\dagger C H_{2k} = C$ for all k. Let this hold up to $k = l-1$. Substitution of (A8) into (A7) then yields

$$\begin{aligned} &C + (\sigma C\zeta)(\sigma C\zeta)^\dagger \\ &= (H_{2n}^\dagger \cdots H_{2l}^\dagger C H_{2l} \cdots H_{2n}) - \frac{2\sigma^2}{1+\sigma}(H_{2n}^\dagger \cdots H_{2l}^\dagger C H_{2l} \cdots H_{2n}) \\ &\quad \times \zeta\zeta^\dagger (H_{2n}^\dagger \cdots H_{2l}^\dagger C H_{2l} \cdots H_{2n}) + \left(\frac{\sigma^2}{1+\sigma}\right)^2 \\ &\quad \times (\zeta^\dagger H_{2n}^\dagger \cdots H_{2l}^\dagger C H_{2l} \cdots H_{2n}\zeta) \\ &\quad \times (H_{2n}^\dagger \cdots H_{2l}^\dagger C H_{2l} \cdots H_{2n})\zeta\zeta^\dagger (H_{2n}^\dagger \cdots H_{2l}^\dagger C H_{2l} \cdots H_{2n}). \end{aligned}$$

To order $2l$ this is equivalent to

$$\begin{aligned} C + (\sigma C\zeta)(\sigma C\zeta)^\dagger &= H_{2l}^\dagger C H_{2l} - \frac{2\sigma^2}{1+\sigma} C\zeta\zeta^\dagger C \\ &\quad + \left(\frac{\sigma^2}{1+\sigma}\right)^2 (\zeta^\dagger C\zeta) C\zeta\zeta^\dagger C. \end{aligned}$$

But this implies that

$$H_{2l}^\dagger C H_{2l} = C + \mathcal{O}(2l+2).$$

Without loss of generality however one can take $H_{2l}^\dagger C H_{2l} = C$ exactly. Indeed, as one can convince oneself of, this simply amounts to a redefinition of H_{2l+2}, which is consistent with Eq. (A9). This in fact completes the proof by induction. The upshot of all this is that A can always be written as in Eq. (A6), H being such that

$$H^\dagger C H = C. \tag{A11}$$

But this in turn means that

$$A = \left(I - \frac{\sigma^2}{1+\sigma} H\zeta\zeta^\dagger H^\dagger C\right) H = H\left(I - \frac{\sigma^2}{1+\sigma}\zeta\zeta^\dagger C\right). \tag{A12}$$

Having parametrized A and E in terms of H and ζ we now turn to B and D. From (A12), (A11), and (2.18) one finds that

$$A^{-1} = \left(I + \frac{\sigma}{1+\sigma}\zeta\zeta^\dagger C\right)(-CH^\dagger C). \tag{A13}$$

Making use of the second part of equation (A4) we obtain through (A5) and (A13)

$$\begin{aligned} B^\dagger &= -D^* E^\dagger A^{-1} C^{-1} \\ &= -D^*(\sigma\zeta^\dagger C^\dagger)\left(I + \frac{\sigma}{1+\sigma}\zeta\zeta^\dagger C\right)(-CH^\dagger C)C^{-1} \\ &= D^*\zeta^\dagger H^\dagger. \end{aligned}$$

Hence

$$B = H\zeta D. \tag{A14}$$

With the help of Eq. (A11) we can eliminate B from (A3) and (A14). We get

$$\begin{aligned} 1 &= D^* D + D^*\zeta^\dagger H^\dagger C H\zeta D \\ &= D^* D(1 + \zeta^\dagger C\zeta) \\ &= D^* D\sigma^{-2}. \end{aligned}$$

This means that

$$D = \sigma u, \tag{A15}$$

where u is an arbitrary complex commuting number of modulus unity. We have now parametrized all four objects A, B, E, and D in terms of H, ζ, and u. Equations (A5), (A12), (A14), and (A15) can be written together as

$$\Omega = \begin{bmatrix} A & B \\ E^\dagger & D \end{bmatrix} = \begin{bmatrix} H\left(I - \frac{\sigma^2}{1+\sigma}\zeta\zeta^\dagger C\right) & \sigma u H\zeta \\ -\sigma\zeta^\dagger C & \sigma u \end{bmatrix}.$$

But, owing to (A11), this is simply equivalent to

$$\Omega = K(H\zeta)\mathcal{U}(H, u), \tag{A16}$$

with K and $\mathcal{U}$ as in (2.15) and (2.3) respectively. This completes the proof of the statement that an arbitrary transformation leaving the bilinear form (2.1) invariant can be written as the product of an odd times an even transformation.

The other result we want to establish here is the following: Given a 4×4 matrix H, made out of complex commuting numbers and satisfying Eq. (2.4a), then it can be written in the form $e^{h_0}e^h$, where h_0 is a 4×4 matrix made out of complex numbers, h is a 4×4 matrix made out of complex commuting numbers of at least second degree in products of anticommuting numbers, and both h_0 and h satisfy Eq. (2.6a). The proof is fairly straightforward, relying on similar known results for ordinary Lie groups and complex numbers, so we only outline it. First it is not difficult to show that any matrix [not necessarily satisfying (2.4a)], provided its "complex part" is not singular, can be written in the form $e^{h_0}e^h$, with h_0 and h as indicated [but not necessarily satisfying (2.6a)]. The matrix h is unique. So let H now satisfy (2.4a) and let $H = e^{h_0}e^h$. We want to show that h_0 and h satisfy (2.6a). We have

$$e^{h^\dagger}e^{h_0^\dagger} C e^{h_0}e^h = C,$$

which implies that

$$e^{h_0^\dagger} C e^{h_0} = C, \tag{A17}$$

$$e^{h^\dagger} C e^h = C. \tag{A18}$$

From Eq. (A17) and well known results of classical Lie group theory, we see that h_0 satisfies Eq. (2.6a). On the other hand (A18) can be rewritten as

$$C^{-1}e^{h^\dagger} C e^h = I,$$

from which we obtain that

$$\exp(C^{-1}h^\dagger C)\exp(h) = I.$$

Multiplying on the right with e^{-h} we now have

$$\exp(C^{-1}h^\dagger C) = \exp(-h).$$

From this and the uniqueness of h we conclude that

$$C^{-1}h^\dagger C = -h. \tag{A19}$$

Equation (A19) means that h also satisfies (2.6a), and this completes the proof.

[1]J. Wess and B. Zumino, Nucl. Phys. B **70**, 39 (1974). See also, D.V. Volkov and V.P. Akulov, Phys. Lett. B **46**, 109 (1973).
[2]J. Wess and B. Zumino, Phys. Lett. B **49**, 52 (1974). See also, J. Iliopoulos and B. Zumino, Nucl. Phys. B **76**, 310 (1974).
[3]D.Z. Freedman, P. van Nieuwenhuizen, and S. Ferrara, Phys. Rev. D **13**, 3214 (1976); S. Deser and B. Zumino, Phys. Lett. B **62**, 335 (1976).
[4]See for instance, F.A. Berezin and G.I. Kac, Math. USSR Sb. **11**, 311 (1970); L. Corwin, Y. Ne'eman, and S. Sternberg, Rev. Mod. Phys. **47**, 573 (1975); V.G. Kac, Commun. Math. Phys. **53**, 31 (1977); M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **17**, 1626, 1640 (1976).
[5]S.W. MacDowell and F. Mansouri, Phys. Rev. Lett. **38**, 739 (1977).
[6]Cf. Ref. 4. See also, B. Kostant, *Graded Manifolds, Graded Lie Theory and Prequantization*, MIT report.
[7]A. Salam and J. Strathdee, Nucl. Phys. B **76**, 477 (1974).
[8]L. Michel and L.A. Radicati, Ann. Phys. (N.Y.) **66**, 758 (1971).
[9]M. Gell-Mann (private communication).
[10]See for instance, R.D. Schafer, *Introduction to Nonassociative Algebras* (Academic, New York, 1966).
[11]See for instance, C. Piron, *Foundations of Quantum Physics* (Benjamin, New York, 1975).
[12]Jordan superalgebras have been considered by V.G Kac (preprint) and I. Kaplansky (preprint).
[13]We use the Dirac matrices conventions of J.D. Bjorken and S.D. Drell. Note however that C is not a charge conjugation matrix.
[14]Strictly speaking the invariance of the bilinear form does not determine the coefficient in U of the monomial of highest degree in the generators of the Grassmann algebra. Since however this term has no effect on ψ, it can always be taken in such a way that U satisfies (2.4a).
[15]It is a real form of Sl(4/1).
[16]If one considers X as an element of a vector space over the complex numbers, then its dimensionality is of course much larger than five, i.e., dimensions are then associated with individual terms in the superfield expansion of the components of X. Nevertheless we shall say that X is of dimension five, keeping in mind the superfield nature of the five components.
[17]If one considers the superfield expansion of Ξ_0, then there are subspaces left invariant by the group action. For instance, if Ξ_0 only contains monomials of degree at least k in products of anticommuting numbers, then it remains so under the group action. The representation is nevertheless not fully reducible, since acting by an arbitrary Ω on certain specific Ξ_0's (namely those of lowest degree) spans the whole space.
[18]The same remark can be made here as in connection with Eq. (2.4a). See Ref. 14.

A Galileian formulation of spin. I. Clifford algebras and spin groups[a)]

J. A. Brooke

Department of Mathematics, University of Alberta, Edmonton, Canada T6G 2G1
(Received 20 September 1977)

By generalizing the concept of spin group to the case where the underlying orthogonal space is degenerate, the spin group associated with the homogeneous Galilei group is calculated. In so doing, the Galilei group and its spin group are clearly displayed as stability subgroups of the de Sitter group and its spin group. A notion of Clifford algebra contraction is introduced in the physical (Galilei) case and its relation to Lie algebra contraction is explored. Both the stated generalization of spin group to cases with degenerate bilinear form and the idea of Clifford algebra contraction appear to be new.

1. INTRODUCTION

Spin groups arise in physics in various contexts.

In nuclear and elementary particle physics for example, the group SU(2) (often in connection with isospin) is ever present. Of course SU(2) is just the spin group associated to the rotation group SO(3). In relativistic quantum mechanics and field theory as well as in general relativity, the group $SL(2;\mathbb{C})$ being the spin group of the Lorentz group, plays a basic role (for example, in the former with respect to representation theory and in the latter with respect to the spinor analysis of curvature tensors and radiation fields). The theory of twistors which has applications to the problem of quantizing null fields in general relativity requires for its formulation, the spin group SU(2, 2) of the conformal group SO(2, 4).[1] The de Sitter group SO(1, 4) often turns up in cosmology and extended theories of the electron[2-4] as well as in some possibly unexpected places such as the Kepler problem.[5] Its spin group naturally finds a place in such applications.

While not generally the case, the twofold and universal coverings of pseudo-orthogonal groups of physical interest quite frequently coincide. In such instances, especially involving quantum theory, a knowledge of the representation theory of the corresponding spin group is essential (for then, one is compelled to admit projective representations and consequently wants to know all about the universal covering ≡ spin group).[6-8]

Interest in the Galilei group is not solely due to its role as kinematical group of nonrelativistic physics, although this is reason enough. (In the literature, to the knowledge of the author, there seems to be no treatment of nonrelativistic spin analogous to and as aesthetically pleasing as that given for the relativistic situation.) Partly due to the somewhat limited successes and many difficulties of relativistic quantum field theory, more and more attention is being focused on Galileian methods and notions. Infinite momentum frame[9] and light-cone quantization[10] schemes typically involve Galilei-like invariances. With the recent upsurge of interest in gravitation as a gauge theory, the Galilei-like inhomogeneous SO(1, n-1) has been suggested as a possible gauge group.[11]

Finally, with the current great popularity of relativistic supersymmetry theories, one may ask what happens in the nonrelativistic situation. For this, one would require a notion of Galilei spinors.

Although the number and type of possible applications of a general theory of Galilei-like spinors is likely to be rather large and diverse (for example whenever one deals with an inhomogeneous pseudo-orthogonal group), the aim of this paper is merely to formulate the concept in a manner reminiscent of the Lorentz construction. After the completion of this paper, the author stumbled onto some old papers of Clifford[12] and others[13,14] which contain ideas very close to what is presented here. (These references are the earliest the author has located and they are concerned only with the group of Euclidean motions and its representation in terms of biquaternions. In such a form, the route to generalization is not at all obvious, but it seems fitting that the general theory of the algebras bearing Clifford's name should offer such an avenue.) Additionally, the newer text[15] puts some of the ideas found in the earlier papers to use in mechanics.

In order to better understand the relations between relativistic and nonrelativistic theory, various concepts of Lie group and Lie algebra contraction have been proposed.[16-19] Motivated by calculations pertaining to the Galilei spin group, one may also devise a type of "Clifford algebra contraction" which it turns out, is intimately connected to the Lie algebra contraction schemes, and which seems not to have appeared previously in the literature.

Part I of the paper is organized as follows. Section 2 deals with arbitrary real orthogonal spaces and embeddings[20-23] and then introduces the notions of Clifford algebras and spin groups.[24-30] Sections 3 and 4 motivate and elaborate on the Galileian case. Section 5 discusses the relationship between the Galilei and de Sitter groups (for de Sitter[31-34] and for Galilei–de Sitter[35,36]). Section 6 involves Lie algebras and contractions (both Lie and Clifford). Section 7 indicates some possible future directions of study along with general remarks.

Part II of this paper is concerned mainly with specific matrix representations of the Galilei spin group and their relation to the de Sitter spin group representations. As one requires for this the use of matrices over noncommutative fields (quaternions) a few of the generalities are developed.

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

 0022-2488/78/1905-0952$1.00

2. REAL ORTHOGONAL SPACES, CLIFFORD ALGEBRAS AND SPIN GROUPS

By $\mathbb{R}^{r,p,q}$ we mean the space $\mathbb{R}^n$, $n=p+q+r$, together with the quadratic form Q (or equivalently its associated bilinear form):

$$Q(a)=-(a^{r+1})^2-\cdots-(a^{r+p})^2+(a^{r+p+1})^2+\cdots+(a^n)^2, \tag{2.1}$$

where $a=(a^1,\ldots,a^r,a^{r+1},\ldots,a^{r+p},a^{r+p+1},\ldots,a^n)$ belongs to $\mathbb{R}^n$.

The symmetric bilinear form B on $\mathbb{R}^n$ associated to Q (as a matrix with respect to the natural bases) is $B=\mathrm{diag}(0,\ldots,0,-1,\ldots,-1,1,\ldots,1)$ with r zeros p minus ones, and q plus ones. The form B is nondegenerate precisely when $r=0$ and in this case one writes $\mathbb{R}^{p,q}$ for $\mathbb{R}^{0,p,q}$.

A simple argument which will be omitted shows that $\mathbb{R}^{p+r,q+r}$ is the nondegenerate real orthogonal space of least dimension which contains the real orthogonal space $\mathbb{R}^{r,p,q}$ as an orthogonal subspace. For example, the degenerate space $\mathbb{R}^{1,0,3}$ (which is the prototype of Galileian space–times) is minimally embedded in the nondegenerate $\mathbb{R}^{1,4}$ (which contains the de Sitter space–time as an hyperboloid).

A real Clifford algebra for the real orthogonal space X is an algebra C, generated as a ring by $\mathbb{R}$ and X, subject to the conditions:

$$x^2+B(x,x)1=0, \tag{2.2}$$

where $x\in X$, $1=$ unity in C identified with $1\in\mathbb{R}$. In addition, C is to be associative and to contain $\mathbb{R}$ and X isomorphically as linear subspaces. All Clifford algebras are to be regarded as real algebras. A constructive definition is the following. Let T be the covariant tensor algebra of X and let I be the ideal of T generated by the elements $x\otimes x+B(x,x)1$, for $x\in X$. The quotient $A=T/I$ is a Clifford algebra for X and in fact $\dim A=2^n$ when $n=\dim X$.

An *orthonormal subset* of C is a linearly independent set $\{\theta^i\}$, $\theta^i\in X$ such that:

$$(\theta^i)^2=-B(\theta^i,\theta^i)=-1 \text{ or } 0 \text{ or } 1 \tag{2.3a}$$

$$\theta^i\theta^j+\theta^j\theta^i=0, \quad i\neq j. \tag{2.3b}$$

It is said to be of type (r,p,q) if r of the squares equal 0, p equal 1, and q equal -1. When $r=0$ we speak of type (p,q).

The following results, stated here without proof, may be found in[37]:

(i) If C is a Clifford algebra for an n-dimensional orthogonal space, then $\dim C\leq 2^n$.

(ii) If C is a Clifford algebra for an n-dimensional nondegenerate orthogonal space X of signature (p,q) then $\dim C=2^n$ or 2^{n-1}, and the second case being possible only if $p-q-1$ is divisible by 4 (in fact, when n is odd and $\theta^1\theta^2\cdots\theta^n=\pm 1$ for any orthonormal basis $\{\theta^i\}$ of X).

(iii) All 2^n-dimensional Clifford algebras for an n-dimensional real orthogonal space are isomorphic.

On account of (iii) Porteous refers to the unique-up-to-isomorphism Clifford algebra of dimension 2^n for X of dimension n, as the *universal* Clifford algebra for X. For $X=\mathbb{R}^{r,p,q}$, we denote the universal Clifford algebra by $\mathbb{R}_{r,p,q}$. Generally C_X will stand for the universal Clifford algebra for the orthogonal space X.

Although not necessary for the present work, (ii) may be generalized. A generalization is stated without proof as:

Proposition 2.1: Let $\theta^0,\theta^1,\ldots,\theta^n$ be an orthonormal basis of type $(1,p,q)$ with $(\theta^0)^2=0$ and $n=p+q$. Let C be the Clifford algebra generated by the θ^i's. Then

(i) $\dim C=2^{n+1}$ if n is even or if n is odd and θ^0, $\theta^0\theta^1\cdots\theta^n$ are linearly independent,

(ii) $\dim C=3\times 2^{n-1}$ if n is odd and θ^0, $\theta^0\theta^1\cdots\theta^n$ are linearly dependent. □

The orthogonal involution $-\mathbb{1}_X$ of X induces on C_X the main involution $\wedge$ defined as follows:

$$1^\wedge=1, \quad (\theta^{\alpha_1}\theta^{\alpha_2}\cdots\theta^{\alpha_r})^\wedge=(-1)^r\theta^{\alpha_1}\theta^{\alpha_2}\cdots\theta^{\alpha_r}, \tag{2.4}$$

with $\alpha_1<\alpha_2<\cdots<\alpha_r$ and $\{\theta^\alpha\}$ an orthonormal basis of X.

The orthogonal involution $-\mathbb{1}_X$ of X also induces on C_X the *conjugation anti-involution*—defined as follows:

$$1^-=1, \quad (\theta^{\alpha_1}\theta^{\alpha_2}\cdots\theta^{\alpha_r})^-=(-1)^r\theta^{\alpha_r}\cdots\theta^{\alpha_2}\theta^{\alpha_1}, \tag{2.5}$$

with the α's and θ's as in (2.4).

The main involution provides a direct sum decomposition of C_X into *even* and *odd* elements (C_X^+, C_X^-, respectively)

$$C_X^+=\{a\in C_X: a^\wedge=a\}, \quad C_X^-=\{a\in C_X: a^\wedge=-a\}. \tag{2.6}$$

One calls C_X^+ the even Clifford algebra of X.

It can be shown that for the universal Clifford algebra $\mathbb{R}_{r,p,q}$

$$\mathbb{R}^+_{r,p,q}\cong\mathbb{R}_{r,p,q-1}\cong\mathbb{R}_{r,q,p-1}. \tag{2.7}$$

For a proof when $r=0$ see Ref. 38; the general case follows in a similar way.

Defining $\Gamma_X=\{g\in C_X: g^{-1}$ exists and $gxg^{\wedge-1}\in X$, for $x\in X\}$ we can see easily[39] that each $g\in\Gamma_X$ defines an orthogonal automorphism $\rho_X(g)$ of X by setting $\rho_X(g)x=gxg^{\wedge-1}$ for $x\in X$. Furthermore, Γ_X is a group uniquely defined up to isomorphism on account of universality of C_X. One calls Γ_X the *Clifford group* and $\Gamma_X^+=\Gamma_X\cap C_X^+$ the even Clifford group of X.

The *norm* on C_X is the mapping $N: C_X\to C_X$ by

$$a\to a^-a. \tag{2.8}$$

In analogy to the nondegenerate case, we define $\mathrm{Pin}(X)=\{g\in\Gamma_X: N(g)=\pm 1\}$ and $\mathrm{Spin}(X)=\{g\in\Gamma_X^+: N(g)=\pm 1\}$. Of course for nondegenerate X, the mappings $\mathrm{Pin}(X)\to\mathrm{O}(X)$ and $\mathrm{Spin}(X)\to\mathrm{SO}(X)$ by $g\to\rho_X(g)$ are surjective homomorphisms whose kernels are isomorphic to $\mathbb{Z}_2$ and $\mathrm{Pin}(X)$, $\mathrm{Spin}(X)$ are the two-fold covering groups of $\mathrm{O}(X)$, $\mathrm{SO}(X)$, respectively.

The extent to which we have similar results for arbitrary degenerate X is at present unsettled. To obtain useful results, it may in fact be necessary to

modify the above definitions in the degenerate situation (for example, use some projective Clifford groups[40]). However, in the degenerate cases of immediate physical interest where $X=\mathbb{R}^{1,p,q}$ the above definitions do suffice as will become clear in the next section.

3. THE PHYSICAL CASE: SPIN(GALILEI)

As briefly noted earlier, $\mathbb{R}^{1,0,3}$ is the simplest model of a Galileian space–time. (To be more precise, one requires on $\mathbb{R}^4$ a symmetric twice-contravariant tensor field γ of rank three and a nowhere zero 1-form θ in the kernel of γ.[41–43]) To see this from a group theoretical viewpoint, let $\gamma=(\gamma^{\alpha\beta})=\mathrm{diag}(0,1,1,1)$ so that

$$O(\gamma)=\{g\in GL(4;\mathbb{R}):g\gamma g^t=\gamma\}$$
$$=\left\{\begin{pmatrix} a & 0\\ \mathbf{b} & R\end{pmatrix}:0\neq a \text{ real},\ \mathbf{b}\in\mathbb{R}^3,\ R\in O(3)\right\},$$

$SO(\gamma)\uparrow$ = time-orientation preserving $SO(\gamma)$

$$=\left\{\begin{pmatrix} 1 & 0\\ \mathbf{b} & R\end{pmatrix}:\mathbf{b}\in\mathbb{R}^3,\ R\in SO(3)\right\}$$
$$\cong G_0 \text{ the homogeneous Galilei group.}$$

Consequently, we identify $\mathrm{Spin}(G_0)$ with $\mathrm{Spin}(\mathbb{R}^{1,0,3})$ which we now proceed to determine. Assume $\{\theta^\alpha\}$, $\alpha=0,1,2,3$, is an orthonormal basis of $X=\mathbb{R}^{1,0,3}$ of type $(1,0,3)$ which moreover generates $C_X=\mathbb{R}_{1,0,3}$, the universal Galilei Clifford algebra. Thus $\mathbb{R}_{1,0,3}$ ≡ algebra generated by $\theta^0,\theta^1,\theta^2,\theta^3$ and $\mathbb{R}^+_{1,0,3}$ $=\mathrm{span}\{1,\theta^0\theta^1,\theta^0\theta^2,\theta^0\theta^3,\theta^1\theta^2,\theta^2\theta^3,\theta^3\theta^1,\theta^0\theta^1\theta^2\theta^3\}$. Conjugation is defined by:

$$1^-=1,\quad \theta^{\alpha-}=-\theta^\alpha,\quad (\theta^\alpha\theta^\beta)^-=-\theta^\alpha\theta^\beta\quad(\alpha\neq\beta),\tag{3.1a}$$
$$(\theta^\alpha\theta^\beta\theta^\gamma)^-=\theta^\alpha\theta^\beta\theta^\gamma\quad(\alpha,\beta,\gamma\text{ distinct}),$$
$$(\theta^0\theta^1\theta^2\theta^3)^-=\theta^0\theta^1\theta^2\theta^3,\tag{3.1b}$$

where $0\leq\alpha,\beta,\gamma\leq 3$.

First we determine those $s\in\mathbb{R}^+_{1,0,3}$ of norm $N(s)=\pm1$. An arbitrary $s\in\mathbb{R}^+_{1,0,3}$ may be written as

$$s=a+b\theta,\tag{3.2a}$$

with

$$a=a^01+a^1\theta^2\theta^3+a^2\theta^3\theta^1+a^3\theta^1\theta^2,\tag{3.2b}$$
$$b=b^01+b^1\theta^2\theta^3+b^2\theta^3\theta^1+b^3\theta^1\theta^2,\quad\theta=\theta^0\theta^1\theta^2\theta^3.\tag{3.2b}$$

Noticing that $a\theta=\theta a$, $b\theta=\theta b$, $\theta^-=\theta$, $(\theta)^2=0$, we have $s^-=a^-+b^-\theta$ so that $N(s)=s^-s=a^-a+(a^-b+b^-a)\,\theta$ with the result that

$$N(s)=\pm1\iff a^-a=\pm1,\quad a^-b+b^-a=0.\tag{3.3}$$

This requires $\pm1=a^-a=(a^0)^2+(a^1)^2+(a^2)^2+(a^3)^2$ and $0=a^-b+b^-a=2(a^0b^0+a^1b^1+a^2b^2+a^3b^3)$ with the result that only $N(s)=1$ is possible. Noting that for such s, $\hat{s}=s$ and $s\theta^\alpha s^{-1}\in\mathrm{span}\{\theta^0,\theta^1,\theta^2,\theta^3\}$; and in fact by tedious calculation:

$$s\theta^0s^{-1}=\theta^0,\tag{3.4a}$$

$$s\theta^1s^{-1}=2(a^0b^1-a^1b^0+a^3b^2-a^2b^3)\theta^0$$
$$+((a^0)^2+(a^1)^2-(a^2)^2-(a^3)^2)\theta^1$$
$$+2(a^1a^2+a^0a^3)\,\theta^2+2(a^1a^3-a^0a^2)\,\theta^3,\tag{3.4b}$$

$$s\theta^2s^{-1}=2(a^0b^2-a^2b^0+a^1b^3-a^3b^1)\,\theta^0+2(a^1a^2-a^0a^3)\,\theta^1$$
$$+((a^0)^2-(a^1)^2+(a^2)^2-(a^3)^2)\,\theta^2+2(a^2a^3+a^0a^1)\theta^3,\tag{3.4c}$$

$$s\theta^3s^{-1}=2(a^0b^3-a^3b^0+a^2b^1-a^1b^2)\,\theta^0+2(a^1a^3+a^0a^2)\,\theta^1$$
$$+2(a^2a^3-a^0a^1)\theta^2+((a^0)^2-(a^1)^2-(a^2)^2+(a^3)^2)\,\theta^3.\tag{3.4d}$$

We have therefore

Theorem 3.1: $\mathrm{Spin(Galilei)}=\{a+b\theta:\ a^-a=1,\ a^-b+b^-a=0\}$, where

$$a=a^01+a^1\theta^2\theta^3+a^2\theta^3\theta^1+a^3\theta^1\theta^2,$$
$$b=b^01+b^1\theta^2\theta^3+b^2\theta^3\theta^1+b^3\theta^1\theta^2,\quad\theta=\theta^0\theta^1\theta^2\theta^3$$

and

$$1=a^-a=(a^0)^2+(a^1)^2+(a^2)^2+(a^3)^2$$
$$0=\tfrac{1}{2}(a^-b+b^-a)=a^0b^0+a^1b^1+a^2b^2+a^3b^3.$$

□

Topologically, $\mathrm{Spin}(G_0)$ is rather simple: $a^-a=1$ can be interpreted by saying that the vector $(a^0,a^1,a^2,a^3)\in\mathbb{R}^4$ lies on the unit sphere S^3, and $a^-b+b^-a=0$ says that the vector (b^0,b^1,b^2,b^3) is perpendicular (in Euclidean $\mathbb{R}^4$) to the radius vector (a^0,a^1,a^2,a^3) which means that the pair (a,b) belongs to TS^3, the tangent bundle of S^3. Thus $\mathrm{Spin}(G_0)$ is homeomorphic to TS^3 which is homeomorphic to $\mathbb{R}^3\times S^3$ because S^3 is also a Lie group.

The group structure can also be analyzed. We have

$$(a+b\theta)\,(a'+b'\theta)=aa'+(ab'+ba')\,\theta,\tag{3.5}$$

where of course,

$$a^-a=1=a'^-a'\ \text{ and }\ a^-b+b^-a=0=a'^-b'+b'^-a'.\tag{3.6}$$

From this, we notice two distinguished subgroups

$$H=\{a:\ a^-a=1\}\cong S^3\cong\mathrm{Spin}(SO(3))$$

and

$$K=\{1+b\theta:\ b^-+b=0\}\cong\mathbb{R}^3\ \text{ as Lie groups.}$$

It is easy to check that

$$\mathrm{Spin}(G_0)=H\cdot K,\tag{3.7a}$$
$$H\cap K=\{1\},\tag{3.7b}$$
$$K\lhd\mathrm{Spin}(G_0)\quad[K\text{ normal in }\mathrm{Spin}(G_0)]\tag{3.7c}$$
$$H\cap C(K)=\{-1,1\},\tag{3.7d}$$

where $C(K)=$ centralizer of $K=\{g\in\mathrm{Spin}(G_0):gk=kg,\ k\in K\}$ and therefore[44]

$$\mathrm{Spin}(G_0)/\{-1,1\}$$
$$=(H\cdot K)/(C(K)\cap H)\cong K\circledS\,(H/(C(K)\cap H))\tag{3.8}$$
$$\cong\mathbb{R}^3\circledS\,(\mathrm{Spin}(SO(3))/\{-1,1\})\cong\mathbb{R}^3\circledS\,SO(3),\tag{3.9}$$

with the semidirect product being unique up to isomorphism. In fact, $\mathrm{Spin}(G_0)\cong\mathbb{R}^3\circledS\,\mathrm{Spin}(SO(3))$ with the $\mathrm{Spin}(SO(3))$ action on $\mathbb{R}^3$ being the obvious one.

It turns out therefore, that $\mathrm{Spin}(G_0)$ is precisely the twofold covering group of G_0 as indeed it should, being a kinematical group for nonrelativistic fermions. The

fact that it is also the universal covering group is something of an accident as the twofold covering group need not be simply connected (for example, $\mathrm{Spin}^+(\mathbb{R}^{1,2}) = \{s \in \mathrm{Spin}(\mathbb{R}^{1,2}): N(s) = 1\}$ is isomorphic to SL(2;$\mathbb{R}$) which has $\mathbb{Z}$ as its first homotopy group).

By way of generalization, one may consider instead of $\mathbb{R}^{1,0,3}$ the space $X = \mathbb{R}^{1,p,q}$. We only state the results here as the proof, which is much like that given for $\mathbb{R}^{1,0,3}$, would take us too far afield.

Theorem 3.2: $\mathrm{Spin}(\mathbb{R}^{1,p,q}) = H \circ K$ (3.10)

where $H \cong \mathrm{Spin}(\mathbb{R}^{p,q})$ and $K \cong \mathbb{R}^n$, $n = p+q$. Furthermore,

$$H \cap K = \{1\}, \tag{3.11a}$$

$$K \lhd \mathrm{Spin}(\mathbb{R}^{1,p,q}), \tag{3.11b}$$

$$H \cap C(K) = \{-1, 1\}, \tag{3.11c}$$

and

$$\mathrm{Spin}(\mathbb{R}^{1,p,q})/\{-1,1\} \cong \mathbb{R}^n \circledS (\mathrm{Spin}(\mathbb{R}^{p,q})/\{-1,1\}) \cong \mathbb{R}^n \circledS \mathrm{SO}(p,q), \tag{3.12}$$

with

$$\{1\} \to \{-1,1\} \to \mathrm{Spin}(\mathbb{R}^{1,p,q}) \xrightarrow{\rho_X} \mathrm{SO}(1,p,q) \uparrow \to \{1\} \tag{3.13}$$

an exact sequence of groups. □

4. EXPLICIT CORRESPONDENCE BETWEEN G_0 AND SPIN (G_0)

Let $\{e_\alpha\}$, $\alpha = 0,1,2,3$, be a basis of $\mathbb{R}^{1,0,3}$ the primitive Galilei space–time. Let us identify $\{\theta^\alpha\}$, $\alpha = 0,1,2,3$, the generators of $\mathbb{R}_{1,0,3}$ with the corresponding basis of $(\mathbb{R}^4)^*$ dual to $\{e_\alpha\}$.[45–47] More formally, one would define a linear mapping $\gamma: (\mathbb{R}^4)^* \to \mathbb{R}_{1,0,3}$ by

$$\theta^\alpha \to \gamma(\theta^\alpha) = \gamma^\alpha \tag{4.1}$$

so that $\gamma^\alpha\gamma^\beta + \gamma^\beta\gamma^\alpha = -2\gamma^{\alpha\beta}$ instead of $\theta^\alpha\theta^\beta + \theta^\beta\theta^\alpha = -2\gamma^{\alpha\beta}$.

Because $G_0 \cong \mathbb{R}^3 \circledS \mathrm{SO}(3)$, an element g of G_0 corresponds to $(\mathbf{v}c, R)$ where $\mathbf{v}$ is a unitless vector in $\mathbb{R}^3$, c is a real constant with units of speed (say, the velocity of light), and $R \in \mathrm{SO}(3)$. Of course, $\mathbf{v}$ defines the boost and R the rotation. Multiplication is as usual:

$$(\mathbf{v}c, R)(\mathbf{v}'c, R') = ((\mathbf{v} + R\mathbf{v}')c, RR'). \tag{4.2}$$

Each $g \in G_0$ defines a transformation $\Lambda(g) = \Lambda$ on frames as follows:

$$e_\alpha \to \hat{e}_\alpha, \tag{4.3}$$

where $e_\alpha = \hat{e}_\beta \Lambda^\beta{}_\alpha$.

Equivalently, it defines a coordinate transformation

$$x^\alpha \to \hat{x}^\alpha = \Lambda^\alpha{}_\beta x^\beta \tag{4.4}$$

(note that $x^0 = ct$, t = time coordinate).

Now $g = (\mathbf{v}c, R)$ corresponds to

$$x^0 \to \hat{x}^0 = x^0, \quad x^A \to \hat{x}^A = v^A x^0 + R^A{}_B x^B \tag{4.5}$$

or

$$\Lambda^0{}_0 = 1, \quad \Lambda^0{}_A = 0, \quad \Lambda^A{}_0 = v^A, \quad \Lambda^A{}_B = R^A{}_B \tag{4.6}$$

with $1 \le A, B \le 3$.

It amounts to the same thing if we regard G_0 as acting on the dual coframe $\{\theta^\alpha\}$ instead of the frame $\{e_\alpha\}$:

$$\theta^\alpha \to \hat{\theta}^\alpha = \Lambda^\alpha{}_\beta \theta^\beta \tag{4.7}$$

with $\{\hat{\theta}^\alpha\}$ dual to $\{\hat{e}_\alpha\}$ because $\{\theta^\alpha\}$ is dual to $\{e_\alpha\}$.

There is however, an obvious action of $\mathrm{Spin}(G_0)$ on coframes:

$$\theta^\alpha \to \hat{\theta}^\alpha = s^{-1}\theta^\alpha s, \quad s \in \mathrm{Spin}(G_0). \tag{4.8}$$

By requiring $s^{-1}\theta^\alpha s = \Lambda(g)^\alpha{}_\beta \theta^\beta$, where s is the unique-up-to-a-sign element of $\mathrm{Spin}(G_0)$ corresponding to $g \in G_0$, we relate the G_0 and $\mathrm{Spin}(G_0)$ coframe actions.

Explicitly, let $s = a + b\theta \in \mathrm{Spin}(G_0)$ have the representation as in Theorem 3.1. With the customary notations for dot and cross product in $\mathbb{R}^3$ we have $|a|^2 = a^\dagger a = (a^0)^2 + (a^1)^2 + (a^2)^2 + (a^3)^2 = (a^0)^2 + \mathbf{a}\cdot\mathbf{a} = 1$ and $\frac{1}{2}(a^\dagger b + b^\dagger a) = a^0b^0 + a^1b^1 + a^2b^2 + a^3b^3 = a^0b^0 + \mathbf{a}\cdot\mathbf{b} = 0$.

Let

$$R = ((a^0)^2 - \mathbf{a}\cdot\mathbf{a})\,\mathbb{1}_3 + 2\mathbf{a}\cdot\mathbf{a}P_{\mathbf{a}} + 2a^0\mathbf{J}_{\mathbf{a}}, \tag{4.9a}$$

$$\mathbf{v} = 2(-a^0\mathbf{b} + b^0\mathbf{a} - \mathbf{a}\times\mathbf{b}), \tag{4.9b}$$

where $\mathbb{1}_3$ = identity map of $\mathbb{R}^3$, $P_{\mathbf{a}}\mathbf{x} = (\mathbf{a}\cdot\mathbf{x})/(\mathbf{a}\cdot\mathbf{a})\,\mathbf{a}$ = projection parallel to $\mathbf{a}$, and $J_{\mathbf{a}}\mathbf{x} = \mathbf{a}\times\mathbf{x}$. A straightforward calculation shows $R \in \mathrm{SO}(3)$ and that

$$\hat{\theta}^0 = s^{-1}\theta^0 s = \theta^0, \tag{4.10a}$$

$$\hat{\theta}^A = s^{-1}\theta^A s = v^A\theta^0 + R^A{}_B\theta^B. \tag{4.10b}$$

Conversely, given $(\mathbf{v}c, R) \in G_0$, R defines an axis of (right-hand) rotation $\mathbf{a}$ and an angle of rotation θ that can be chosen so as to satisfy along with $R\mathbf{a} = \mathbf{a}$,

$$\mathbf{a}\cdot\mathbf{a} = \sin^2(\theta/2) \quad \text{and} \quad (a^0)^2 = \cos^2(\theta/2). \tag{4.11}$$

Now setting

$$b^0 = \tfrac{1}{2}(\mathbf{a}\cdot\mathbf{v}) \quad \text{and} \quad \mathbf{b} = -\tfrac{1}{2}(a^0\mathbf{v} - \mathbf{a}\times\mathbf{v}) \tag{4.12}$$

we find that $s = a + b\theta \in \mathrm{Spin}(G_0)$ corresponds to $(\mathbf{v}c, R) \in G_0$.

The 2–1 relationship between the $\mathrm{Spin}(G_0)$ and G_0 actions on coframes is now apparent and renders the generalization to arbitrary (possibly curved) Galilei space–times almost trivial.

5. RELATION BETWEEN G_0 AND THE DE SITTER GROUP

As noted in Sec. 2, the orthogonal space $\mathbb{R}^{1,4}$ is the smallest nondegenerate space containing $\mathbb{R}^{1,0,3}$. The group preserving the structure of $\mathbb{R}^{1,4}$ is just the de Sitter group $S = \mathrm{SO}(1,4)$, and its associated spin group is $\mathrm{Spin}(\mathbb{R}^{1,4})$. Thanks to the embedding of $\mathbb{R}^{1,0,3}$ in $\mathbb{R}^{1,4}$ we also have subgroup embeddings of G_0 and $\mathrm{Spin}(G_0)$ into S and $\mathrm{Spin}(S)$, respectively.

It is well known that the Galilei group G_0 can be thought of as a stability subgroup of the de Sitter group $\mathrm{SO}(1,4)\uparrow$ (namely, that associated to a de Sitter null vector). A similar result holds for the spin groups.

To describe the situation, we require a representation of G_0 on $\mathbb{R}^5$, the coordinates of which are identified with the physical quantities of energy, momentum and mass. For $(\mathbf{v}c, R) \in G_0$

$$\hat{E} = E + \mathbf{v}c \cdot R\mathbf{p} + \tfrac{1}{2} mc^2 \mathbf{v} \cdot \mathbf{v}, \tag{5.1a}$$

$$\hat{\mathbf{p}} = R\mathbf{p} + mc\mathbf{v}, \tag{5.1b}$$

$$\hat{m} = m, \tag{5.1c}$$

defines the G_0 transformation law on $\mathbb{R}^5$. Consider the quadratic form U on $\mathbb{R}^5$ defined by

$$(E, \mathbf{p}, m) \to U(E, \mathbf{p}, m) = \mathbf{p} \cdot \mathbf{p} - 2mE. \tag{5.2}$$

It is trivial to check that this form is G_0 invariant

$$\hat{\mathbf{p}} \cdot \hat{\mathbf{p}} - 2\hat{m}\hat{E} = \mathbf{p} \cdot \mathbf{p} - 2mE \tag{5.3}$$

and that the form when diagonalized is essentially diag$(-1, 1, 1, 1, 1)$, precisely that associated to the de Sitter group. Hence G_0 is seen to be isomorphic to a subgroup of $SO(1,4)^\uparrow$ which leaves invariant the function on $\mathbb{R}^5$, $m(E, \mathbf{p}, m) = m$. Expressing this slightly differently by noticing that

$$U = \mathbf{p} \cdot \mathbf{p} - 2mE = -2m(E - \mathbf{p} \cdot \mathbf{p}/2m)$$

$$= -2 \times (\text{mass}) \times (\text{internal energy}). \tag{5.4}$$

G_0 is characterized as the subgroup of $SO(1,4)^\uparrow$ coordinate transformations which, as far as an observer is concerned, leave mass and internal energy constant.

To see the corresponding situation in terms of spin groups, let us suppose that $\{\Sigma^a\}$, $a = 0, 1, 2, 3, 4$, is an orthonormal basis generating $\mathbb{R}_{1,4}$ and that $\mathrm{Spin}^+(\mathbb{R}^{1,4})$ is the subgroup of $\mathrm{Spin}(\mathbb{R}^{1,4})$ of "time-orientation" preserving elements.[48] Performing the coordinate transformation on $\mathbb{R}^5$,

$$\epsilon = (1/\sqrt{2})(E + m), \tag{5.5a}$$

$$\pi^A = p^A, \tag{5.5b}$$

$$\mu = (1/\sqrt{2})(-E + m), \tag{5.5c}$$

we find $U = \mathbf{p} \cdot \mathbf{p} - 2mE = -\epsilon^2 + \boldsymbol{\pi} \cdot \boldsymbol{\pi} + \mu^2$. To the 1-forms $d\epsilon$, $d\pi^A$, $d\mu$ on $\mathbb{R}^5$ we associate (see Sec. 4) the $\mathbb{R}_{1,4}$ generators Σ^0, Σ^A, Σ^4, respectively, and then define

$$\Theta^0 \equiv (1/\sqrt{2})(\Sigma^0 + \Sigma^4), \quad \Theta^A \equiv \Sigma^A, \quad \Theta^4 \equiv (1/\sqrt{2})(\Sigma^0 - \Sigma^4) \tag{5.6}$$

so that the 1-forms dm, dp^A, dE are associated to Θ^0, Θ^A, Θ^4, respectively.

The thing to notice is that

$$\Theta^0 = (1/\sqrt{2})(\Sigma^0 + \Sigma^4), \quad \Theta^A = \Sigma^A, \quad A = 1, 2, 3 \tag{5.7}$$

defines an embedding of dual spaces $(\mathbb{R}^{1,0,3})^*$ into $(\mathbb{R}^{1,4})^*$, and in fact

$$\Theta^\alpha \Theta^\beta + \Theta^\beta \Theta^\alpha = -2\gamma^{\alpha\beta}, \quad 0 \le \alpha, \beta \le 3 \tag{5.8}$$

so in addition we have the embedding $\mathbb{R}_{1,0,3} \subset \mathbb{R}_{1,4}$. (Because Σ^a, $0 \le a \le 4$, generate $\mathbb{R}_{1,4}$ it follows that Θ^α, $0 \le \alpha \le 3$, generate a $16 = 2^4$ dimensional Clifford algebra hence it must, by universality, be $\mathbb{R}_{1,0,3}$.) Invariance of m becomes invariance of the 1-form dm which by association, becomes invariance of Θ^0 [a manifestation of (4.10) when we identify θ^α with Θ^α, $\alpha = 0, 1, 2, 3$]. All this comes about because $\mathrm{Spin}^+(\mathbb{R}^{1,4})$ acts on $\mathbb{R}^5$ coframes $\{\Theta^a\}$ by $\Theta^a \to \hat{\Theta}^a = s^{-1}\Theta^a s$s. Thus $\mathrm{Spin}(G_0) = \mathrm{Spin}^+(\mathbb{R}^{1,4})_{\Theta^0} \equiv$ stabilizer of Θ^0, is the spin analogue of the vector result mentioned earlier.

6. LIE ALGEBRAS AND LIE-CLIFFORD CONTRACTIONS

Pseudo-orthogonal Lie algebras are always realizable within the associated even Clifford algebra, the Lie algebra product being the commutator. The same situation is true for the Lie algebras $\mathfrak{so}(1, p, q)$ and in particular for the Galilei Lie algebra $\mathfrak{g}_0 = \mathfrak{so}(1, 0, 3)$. This is made explicit as follows by letting:

$$J_G^A = \tfrac{1}{4}\epsilon^A{}_{BC}\theta^B\theta^C, \quad K_G^A = \tfrac{1}{2}\theta^0\theta^A, \quad A = 1, 2, 3 \tag{6.1}$$

where ϵ = completely skew tensor on $\mathbb{R}^3$. The Galilei Lie algebra relations are:

$$[J_G^A, J_G^B] = \epsilon^{AB}{}_C J_G^C, \tag{6.2a}$$

$$[J_G^A, K_G^B] = \epsilon^{AB}{}_C K_G^C, \tag{6.2b}$$

$$[K_G^A, K_G^B] = 0, \tag{6.2c}$$

or symbolically

$$[\mathbf{J}_G, \mathbf{J}_G] = \mathbf{J}_G, \tag{6.3a}$$

$$[\mathbf{J}_G, \mathbf{K}_G] = \mathbf{K}_G, \tag{6.3b}$$

$$[\mathbf{K}_G, \mathbf{K}_G] = 0. \tag{6.3c}$$

A rotation through angle ω about an axis $\mathbf{n}$ $(\mathbf{n} \cdot \mathbf{n} = 1)$ is represented up-to-a-sign in $\mathrm{Spin}(G_0)$ by

$$s = \exp\omega(\mathbf{n} \cdot \mathbf{J}_G) = \cos(\omega/2) \cdot 1 + \sin(\omega/2) \cdot n, \tag{6.4}$$

with $n = 2\mathbf{n} \cdot \mathbf{J}_G = n^1\theta^2\theta^3 + n^2\theta^3\theta^1 + n^3\theta^1\theta^2$ and a boost by $\mathbf{v}c$ is represented by

$$\pm s = \exp(\mathbf{v}c \cdot \mathbf{K}_G) = 1 + \mathbf{v}c \cdot \mathbf{K}_G$$

$$= 1 + (c/2)(-v^1\theta^2\theta^3 - v^2\theta^3\theta^1 - v^3\theta^1\theta^2) \tag{6.5}$$

[see Theorem 3.1 and (4.9), (4.11), (4.12)].

Turning to the de Sitter Lie algebra $\mathfrak{s}$, we define

$$J_S^A = \tfrac{1}{4}\epsilon^A{}_{BC}\Sigma^B\Sigma^C, \tag{6.6a}$$

$$K_S^A = \tfrac{1}{2}\Sigma^0\Sigma^A, \tag{6.6b}$$

$$P_S^A = \tfrac{1}{2}\Sigma^A\Sigma^4, \tag{6.6c}$$

$$E_S = \tfrac{1}{2}\Sigma^0\Sigma^4, \tag{6.6d}$$

and note that symbolically as before

$$[\mathbf{J}_S, \mathbf{J}_S] = \mathbf{J}_S, \quad [\mathbf{J}_S, \mathbf{K}_S] = \mathbf{K}_S, \quad [\mathbf{J}_S, \mathbf{P}_S] = \mathbf{P}_S, \tag{6.7a}$$

$$[\mathbf{K}_S, \mathbf{K}_S] = -\mathbf{J}_S, \quad [\mathbf{K}_S, \mathbf{P}_S] = -E_S \ ([K_S^A, P_S^B] = -\delta^{AB}E_S),$$

$$[\mathbf{P}_S, \mathbf{P}_S] = \mathbf{J}_S, \tag{6.7b}$$

$$[\mathbf{J}_S, E_S] = 0, \quad [\mathbf{K}_S, E_S] = -\mathbf{P}_S, \quad [\mathbf{P}_S, E_S] = -\mathbf{K}_S. \tag{6.7c}$$

We are now in a position to study the Lie algebra embedding of $\mathfrak{g}_0$ in $\mathfrak{s}$. As $\mathrm{Spin}(G_0)$ is the stabilizer under conjugation of $\Theta^0 = (1/\sqrt{2})(\Sigma^0 + \Sigma^4)$ then $\mathfrak{g}_0$ should be the maximal Lie subalgebra $\mathfrak{a}$ of $\mathfrak{s}$ satisfying $[\mathfrak{a}, \Theta^0] = 0$. This is easily seen to be the case because

$$[J_S^A, \Theta^0] = 0, \quad [K_S^A, \Theta^0] = \frac{-\Sigma^A}{\sqrt{2}} = [P_S^A, \Theta^0], \quad [E_S, \Theta^0] = -\Theta^0 \tag{6.8}$$

with the result that $\langle \mathbf{J}_S, \mathbf{K}_S - \mathbf{P}_S \rangle \subset \mathfrak{a}$. But $J_S^A = \tfrac{1}{4}\epsilon^A{}_{BC}\Theta^B\Theta^C$, $K_S^A - P_S^A = \tfrac{1}{2}\Theta^0\Theta^A$ are identified with J_G^A, K_G^A respectively; therefore, $\mathfrak{g}_0 \subset \mathfrak{a}$. Dimensional considera-

tions however, show the reverse inclusion and hence $\mathcal{I}_0 \cong \langle \mathbf{J}_S, \mathbf{K}_S - \mathbf{P}_S \rangle$ displays the embedding in a particularly transparent form. [Note: $\langle x, y, \cdots \rangle$ in the foregoing and the following means the algebra (Lie or Clifford) or group generated by $x, y, \cdots$.]

Before introducing the Clifford algebra contraction, we consider some Lie algebra contractions of a slightly modified Inönü, Wigner, Saletan type.

For real σ let

$$J_S^A(\sigma) = J_S^A, \quad K_S^A(\sigma) = \cosh\sigma\, K_S^A - \sinh\sigma\, P_S^A, \tag{6.9a}$$

$$P_S^A(\sigma) = -\sinh\sigma\, K_S^A + \cosh\sigma\, P_S^A, \quad E_S(\sigma) = E_S, \tag{6.9b}$$

define a mapping $\mathbf{J}_S \to \mathbf{J}_S(\sigma)$, $\mathbf{K}_S \to \mathbf{K}_S(\sigma)$, $\mathbf{P}_S \to \mathbf{P}_S(\sigma)$, $E_S \to E_S(\sigma)$. The $\mathbf{J}_S(\sigma)$, $\mathbf{K}_S(\sigma)$, $\mathbf{P}_S(\sigma)$, $E_S(\sigma)$ then satisfy the Lie relations (6.7) so that the mapping $\nu(\sigma)$ so defined is an automorphism of $\mathcal{A}$. Now the Lorentz Lie algebra $\mathcal{L}_0$ is naturally embedded in $\mathcal{A}$ as a subalgebra with generators $\mathbf{J}_L = \mathbf{J}_S$, $\mathbf{K}_L = \mathbf{K}_S$. Letting $\mathbf{J}_L(\sigma) = \mathbf{J}_S(\sigma)$, $\mathbf{K}_L(\sigma) = \mathbf{K}_S(\sigma)$, we find that $\nu(\sigma)\mathcal{L}_0 = \langle \mathbf{J}_L(\sigma), \mathbf{K}_L(\sigma) \rangle$ is isomorphic to $\mathcal{L}_0$. Finally, define a linear map $W_\sigma : \mathcal{L}_0 \to \nu(\sigma)\mathcal{L}_0$ of vector spaces by

$$\mathbf{J}_L \to \nu(\sigma)\,\mathbf{J}_L, \tag{6.10a}$$

$$\mathbf{K}_L \to \sqrt{2}\, e^{-\sigma}\nu(\sigma)\,\mathbf{K}_L. \tag{6.10b}$$

A new Lie product $[\ ,\ ]$ is defined on $\mathcal{L}_0$ by

$$[x, y]_\sigma \equiv W_\sigma^{-1}[W_\sigma x, W_\sigma y] \tag{6.11}$$

and $\mathcal{L}_0^\sigma = (\mathcal{L}_0, [\ ,\ ]_\sigma)$ becomes a Lie algebra isomorphic to $\mathcal{L}_0$. The limit

$$\mathcal{L}_0^\infty = \lim_{\sigma \to \infty} \mathcal{L}_0^\sigma$$

can be calculated. In fact

$$[\mathbf{J}_L, \mathbf{J}_L]_\sigma = \mathbf{J}_L, \tag{6.12a}$$

$$[\mathbf{J}_L, \mathbf{K}_L]_\sigma = \mathbf{K}_L, \tag{6.12b}$$

$$[\mathbf{K}_L, \mathbf{K}_L]_\sigma = -2e^{-2\sigma}\mathbf{J}_L, \tag{6.12c}$$

showing that

$$[\mathbf{J}_L, \mathbf{J}_L]_\infty = \mathbf{J}_L, \tag{6.13a}$$

$$[\mathbf{J}_L, \mathbf{K}_L]_\infty = \mathbf{K}_L, \tag{6.13b}$$

$$[\mathbf{K}_L, \mathbf{K}_L]_\infty = 0, \tag{6.13c}$$

which says that

$$\mathcal{I}_0 = \mathcal{L}_0^\infty = \lim_{\sigma \to \infty} \mathcal{L}_0^\sigma.$$

The rather peculiar choice for $W_\sigma(\mathbf{K}_L)$ is made so that

$$\lim_{\sigma \to \infty} W_\sigma(\mathbf{K}_L) = \mathbf{K}_G.$$

Of course we also have

$$\lim_{\sigma \to \infty} W_\sigma(\mathbf{J}_L) = \mathbf{J}_G.$$

Considering how one may obtain the Galilei Lie algebra from the Lorentz Lie algebra as a limit, the question arises whether such a limiting procedure can be defined for the Clifford algebras which induces the given Lie algebra contraction. While a general discussion of such limits is not carried out here, we do treat the physical case at hand, giving an idea of what might be done in general.

For this purpose let $\mu(\sigma)$, for real σ, be the linear mapping:

$$\Sigma^0 \to \Sigma^0(\sigma) = \cosh\sigma\, \Sigma^0 + \sinh\sigma\, \Sigma^4, \tag{6.14a}$$

$$\Sigma^4 \to \Sigma^4(\sigma) = \sinh\sigma\, \Sigma^0 + \cosh\sigma\, \Sigma^4, \tag{6.14b}$$

$$\Sigma^A \to \Sigma^A(\sigma) = \Sigma^A. \tag{6.14c}$$

It happens that $\Sigma^a(\sigma)\Sigma^b(\sigma) + \Sigma^b(\sigma)\Sigma^a(\sigma) = -2g^{ab}$ with $g = \mathrm{diag}(-1, 1, 1, 1, 1)$ so that $\mu(\sigma)$ extends uniquely to an automorphism of $\mathbb{R}_{1,4}$. Denoting by $\mu(\sigma)\mathbb{R}_{1,3}$ the algebra $\langle \Sigma^\alpha(\sigma) : \alpha = 0, 1, 2, 3 \rangle$, we define a linear mapping $U_\sigma : \mathbb{R}_{1,3} \to \mu(\sigma)\mathbb{R}_{1,3}$ between vector spaces by

$$1 \to 1, \quad \Sigma^0 \to U_\sigma(\Sigma^0) = \sqrt{2}\, e^{-\sigma}\mu(\sigma)\Sigma^0,$$
$$\Sigma^A \to U_\sigma(\Sigma^A) = \mu(\sigma)\Sigma^A, \tag{6.15a}$$

$$\Sigma^{\alpha_1}\Sigma^{\alpha_2}\cdots\Sigma^{\alpha_r} \to U_\sigma(\Sigma^{\alpha_1})\,U_\sigma(\Sigma^{\alpha_2})\cdots U_\sigma(\Sigma^{\alpha_r}), \tag{6.15b}$$

with $\alpha_1 < \alpha_2 < \cdots < \alpha_r$ and $0 \le \alpha_i \le 3$. It is important to note that U_σ is not a Clifford algebra homomorphism (for example, $1 = U_\sigma(1) = U_\sigma(\Sigma^0\Sigma^0)$ whereas $U_\sigma(\Sigma^0)\,U_\sigma(\Sigma^0) = 2e^{-2\sigma}$), but that U_σ is a vector space isomorphism. We define a new associative multiplication $*_\sigma$ on $\mathbb{R}_{1,3}$:

$$\Sigma^{\alpha_1} *_\sigma \Sigma^{\alpha_2} *_\sigma \cdots *_\sigma \Sigma^{\alpha_r} \equiv U_\sigma^{-1}(U_\sigma(\Sigma^{\alpha_1})\,U_\sigma(\Sigma^{\alpha_2})\cdots U_\sigma(\Sigma^{\alpha_r})) \tag{6.16}$$

with $0 \le \alpha_i \le 3$ and calculate

$$\Sigma^0 *_\sigma \Sigma^0 = 2e^{-2\sigma}, \quad \Sigma^0 *_\sigma \Sigma^A = \Sigma^0\Sigma^A, \quad \Sigma^A *_\sigma \Sigma^B = \Sigma^A\Sigma^B. \tag{6.17}$$

Thus $(\mathbb{R}_{1,3}, *_\sigma)$ is a Clifford algebra associated to the bilinear form $\eta_\sigma = \mathrm{diag}(-2e^{-2\sigma}, 1, 1, 1)$ and is therefore isomorphic by U_σ to the usual $\mathbb{R}_{1,3}$. For convenience we let $\mathbb{R}_{1,3}^\sigma$ stand for $(\mathbb{R}_{1,3}, *_\sigma)$.

The limit $\lim_{\sigma\to\infty}\mathbb{R}_{1,3}^\sigma = \mathbb{R}_{1,3}^\infty$ is now of interest and it may be computed by first noting that for $0 \le \alpha_i \le 3$

$$\Sigma^{\alpha_1} *_\sigma \Sigma^{\alpha_2} *_\sigma \cdots *_\sigma \Sigma^{\alpha_r} = (\sqrt{2}\, e^{-\sigma})^{2[p/2]}\Sigma^{\alpha_1}\Sigma^{\alpha_2}\cdots\Sigma^{\alpha_r}, \tag{6.18}$$

where p = number of zeros in $\alpha_1, \alpha_2, \ldots, \alpha_r$ and $[p/2]$ = greatest integer not exceeding $p/2$. As a result of this,

$$\lim_{\sigma\to\infty}\Sigma^{\alpha_1} *_\sigma \cdots *_\sigma \Sigma^{\alpha_r} = \Sigma^{\alpha_1}\cdots\Sigma^{\alpha_r} \text{ or } 0 \tag{6.19}$$

depending whether zero occurs at most once or more than once in $\alpha_1, \alpha_2, \ldots, \alpha_r$. Thus $\mathbb{R}_{1,3}^\infty = \mathbb{R}_{1,0,3}$ the Galilei Clifford algebra. The choice of U_σ was made to ensure that

$$\Sigma^\alpha \to \Theta^\alpha, \quad \alpha = 0, 1, 2, 3, \quad \text{as } \sigma \to \infty.$$

The point to notice now is that the mapping W_σ defined earlier is just the restriction of U_σ to the Lorentz Lie algebra $\mathcal{L}_0$, which sits naturally within $\mathbb{R}_{1,3}$. And in fact after a few computations we see that

$$[J_L^A, J_L^B]_\sigma = J_L^A *_\sigma J_L^B - J_L^B *_\sigma J_L^A = [J_L^A, J_L^B] = \epsilon^{AB}{}_C J_L^C, \tag{6.20a}$$

$$[J_L^A, K_L^B]_\sigma = J_L^A *_\sigma K_L^B - K_L^B *_\sigma J_L^A = [J_L^A, K_L^B] = \epsilon^{AB}{}_C K_L^C, \tag{6.20b}$$

$$[K_L^A, K_L^B]_\sigma = K_L^A *_\sigma K_L^B - K_L^B *_\sigma K_L^A = 2e^{-2\sigma}[K_L^A, K_L^B] = -2e^{-2\sigma}\epsilon^{AB}{}_C J_L^C. \tag{6.20c}$$

(A sample calculation follows:

$$
\begin{aligned}
[K_L^A, K_L^B]_\sigma &\equiv W_\sigma^{-1}[W_\sigma(K_L^A), W_\sigma(K_L^B)] \\
&= U_\sigma^{-1}[U_\sigma(\tfrac{1}{2}\Sigma^0\Sigma^A), U_\sigma(\tfrac{1}{2}\Sigma^0\Sigma^B)] \\
&= U_\sigma^{-1}(\tfrac{1}{4}U_\sigma(\Sigma^0)\,U_\sigma(\Sigma^A)\,U_\sigma(\Sigma^0)\,U_\sigma(\Sigma^B) \\
&\quad - \tfrac{1}{4}U_\sigma(\Sigma^0)\,U_\sigma(\Sigma^B)\,U_\sigma(\Sigma^0)\,U_\sigma(\Sigma^A)) \\
&= K_L^A \overset{*}{\sigma} K_L^B - K_L^B \overset{*}{\sigma} K_L^A \text{ from (6.16) and (6.17)} \\
&= U_\sigma^{-1}(U_\sigma(\tfrac{1}{4}\Sigma^0 \overset{*}{\sigma} \Sigma^A \overset{*}{\sigma} \Sigma^0 \overset{*}{\sigma} \Sigma^B \\
&\quad - \tfrac{1}{4}\Sigma^0 \overset{*}{\sigma} \Sigma^B \overset{*}{\sigma} \Sigma^0 \overset{*}{\sigma} \Sigma^A)) \text{ from (6.17)} \\
&= \tfrac{1}{2}e^{-2\sigma}(\Sigma^0\Sigma^A\Sigma^0\Sigma^B - \Sigma^0\Sigma^B\Sigma^0\Sigma^A) \text{ from (6.18)} \\
&= 2e^{-2\sigma}[K_L^A, K_L^B]\cdot)
\end{aligned}
$$

Thus $[\mathbf{J}_L, \mathbf{J}_L]_\infty = \mathbf{J}_L$, $[\mathbf{J}_L, \mathbf{K}_L]_\infty = \mathbf{K}_L$, $[\mathbf{K}_L, \mathbf{K}_L]_\infty = 0$ defining the Galilei Lie algebra, and the relation between $[\ ,\ ]_\sigma$ and $\overset{*}{\sigma}$ is explicit and shows that the Clifford algebra contraction $\mathbb{R}_{1,3} \to \mathbb{R}_{1,0,3}$ induces the Lie algebra contraction $\mathcal{l}_0 \to \mathcal{g}_0$.

We make one last remark on the connection between the two types of contraction. This involves a contraction of the de Sitter Lie algebra induced by an obvious contraction of $\mathbb{R}_{1,4}$. We define a linear mapping $U_\sigma : \mathbb{R}_{1,4} \to \mu(\sigma)\,\mathbb{R}_{1,4}$ by

$$\Sigma^0 \to \sqrt{2}\,e^{-\sigma}\Sigma^0(\sigma), \quad \Sigma^4 \to \sqrt{2}\,e^{-\sigma}\Sigma^4(\sigma), \quad \Sigma^A \to \Sigma^A(\sigma), \tag{6.21a}$$

$$\Sigma^{a_1}\Sigma^{a_2}\cdots\Sigma^{a_r} \to U_\sigma(\Sigma^{a_1})\,U_\sigma(\Sigma^{a_2})\cdots U_\sigma(\Sigma^{a_r}), \tag{6.21b}$$

with $a_1 < a_2 < \cdots < a_r$, $a_i = 0, 1, 2, 3, 4$. A new product $\overset{*}{\sigma}$ is defined on $\mathbb{R}_{1,4}$ as before:

$$\Sigma^{a_1} \overset{*}{\sigma} \cdots \overset{*}{\sigma} \Sigma^{a_r} \equiv U_\sigma^{-1}(U_\sigma(\Sigma^{a_1})\cdots U_\sigma(\Sigma^{a_r})) \tag{6.22}$$

and a new Clifford algebra $\mathbb{R}_{1,4}^\sigma = (\mathbb{R}_{1,4}, \overset{*}{\sigma})$ corresponding to the bilinear form $g_\sigma = \mathrm{diag}(-2e^{-2\sigma}, 1, 1, 1, 2e^{-2\sigma})$ is obtained as may be shown by noting that

$$\Sigma^{a_1} \overset{*}{\sigma} \Sigma^{a_2} \overset{*}{\sigma} \cdots \overset{*}{\sigma} \Sigma^{a_r} = (\sqrt{2}e^{-\sigma})^{2[p/2]+2[q/2]}\Sigma^{a_1}\Sigma^{a_2}\cdots\Sigma^{a_r}, \tag{6.23}$$

where 0 occurs p times and 4 occurs q times in $a_1, a_2, \ldots, a_r$. We find therefore that $\mathbb{R}_{1,4}^\infty = \lim_{\sigma\to\infty}\mathbb{R}_{1,4}^\sigma = \mathbb{R}_{2,0,3}$.

By using the relation $[x, y]_\sigma \equiv x \overset{*}{\sigma} y - y \overset{*}{\sigma} x$ we find [in the notation of (6.7)]

$$[\mathbf{J}_S, \mathbf{J}_S]_\sigma = \mathbf{J}_S, \quad [\mathbf{J}_S, \mathbf{K}_S]_\sigma = \mathbf{K}_S, \quad [\mathbf{J}_S, \mathbf{P}_S]_\sigma = \mathbf{P}_S, \tag{6.24a}$$

$$[\mathbf{K}_S, \mathbf{K}_S]_\sigma = -2e^{-2\sigma}\mathbf{J}_S, \quad [\mathbf{K}_S, \mathbf{P}_S]_\sigma = -E_S, \quad [\mathbf{P}_S, \mathbf{P}_S]_\sigma = 2e^{-2\sigma}\mathbf{J}_S, \tag{6.24b}$$

$$[\mathbf{J}_S, E_S]_\sigma = 0, \quad [\mathbf{K}_S, E_S]_\sigma = -2e^{-2\sigma}\mathbf{P}_S, \quad [\mathbf{P}_S, E_S]_\sigma = -2e^{-2\sigma}\mathbf{K}_S, \tag{6.24c}$$

so that, taking the limit $\sigma \to \infty$,

$$[\mathbf{J}_S, \mathbf{J}_S]_\infty = \mathbf{J}_S, \quad [\mathbf{J}_S, \mathbf{K}_S]_\infty = \mathbf{K}_S, \quad [\mathbf{J}_S, \mathbf{P}_S]_\infty = \mathbf{P}_S, \tag{6.25a}$$

$$[\mathbf{K}_S, \mathbf{K}_S]_\infty = 0, \quad [\mathbf{K}_S, \mathbf{P}_S]_\infty = -E_S, \quad [\mathbf{P}_S, \mathbf{P}_S]_\infty = 0, \tag{6.25b}$$

$$[\mathbf{J}_S, E_S]_\infty = 0, \quad [\mathbf{K}_S, E_S]_\infty = 0, \quad [\mathbf{P}_S, E_S]_\infty = 0. \tag{6.25c}$$

This Lie algebra (the so-called Carroll Lie algebra[49] is a generalized contraction[50] of $\mathcal{s}$ as can be seen by defining

$$W_\sigma(\mathbf{J}_S) = \mathbf{J}_S(\sigma), \quad W_\sigma(\mathbf{K}_S) = \sqrt{2}\,e^{-\sigma}\mathbf{K}_S(\sigma), \tag{6.26a}$$

$$W_\sigma(\mathbf{P}_S) = \sqrt{2}\,e^{-\sigma}\mathbf{P}_S(\sigma), \quad W_\sigma(E_S) = 2e^{-2\sigma}E_S(\sigma) \tag{6.26b}$$

and noticing that $[x, y]_\sigma = W_\sigma^{-1}[W_\sigma x, W_\sigma y]$.

7. CONCLUSION AND PROSPECTS

In summary, the Galilei group possesses a spin group consisting of invertible elements of the Clifford algebra corresponding to the degenerate Galilei "metric." This group which coincides with the twofold (and in fact the universal) covering group $\mathbb{R}^3 \circledS SU(2)$ was obtained by an approach novel at least in its generality and which at the same time covers the more usual relativistic situation. There is a form of limit or contraction of Clifford algebras which induces a Lie algebra contraction between the Lorentz and Galilei Lie algebras. This idea which seems to be new, might shed light on the various older Lie algebra contraction schemes.

The present formulation permits one to write down[51] a Galilei-covariant Dirac-like equation on a general curved Newtonian space—time. One should thereby be able to describe a nonrelativistic electron in an external electromagnetic field within the framework of Newtonian relativity, illustrating a close anlogy with the situation in general relativity.

For future consideration there are several problems which may be tackled.

One could study the definition given for Spin(X) and determine whether or not it is adequate when X is an arbitrary degenerate space. Some preliminary calculations involving O(X) suggest that it is not adequate.

There is reason to believe that the extended Galilei group may find its way into the Clifford algebra formulation of the Galilei group. This remains to be investigated.

A complete representation theory of Galilei Clifford algebras seems not to have been carried out. Such would likely prove informative in its relation to the representation theory of the Galilei group.

It would be of interest to examine in greater detail the idea of Clifford algebra contraction especially as regards its Lie algebra connections. Questions as to what extent one type of contraction determines the other remain open.

ACKNOWLEDGMENTS

The author is grateful to J.-F. Dumais and H. P. Künzle for a number of interesting conversations related to this work.

[1] R. Penrose and M. MacCallum, Phys. Rep. C 6, 242 (1972).
[2] P. M. Dirac, Ann. Math. 36, 657 (1935).
[3] J. Hilgevoord and J. de Vos, Nucl. Phys. B 1, 494 (1967).
[4] W. Drechsler, Fortschr. Phys. 23, 607 (1975).
[5] J.-M. Souriau, in *Symposia Mathematica* (Academic, New York, 1973), Vol. 14.

[6]V. Bargmann, Ann. Math. **59**, 1 (1954).
[7]J.-M. Lévy-Leblond, in *Group Theory and Its Applications*, edited by E. Loebl (Academic, New York, 1971), Vol. 2.
[8]U. Niederer and L. O'Raifeartaigh, Fortschr. Phys. **22**, 111, 131 (1974).
[9]K. Bardakci and M. Halpern, Phys. Rev. **176**, 1686 (1968).
[10]E. Elizalde and J. Gomis, Nuovo Cimento A **35**, 336, 347, 367 (1976).
[11]F. Mansouri and L. Chang, Phys. Rev. D **13**, 3192 (1976).
[12]W.K. Clifford, Proc. Lond. Math. Soc. **4**, 381 (1873).
[13]A. Buchheim, Am. J. Math. **7**, 293 (1885).
[14]E. Study, Monatsschr. Math. Phys. **1**, 283 (1890), p. 350; and Math. Ann. **39**, 441 (1891), p. 514.
[15]W. Blaschke and H. Müller, *Ebene Kinematik* (Oldenbourg, München, 1956); Chap. 4.
[16]E. Inönü and E. Wigner, Proc. Natl. Acad. Sci. (US) **39**, 510 (1953).
[17]E. Saletan, J. Math. Phys. **2**, 1 (1961).
[18]R. Hermann, *Lie Groups for Physicists* (Benjamin, New York, 1966).
[19]H. Bacry and J.-M. Lévy-Leblond, J. Math. Phys. **9**, 1605 (1968).
[20]I. Porteous, *Topological Geometry* (van Nostrand, Princeton, N.J., 1969).
[21]C. Chevalley, *Algebraic Theory of Spinors* (Columbia University, New York, 1954).
[22]E. Artin, *Geometric Algebra* (Interscience, New York, 1957).
[23]J. Dieudonné, *Géométrie des Groupes Classiques* (Springer, Berlin, 1963).
[24]E. Cartan, *Theory of Spinors* (MIT, Cambridge, 1966).
[25]See Ref. 20, Chap. 13.
[26]See Ref. 21.
[27]R. Brauer and H. Weyl, Am. J. Math. **57**, 425 (1935).
[28]H. Weyl, *Classical Groups* (Princeton University, Princeton, N.J., 1939).
[29]See Ref. 22.
[30]See Ref. 23.
[31]R. Takahashi, Bull. Soc. Math. Fr. **91**, 289 (1963).
[32]S. Ström, in *Lectures in Theoretical Physics*, edited by A. Barut and W. Britten (Colorado University, Boulder, 1971); Vol. 13 and others therein.
[33]F. Gürsey, in *Group Theoretical Concepts, Methods in Elementary Particle Physics*, edited by F. Gürsey (Gordon-Breach, New York, 1964).
[34]T. Philips and E. Wigner, in *Group Theory and Applications*, edited by E. Loebl (Academic, New York, 1968), Vol. 1.
[35]See Ref. 3.
[36]J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **17**, 717 (1976).
[37]See Ref. 25, Cor. 13.9; Thm. 13.10; Cor. 13.14
[38]See Ref. 25, Cor. 13.34.
[39]See Ref. 26, Prop. 13.37; Prop. 13.38.
[40]See Ref. 25, see p. 255.
[41]H. Weyl, *Space, Time, Matter* (Dover, New York, 1952), p. 156.
[42]A. Trautman, in *Brandeis Summer Institute* 1964 (Prentice-Hall, Englewood Cliffs, N.J., 1965), Chap. 5.
[43]H.P. Künzle, Ann. Inst. Henri Poincare **17**, 337 (1972).
[44]L. Jansen and M. Boon, *Theory of Finite Groups: Applications in Physics* (North-Holland, Amsterdam, 1967), problem 31, p. 75.
[45]J.-M. Souriau, *Géométrie et Relativité* (Hermann, Paris, 1964), p. 439.
[46]H. Bacry, *Leçons sur la Théorie des Groupes et les Symétries des Particules Elémentaires* (Gordon-Breach, Paris, 1967), p. 211.
[47]Y. Kosmann, Anal. Math. Pura Appl. **91**, 317 (1972), pp. 319–322.
[48]See Ref. 25, p. 268.
[49]See Ref. 19.
[50]See Ref. 18, p. 100.
[51]J.A. Brooke (paper in preparation).

Solitonlike solutions of the elliptic sine–cosine equation by means of harmonic functions

G. Leibbrandt

Lyman Laboratory of Physics, Harvard University, Cambridge, Massachusetts 02138
(Received 2 November 1977)

Exact solutions of the elliptic sine–cosine equation $\partial^2\psi/\partial x^2+\partial^2\psi/\partial y^2=\sin(\psi+g)$ are derived in two space dimensions with the aid of a new Bäcklund transformation and by exploiting the properties of the *harmonic* function $g(x,y)$. Two generating formulas are developed which allow us to generate *without additional quadratures* an infinite number of real solutions α and infinitely many imaginary solutions $i\beta$. Some α solutions behave like solitons and can be labeled by a topological quantum number. Which solutions are solitonlike and which are not depends decisively on the analytic structure of g and its domain of harmonicity in the R^2 plane.

I. INTRODUCTION

Nonlinear partial differential equations[1,2] continue to challenge the imagination and mathematical ingenuity of a wide spectrum of physicists both at the classical level and in quantum theory and in such diverse branches as plasma physics,[3] nonlinear optics,[4,5] hydrodynamics,[6,7] and particle physics.[8,9] Especially prominent among these partial differential equations are the Korteweg–de Vries equation,[10] the nonlinear Schrödinger equation,[5,11,12] and the sine-Gordon system.[9,11,13]

Investigation of these and related equations[14] has contributed significantly to our understanding of solitary waves and solitons[2] and has led to the development of several new techniques, known as nonperturbative techniques,[15] such as the inverse scattering method, for example.[16] The study of nonlinear dispersive phenomena also involves, to a greater or lesser extent, the study of Bäcklund transformations[17] and Bianchi diagrams, of topological quantum numbers,[18] integrable Hamiltonian systems[7,19] and of the existence of an infinite number of conserved currents.[20,21]

Encouraging as recent developments in this field have been, one must not forget that most nonlinear partial differential equations of physical importance have only been solved in one space and one time dimension, a stark reminder that *exact* solutions of four-dimensional nonlinear equations of second order are hard to come by.

The purpose of this article is to exploit the powerful method of Bäcklund transformations to derive in two space dimensions exact solutions of the *elliptic sine-cosine equation*

$$\nabla^2\psi=\sin\psi\cos g+\cos\psi\sin g=\sin(\psi+g),\qquad \nabla^2\equiv\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}, \tag{1.1}$$

where $\psi(x,y)$ is a massless scalar field, $g(x,y)$ is a *harmonic* function, x and y are space variables, and $\partial_x=\partial/\partial x$, $\partial_y=\partial/\partial y$. For constant g, system (1.1) reduces to the well-known equations

$$\nabla^2\psi=\begin{cases}\sin\psi & (1.2a)\\ \text{or}\\ \cos\psi & (1.2b)\end{cases}$$

and their hyperbolic versions

$$\nabla^2\varphi=\begin{cases}\sinh\varphi & (1.3a)\\ \text{or}\\ \cosh\varphi. & (1.3b)\end{cases}$$

The case $g=0$, Eq. (1.2a), has already been treated in Ref. 22 and we shall frequently refer to it for both method and content.

The article is organized thus. After some initial comments on harmonic functions in Sec. II, we discuss the Bäcklund transformation and associated Bianchi diagram for system (1.1). Some simple, but exact, solutions of the elliptic sine-cosine equation are derived in Sec. III and illustrated there with several examples. In the first part of Sec. IV we present, without proof, a generating formula for real solutions $\alpha(x,y)$; in Sec. IV B we prove a second formula which permits us to generate *without additional quadratures* an infinite number of imaginary solutions $i\beta(x,y)$. The article concludes in Sec. V with a brief summary and discussion.

II. HARMONIC FUNCTIONS AND A BÄCKLUND TRANSFORMATION

A. Harmonic functions in the plane

As stated in the introduction, our aim is to find exact solutions of the elliptic sine-cosine equation

$$\nabla^2\psi=\sin\psi\cos g+\cos\psi\sin g=\sin(\psi+g), \tag{2.1}$$

where ψ is a massless scalar field and $g(x,y)$ is a *harmonic function* of the space variables x and y. We recall[23,24] that a real-valued function $u(x,y)$ of two real variables x,y is said to be *harmonic* in a domain D of the R^2 plane, if it has continuous first and second partial derivatives in D and if it satisfies Laplace's equation

$$\nabla^2u(x,y)=0 \tag{2.2}$$

in D.

There exists an intimate connection between harmonic and analytic functions which is expressed in the following two lemmas:

Lemma 1: Let $f(z)=u(x,y)+iv(x,y)$ be *analytic* in a domain D, with $z=x+iy$. Then the component functions $u(x,y)$ and $v(x,y)$ are *harmonic* in D.

 0022-2488/78/1905-0960$1.00

Lemma 2: If two harmonic functions u and v satisfy in a domain D the Cauchy–Riemann conditions

$$\frac{\partial u}{\partial x}=\frac{\partial v}{\partial y} \quad \text{and} \quad \frac{\partial u}{\partial y}=-\frac{\partial v}{\partial x}, \tag{2.3}$$

then v is called the *harmonic conjugate* of u. Conversely one may call u the harmonic conjugate of $-v$.

The importance of these two lemmas will become apparent during the course of the discussion. For other properties of harmonic functions, which play a significant role in complex variable theory and in many branches of mathematical physics, we refer the reader to the extensive literature on this subject.

B. Bäcklund transformations and Bianchi diagrams

It turns out that the most general Bäcklund transformation associated with the elliptic sine-cosine equation (2.1) is of the form

$$(\partial_x+i\partial_y)\left(\frac{\alpha+u-i(\beta+v)}{2}\right)=\exp(i\phi) \times \sin\left(\frac{\alpha+u+i(\beta+v)}{2}\right), \tag{2.4}$$

or, when complex conjugated,

$$(\partial_x-i\partial_y)\left(\frac{\alpha+u+i(\beta+v)}{2}\right)=\exp(-i\phi)\sin\left(\frac{\alpha+u-i(\beta+v)}{2}\right), \tag{2.5}$$

where α and β are both real, $u(x,y)$ and $v(x,y)$ are harmonic functions satisfying

$$\nabla^2 u(x,y)=0=\nabla^2 v(x,y), \tag{2.6}$$

and ϕ is the Bäcklund transformation parameter.[25] Equation (2.4) represents a transformation from the "old" solution $(\alpha+u)$ to the "new" solution $i(\beta+v)$,

$$i(\beta+v)=B_\phi(\alpha+u), \tag{2.7a}$$

which can, in turn, be represented by the Bianchi diagram shown in Fig. 1. B_ϕ is called the Bäcklund transformation operator and is characterized by ϕ. The "inverse" of (2.7a) reads

$$u+\alpha=i(B_\phi)^{-1}(\beta+v), \tag{2.7b}$$

where

$$(B_\phi)^{-1}=(-1)^m B_\phi, \quad m=1,3,5,\cdots.$$

To verify that (2.4) is indeed the correct Bäcklund transformation for (2.1), multiply (2.4) from the left by $(\partial_x-i\partial_y)$ and first apply Eq. (2.5), then E_1. (2.6) to get

$$\nabla^2(\alpha-i\beta)=\sin\left(\frac{\alpha+u-i(\beta+v)}{2}\right)\cos\left(\frac{\alpha+u+i(\beta+v)}{2}\right).$$

Separating this equation into real and imaginary components, we find that $(\alpha+u)$ and $(\beta+v)$ satisfy

$$\nabla^2\alpha=\sin(\alpha+u) \tag{2.8}$$

and

$$\nabla^2\beta=\sinh(\beta+v), \tag{2.9}$$

respectively. As a starter we see that Eqs. (2.8) and (2.9) possess infinitely many "vacuum" solutions of the form

$$\alpha(x,y)=n\pi-u(x,y), \quad n=0,\pm1,\pm2,\cdots, \tag{2.10}$$

and

$$\beta(x,y)=im\pi-v(x,y), \quad m=0,\pm1,\pm2,\cdots. \tag{2.11}$$

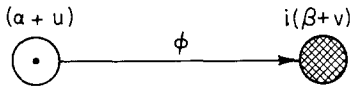

FIG. 1. Bianchi diagram for the Bäcklund transformation from $(\alpha+u)$ to $i(\beta+v)$, characterized by the real parameter ϕ.

Before finding their general solutions, we should like to make the following observation.

It is known[17] that a Bäcklund transformation may be regarded, in geometrical language, as a transformation of a surface S into a new surface S', where S is a solution of a given partial differential equation, but where the transformed surface S' may either be a solution of the original partial differential equation or of some other differential equation.[26] Suppose we elect to call (2.8) the *original* equation. Then the Bäcklund transformation (2.4) is seen to generate the new surface $i(\beta+v)$ which is a solution, *not* of (2.8), but rather of the different differential equation (2.9). We assume, of course, that $u\neq v$.

III. SIMPLE SOLUTIONS FOR α AND β

A. First generation of nontrivial solutions

It is straightforward now, with the help of the Bäcklund transformation (2.4), to derive the first generation of α and β solutions. Unless otherwise stated, the reader may assume that the harmonic functions u and v are distinct and different from zero.

We first decompose (2.4) into two real differential equations:

$$\partial_x\left(\frac{\alpha+u}{2}\right)+\partial_y\left(\frac{\beta+v}{2}\right)=\cos\phi\sin\left(\frac{\alpha+u}{2}\right)\cosh\left(\frac{\beta+v}{2}\right) -\sin\phi\cos\left(\frac{\alpha+u}{2}\right)\sinh\left(\frac{\beta+v}{2}\right), \tag{3.1}$$

$$\partial_y\left(\frac{\alpha+u}{2}\right)-\partial_x\left(\frac{\beta+v}{2}\right)=\cos\phi\cos\left(\frac{\alpha+u}{2}\right)\sinh\left(\frac{\beta+v}{2}\right) +\sin\phi\sin\left(\frac{\alpha+u}{2}\right)\cosh\left(\frac{\beta+v}{2}\right). \tag{3.2}$$

1. *α solutions*: We set $\beta+v=0\equiv\beta_0+v$ (a "vacuum" solution) and note that $\beta=0$ does *not* solve (2.9). Substituting $\beta_0=-v$ into Eq. (3.1), so that

$$\partial_x(\alpha+u)/2=\cos\phi\sin[(\alpha+u)/2], \tag{3.3}$$

and integrating partially with respect to x, we obtain (call $\alpha\equiv\alpha_1$):

$$\tan\left(\frac{\alpha_1+u}{4}\right)=c\exp(x\cos\phi+y\sin\phi), \quad c \text{ const}, \tag{3.4a}$$

or

$$\alpha_1(x, y, \phi) = 4 \arctan\left(\frac{c \exp(x\cos\phi + y\sin\phi) - \tan[\frac{1}{4}u(x,y)]}{1 + c\tan[\frac{1}{4}u(x,y)]\exp(x\cos\phi + y\sin\phi)}\right), \tag{3.4b}$$

where

$$\nabla^2\alpha_1 = \sin(\alpha_1 + u), \quad \nabla^2 u = 0, \tag{3.5a}$$

$$\nabla^2\beta_0 = \sinh(\beta_0 + v), \quad \nabla^2 v = 0. \tag{3.5b}$$

The expression for α_1 is a genuine two-dimensional solution of (2.8), since there is no value of the Bäcklund parameter ϕ which eliminates completely, for general u, either x or y.

2. *β solutions*: Setting $\alpha + u = 0 \equiv \alpha_0 + u$ (a "vacuum" solution) in Eq. (3.1), we get the first generation of nontrivial β solutions (call $\beta \equiv \beta_1$):

$$\left.\begin{matrix}\tanh\frac{1}{4}(\beta_1 + v) \\ \text{or} \\ \operatorname{cotanh}\frac{1}{4}(\beta_1 + v)\end{matrix}\right\} = (\text{constant factor})\exp(x\cos\phi_1 + y\sin\phi_1), \tag{3.6}$$

or in detail,

$$\beta_1(x, y, \phi_1) = 4\,\operatorname{arctanh}\left(\frac{\bar{c}\exp(x\cos\phi_1 + y\sin\phi_1) - \tanh(\frac{1}{4}v)}{1 - \bar{c}\tanh(\frac{1}{4}v)\exp(x\cos\phi_1 + y\sin\phi_1)}\right),$$

$$\text{if } (x\cos\phi_1 + y\sin\phi_1) \leq 0; \tag{3.7a}$$

or

$$\beta_1(x, y, \phi_1) = 4\,\operatorname{arccotanh}\left(\frac{1 - \bar{c}'\operatorname{cotanh}(\frac{1}{4}v)\exp(x\cos\phi_1 + y\sin\phi_1)}{\bar{c}'\exp(x\cos\phi_1 + y\sin\phi_1) - \operatorname{cotanh}(\frac{1}{4}v)}\right),$$

$$\text{if } (x\cos\phi_1 + y\sin\phi_1) > 0. \tag{3.7b}$$

Here $\bar{c}$, $\bar{c}'$ are integration constants and

$$\nabla^2\alpha_0 = \sin(\alpha_0 + u), \quad \nabla^2 u = 0, \tag{3.8a}$$

$$\nabla^2\beta_1 = \sinh(\beta_1 + v), \quad \nabla^2 v = 0. \tag{3.8b}$$

For general values of the harmonic function $v(x, y)$, Eqs. (3.7) again truly represent two-space dimensional solutions of Eq. (2.9).

B. Examples

Since there are infinitely many analytic functions, the number of harmonic and conjugate harmonic functions is equally large (cf. Lemma 1, Sec. IIA). In the following examples we shall concentrate on a few particular cases to illustrate our method. The reader will notice that most of the examples deal with solutions of $\nabla^2\alpha = \sin(\alpha + u)$, for the simple reason that these are physically more relevant than the β solutions of Eq. (2.9).[27,28]

Example 1: For the *constant* function

$$u(x, y) = A, \quad \mathcal{D} = R^2, \tag{3.9}$$

where $\mathcal{D}$ specifies the domain of harmonicity, Eq. (2.8) reads

$$\nabla^2\alpha_1 = \cos A\,\sin\alpha_1 + \sin A\,\cos\alpha_1, \tag{3.10}$$

with the solution [cf. Eq. (3.4a)]

$$\alpha_1(x, y, \phi) = -A + 4\arctan[c\exp(x\cos\phi + y\sin\phi)]. \tag{3.11}$$

To study its asymptotic behavior, we express α_1 in polar coordinates $x = r\cos\theta$, $y = r\sin\theta$ and choose $c > 0$:

$$\alpha_1(r, \theta, \phi) = -A + 4\arctan\{c\exp[r\cos(\theta - \phi)]\}, \tag{3.12}$$

so that

$$\lim_{r\to\infty}\alpha_1 = 2\pi - A, \quad \text{if } -\tfrac{1}{2}\pi < \theta - \phi < \tfrac{1}{2}\pi, \tag{3.13a}$$

$$\lim_{r\to\infty}\alpha_1 = -A, \quad \text{if } \tfrac{1}{2}\pi < \theta - \phi < \tfrac{3}{2}\pi. \tag{3.13b}$$

It follows from (3.13) that the solution (3.12) can be characterized by a topological charge Q,[18] where

$$Q = \frac{1}{2\pi}[(2\pi - A) - (-A)] = +1. \tag{3.14}$$

The system (3.9)—(3.12) admits, therefore, not only a single solitonlike solution, but also multiple solitonlike solutions which follow quite readily from the generating formula (4.1) in Sec. IV. Finally, a comment on terminology:[29] We speak, in R^2 space, of "solitonlike solutions" rather than "soliton solutions," since the asymptotic behavior of α_1 depends, according to (3.13), on the polar angle θ.

Example 2: Consider the harmonic function

$$u(x, y) = A + B\ln(x^2 + y^2), \quad A, B \text{ constants}, \tag{3.15}$$

whose domain of harmonicity is the punctured R^2 plane: $\mathcal{D} = R^2 - \{0\}$. According to Eqs. (2.8) and (3.4a), the differential equation

$$\nabla^2\alpha_1 = \sin[\alpha_1 + A + B\ln(x^2 + y^2)] \tag{3.16}$$

admits first-generation solutions of the type

$$\alpha_1(x, y, \phi) = -A - B\ln(x^2 + y^2) + 4\arctan[c\exp(x\cos\phi + y\sin\phi)], \tag{3.17}$$

which, it would seem, *cannot* be labeled by a topological quantum number Q.

Example 3: For the harmonic function

$$u(x, y) = \arctan(y/x)$$

$$\mathcal{D} = \{(x, y)\,|\,(x^2 + y^2)^{1/2} > 0, \quad 0 < \arctan(y/x) < 2\pi\}, \tag{3.18}$$

the equation

$$\nabla^2\alpha_1 = \sin(\alpha_1 + \arctan(y/x)) \tag{3.19}$$

leads to the solution

$$\alpha_1(x, y, \phi) = -\arctan(y/x) + 4\arctan[c\exp(x\cos\phi + y\sin\phi)], \quad c > 0, \tag{3.20}$$

or, in polar coordinates, to

$$\alpha_1(r, \theta, \phi) + \theta = +4\arctan\{c\exp[r\cos(\theta - \phi)]\}, \quad c > 0,$$

$$\mathcal{D} = \{(r, \theta)\,|\,r > 0, \quad 0 < \theta < 2\pi\}. \tag{3.21}$$

To determine whether or not this solution carries

topological charge, implying a certain solitonlike behavior, we examine the asymptotic structure of (3.21) in specific sectors of the $r-\theta$ plane.[30] Clearly

$$\lim_{r\to\infty}(\alpha_1+\theta)=2\pi,$$

if

$$-\tfrac{1}{2}\pi<\theta-\phi<0 \text{ and } 0<\theta-\phi<\tfrac{1}{2}\pi, \tag{3.22a}$$

and

$$\lim_{r\to\infty}(\alpha_1+\theta)=0, \quad \text{if } \tfrac{1}{2}\pi<\theta-\phi<\tfrac{3}{2}\pi, \tag{3.22b}$$

so that the solution $[\alpha_1(r,\theta,\phi)+\theta]$, or, equivalently, $[\alpha_1(x,y,\phi)+\arctan(y/x)]$ can indeed be labeled by the topological charge $Q=+1$. Some of the solutions of Eq. (3.19) are, therefore, solitonlike in nature.

The next example illustrates the method for β solutions.

Example 4: Consider the harmonic function

$$v(x,y)=e^x\cos y, \quad \mathcal{D}\subset\mathcal{R}^2, \tag{3.23}$$

for which Eq. (2.9) reads

$$\nabla^2\beta_1=\sinh(\beta_1+e^x\cos y). \tag{3.24}$$

According to Eqs. (3.6) and (3.7), the solutions are of the form

$$\beta_1(x,y,\phi_1)$$
$$=-e^x\cos y+4\,\mathrm{arctanh}[\bar{c}\exp(x\cos\phi_1+y\sin\phi_1)],$$
$$\text{if } (x\cos\phi_1+y\sin\phi_1)\le 0, \quad \bar{c}>0, \tag{3.25}$$

and

$$\beta_1(x,y,\phi_1)$$
$$=-e^x\cos y+4\,\mathrm{arccotanh}[\bar{c}'\exp(x\cos\phi_1+y\sin\phi_1)],$$
$$\text{if } (x\cos\phi_1+y\sin\phi_1)>0, \quad \bar{c}'>0. \tag{3.26}$$

The behavior of these expressions for large $|x|$ and large $|y|$ is, again, most easily studied in the polar representation, but we shall not pursue this problem any further.

In the next section we demonstrate the existence of a whole hierarchy of α and β solutions.

IV. TWO GENERATING FORMULAS

A. Generating formula for α solutions

In this section we give a formula for generating an infinite number of real solutions of the equation $\nabla^2\alpha=\sin(\alpha+u)$, Eq. (2.8). The result is contained in the following theorem.

Theorem 1: Let (α_0+u) be a solution of Eq. (2.8) and let $\beta_1^{(1)}$ and $\beta_1^{(2)}$ be two distinct solutions of Eq. (2.9), the three solutions being related by $i(\beta_1^{(1)}+v)=B_{\phi_1}(\alpha_0+u)$ and $i(\beta_1^{(2)}+v)=B_{\phi_2}(\alpha_0+u)$. A new α solution, called α_2, is then given by

$$\tan\left(\frac{(\alpha_2+u)-(\alpha_0+u)}{4}\right)$$
$$=\cot\left(\frac{\phi_1-\phi_2}{2}\right)\tanh\left(\frac{(\beta_1^{(1)}+v)-(\beta_1^{(2)}+v)}{4}\right), \tag{4.1}$$

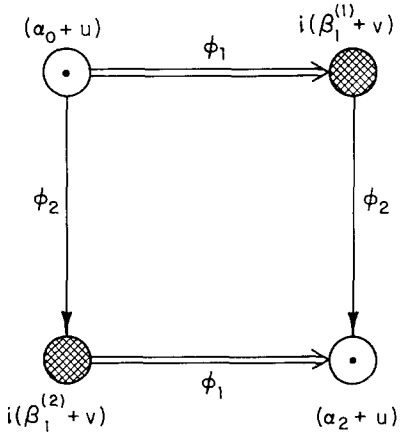

FIG. 2. Bianchi diagram associated with the generating formula (4.1), Theorem 1.

$$\phi_2\neq\phi_1\pm\pi N, \quad N=0,1,2,\cdots,$$

where α_2 satisfies Eq. (2.8), u and v are harmonic in a certain domain $\mathcal{D}$ of the $\mathcal{R}^2$ plane, ϕ_1 and ϕ_2 are Bäcklund transformation parameters and $\beta_1^{(j)}\equiv\beta_1(\phi_j)$, $j=1,2$. (See Fig. 2).

Since the proof of Theorem 1 follows closely our derivation of Eq. (2.12) in Ref. 22, we shall omit it here and instead prove Theorem 2 below. However, before leaving Theorem 1, we should like to make the following remarks about formula (4.1).

(a) One can verify, by explicit differentiation of (4.1), that α_2 does indeed satisfy $\nabla^2\alpha_2=\sin(\alpha_2+u)$ for u and v harmonic, provided $\nabla^2\alpha_0=\sin(\alpha_0+u)$ and $\nabla^2\beta_1^{(j)}=\sinh(\beta_1^{(j)}+v)$, $j=1,2$.

(b) Since (4.1) holds not just for the "vacuum" solution $\alpha_0+u=0$, but for any solution (α_0+u) of (2.8) and its Bäcklund-generated solutions $(\beta_1^{(1)}+v)$ and $(\beta_1^{(2)}+v)$, we can easily generalize (4.1) to read

$$\tan\left(\frac{(\alpha_{2n}+u)-(\alpha_{2n-2}+u)}{4}\right)$$
$$=\cot\left(\frac{\phi_{2n-1}-\phi_{2n}}{2}\right)\tanh\left(\frac{(\beta_{2n-1}^{(1)}+v)-(\beta_{2n-1}^{(2)}+v)}{4}\right),$$
$$n=1,2,3,\cdots \text{ and } \phi_{2n}\neq\phi_{2n-1}\pm\pi N, \quad N=0,1,2,\cdots. \tag{4.2}$$

The last formula can be written symbolically as

$$(\alpha_{2n}+u)=(B_{\phi_{2n-1}}B_{\phi_{2n}})(\alpha_{2n-2}+u)$$
$$=(B_{\phi_{2n}}B_{\phi_{2n-1}})(\alpha_{2n-2}+u), \tag{4.3}$$

and allows us to generate *without additional quadratures* an infinite number of real solutions α.

B. Generating formula for β solutions

Our next task is to derive a generating formula for solutions of the equation $\nabla^2\beta=\sinh(\beta+v)$, Eq. (2.9).

Theorem 2: Let (β_0+v) be a solution of Eq. (2.9) and let $\alpha_1^{(1)}$ and $\alpha_1^{(2)}$ be two distinct solutions of Eq. (2.8), the three solutions being related by $(\alpha_1^{(1)}+u)$

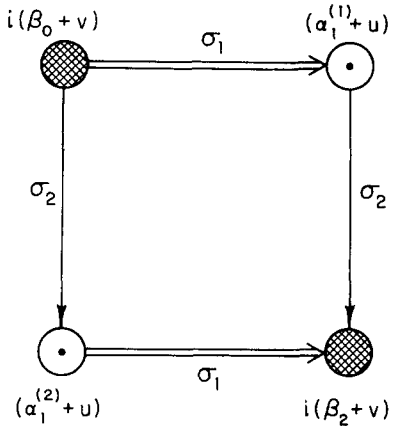

FIG. 3. Bianchi diagram used in the proof of the generating formula (4.9), Theorem 2.

$=iB_{\sigma_1}(\beta_0+v)$ and $(\alpha_1^{(2)}+u)=iB_{\sigma_2}(\beta_0+v)$. A new β solution, called β_2, is then given by

$$\tanh\left(\frac{(\beta_2+v)-(\beta_0+v)}{4}\right) = \cot\left(\frac{\sigma_1-\sigma_2}{2}\right)\tan\left(\frac{(\alpha_1^{(2)}+u)-(\alpha_1^{(1)}+u)}{4}\right),$$

$\sigma_2\neq\sigma_1\pm\pi N,\quad N=0,\ 1,\ 2,\cdots,$

where β_2 satisfies $\nabla^2\beta_2=\sinh(\beta_2+v)$, u and v are again harmonic in a certain domain $\mathcal{D}$, σ_1 and σ_2 are real Bäcklund parameters, and $\alpha_1^{(j)}\equiv\alpha_1(\sigma_j)$, $j=1,\ 2$.

Proof: From the Bianchi diagram shown in Fig. 3 and the basic Bäcklund transformation (2.4) we first deduce the transformations

$$\alpha_1^{(1)}+u=iB_{\sigma_1}(\beta_0+v), \tag{4.4a}$$

$$\alpha_1^{(2)}+u=iB_{\sigma_2}(\beta_0+v), \tag{4.4b}$$

$$i(\beta_2+v)=B_{\sigma_2}(\alpha_1^{(1)}+u), \tag{4.4c}$$

$$i(\beta_2+v)=B_{\sigma_1}(\alpha_1^{(2)}+u), \tag{4.4d}$$

which are equivalent to the first-order differential equations

$$(\partial_x+i\partial_y)\left(\frac{i(\beta_0+v)-(\alpha_1^{(1)}+u)}{2}\right) = \exp(i\sigma_1)\sin\left(\frac{i(\beta_0+v)+\alpha_1^{(1)}+u}{2}\right), \tag{4.5a}$$

$$(\partial_x+i\partial_y)\left(\frac{i(\beta_0+v)-(\alpha_1^{(2)}+u)}{2}\right) = \exp(i\sigma_2)\sin\left(\frac{i(\beta_0+v)+\alpha_1^{(2)}+u}{2}\right), \tag{4.5b}$$

$$(\partial_x+i\partial_y)\left(\frac{\alpha_1^{(1)}+u-i(\beta_2+v)}{2}\right) = \exp(i\sigma_2)\sin\left(\frac{\alpha_1^{(1)}+u+i(\beta_2+v)}{2}\right), \tag{4.5c}$$

$$(\partial_x+i\partial_y)\left(\frac{\alpha_1^{(2)}+u-i(\beta_2+v)}{2}\right) = \exp(i\sigma_1)\sin\left(\frac{\alpha_1^{(2)}+u+i(\beta_2+v)}{2}\right), \tag{4.5d}$$

respectively. Addition of Eqs. (4.5) according to the prescription [(4.5a)–(4.5b)+(4.5c)–(4.5d)] leads to

$$0=\exp(i\sigma_1)\left[\sin\left(\frac{i(\beta_0+v)+\alpha_1^{(1)}+u}{2}\right)-\sin\left(\frac{\alpha_1^{(2)}+u+i(\beta_2+v)}{2}\right)\right] -\exp(i\sigma_2)\left[\sin\left(\frac{i(\beta_0+v)+\alpha_1^{(2)}+u}{2}\right)-\sin\left(\frac{\alpha_1^{(1)}+u+i(\beta_2+v)}{2}\right)\right],$$

or

$$\exp(i\sigma_1)\sin\left(\frac{i(\beta_0-\beta_2)+(\alpha_1^{(1)}-\alpha_1^{(2)})}{4}\right) = \exp(i\sigma_2)\sin\left(\frac{i(\beta_0-\beta_2)-(\alpha_1^{(1)}-\alpha_1^{(2)})}{4}\right), \tag{4.6}$$

provided

$$\cos\tfrac{1}{4}[\alpha_1^{(1)}+\alpha_1^{(2)}+2u+i(\beta_0+\beta_2+2v)]\neq 0. \tag{4.7}$$

Noting that

$$\frac{\exp(i\sigma_1)+\exp(i\sigma_2)}{\exp(i\sigma_1)-\exp(i\sigma_2)}=-i\cot\left(\frac{\sigma_1-\sigma_2}{2}\right), \tag{4.8}$$

we may manipulate (4.6) into the form

$$\tanh\left(\frac{(\beta_2+v)-(\beta_0+v)}{4}\right) = \cot\left(\frac{\sigma_1-\sigma_2}{2}\right)\tan\left(\frac{(\alpha_1^{(2)}+u)-(\alpha_1^{(1)}+u)}{4}\right), \tag{4.9}$$

$\sigma_2\neq\sigma_1\pm\pi N,\quad N=0,\ 1,\ 2,\cdots,$

which is the desired result.

Remarks similar to those following Theorem 1 also apply to the generating formula (4.9). Thus (β_0+v) may be any solution of $\nabla^2\beta_0=\sinh(\beta_0+v)$, with the corresponding Bäcklund-generated α solutions satisfying $\nabla^2\alpha_1^{(j)}=\sin(\alpha_1^{(j)}+u)$, $j=1,\ 2$. Moreover, we see from the Bianchi diagram Fig. 3 that

$$i(\beta_2+v)=i(B_{\sigma_1}B_{\sigma_2})(\beta_0+v)=i(B_{\sigma_2}B_{\sigma_1})(\beta_0+v). \tag{4.10}$$

The generalization of formula (4.9) to multiple β solutions is straightforward and yields

$$\tanh\left(\frac{(\beta_{2n}+v)-(\beta_{2n-2}+v)}{4}\right) = \cot\left(\frac{\sigma_{2n-1}-\sigma_{2n}}{2}\right)\tan\left(\frac{(\alpha_{2n-1}^{(2)}+u)-(\alpha_{2n-1}^{(1)}+u)}{4}\right), \tag{4.11}$$

$n=1,\ 2,\ 3,\cdots$ and $\sigma_{2n}\neq\sigma_{2n-1}\pm\pi N,\quad N=0,\ 1,\ 2,\cdots,$

for any harmonic functions $u(x,y)$ and $v(x,y)$ defined in their appropriate domains.

V. CONCLUDING REMARKS

In this paper we have exploited the properties of harmonic functions to derive, in two space dimensions, exact solutions of the elliptic sine-cosine equation

$$\nabla^2\psi(x,y)=\sin[\psi(x,y)+g(x,y)],\quad \nabla^2 g=0, \tag{5.1}$$

which includes, for g constant, the well-known equations

$$\nabla^2\psi = \begin{pmatrix}\sin\psi\\ \cos\psi\end{pmatrix} \text{ and } \nabla^2\varphi = \begin{pmatrix}\sinh\varphi\\ \cosh\varphi\end{pmatrix}. \tag{5.2}$$

For general forms of the harmonic function g, new and truly two-space dimensional solutions of (5.1) are obtained. Some of the expressions, such as (3.12) and (3.21), behave like solitons or multiple solitons [the latter can be deduced from Eq. (4.1), for example] in that they carry a topological quantum number Q.[18] Which solutions of Eq. (5.1) are solitonlike and which are not, depends decisively on the analytic structure of the harmonic function $g(x,y)$.

Our treatment of the elliptic sine-cosine system is based on the powerful technique of Bäcklund transformations and Bianchi diagrams which has, in our opinion, three distinct advantages.

In the first place, the Bäcklund transformation (2.4) enables us to replace the second-order *partial* differential equation (5.1) by a system of real first-order *ordinary* differential equations, Eqs. (3.1) and (3.2), which are easy to solve.

Secondly, the existence of a Bäcklund transformation—and its associated Bianchi diagram—virtually guarantees the existence of a generating formula and through it, the presence of an infinite set of solutions. The elliptic case, treated in this communication, is characterized by *two* such generating formulas (in contrast to the corresponding hyperbolic equation in 1 + 1 dimensions which possesses only *one* generating formula). The first formula (4.2) generates infinitely many real solutions $\alpha(x,y)$, while the second one, given by Eq. (4.11), yields *without additional quadratures* an infinite number of imaginary solutions $i\beta(x,y)$.

Thirdly, the Bäcklund transformation (2.4) should simplify the investigation of conserved currents,[21] for example, of the equation $\nabla^2\alpha = \sin(\alpha + u)$. The search for conservation laws is, in view of the functional nature of u, somewhat more complicated here than in the case of the sine-Gordon equation in one space and one time dimension.[2] It would seem logical to expect, however, that those solutions which can be labeled by a topological quantum number, would also yield an infinite number of conserved currents. But then nature is not always logical!

We remark in closing that the "procedure of harmonic functions" is also applicable to the *hyperbolic* equation (c is the speed of light in a vacuum)

$$\left(\frac{\partial^2}{\partial x^2} - \frac{\partial^2}{c^2\partial t^2}\right)\psi(x,t) = \sin[\psi(x,t) + k(x,t)], \tag{5.3a}$$

where

$$\left(\frac{\partial^2}{\partial x^2} - \frac{\partial^2}{c^2\partial t^2}\right)k(x,t) = 0, \tag{5.3b}$$

and t is the time variable. Condition (5.3b) for $k(x,t)$ replaces Laplace's condition (2.2) and is obviously satisfied for $k = x$, $k = ct$ as well as for $k = \cos x\cos(ct)$ and $k = \ln(x^2 - c^2t^2)$, $(x^2 - c^2t^2) > 0$. Details, including the appropriate Bäcklund transformation and generating formula for system (5.3), will appear elsewhere.

ACKNOWLEDGMENTS

We should like to acknowledge several illuminating disucssions with Tri-Nang Pham and thank Sidney Coleman and Howard Georgi, as well as the secretarial staff, for their hospitality and assistance during our stay in the Theoretical Physics Group. This research was supported in part by the National Science Foundation, under Grant No. PHY 75-20427, and by the National Research Council of Canada, under Grant No. A8063, while the author was on leave from the Department of Mathematics and Statistics, University of Guelph, Guelph, Ontario.

[1]W. Ames, *Nonlinear Partial Differential Equations* (Academic, New York, 1967); G.B. Whitham, *Linear and Nonlinear Waves* (Wiley, New York, 1974); *Lecture Notes in Physics*, Vol. 38, edited by J. Moser (Springer-Verlag, Berlin, 1975) (Battelle Rencontres, Seattle, 1974).
[2]A.C. Scott, F.Y.F. Chu, and D.W. McLaughlin, Proc. IEEE **61**, 1443 (1973).
[3]H. Washimi and T. Taniuti, Phys. Rev. Lett. **17**, 996 (1966); H. Kever and G.K. Morikawa, Phys. Fluids **12**, 2090 (1969); F.D. Tappert, Phys. Fluids **15**, 2446 (1972).
[4]A.J. De Maria, D.A. Stetser, and W.H. Glenn, Jr., Science **156**, 1557 (1967); G.L. Lamb, Jr., Phys. Lett. A **25**, 181 (1967); Rev. Mod. Phys. **43**, 99 (1971); Phys. Rev. Lett. **31**, 196 (1973).
[5]V.I. Karpman and E.M. Kruskal, Sov. Phys. JETP **28**, 277 (1969).
[6]G.B. Whitham, J. Fluid Mech. **22**, 273 (1965); **44**, 373 (1970); N.J. Zabusky and C.J. Galvin, J. Fluid Mech. **47**, 811 (1971).
[7]V.E. Zakharov and L.D. Faddeev, Funct. Anal. Appl. **5**, 280 (1971).
[8]R. Dashen, B. Hasslacher, and A. Neveu, Phys. Rev. D **10**, 4114, 4130, 4138 (1974); **11**, 3424 (1975); R. Rajaraman, Phys. Rep. C **21**, 227 (1975); U. Enz, Phys. Rev. **131**, 1392 (1963); T.H.R. Skyrme, Proc. Roy. Soc. A **247**, 260 (1958); **262**, 237 (1961).
[9]S. Coleman, Phys. Rev. D **11**, 2088 (1975), and Lectures delivered at the 1975 International School of Subnuclear Physics "Ettore Majorana," Harvard report, 1975 (unpublished).
[10]D.J. Korteweg and G. de Vries, Phil. Mag. **39**, 422 (1895); N.J. Zabusky and M.D. Kruskal, Phys. Rev. Lett. **15**, 240 (1965); R.M. Miura, J. Math. Phys. **9**, 1202, 1204 (1968); C.S. Gardner, J. Math. Phys. **12**, 1548 (1971); M. Wadati and M. Toda, J. Phys. Soc. Jpn. **32**, 1403 (1972).
[11]D.J. Benney and A.C. Newell, J. Math. Phys. **46**, 133 (1967).
[12]A. Hasegawa and F. Tappert, Appl. Phys. Lett. **23**, 142 (1973); K. Shimizu and Y.H. Ichikawa, J. Phys. Soc. Jpn. **33**, 789 (1972).
[13]A. Barone, F. Esposito, and C.J. Magee, Riv. Nuovo Cimento **1**, 227 (1971); B. Hu, Nuovo Cimento A **38**, 441 (1977).
[14]G. Leibbrandt and G. McKeon, Phys. Lett. A **61**, 83 (1977); A.C. Newell, J. Math. Phys. **18**, 922 (1977).
[15]See "Extended Systems in Field Theory," edited by J.L. Gervais and A. Neveu, Phys. Rep. C **23**, 236 (1976).
[16]M.J. Ablowitz, D.J. Kaup, A.C. Newell, and H. Segur, Phys. Rev. Lett. **30**, 1462 (1973); **31**, 125 (1973).
[17]L.P. Eisenhart, *Differential Geometry of Curves and Surfaces* (Dover, New York, 1960); A.R. Forsyth, *Theory of Differential Equations*, Vol. 6 (Dover, New York, 1959), Chap. 21.

[18]D. Finkelstein and C. Misner, Ann. Phys. (N.Y.) **6**, 230 (1959); J. Arafune, P.G.O. Freund, and C. Goebel, J. Math. Phys. **16**, 433 (1975).
[19]K. Pohlmeyer, Commun. Math. Phys. **46**, 207 (1976).
[20]See Ref. 2, Sec. VI.
[21]G. Leibbrandt, Phys. Rev. D **16**, 970 (1977).
[22]G. Leibbrandt, Phys. Rev. B **15**, 3353 (1977).
[23]L.V. Ahlfors, *Complex Analysis* (McGraw-Hill, New York, 1966); R.V. Churchill, J.W. Brown, and R.F. Verhey, *Complex Variables and Applications* (McGraw-Hill, New York, 1974).
[24]F.J. Flanigan, *Complex Variables: Harmonic and Analytic Functions* (Allyn and Bacon, Boston, 1972).
[25]For a comparison with the Bäacklund transformation of the elliptic sine equation, see Eq. (2.2) of Ref. 22.
[26]See in this connection the article by A.C. Scott and D.W. McLaughlin, J. Math. Phys. **14**, 1818 (1973).
[27]See, for example, the discussion on the propagation of magnetic flux through a two-dimensional Josephson junction, as presented in Sec. IV of Ref. 22, especially Eq. (4.11), and also A.C. Scott, Nuovo Cimento B **69**, 241 (1970).
[28]The author has benefitted from several conversations with M. Tinkham on the Josephson effect, especially on Josephson junctions.
[29]The same terminology was used in Ref. 22.
[30]See the comment at the end of Example 1.

Stability of Weiss Ising model

Masanori Ohya

The Science University of Tokyo, Department of Information Science, 278, Noda, Chiba, Japan [a)]
and Institute of Fundamental Studies, Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627
(Received 21 June 1977)

The problems of stability and approach to equilibrium of the Weiss Ising model are studied. Our investigations are performed in the exact and linear response senses in order to compare both theories. The change of a metastable state of the Weiss Ising model is discussed under local perturbations.

I. INTRODUCTION

The problem of stability of a dynamical system with infinitely many degrees of freedom encountered in quantum statistical mechanics has been recently paid much attention by several authors.[1-4] Verbeure and Weder[3] have studied this problem in the linear response theory. Ohya[4] has considered similar problems in the linear response and exact theories. In this paper, we apply some of the general results obtained in Refs. 3, 4 to the Weiss Ising model (WIM for brevity in the following) formulated in the operator algebraic framework.[5]

In Sec. II, we review the WIM briefly. In Sec. III, we show that the WIM provides us a nontrivial example of the work by Verbeure and Weder[3] concerning the stability of a dynamical system in the linear response sense (LRS for brevity). In Sec. IV, we study the dynamical behavior of locally perturbed states of the WIM by bounded self-adjoint operators belonging to the quasilocal C^*-algebra of this model. Namely, we look at the time development of the perturbed states after removing the perturbations dynamically. The discussion of that section is of the approach to equilibrium in both linear response and exact dynamics. We also discuss the dynamical change of metastable states in the WIM in Sec. IV. The change of a metastable state to a stable state under some local perturbation is extremely interesting to be investigated in quantum statistical mechanics and quantum measuring processes.[6-8] Although we first have to show the existence of metastable states in a dynamical system considered, we fortunately know[5] that there exist metastable states in the WIM, provided that the temperature is below a critical point. We are thus at the stage in the WIM to consider the dynamical change of the metastable states. We want to know whether there exists a local perturbation under which a metastable state goes to a stable state or not.

We take the inverse temperature of the WIM $\beta=1$ throughout this paper.

II. WEISS ISING MODEL

When we treat a physical system composed of many interacting particles, we have a powerful approximation, the so-called mean field method, which will be considered as follows: Observe a particle in the system, then we can regard the effect to this representative particle from other particles surrounding it as the field determined by an averaging procedure over the system of interest. Thus the representative particle can be statistically treated in the mean field. Conversely, we may determine the average field which the representative particle exerts on its neighbors. If all particles are identical, then every mean field should be coincident (requirement of self-consistency of mean field). This self-consistency requirement determines the mean field of the system precisely. We can study statistical properties of the system by this mean field. In this section, we briefly review one of the mean fields: the Weiss Ising model, in the algebraic framework.

Following Emch and Knops,[5] we consider a one-dimensional lattice Z (this restriction can be lifted, but we take it for simplicity). The quasilocal C^*-algebra of this lattice system Z is given by

$$\mathfrak{A}=\tilde{\mathfrak{A}}_0^n$$

$$\mathfrak{A}_0=\bigcup_{\Lambda\subset Z}\mathfrak{A}_\Lambda$$

$$\mathfrak{A}_\Lambda=\bigotimes_{k\in\Lambda}\mathfrak{A}_k \quad (\Lambda \text{ is a finite region of } Z)$$

$$\mathfrak{A}_k=\{\lambda_1 I_k+\lambda_2\sigma_k^x+\lambda_3\sigma_k^y+\lambda_4\sigma_k^z:\lambda_j\in C\ (j=1,2,3,4)\},$$

where $\sigma_k^{x,y,z}$ are the Pauli matrices and I_k is the identity matrix at the site k of the lattice Z. The local C^*-algebra $\mathfrak{A}_\Lambda$ enjoys two properties, named "isotony" and "locality."[9]

The local Hamiltonian of a ferromagnetic system considered here is given by

$$H(\Lambda)=-B\sum_{k\in\Lambda}\sigma_k^z-\sum_{k,i\in\Lambda}J_{k,i}(\Lambda)\sigma_k^z\sigma_i^z,$$

where B is an external magnetic field on the system along the z axis and $J_{k,i}(\Lambda)$ are real coupling constants on Λ with $J_{k,i}(\Lambda)=J_{i,k}(\Lambda)=J_{|k-i|}(\Lambda)$, $J_{k,k}(\Lambda)=0$. In order to study more about our system, let us introduce the following notations:

$$H(\Lambda)=\sum_{k\in\Lambda}H_k,$$

$$H_k=-[B+B_k(\Lambda)]\sigma_k^z,$$

$$B_k(\Lambda)=2\sum_{i\in\Lambda}J_{k,i}(\Lambda).$$

We can assume $B=0$ in the sequel discussions without loss of generality.

The idea of the mean field approximation of the ferromagnetic system comes from that we replace the above volume dependent $B_k(\Lambda)$ with some averaged volume in-

[a)]Mailing address.

 0022-2488/78/1905-0967$1.00

dependent one. Since the above local Hamiltonian $H(\Lambda)$ is bounded, it defines a time evolution automorphism of $\mathfrak{A}$: $\tau_t^\Lambda(A)=\exp[itH(\Lambda)]A\exp[-itH(\Lambda)]$ for any $A\in\mathfrak{A}$. We then ask whether and in which sense the time evolution τ_t after taking the limit $\Lambda\to\infty$ exists. Emch and Knops[5] have studied this question under the following conditions: (i) Stability condition: $\sum_{i\in\Lambda}|J_{k,i}(\Lambda)|\leq\frac{1}{2}c_k$ for any finite $\Lambda\subset Z$ containing k, where c_k is a Λ-independent constant. (ii) Van der Waals condition: $\lim_{\Lambda\to\infty}J_{k,i}(\Lambda)=0$ for all $k,i\in\Lambda$. (iii) Periodic condition: $J_{|k-i|}(\Lambda)=J_{|k-i|+p}(\Lambda)$, where p is a period of the lattice.

It is then shown that the volume infinite limit of $\tau_t^\Lambda(A)$ for any $A\in\mathfrak{A}$ exists for each representation Π such that its domain is restricted to the quasilocal C^*-algebra $\mathfrak{A}$, but its range cannot be in general the representation space $\Pi(\mathfrak{A})$ but the weak closure $\Pi(\mathfrak{A})''$. After taking the limit $\Lambda\to\infty$, the representation dependent time evolution τ_t can be extended to the automorphism $\tilde{\tau}_t$, the canonical extension of τ_t (see the remark at the end of this section), on the von Neumann algebra $\Pi(\mathfrak{A})''$. Here the time evolutions τ_t and $\tilde{\tau}_t$ should be written as τ_t^Π and $\tilde{\tau}_t^\Pi$ if we want to show their representation dependence explicitly, but we take the former expressions for simplicitly.

We now notice that if the von Neumann algebra $\Pi(\mathfrak{A})''$ is a factor, then the operator $\Pi(B_k)$, in the appropriate limit[5] of $\Pi(B_k(\Lambda))$ as $\Lambda\to\infty$, becomes a multiple of identity because of $\Pi(B_k)\in\Pi(\mathfrak{A})''\cap\Pi(\mathfrak{A})'$. We next have to construct an extremal KMS state φ in the following sense[5,10]:

(a) $\Pi_\varphi(\mathfrak{A})''\cap\Pi_\varphi(\mathfrak{A})'=CI$,

(b) $\int dt\langle\tilde{\varphi};\tilde{\tau}_t(\Pi_\varphi(A))\Pi_\varphi(B)\rangle f(t)$

$=\int dt\langle\tilde{\varphi};\Pi_\varphi(B)\tilde{\tau}_t(\Pi_\varphi(A))\rangle f(t-i)$

for any finite regions Λ, Λ' of Z, any $A\in\mathfrak{A}_\Lambda$, $B\in\mathfrak{A}_{\Lambda'}$, and any $f\in\hat{\mathcal{D}}$, the set of the Fourier-transforms of functions in the set $\mathcal{D}$ of infinitely differentiable functions with compact supports. The above $\tilde{\varphi}$ in (b) is the canonical extension of φ to $\Pi_\varphi(\mathfrak{A})''$ (see the remark below) and Π_φ is the GNS representation associated with φ. This construction can be carried out by the following facts:

(1) φ is a locally normal state; that is, for any finite region Λ of Z

$$\varphi\restriction\mathfrak{A}_\Lambda=\exp[-H(\Lambda)]/\mathrm{Tr}[-\exp H(\Lambda)], \tag{2.1}$$

where, from $\Pi_\varphi(B_k)=B_{k,\varphi}I_k(B_{k,\varphi}\in C)$, $H(\Lambda)$ is redefined as

$$H(\Lambda)=-\sum_{k\in\Lambda}B_{k,\varphi}\sigma_k^z.$$

(2) φ is a product state; that is,

$$\langle\varphi;AB\rangle=\langle\varphi;A\rangle\langle\varphi;B\rangle \tag{2.2}$$

for any A,B in $\mathfrak{A}_\Lambda$ and $\mathfrak{A}_{\Lambda'}$ respectively with $\Lambda\cap\Lambda'=\phi$.

(3) We have the self-consistency equations such that

$$\langle\varphi;\sigma_k^x\rangle=\langle\varphi;\sigma_k^y\rangle=0, \tag{2.3}$$

$$\langle\varphi;\sigma_k^z\rangle=\tanh B_{k,\varphi}, \tag{2.4}$$

$$B_{k,\varphi}=2\tilde{J}_k[\langle\varphi;\sigma_k^z\rangle], \tag{2.5}$$

where $\tilde{J}_k$ is the extension of a positive bounded linear functional J_k on a C^*-algebra $C(Z)$ of all bounded functions on Z. Here J_k is defined on a subspace $C_J(Z)$, for which $J_k[f]=\lim_{\substack{\Lambda\to\infty\\ k\in\Lambda}}\sum_{i\in\Lambda}J_{k,i}(\Lambda)f(i)$ exists. The extension from J_k to $\tilde{J}_k$ is always possible by the Hahn–Banach theorem. We can thus construct an extremal KMS locally normal state φ. For this state φ, the following important properties hold:

(4) $\tilde{\tau}_t$ is an automorphism of $\Pi_\varphi(\mathfrak{A})$. (2.6)

(5) $\tilde{\tau}_t$ is an automorphism of $\Pi_\varphi(\mathfrak{A}_\Lambda)$ for any finite $\Lambda\subset Z$. (2.7)

Let us finally notice that the canonical extension $\tilde{\varphi}$ of φ from $\mathfrak{A}$ to $\Pi_\varphi(\mathfrak{A})''$ is a faithful normal extremal KMS state with respect to $\tilde{\tau}_t$.[11]

Remark: (1) The canonical extension $\tilde{\tau}_t$ of τ_t to $\Pi_\varphi(\mathfrak{A})''$ is defined by $\tilde{\tau}_t(Q)=U_tQU_{-t}$ for any $Q\in\Pi_\varphi(\mathfrak{A})''$, where U_t is a strongly continuous one-parameter unitary group implementing $\tilde{\tau}_t$ on $\Pi_\varphi(\mathfrak{A})''$.

(2) The canonical extension $\tilde{\varphi}$ of φ to $\Pi_\varphi(\mathfrak{A})''$ is defined by $\langle\varphi;Q\rangle=(\Phi,Q\Phi)$ for any $Q\in\Pi_\varphi(\mathfrak{A})''$, where Φ is the GNS cyclic vector induced by φ.

III. STABILITY OF WEISS ISING MODEL

Verbeure and Weder[3] have studied the stability of dynamical systems in the linear response theory. Let a dynamical system be described by $(\mathcal{A},\phi,\alpha)$, where $\mathcal{A}$ is a C^*-algebra with an identity operator I, ϕ is a KMS state with respect to the time evolution α_t ($t\in R$), a strongly continuous one-parameter group of automorphisms of $\mathcal{A}$. Verbeure and Weder introduced the linear response time evolution as follows: For any $A\in\mathcal{A}$ and a perturbation $\lambda V=\lambda V^*\in\mathcal{A}$ with $\lambda\in[0,1]$,

$$\alpha_t^{V,\lambda}(A)=\alpha_t(A)+i\lambda\int_0^t ds[\alpha_s(V),\alpha_t(A)] \tag{3.1}$$

for $t\geq 0$, and

$$\alpha_t^{V,\lambda}(A)=\alpha_t(A)+i\lambda\int_t^0 ds[\alpha_s(V),\alpha_t(A)] \tag{3.2}$$

for $t<0$. These are obtained by linearizations w.r.t. λ of the exact time evolution defined by

$$\alpha_t^V(A)=\sum_{n\geq 0}(i\lambda)^n\int dt_1\cdots\int_{0\leq t_1\leq\cdots\leq t_n\leq t}dt_n[\alpha_{t_1}(V),[\cdots[\alpha_{t_n}(V),\alpha_t(A)]\cdots]] \tag{3.3}$$

for $t\geq 0$, and

$$\alpha_t^V(A)=\sum_{n\geq 0}(i\lambda)^n\int dt_1\cdots\int_{t\leq t_1\leq\cdots\leq t_n\leq 0}dt_n\times[\alpha_{t_1}(V),[\cdots[\alpha_{t_n}(V),\alpha_t(A)]\cdots]] \tag{3.4}$$

for $t<0$.

The C^*-algebraic versions of the definitions of stability in LRS of the dynamical system $(\mathcal{A},\phi,\alpha)$ due to Verbeure and Weder are as followings.

Definition III.1: A dynamical system $(\mathcal{A},\phi,\alpha)$ is said to be stable in LRS under a perturbation $V=V^*\in\mathcal{A}$ if the weak* limits $\phi_{\pm,\lambda}=\mathrm{w}^*\text{-}\lim_{t\to\pm\infty}\phi\circ\alpha_t^{V,\lambda}$ exist under V.

Definition III.2: A dynamical system $(\mathcal{A}, \phi, \alpha)$ is said to be stable in LRS if it is stable in LRS under all perturbations $V = V^* \in \mathcal{A}$.

As we have seen in the previous section, the WIM should be originally treated by the triple $(\mathfrak{A}, \varphi, \tau)$, but since τ_t does not exist as an automorphism of $\mathfrak{A}$ for a general state φ on $\mathfrak{A}$, the WIM has to be considered through the von Neumann algebraic triple $(\Pi_\varphi(\mathfrak{A})'', \tilde{\varphi}, \tilde{\tau})$. However, for an extremal KMS state φ, the time evolution $\tilde{\tau}_t$ can be an automorphism not only of $\Pi_\varphi(\mathfrak{A})''$ but also of $\Pi_\varphi(\mathfrak{A})$, and hence of $\mathfrak{A}$ (since all representations are faithful). We thus define the linear response time evolution $\tilde{\tau}_t^{\Pi_\varphi(V),1}$ by replacing $\alpha_t, A, \lambda V$ with $\tilde{\tau}_t$, $\Pi_\varphi(A)$, $\lambda\Pi_\varphi(V)$ in (3.1) and (3.2). It is readily shown by merely extending the von Neumann algebraic results obtained by Verbeure and Weder[3] to the C^*-algebraic dynamical system $(\mathcal{A}, \phi, \alpha)$ that the dynamical system $(\mathcal{A}, \phi, \alpha)$ is clustering [i.e., $\lim_{t\to\pm\infty}\langle\phi; \alpha_t(A)B\rangle = \langle\phi; A\rangle\langle\phi; A\rangle$ for any $A, B \in \mathcal{A}$] if and only if it is η_t clustering [i.e., $\eta_t\langle\phi; \alpha_t(A)B\rangle = \langle\phi; A\rangle\langle\phi; B\rangle$ for any $A, B \in \mathcal{A}$, where η_t is a mean over $t \in R$] and stable in LRS. This fact implies that if a dynamical system is clustering for its KMS state (e.g., X–Y model, free Fermi gas), Definitions III.1 and III.2 become equivalent. Therefore, for such systems, the problem of stability in LRS is not so attractive. However, the WIM is even not η_t clustering, so that it may be interesting to study the problem of stability for this model. We ask a following question: Under which conditions on $V = V^* \in \mathcal{A}$ is the WIM stable in LRS? In other words, can we determine the set of perturbations under which the WIM is stable in LRS?

Before answering this question, let us introduce a C^*-subalgebra $\mathfrak{A}^z$ of $\mathfrak{A}$ such that

$$\mathfrak{A}^z = \overline{\mathfrak{A}_0^z}^n$$

$$\mathfrak{A}_0^z = \bigcup_{\Lambda\subset Z} \mathfrak{A}_\Lambda^z ,$$

$$\mathfrak{A}_\Lambda^z = \bigotimes_{k\in\Lambda} \mathfrak{A}_k^z,$$

$$\mathfrak{A}_k^z = \{\mu_1 I_k + \mu_2 \sigma_k^z : \mu_1, \mu_2 \in C\}.$$

We then have

Theorem III.1: The WIM is stable in LRS under a perturbation $V = V^* \in \mathfrak{A}$ if and only if $V \in \mathfrak{A}^z$.

Proof: According to Verbeure and Weder,[3] the WIM is stable in LRS under a perturbation $V = V^* \in \mathfrak{A}$ if and only if $U_t\Pi_\varphi(V)\Phi = \exp(itH)\Pi_\varphi(V)\Phi$ weakly converges in the GNS Hilbert space $\mathcal{H}_\varphi$ as time tends to infinite, where H is the infinitesimal generator of the unitary one-parameter group U_t implementing $\tilde{\tau}_t$. Let us first consider the case when the perturbation $V = V^* \in \mathfrak{A}$ is local; that is, there exists a finite region Λ of Z such that $V = V^* \in \mathfrak{A}_\Lambda$. As we have discussed, the time evolution $\tilde{\tau}_t$ is an automorphism of $\Pi_\varphi(\mathfrak{A}_\Lambda)$, hence of $\mathfrak{A}_\Lambda$. The restriction $\tilde{\tau}_t^\Lambda$ of $\tilde{\tau}_t$ to $\mathfrak{A}_\Lambda$ is generated by the local Hamiltonian $H(\Lambda) = -\sum_{k\in\Lambda} B_{k,\varphi}\sigma_k^z$. The extremal KMS product state $\tilde{\varphi}$ is identical to $\exp[-H(\Lambda)]/\mathrm{Tr}\exp[-H(\Lambda)]$ on the local algebra $\mathfrak{A}_\Lambda$. Therefore, the restriction $\tilde{\varphi}^\Lambda$ of $\tilde{\varphi}$ to $\mathfrak{A}_\Lambda$ is a KMS state with respect to the time evolution $\tilde{\tau}_t^\Lambda$. If the WIM is stable in LRS under the local perturbation V, $(\chi, U_t\Pi_\varphi(V)\Phi)$ converges to a definite value for any $\chi \in \mathcal{H}_\varphi$ as time tends to infinite. Moreover,

$$U_t\Pi_\varphi(V)\Phi = \exp(itH)\Pi_\varphi(V)\Phi$$
$$= \exp[itH(\Lambda)]\Pi_\varphi(V)\Phi = \tilde{\tau}_t(\Pi_\varphi(V))\Phi.$$

As φ is a product state, the Hilbert space $\mathcal{H}_\varphi$ can be an incomplete direct produce space denoted by $\otimes_{k\in Z}\mathcal{H}_k$,[9] where $\mathcal{H}_k$ is isomorphic to C^2 generated by four-dimensional vectors at the site k. Hence Φ can be written as $\otimes_{k\in Z}\Phi_k$ by cyclic vectors Φ_k for $B(\mathcal{H}_k)$. Choosing $\chi = \chi_\Lambda \otimes_{k\in Z\setminus\Lambda}\Phi_k (Z\setminus\Lambda \equiv \{k \in Z : k \notin \Lambda\})$ for any χ_Λ in $\mathcal{H}_\Lambda = \otimes_{k\in\Lambda}\mathcal{H}_k$, $(\chi, U_t\pi_\varphi(V)\Phi)$ is equal to $(\chi_\Lambda, \exp[itH(\Lambda)] \times \Pi_\varphi(V)\Phi_\Lambda)$, where $\Phi_\Lambda = \otimes_{k\in\Lambda}\Phi_k$. Since Λ is a finite region of Z and χ_Λ is arbitrary element in $\mathcal{H}_\Lambda$, the function $t \in R \to (\chi_\Lambda, \exp[itH(\Lambda)]\Pi_\varphi(V)\Phi_\Lambda)$ is almost periodic in t. This fact implies[12] that if the weak limit $\text{w-}\lim_{t\to\pm\infty}U_t \times \Pi_\varphi(V)\Phi$ exists, then $U_t\Pi_\varphi(V)\Phi$ is identical to $\Pi_\varphi(V)\Phi$. Since $U_t\Pi_\varphi(V)\Phi = \tilde{\tau}_t^\Lambda(\Pi_\varphi(V))\Phi$ and Φ is separating for the von Neumann algebra $\Pi_\varphi(\mathfrak{A})''$, we have $\tilde{\tau}_t^\Lambda(\Pi_\varphi(V)) = \Pi_\varphi(V)$. As the quasilocal C^*-algebra $\mathfrak{A}$ is simple, the representation Π_φ is faithful; hence $[\exp[itH(\Lambda)], V] = 0$. Thus we immediately conclude that V is in $\mathfrak{A}_\Lambda^z$ according to the forms of $H(\Lambda)$ and our local C^*-subalgebra $\mathfrak{A}_\Lambda^z$. Conversely, if the local perturbation $V = V^* \in \mathfrak{A}_\Lambda$ is in $\mathfrak{A}_\Lambda^z$, then it is obvious that $U_t\Pi_\varphi(V)\Phi$ weakly converges in $\mathcal{H}_\varphi$ as time tends to infinite.

Next we have to consider for a general perturbation $V = V^* \in \mathfrak{A}$. Namely, we have to show that the weak limit $\text{w-}\lim_{t\to\pm\infty}U_t\Pi_\varphi(V)\Phi$ exists in $\mathcal{H}_\varphi$ if and only if $V = V^* \in \mathfrak{A}^z$. The "if" part is shown as follows: For any $V \in \mathfrak{A}^z$ and any $\epsilon > 0$, there exists $V_0 = V_0^* \in \mathfrak{A}_0^z$ such that $\|V - V_0\| < \epsilon$. Hence we obtain

$$\|U_t\Pi_\varphi(V)\Phi - \Pi_\varphi(V)\Phi\| \le \|U_t\Pi_\varphi(V)\Phi - U_t\Pi_\varphi(V_0)\Phi\|$$
$$+ \|U_t\Pi_\varphi(V_0)\Phi - \Pi_\varphi(V_0)\Phi\| + \|\Pi_\varphi(V_0)\Phi - \Pi_\varphi(V)\Phi\|$$
$$= 2\|\Pi_\varphi(V) - \Pi_\varphi(V_0)\| = 2\|V - V_0\| < 2\epsilon.$$

Conversely, let us prove the "only if" part. If there exists $V = V^* \in \mathfrak{A}$ such that $\text{w-}\lim_{t\to\pm\infty}U_t\Pi_\varphi(V)\Phi$ exists in $\mathcal{H}_\varphi$, then as the local algebra $\mathfrak{A}_0 = \cup_{\Lambda\subset Z}\mathfrak{A}_\Lambda$ is norm dense in $\mathfrak{A}$, for any $V = V^* \in \mathfrak{A}$ and any $\epsilon > 0$, there exists $V_0 = V_0^* \in \mathfrak{A}_0$ such that $\|V - V_0\| < \epsilon$; hence it is easily seen that

$$\|U_t\Pi_\varphi(V)\Phi - U_t\Pi_\varphi(V_0)\Phi\| < \epsilon.$$

Therefore, if $\text{w-}\lim_{t\to\pm\infty}U_t\Pi_\varphi(V)\Phi$ exists in $\mathcal{H}_\varphi$, then $\text{w-}\lim_{t\to\pm\infty}U_t\Pi_\varphi(V_0)\Phi$ exists in $\mathcal{H}_\varphi$ too. Since V_0 is a local element of $\mathfrak{A}$, this V_0 should be in the algebra $\mathfrak{A}_0^z$ as seen before. This fact implies that the perturbation $V = V^* \in \mathfrak{A}$ is in the algebra $\mathfrak{A}^z$ because $\mathfrak{A}_0^z$ is norm dense in $\mathfrak{A}$[3]. Q.E.D.

From this theorem, we can divide the perturbations into two classes. The first class is the set of perturbations under which the WIM is stable in LRS. Under the perturbations of the second class, the WIM is not stable in LRS.

The following theorem tells us that when $V = V^* \in \mathfrak{A}^z$, the linear response approximation is exact as far as the stability of the WIM is concerned.

Theorem III.2: For any perturbation $V = V^* \in \mathfrak{A}^z$, we have

(1) $\mathrm{w}^*\text{-}\lim_{t\to\pm\infty} \tilde{\varphi}\circ\tilde{\tau}_t^{\Pi_\varphi(V),1} = \varphi$ on ,

(2) $\mathrm{w}^*\text{-}\lim_{t\to\pm\infty} \tilde{\varphi}\circ\tilde{\tau}_t^{\Pi_\varphi(V)} = \varphi$ on $\mathfrak{A}$.

Remark: (a) The above exact perturbed time evolution $\tilde{\tau}_t^{\Pi_\varphi(V)}$ is given by (3.3) and (3.4) under the replacements of α_t and λV with $\tilde{\tau}_t$ and $\lambda\Pi_\varphi(V)$. (b) This theorem can be read that if the WIM is stable in LRS under a perturbation $V = V^* \in \mathfrak{A}$, then it is stable in exact sense under the same V.

Proof: According to Verbeure and Weder,[3] we have

$$\lim_{t\to\pm\infty}\langle\tilde{\varphi}; \tilde{\tau}_t^{\Pi_\varphi(V),1}(\Pi_\varphi(V))\rangle$$
$$= \langle\tilde{\varphi}; \Pi_\varphi(A)\rangle - \lambda \lim_{t\to\pm\infty} (T\Pi_\varphi(A^*)\Phi, T[1-\exp(itH)]\Pi_\varphi(V)\Phi),$$

where $T = [(\Delta-1)/\ln\Delta]^{1/2}$ and $\Delta = \exp(-H)$. We have seen in the proof of the theorem III.1

$$\exp(itH)\Pi_\varphi(V)\Phi = \Pi_\varphi(V)\Phi,$$

for any $V = V^* \in \mathfrak{A}^z$. We hence have

$$(T\Pi_\varphi(A^*)\Phi,\ T[1-\exp(itH)]\Pi_\varphi(V)\Phi) = 0,$$

which implies $\lim_{t\to\pm\infty}\langle\tilde{\varphi}; \tilde{\tau}_t^{\Pi_\varphi(V),1}(\Pi_\varphi(A))\rangle = \langle\tilde{\varphi}; \pi_\varphi(A)\rangle = \langle\varphi; A\rangle$ for any $A \in \mathfrak{A}$. Let us prove (2). Simple computation tells us

$$\frac{d}{dl}\tilde{\tau}_{-t}\tilde{\tau}_t^{\Pi_\varphi(V)}(\Pi_\varphi(A)) = i\lambda\tilde{\tau}_{-t}([\Pi_\varphi(V), \tilde{\tau}_t^{\Pi_\varphi(V)}(\Pi_\varphi(A))]).$$

Hence

$$\langle\tilde{\varphi}; \tilde{\tau}_t^{\Pi_\varphi(V)}(\Pi_\varphi(A))\rangle = \langle\tilde{\varphi}; \tilde{\tau}_{-t}\tilde{\tau}_t^{\Pi_\varphi(V)}(\Pi_\varphi(A))\rangle$$
$$= \langle\varphi; A\rangle + i\lambda\int_0^t ds\,\langle\tilde{\varphi}; [\Pi_\varphi(V), \tilde{\tau}_t^{\Pi_\varphi(V)}(\Pi_\varphi(A))]\rangle.$$

Therefore, we have only to show that the limits $\lim_{t\to\pm\infty}\int_0^t ds\langle\tilde{\varphi}; [\Pi_\varphi(V), \tilde{\tau}_t^{\Pi_\varphi(V)}(\Pi_\varphi(A))]\rangle$ exist for any A in $\mathfrak{A}_0 = \cup_{\Lambda\subset Z}\mathfrak{A}_\Lambda$. Since $\tilde{\tau}_t(\Pi_\varphi(V)) = \Pi_\varphi(V)$ for any $V = V^* \in \mathfrak{A}^z$, the perturbed time evolution $\tilde{\tau}_t^{\Pi_\varphi(V)}$ is given by

$$\tilde{\tau}_t^{\Pi_\varphi(V)}(\Pi_\varphi(A)) = \exp[it\Pi_\varphi(V)]\,\tilde{\tau}_t(\Pi_\varphi(A))\exp[-it\Pi_\varphi(V)].$$

Let us take any $V_0 = V_0^*$ from $\mathfrak{A}_{\Lambda_0}^z$, where Λ_0 is a finite region of Z. For a local element A of $\mathfrak{A}$, there exists a finite region Λ of Z such that $A \in \mathfrak{A}_\Lambda$. According to the locally normality of the state φ, we obtain

$$\langle\tilde{\varphi}; [\Pi_\varphi(V_0), \tilde{\tau}_t^{\Pi_\varphi(V)}(\Pi_\varphi(A))]\rangle$$
$$= \mathrm{Tr}\exp[-H(\Lambda\cup\Lambda_0)][V_0, \exp(itV_0)\exp[itH(\Lambda)]A$$
$$\times\exp[-itH(\Lambda)]\exp(-itV_0)]/\mathrm{Tr}\exp[-H(\Lambda\cup\Lambda_0)], \tag{3.5}$$

where $H(\Omega) = -\sum_{k\in\Omega} B_{k,\varphi}\sigma_k^z$ with $\Omega = \Lambda$ or $\Lambda\cup\Lambda_0$. The above (3.5) is equal to zero due to the forms of $H(\Omega)$ and V_0. As $\mathfrak{A}_0^z$ is norm dense in $\mathfrak{A}^z$, we conclude (2) by continuity. Q.E.D.

IV. APPROACH TO EQUILIBRIUM

One way of analyzing the structure of statistical mechanics is to study the time development of states which have been perturbed from equilibrium. If a perturbed state relaxes to another state as time tends to infinite, we may say that the approach to equilibrium occurs under that perturbation. In this section, we will study the time development of such perturbed states in the WIM. As discussed in the previous section, the WIM is described by a triple $(\Pi_\varphi(\mathfrak{A}), \tilde{\varphi}, \tilde{\tau})$ for an extremal KMS state φ. For a perturbation $\lambda V = \lambda V^* \in \mathfrak{A}$ with $\lambda \in [0,1]$, we denote the exact perturbed WIM by a triple $(\Pi_\varphi(\mathfrak{A}), \tilde{\varphi}^{\Pi_\varphi(V)}, \tilde{\tau}^{\Pi_\varphi(V)})$ and the linear response WIM by $(\Pi_\varphi(\mathfrak{A}), \tilde{\varphi}^{\Pi_\varphi(V),1}, \tilde{\tau}^{\Pi_\varphi(V),1})$. They are given as follows[4,13]: The exact perturbed and linear response time evolutions $\tilde{\tau}_t^{\Pi_\varphi(V)}, \tilde{\tau}_t^{\Pi_\varphi(V),1}$ respectively are given in Sec. III. Here, let us introduce the exact perturbed state $\tilde{\varphi}^{\Pi_\varphi(V)}$ and the linear response functional $\tilde{\varphi}^{\Pi_\varphi(V),1}$. For any $A \in \mathfrak{A}$,

$$\langle\tilde{\varphi}^{\Pi_\varphi(V)}; \Pi_\varphi(A)\rangle = (\Phi^V, \Pi_\varphi(A)\Phi^V), \tag{4.1}$$

$$\Phi^V = \Pi_\varphi(W_{i/2}^V)\Phi/\|\Pi_\varphi(W_{i/2}^V)\Phi\|, \tag{4.2}$$

$$W_{i/2}^V = \sum_{n\geq 0}(-\lambda)^n\int_{0\leq t_1}dt_1\cdots\int_{\cdots\leq t_n\leq 1/2}dt_n\,\tilde{\tau}_{it_1}(\Pi_\varphi(V))\cdots \tilde{\tau}_{it_n}(\Pi_\varphi(V)), \tag{4.3}$$

$$\langle\tilde{\varphi}^{\Pi_\varphi(V),1}; \Pi_\varphi(A)\rangle = \langle\tilde{\varphi}; \Pi_\varphi(A)\rangle - \lambda\langle\tilde{\varphi}; \Pi_\varphi(A)\int_0^1 ds\,\tilde{\tau}_{is}(\Pi_\varphi(V))\rangle + \lambda\langle\tilde{\varphi}; \Pi_\varphi(A)\rangle\langle\tilde{\varphi}; \Pi_\varphi(V)\rangle, \tag{4.4}$$

where $\tilde{\varphi}^{\Pi_\varphi(V),1}$ is obtained by linearizing $\tilde{\varphi}^{\Pi_\varphi(V)}$ w.r.t. λ and some computations.[4]

The motivation of this section is the following:

(i) Under which conditions on $V = V^* \in \mathfrak{A}$, do the limits $\tilde{\varphi}_\pm^{\Pi_\varphi(V),1} = \mathrm{w}^*\text{-}\lim_{t\to\pm\infty}\tilde{\varphi}^{\Pi_\varphi(V),1}\circ\tilde{\tau}_t$ exist on $\Pi_\varphi(\mathfrak{A})$? (Approach to equilibrium in LRS under V.)

(ii) Under which conditions on $V = V^* \in \mathfrak{A}$, do the limits $\tilde{\varphi}_\pm^{\Pi_\varphi(V)} = \mathrm{w}^*\text{-}\lim_{t\to\pm\infty}\tilde{\varphi}^{\Pi_\varphi(V)}\circ\tilde{\tau}_t$ exist on $\Pi_\varphi(\mathfrak{A})$? [Approach to equilibrium in ES (exact sense) under V.]

(iii) Is the linear response approximation useful to study some dynamical processes?

We study these questions through the WIM. The answer concerning the first two question is

Theorem IV.1: For the WIM initially in equilibrium described by an extremal KMS state φ, the approach to equilibrium in LRS occurs under a perturbation $V = V^* \in \mathfrak{A}$ if and only if the approach to equilibrium in ES occurs under the same V.

Proof: As we have shown[4] that for a dynamical system satisfying the KMS condition, the approach to equilibrium in LRS occurs under a perturbation $V = V^*$ belonging to the C^*-algebra of the dynamical system if and only if the dynamical system is stable in LRS under the same V. Therefore, for the WIM, when the approach to equilibrium in LRS occurs under a perturbation $V = V^* \in \mathfrak{A}$, the WIM is stable in LRS under this perturbation V. According to Theorem III.1, this perturbation V is in the C^*-algebra $\mathfrak{A}^z$. As seen in the proof of Theorem III.1, the equality $\tilde{\tau}_t(\Pi_\varphi(V)) = \Pi_\varphi(V)$ holds for all $t \in R$. It is easily checked that $\Pi_\varphi(W_{i/2}^V)\Phi = \exp[-\frac{1}{2}\lambda\Pi_\varphi(V)]\Phi$ is satisfied for $V = V^* \in \mathfrak{A}^z$. Thus we obtain

$$\langle\tilde{\varphi}^{\Pi_\varphi(V)}; \tilde{\tau}_t(\Pi_\varphi(A))\rangle = (\exp[-\tfrac{1}{2}\lambda\Pi_\varphi(V)]\Phi,\ \tilde{\tau}_t(\Pi_\varphi(A))\exp[-\tfrac{1}{2}\lambda\Pi_\varphi(V)]\Phi)/\|\exp[-\tfrac{1}{2}\lambda\Pi_\varphi(V)]\Phi\|^2,$$

which is equal to $\langle\tilde{\varphi}^{\Pi_\varphi(V)};\Pi_\varphi(A)\rangle$ because of $[\Pi_\varphi(V)]^n \in \Pi_\varphi(\mathfrak{A}^z)$ for any positive integer n and $[U_t,R]=0$ for any $R\in\Pi_\varphi(\mathfrak{A}^z)$. Hence the approach to equilibrium in ES occurs under $V=V^*\in\mathfrak{A}^z$. Conversely, if the approach to equilibrium in ES occurs under a perturbation $V=V^*\in\mathfrak{A}$, then the limits $\lim_{t\to\pm\infty}\langle\tilde{\varphi}^{\Pi_\varphi(V)};\tilde{\tau}_t(\Pi_\varphi(A))\rangle$ exist uniformly in $\lambda\in[0,1]$ for any $A\in\mathfrak{A}$. According to the definition of $\tilde{\varphi}^{\Pi_\varphi(V)}$, $\langle\tilde{\varphi}^{\Pi_\varphi(V)};\tilde{\tau}_t(\Pi_\varphi(A))\rangle$ is expressed by a power series of λ. Moreover, it is readily seen that the function $\lambda\in[0,1]\to\langle\tilde{\varphi}^{\Pi_\varphi(V)};\tilde{\tau}_t(\Pi_\varphi(A))\rangle$ is differentiable at all λ for any $A\in\mathfrak{A}$ and each term of the power series of λ is uniformly bounded in t. Hence

$$\lim_{t\to\pm\infty}\langle\tilde{\varphi}^{\Pi_\varphi(V),1};\tilde{\tau}_t(\Pi_\varphi(A))\rangle-\langle\tilde{\varphi};\Pi_\varphi(A)\rangle$$

$$=\lambda\lim_{t\to\pm\infty}\left[\frac{d}{d\lambda}\langle\tilde{\varphi}^{\Pi_\varphi(V)};\tilde{\tau}_t(\Pi_\varphi(A))\rangle\right]_{\lambda=0}$$

$$=\lambda\left[\frac{d}{d\lambda}\lim_{t\to\pm\infty}\langle\tilde{\varphi}^{\Pi_\varphi(V)};\tilde{\tau}_t(\Pi_\varphi(A))\rangle\right]_{\lambda=0}$$

exists for any $A\in\mathfrak{A}$. This concludes that the approach to equilibrium in LRS occurs under V. Q. E. D.

The above theorem tells us that the linear response method is a good approximation of the exact method as far as the approach to equilibrium is concerned for the WIM.

Let us consider the dynamical change of metastable states in the WIM. In physical systems, we often encounter metastable states which satisfy the same equilibrium condition of stable states. A stable state gives an absolute minimum to the free energy density of the system. On the contrary, a metastable state gives a relative minimum to the free energy density. A metastable state however easily changes to a stable state (or mixture of stable states) by a local external disturbance. The rigorous interpretation of the existence of metastable states and the transition from a metastable state to a stable state has not much done yet. In the WIM, we, however, know[5] that the metastable states exist below some critical temperature. These metastable states are extremal KMS states as the stable state. Let us discuss the above transition of a metastable state appeared in the WIM under the effect of local perturbations. We ask whether there exists a perturbation which causes such transition. Let $\tilde{\varphi}$ be a metastable state satisfying the KMS condition with respect to $\tilde{\tau}_t$. The locally perturbed state $\tilde{\varphi}^{\Pi_\varphi(V)}$ of $\tilde{\varphi}$ under a perturbation $\lambda V=\lambda V^*\in\mathfrak{A}$ is given by (4.1). We then switch off the perturbation at $t=0$, that is, we look at the time development of the perturbed state $\tilde{\varphi}^{\Pi_\varphi(V)}$ under $\tilde{\tau}_t$. It is interesting to know when the limit $\tilde{\varphi}_+^{\Pi_\varphi(V)}=\mathrm{w}^*\text{-}\lim_{t\to\infty}\tilde{\varphi}^{\Pi_\varphi(V)}\circ\tilde{\tau}_t$ exists and whether $\tilde{\varphi}$ and $\tilde{\varphi}_+^{\Pi_\varphi(V)}$ are disjoint. The first question has been answered by Theorem IV.1. We are interested in the second question here. If these states $\tilde{\varphi}$ and $\tilde{\varphi}_+^{\Pi_\varphi(V)}$ are disjoint, then there is possibility of having the desired transition from the metastable state to a stable state. Due to Theorem IV.1, the limit $\tilde{\varphi}_+^{\Pi_\varphi(V)}=\mathrm{w}^*\text{-}\lim_{t\to\infty}\tilde{\varphi}^{\Pi_\varphi(V)}\circ\tilde{\tau}_t$ exists if and only if the perturbation V is in the C^*-algebra $\mathfrak{A}^z$. As shown in the proof of that theorem, $\tilde{\varphi}_+^{\Pi_\varphi(V)}$ is identical to $\tilde{\varphi}^{\Pi_\varphi(V)}$. It is easily seen that the state $\tilde{\varphi}^{\Pi_\varphi(V)}$ is faithful normal on the von Neumann algebra $\Pi_\varphi(\mathfrak{A})''$. Thus the states $\tilde{\varphi}$ and $\tilde{\varphi}^{\Pi_\varphi(V)}$ cannot be disjoint. This fact implies that there does not exist a perturbation $V=V^*\in\mathfrak{A}$ under which a metastable state of the WIM goes to a stable state.

ACKNOWLEDGMENTS

The author thanks Professor G. G. Emch for valuable discussions and suggestions of this work. He also thanks Professor H. Umegaki for his interest to the author's work and his kindness at Tokyo Institute of Technology.

[1]R. Haag, D. Kastler, and E. Trych-Pohlymeyer, Commun. Math. Phys. 38, 173 (1974).
[2]J. L. Lebowitz, M. Aizemann, and S. Goldstein, J. Math. Phys. 6, 1284 (1975).
[3]A. Verbeure and R. Weder, Commun. Math. Phys. 44, 101 (1975).
[4]M. Ohya, "Dynamical Process in Linear Response Theory," preprint.
[5]G. G. Emch and J. F. Knops, J. Math. Phys. 12, 2043 (1971).
[6]G. G. Emch, J. Math. Phys. 7, 1198 (1966).
[7]G. L. Sewell, Lett. Nuovo Cimento 10, 430 (1974).
[8]A. Daneri, A. Loinger, and G. M. Prosperi, Nucl. Phys. 33, 297 (1962).
[9]G. G. Emch, *Algebraic Methods in Statistical Mechanics and Quantum Field Theory* (Wiley, New York, 1972).
[10]R. Haag, N. Hugenholtz, and M. Winnik, Commun. Math. Phys. 5, 215 (1967).
[11]M. Takesaki, *Tomita's Theory of Modular Hilbert Algebra and Its Applications* (Springer-Verlag, Berlin, 1970).
[12]M. Winnik, in *Fundamental Poblems in Statistical Mechanics*, edited by E. D. G. Cohen (North-Holland, Amsterdam, 1975).
[13]D. W. Robinson, lecture at Varenna School, 1973.

The asymptotic reduction of products of representations of the universal covering group of SU(1,1)

W. Rühl and B. C. Yunn

Fachbereich Physik, Universität Kaiserslautern, 6750 Kaiserlauten, F. R. Germany
(Received 21 March 1977)

Irreducible representations are realized on nuclear spaces. Products of certain elements of two such spaces can be expanded into an almost everywhere convergent series that possesses rudimentary covariance properties. This unique expansion is therefore called "asymptotic reduction." Products of elements of three such spaces possess a simultaneous expansion only in an asymptotic sense. We define recoupling coefficients for this reduction and give them explicitly.

1. INTRODUCTION

Products of unitary representations of a noncompact Lie group can be decomposed into a direct integral of irreducible unitary representations following a general scheme based on the concept of Hilbert spaces. Using realizations on L^2-spaces of functions, this problem is a special case of harmonic analysis for the group under consideration. Its solution can be formulated in terms of a Plancherel theorem that expresses the one-to-one nature of the mapping of the product Hilbert space onto the direct integral of Hilbert spaces each of which carries an irreducible unitary representation. Technically this mapping is achieved by integral kernels ("covariant kernels") and their inverses ("dual covariant kernels") that are applied to nuclear subspaces first and then extended to the whole Hilbert spaces. In the case of the universal covering group of SU(1,1), $SU(1,1)^{uc}$, a detailed description of this problem including a study of the covariant and dual covariant kernels has been given by the authors in Ref. 1. The method used in that article goes back as far as Naimark's investigation of the same problem for the group $SL(2,C)$.[2]

In this article we apply the method of "asymptotic reduction" to the problem of product representations. The general scheme into which this method falls is still rather obscure. Let us describe it therefore in technical terms and in the way it has been developed historically. The first systematic use of it was made by Toller[3] in the analysis of two-particle relativistic scattering amplitudes. This scattering amplitude was considered as a function on a homogeneous space of SU(1,1). Because of Lorentz covariance it transforms as a representation of SU(1,1) carried by this space. However, Toller did not submit it to an L^2 harmonic analysis (in fact it is not L^2 but rather a distribution) but tried instead to derive an asymptotic expansion.

This idea can best be illustrated in terms of standard Fourier transforms, though crucial modifications are necessary when dealing with harmonic analysis on non-Abelian groups. Let

$$f(x)\in L^2(\mathbb{R}),\quad g(t)=\int_{-\infty}^{+\infty} f(x)\exp(-ixt)\,dx,$$

$$f(x)=\frac{1}{2\pi}\int_{-\infty}^{+\infty} g(t)\exp(ixt)\,dt.$$

If we consider $\mathbb{R}$ as a homogeneous space for the translation group

$$T_h f(x)=f(x+h),$$

then t describes a unitary irreducible representation of it, and $\exp(iht)$ is the coordinate function corresponding to the group element h and the representation t

$$T_h^t g(t)=\exp(iht)g(t),$$

(for almost all t). The Fourier transform $g(t)$ in this sense is the L^2 reduction of the representation on $\mathbb{R}$.

If $g(t)$ is the boundary value of a meromorphic function in the upper half t plane and tends to zero sufficiently fast for $\mathrm{Re}\,t\to\pm\infty$, then $f(x)$ can be expanded asymptotically for $x\to+\infty$,

$$f(x)=i\sum_{n=1}^{N}\gamma_n\exp(ixt_n)+R_N(x).$$

Here γ_n is the residue of $g(t)$ at $t=t_n$, $\mathrm{Im}\,t_n$ is assumed to increase with n and $R_N(x)$ is $O(\exp(-x\cdot\mathrm{Im}\,t_N))$. We may define nonunitary representations

$$T_h^{(n)}\gamma_n=\exp(iht_n)\gamma_n,$$

and then regard the asymptotic expansion of $f(x)$ as an "asymptotic reduction" into nonunitary representations. It is obvious that this procedure can be extended to distributions f belonging to test functions with faster than exponential decrease on a half axis. The premises of meromorphy and sufficient decrease for the Fourier transform g remain essential also in this case.

In carrying over this idea to noncompact nonabelian Lie groups G [such as SU(1,1)] typical new features arise. Consider some homogeneous space $\mathbb{H}$ of G with measure $d\mu(x)$ and a function $f(x)\in L^2_\mu(\mathbb{H})$. Define the Fourier transform

$$g_\lambda(\chi)=\int f(x)\,P^\chi_\lambda(x)\,d\mu(x)$$

where χ denotes a principal or discrete series representation of G and λ is some discrete label. Among the functions $P^\chi_\lambda(x)$ those with a fixed χ belong to a representation χ of G, namely these functions satisfy the covariance constraint

$$T_h f(x)=\alpha(x,h)f(x_h),$$

$$\alpha(x,h)\,P^\chi_\lambda(x_h)=\sum_{\lambda'}D^\chi_{\lambda',\lambda}(h)\,P^\chi_{\lambda'}(x).$$

Here $\alpha(x,h)$ is a multiplier and $D^\chi_{\lambda',\lambda}(h)$ are the coordinate functions. $P^\chi_\lambda(x)$ is fixed by this constraint up to a normalization.

The inverse Fourier transform is

 0022-2488/78/1905-0972$1.00

$$f(x)=\sum_{\lambda}\int g_{\lambda}(\chi)\,P_{\lambda}^{\chi}(x)\,d\mu(\chi),$$

where the measure $d\mu(\chi)$ is generally related with the Plancherel measure of the group. The Fourier transform $g_\lambda(\chi)$ is L^2 with respect to this measure and summation over λ. Again this L^2 harmonic analysis can easily be extended to distributions f so that $g_\lambda(\chi)$ appears as a distribution over a manifold

$$X=\{\lambda,\chi\}.$$

Though it is easy to define a subspace X_+ of X on which $g_\lambda(\chi)$ is assumed to be meromorphic and to decrease, this is not sufficient yet to derive an asymptotic expansion of $f(x)$ from the inverse Fourier transformation for x tending to infinity.

The reason for this lies in the geometry of $\mathbb{H}$. Typically it has the structure of one shell of a two-shell hyperboloid. $x\to\infty$ means any direction on this shell, contrary to the two directions $x\to\pm\infty$ on $\mathbb{R}$ that were treated independently in the Abelian case. There exist discrete reflections c in $X:\chi_c\to\chi^c$, which if applied to X_+ cover all X and are such that

$$P_{\lambda}^{\chi}(x)=P_{\lambda}^{\chi^c}(x),\quad D_{\lambda\lambda'}^{\chi}(h)=D_{\lambda\lambda'}^{\chi^c}(h)$$

and consequently

$$g_\lambda(\chi)=g_\lambda(\chi^c).$$

The functions $P_\lambda^\chi(x)$ have an exponentially increasing behavior for $x\to\infty$ and any nonunitary representation χ, similar as $\cos xt$ (contrary to $\exp ixt$). However, it is possible to split them up

$$P_{\lambda}^{\chi}(x)=Q_{\lambda}^{\chi}(x)+\sum_{c}Q_{\lambda}^{\chi^c}(x),$$

such that $Q_\lambda^\chi(x)$ has the desired asymptotic behavior, namely exponential decrease for $x\to\infty$ in X_+. The Q functions are defined by this property and the covariance requirement

$$\alpha(x,h)Q_{\lambda}^{\chi}(x_h)=\sum_{\lambda'}D_{\lambda',\lambda}^{\chi}(h)Q_{\lambda'}^{\chi}(x),$$

which, contrary to the P-functions, cannot be satisfied for all G but at least for infinitesimal elements. In the case of the group SU(1,1) the P and Q functions are Legendre functions of the first and second kind, respectively, and the covariance constraint for infinitesimal elements reduces to the Legendre differential equation. Thus the asymptotic reduction on non-Abelian noncompact groups is faced with the problem of defining "second kind" functions. Having found them, we easily arrive at

$$\begin{aligned}f(x)&=k\sum_{\lambda}\int g_{\lambda}(\chi)Q_{\lambda}^{\chi}(x)\,d\mu(\chi)\\&=ki\sum_{\lambda}\sum_{n=1}^{N}\gamma_{\lambda,n}Q_{\lambda}^{\chi^n}(x)+R_N(x),\end{aligned}$$

where k is the order of the group of reflections in the space X.

In this article we extend this method to the decomposition of tensor products. Tensor products are studied in this context because we have a physical application in mind, namely operator product expansions in conformally covariant quantum field theory.[4] Let $\phi^\chi(x)$ denote a field operator in Minkowski space transforming as the representation χ of the universal covering of the conformal group.[5] Then one wants to derive expansions

$$\begin{aligned}\phi^{\chi_1}(x)\,\phi^{\chi_2}(y)=\sum_{n=1}^{N}\int Q(x,y,z\,|\chi_1,\chi_2,\chi(n))\\ \phi^{\chi(n)}(z)\,d\mu(z)+R_N(x,y),\end{aligned}$$

that are asymptotic for the arguments x,y approaching each other. The conformal group of two-dimensional space–time has the structure of a direct product $SO(2,1)\otimes SO(2,1)$. So product representations of $SU(1,1)^{uc}$ are the most elementary examples of interest in this context.

Our notations follow those of Ref. 1. All normalization constants N, N^d are chosen as $(2\pi)^{-3}$.

2. THE MAIN THEOREM

According to the remarks in the Introduction we are not interested in Hilbert spaces and thus define representations on certain nuclear spaces $\mathcal{D}_\tau$, $0\le\tau<1$,

$$\mathcal{D}_\tau=\{g(\varphi)\in C^\infty(-\infty,+\infty)\,|\,g(\varphi+2\pi)=\exp(2\pi i\tau)g(\varphi)\},\tag{1}$$

of complex valued functions. Let a group element g of $SU(1,1)^{uc}$ be defined by the matrix

$$v=\begin{pmatrix}\bar\alpha, & -\beta\\ -\bar\beta, & \alpha\end{pmatrix},\quad \det v=1,$$

and by $\xi=\arg\alpha$, $g=(v,\xi)$. Then we define a continuous representation $\chi=(j,\tau)$ in $\mathcal{D}_\tau$ by

$$T_g^{\chi}g(\varphi)=|\alpha+\beta\exp(-i\varphi)|^{2j-1}g(\varphi_g),\tag{2}$$

$$\varphi_g=\varphi+2\arg(\alpha+\beta\exp(-i\varphi)),\tag{3}$$

such that φ_g is C^∞ in g and φ_g reducing to $\varphi_g=\varphi$ at the group unit. j may be any complex number. Purely imaginary j lead to the principal series of $SU(1,1)^{uc}$ by completing $\mathcal{D}_\tau$ with the sesquilinear form

$$(g_1,g_2)=\frac{1}{2\pi}\int_0^{2\pi}\overline{g_1(\varphi)}\,g_2(\varphi)\,\alpha\varphi,\tag{4}$$

and a corresponding extension of (2), (3). The functions

$$\{g_q(\varphi)\,|\,g_q(\varphi)=\exp(iq\varphi),\ q\hat{=}\tau\bmod 1\},\tag{5}$$

form the canonical basis of $\mathcal{D}_\tau$ that is orthonormal with respect to the form (4). In this article we shall assume that $\mathcal{D}_\tau$ does not possess an invariant subspace of the type $\mathcal{J}_\tau^{(\pm)}$.[1] Such subspaces can be made to carry representation of the discrete series. By a straightforward generalization of our results such cases can be included (see, e.g., Ref. 4).

Given two representations χ_1,χ_2 we consider the product representation carried by the linear span of functions

$$g_1(\varphi_1)g_2(\varphi_2),\quad g_i(\varphi_i)\in\mathcal{D}_{\tau_i}.\tag{6}$$

In our main theorem we will establish a formula

$$\begin{aligned}g_1(\varphi_1)g_2(\varphi_2)=\sum_{k=0}^{\infty}\gamma_k^{j_1j_2}\int_0^{2\pi}Q(\chi_1,\varphi_1;\chi_2,\varphi_2\,|\chi(k),\varphi)\\ F(\chi(k),\varphi)\,d\varphi,\end{aligned}\tag{7}$$

where $\gamma_k^{j_1 j_2}$ are normalization constants, $F(\chi, \varphi)$ for fixed χ belongs to a space $\mathcal{D}_\tau$ and transforms as a representation $\chi = \chi(k) = (j(k), \tau)$, $j(k) = -\frac{1}{2} + j_1 + j_2 - k$, $\tau \hat{=} \tau_1 + \tau_2 \bmod 1$ and Q is a "semicovariant kernel" in the sense that

$$T_g^{\chi_1} T_g^{\chi_2}\, Q(\chi_1, \varphi_1; \chi_2 \varphi_2 | \chi(k), \varphi) = Q(\chi_1, \varphi_1; \chi_2, \varphi_2 | \chi(k), \varphi)\, T_g^{\chi}, \tag{8}$$

holds for a certain subset of the group $SU(1,1)^{uc}$. This series (7) is asymptotic in the sense of the limit $\varphi_1 - \varphi_2 - 2\pi n \to 0$, the rhs is an expansion in increasing powers of $\varphi_1 - \varphi_2 - 2\pi n$. Before we formulate the expansion (7) as our main theorem, we describe the quantities F and Q in detail, introduce some further notations, and prepare the proof of the main theorem by establishing some lemmas.

Define functions $F(\chi(k), \varphi)$ in terms of $g_1(\varphi_1) \in \mathcal{D}_{\tau_1}$, $g_2(\varphi_2) \in \mathcal{D}_{\tau_2}$ by

$$F(\chi(k), \varphi) = \int_0^{2\pi} \int_0^{2\pi} P(\chi(k), \varphi | \chi_1, \varphi_1; \chi_2, \varphi_2) \times g_1(\varphi_1) g_2(\varphi_2)\, d\varphi_1\, d\varphi_2, \tag{9}$$

where $P(\chi(k), \varphi | \chi_1, \varphi_1; \chi_2, \varphi_2)$ is the differential operator

$$P(\chi(k), \varphi | \chi_1, \varphi_1; \chi_2, \varphi_2) = P_k^{j_1 j_2}\left(+i\frac{\partial}{\partial \varphi_1}, +i\frac{\partial}{\partial \varphi_2}\right) \delta_{\tau_1}(\varphi - \varphi_1)\, \delta_{\tau_2}(\varphi - \varphi_2) \tag{10}$$

with the polynomial

$$P_k^{j_1 j_2}(q_1, q_2) = \sum_{m=0}^{k} (-1)^m \binom{k}{m} (2j_1 - k)_m \left(\tfrac{1}{2} - j_2 - q_2\right)_m \times (2j_2 - k)_{k-m} \left(\tfrac{1}{2} - j_1 - q_2\right)_{k-m}, \tag{11}$$

and the "τ-periodic" delta functions

$$\delta_\tau(\varphi) = \frac{1}{2\pi} \sum_q \exp(iq\varphi) = \sum_{n=-\infty}^{+\infty} \exp(2\pi i n \tau)\, \delta(\varphi - 2\pi n). \tag{12}$$

Lemma 1: The integral transform (9) establishes a linear continuous map of $\mathcal{D}_{\tau_1} \times \mathcal{D}_{\tau_2}$ (continuity in separate variables) into $\mathcal{D}_\tau$ that is covariant in the sense

$$T_g^{\chi(k)} F(\chi(k), \varphi) = \int_0^{2\pi} \int_0^{2\pi} P(\chi(k), \varphi | \chi_1, \varphi_1; \chi_2, \varphi_2)\, T_g^{\chi_1} g_1(\varphi_1) \times T_g^{\chi_2} g_2(\varphi_2)\, d\varphi_1\, d\varphi_2 \tag{13}$$

for all $g \in SU(1,1)^{uc}$.

Proof: By construction we have the first part of the lemma immediately. It remains to establish (13). The most elegant way to do this is to remember[1] that for the covariant kernels K_1, K_3 applied to $g_1 \in \mathcal{D}_{\tau_1}$, $g_2 \in \mathcal{D}_{\tau_2}$ the integral transforms

$$\int_0^{2\pi} \int_0^{2\pi} K_{1,3}(\chi, \varphi | \chi_1, \varphi_1; \chi_2, \varphi_2) g_1(\varphi_1) g_2(\varphi_2)\, d\varphi_1\, d\varphi_2 \tag{14}$$

are meromorphic in j if $\chi = (j, \tau)$ and, at points of holomorphy, exhibit the same properties as asserted in the lemma, namely continuity of $\mathcal{D}_{\tau_1} \times \mathcal{D}_{\tau_2} \to \mathcal{D}_\tau$ and covariance. At $j = j(k)$ the K_1 transform has a pole with residue

$$\operatorname*{Res}_{j=j(k)} \int_0^{2\pi} \int_0^{2\pi} K_1(\chi, \varphi | \chi_1, \varphi_1; \chi_2, \varphi_2) g_1(\varphi_1) g_2(\varphi_2)\, d\varphi_1\, d\varphi_2 = \text{const} \int_0^{2\pi} \int_0^{2\pi} P(\chi(k), \varphi | \chi_1, \varphi_1; \chi_2, \varphi_2) g_1(\varphi_1) g_2(\varphi_2) \times d\varphi_1\, d\varphi_2, \tag{15}$$

where const has been calculated explicitly.[4] Thus (13) follows from the covariance of K_1 by analytic continuation in j.

However, (13) can also be proven directly. Any element $g \in SU(1,1)^{uc}$ can be decomposed as

$$g = g_3(\lambda_1) g_1(\vartheta) g_3(\lambda_2) \tag{16}$$

with

$$g_3(\lambda) = [\exp(-i\lambda\sigma_{3/2}),\ \xi = \tfrac{1}{2}\lambda], \quad -\infty < \lambda < \infty, \tag{17}$$

$$g_1(\vartheta) = [\exp(-\vartheta\sigma_{1/2}),\ \xi = 0], \quad -\infty < \vartheta < +\infty. \tag{18}$$

It suffices therefore to prove (13) for these one-parameter subgroups. The elements $g_3(\lambda)$ are also called translations, since

$$T_{g_3(\lambda)}^{\chi} g(\varphi) = g(\varphi + \lambda). \tag{19}$$

Thus (13) is trivial for this subgroup of translations.

In order to prove (13) for the subgroup (18), we note that by continuity it suffices to prove it for the canonical basis elements. A product of basis elements of $\mathcal{D}_{\tau_1}$ and $\mathcal{D}_{\tau_2}$ goes into a basis element of $\mathcal{D}_\tau$ times $P_k^{j_1 j_2}(q_1, q_2)$. For these basis vectors

$$T_{g_1(\vartheta)}^{\chi} g_q(\varphi)$$

is analytic in ϑ for $|\mathrm{Im}\,\vartheta| < \pi/2$. It suffices therefore to study infinitesimal transformations

$$T_{g_1(\vartheta)}^{\chi} = 1 + \tfrac{1}{2}\vartheta A^{\chi} + O(\vartheta^2),$$

$$A^{\chi} g(\varphi) = \left[\exp[i\varphi]\left(j - \tfrac{1}{2} + i\frac{\partial}{\partial\varphi}\right) + \exp(-i\varphi) \times \left(j - \tfrac{1}{2} - i\frac{\partial}{\partial\varphi}\right)\right] g(\varphi). \tag{20}$$

Acting on canonical basis elements, (13) and (20) reduce to two identities

$$\left(\tfrac{1}{2} - j_1 \pm q_1\right) P_k^{j_1 j_2}(q_1 \pm 1, q_2) + \left(\tfrac{1}{2} - j_2 \pm q_2\right) P_k^{j_1 j_2}(q_1, q_2 \pm 1) = (1 - j_1 - j_2 \pm q_1 \pm q_2 + k) P_k^{j_1 j_2}(q_1, q_2). \tag{21}$$

One of these can be reduced to the other one by means of

$$P_k^{j_1 j_2}(q_1, q_2) = (-1)^k P_k^{j_1 j_2}(-q_1, -q_2). \tag{22}$$

The lower sign identity (21) is easily proven inserting (11). The relation (22) is a known identity for ${}_3F_2$-series (Ref. 6, Sec. 4.3.3). This completes the proof of the lemma.

We remark that continuity and the covariance constraint (13) determine $P_k^{j_1 j_2}(q_1, q_2)$ uniquely up to a normalization factor depending at most on χ_1, χ_2, and k.

Next we define the kernels Q which we denote "second kind kernel" or "semicovariant kernel."[4] We set

$$Q(\chi_1, \varphi_1; \chi_2, \varphi_2 | \chi, \varphi) = \frac{-\pi\Gamma(1-2j)}{\Gamma(\frac{1}{2}+j_1-j_2-j)\Gamma(\frac{1}{2}-j_1+j_2-j)}$$
$$\times\{\exp[i\pi(\tfrac{1}{2}-j_2+\tau_2)\epsilon(\varphi_1-\varphi_2)]$$
$$\times K_1^d(\chi_1, \varphi_1; \chi_2, \varphi_2 | \chi, \varphi)$$
$$-\exp[i\pi(\tfrac{1}{2}-j_1-\tau_1)\epsilon(\varphi_1-\varphi_2)]$$
$$\times K_3^d(\chi_1, \varphi_1; \chi_2, \varphi_2 | \chi, \varphi)\}, \quad (23)$$

where

$$\epsilon(\varphi) = \mathrm{sgn}(\sin\varphi). \quad (24)$$

First note that Q maps $\mathcal{D}_\tau$, $\tau \hat{=} \tau_1 + \tau_2$ mod1, into the space $\mathcal{D}'_{\tau_1\tau_2}$ of distributions continuously. Here $\mathcal{D}'_{\tau_1\tau_2}$ denotes the distribution space over test functions $g(\varphi_1, \varphi_2)$ that are C^∞ and lie in $\mathcal{D}_{\tau_1}(\mathcal{O}_{\tau_2})$ for fixed $\varphi_2(\varphi_1)$, i.e.,

$$\delta \in \mathcal{D}'_{\tau_1\tau_2} : (\delta, g) = \int_0^{2\pi}\int_0^{2\pi} \delta(\varphi_1, \varphi_2) \times \overline{g(\varphi_1, \varphi_2)}\, d\varphi_1\, d\varphi_2. \quad (25)$$

Moreover, the operation of K_1^d and K_3^d on $g \in \mathcal{D}_\tau$ is defined by analytic continuation in χ_1, χ_2, χ: First take j_1, j_2, j pure imaginary and then continue analytically in these parameters. If applied to elements of the polynomial subspace $\mathcal{P}_\tau \subset \mathcal{D}_\tau$ that consists of finite linear combinations of the canonical basis (5) K_1^d and K_3^d both behave asymptotically as

$$|K^d| \sim \mathrm{const}\,|\sin\tfrac{1}{2}(\varphi_1-\varphi_2)|^{-1/2+j_1+j_2-|\mathrm{Re}j|}, \quad (26)$$

in the neighborhood of a zero of $\sin\frac{1}{2}(\varphi_1-\varphi_2)$. Otherwise both kernels are C^∞ in both arguments. Thus $Q(\chi_1, \varphi_1; \chi_2, \varphi_2 | \chi, \varphi)$ (23) can be defined on $\mathcal{P}_\tau$ by (23) in the domain

$$\mathrm{Re}(j_1+j_2) - |\mathrm{Re}j| > -\tfrac{1}{2}, \quad (27)$$

and outside this domain by analytic continuation. The image of $\mathcal{P}_\tau$ lies in $\mathcal{D}'_{\tau_1\tau_2}$.

Lemma 2: If applied to elements of $\mathcal{P}_\tau$, Q behaves asymptotically for $\varphi_1 - \varphi_2 \to 2\pi n$, n an integer, as $O((\varphi_1-\varphi_2-2\pi n)^{-1/2+j_1+j_2-j})$ and is C^∞ except $\varphi_1-\varphi_2$ a multiple of π.

Proof: Inserting the explicit forms of K_1^d and K_3^d into (23) yields

$$L(\chi_1, \varphi_1; \chi_2, \varphi_2 | \chi, q) = \int_0^{2\pi} Q(\chi_1, \varphi_1; \chi_2, \varphi_2 | \chi, \varphi)\exp(iq\varphi)\, d\varphi$$
$$= (2\pi)^{-2}\exp[i(\tfrac{1}{2}-j_2+\tau_2)S(\varphi_1-\varphi_2) + iq\varphi_1 - i\tau_2(\varphi_1-\varphi_2)]$$
$$\times\{1-\exp[i(\varphi_1-\varphi_2)]\}^{-1/2+j_1+j_2-j}$$
$$\times {}_2F_1(\tfrac{1}{2}+j_1-j_2-j,\ \tfrac{1}{2}-j+q;\ 1-2j;\ 1-\exp[i(\varphi_1-\varphi_2)]). \quad (28)$$

Here $S(\varphi)$ is the sawtooth function

$$S(\varphi) = \begin{cases} \varphi, & -\pi < \varphi \le +\pi, \\ \text{periodic otherwise with period } 2\pi. \end{cases} \quad (29)$$

Moreover, ${}_2F_1(\cdots; z)$ denotes that branch of the hypergeometric function that is obtained from the hypergeometric series around $z=0$ by analytic continuation with cut at $1 \le z < \infty$. The power in front of it, $z^{-1/2+j_1+j_2-j}$, is defined on the principal sheet with cut along $-\infty < z \le 0$. Thus both factors are periodic in φ_1 and φ_2. The assertion of the lemma follows immediately from (28).

We remark that for $j = j(k)$ as in (7) with $k = 0, 1, 2, \cdots$ the cut in z at $-\infty < z \le 0$ cancels. Q becomes C^∞ except at values of $\varphi_1 - \varphi_2$ equal to odd multiples of π.

We observe that L [(28)] is τ_1-periodic in φ_1 and τ_2-periodic in φ_2 but does not lie in $\mathcal{D}_{\tau_1}(\mathcal{O}_{\tau_2})$ for fixed $\varphi_2(\varphi_1)$ due to the singularities at $\varphi_1 - \varphi_2 = \pi m$. If

$$\mathrm{Re}(-\tfrac{1}{2}+j_1+j_2-j) \ge l \text{ integral}, \quad l \ge 0,$$

then L is l times differentiable at $\varphi_1 - \varphi_2 = 2\pi m$. Define

$$S = \{(\varphi_1, \varphi_2) \,|\, \varphi_1 - \varphi_2 = (2n+1)\pi,\ n \text{ integer}\}, \quad (30)$$

and consider the difference

$$\mathbb{R}_2 - S = \mathbb{R}'_2.$$

For any $(\varphi_1, \varphi_2) \in \mathbb{R}'_2$ we can define an open connected subset $G(\varphi_1, \varphi_2) \subset \mathrm{SU}(1,1)^{\mathrm{uc}}$ containing the group unit so that

$$(\varphi_{1g}, \varphi_{2g}) \in \mathbb{R}'_2 \text{ for all } g \in G(\varphi_1, \varphi_2).$$

For example decompose g as in (16) and let d denote the distance of (φ_1, φ_2) from S. Then $G(\varphi_1, \varphi_2)$ can be defined as the subset with

$$|\vartheta| < 2\,\mathrm{arth}(\sin d/\sqrt{8}). \quad (31)$$

Using the same multipliers and substitutions as in (2), (3), we can thus define, for each $g \in G(\varphi_1, \varphi_2)$,

$$T_g^{\chi_1} \times T_g^{\chi_2} L(\chi_1, \varphi_1; \chi_2, \varphi_2 | \chi, q)$$

Lemma 3: The kernel Q establishes a linear map of $\mathcal{P}_\tau$ into $\mathcal{D}'_{\tau_1\tau_2}$ satisfying the covariance relation

$$T_g^{\chi_1} \times T_g^{\chi_2} \int_0^{2\pi} Q(\chi_1, \varphi_1; \chi_2, \varphi_2 | \chi, \varphi) g(\varphi)\, d\varphi$$
$$= \int_0^{2\pi} Q(\chi_1, \varphi_1; \chi_2, \varphi_2 | \chi, \varphi)\, T_g^{\chi} g(\varphi)\, d\varphi, \quad (32)$$

if g is restricted to $G(\varphi_1, \varphi_2)$ and $(\varphi_1, \varphi_2) \in \mathbb{R}'_2$ is kept fixed.

Proof: It remains to prove (32). We replace the functions $\epsilon(\varphi_1 - \varphi_2)$ in (23) by

$$[\epsilon(\varphi_1-\varphi_2) - (2/\pi)\sigma(\varphi_1-\varphi_2)] + (2/\pi)\sigma(\varphi_1-\varphi_2), \quad (33)$$

with $\sigma(\varphi)$ defined by

$$\sigma(\varphi) = \begin{cases} \arg(\sin[\varphi/2 - i0]) + \pi/2), \\ \pi/2 \quad \text{for } 0 < \varphi < 2\pi. \end{cases} \quad (34)$$

This function is known[1] to obey

$$\sigma(\varphi_{1g} - \varphi_{2g}) = \sigma(\varphi_1 - \varphi_2), \quad (35)$$

for all $g \in \mathrm{SU}(1,1)^{\mathrm{uc}}$. The difference $\epsilon - (2/\pi)\sigma$ can be made continuous on $\mathbb{R}_2$ (see Fig. 1); it is then constant on each connected subset of $\mathbb{R}'_2$. On the other hand $\exp[-i\xi\sigma(\varphi_1-\varphi_2)]\, K_{1,3}^d$ can be expressed each as a linear combination of K_1^d and K_3^d with constant coefficients as follows from the construction of these kernels.[1] Thus Q is a linear combination of K_1^d and K_3^d with constant coefficients on connected domains of $\mathbb{R}'_2$. This completes the proof.

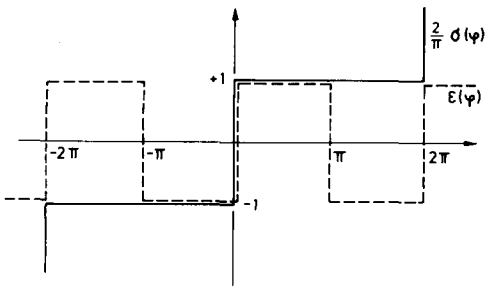

FIG. 1. The functions $(2/\pi)\sigma(\varphi)$ and $\epsilon(\varphi)$.

In turn one would like to use appropriately weakened assertions of Lemmas 2 and 3 to prove the uniqueness of the kernel Q. This, however, leads us to the core of the whole problem, namely to find a mathematical framework in which assumptions that guarantee the uniqueness of Q are "natural." Since this is an unsolved problem, we refrain from discussing this uniqueness question here.

Theorem 1 (*the main theorem*): Consider $g_1(\varphi_1) \in \mathcal{D}_{\tau_1}$, $g_2(\varphi_2) \in \mathcal{D}_{\tau_2}$ and construct $F(\chi(k), \varphi)$ from g_1, g_2 as in (9). Then

$$\sum_{k=0}^{\infty} \gamma_k^{j_1 j_2} \int_0^{2\pi} Q(\chi_1, \varphi_1; \chi_2, \varphi_2 | \chi(k), \varphi) F(\chi(k), \varphi)\, d\varphi, \tag{36}$$

with

$$\gamma_k^{j_1 j_2} = (2\pi)^2 \frac{(-1)^k}{k!} \frac{\Gamma(-2j(k)-k)}{\Gamma(-2j(k))}, \tag{37}$$

converges uniformly together with all its derivatives towards $g_1(\varphi_1) g_2(\varphi_2)$ on any compact subset of $\mathbb{R}_2'$.

Proof: Obviously it suffices to prove the theorem for elements on the canonical basis, where it assumes the form

$$\exp(iq_1\varphi_1 + iq_2\varphi_2) = \sum_{k=0}^{\infty} \gamma_k^{j_1 j_2} P_k^{j_1 j_2}(q_1, q_2) \times L(\chi_1, \varphi_1; \chi_2, \varphi_2 | \chi(k), q_1 + q_2).$$

Inserting the shorthand

$$z = 1 - \exp[i(\varphi_1 - \varphi_2)], \tag{39}$$

we have to prove

$$(1-z)^{-(1/2 - j_2 + q_2)} = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \frac{\Gamma(-2j(k)-k)}{\Gamma(-2j(k))} \times P_k^{j_1 j_2}(q_1, q_2) z^k {}_2F_1(\tfrac{1}{2} + j_1 - j_2 - j(k), \tfrac{1}{2} - j(k) + q_1 + q_2; 1 - 2j(k); z). \tag{40}$$

The left-hand side is defined on the z plane for all z except the cut at $1 \le z < \infty$ by analytic continuation from the binomial expansion.

Our proof establishes two things: First uniform convergence of the rhs of (40) in all compact subsets of the cut z plane; second the equality of the Taylor expansions of both sides of (40) around $z = 0$ (that converge for $|z| < 1$).

For the first step we apply a conformal mapping of the cut z plane on the unit circle in the w plane by (Fig. 2)

$$z = 4w/(1+w)^2. \tag{41}$$

Using Watson's results on the asymptotic behavior of ${}_2F_1$ functions in a parameter,[7] we estimate the ${}_2F_1$ function for $k \to \infty$ and any $|w| < 1$. The polynomial $P_k^{j_1 j_2}(q_1, q_2)$ is of the ${}_3F_2$ type and can be estimated after application of a three-term relation[6] with the Lemma of Ref. 1, Appendix. It follows that each term in the expansion (40) behaves as

$$+\pi^{1/2} 2^{2j_1 + 2j_2} (1+w)^{3/2 - j_1 - 3j_2 + q_1 + q_2} (1-w)^{-1/2 - j_1 + j_2 - q_1 - q_2} \times w^k \left[\frac{k^{-1/2 - j_1 - j_2 - q_1 + q_2}}{\Gamma(\frac{1}{2} - j_2 + q_2)\Gamma(\frac{1}{2} - j_1 - q_1)} \left(1 + \sum_{n=1}^{\infty} a_n k^{-n}\right) + (-1)^k \frac{k^{-1/2 - j_1 - j_2 + q_1 - q_2}}{\Gamma(\frac{1}{2} - j_2 - q_2)\Gamma(\frac{1}{2} - j_1 + q_1)} \left(1 + \sum_{n=1}^{\infty} b_n k^{-n}\right)\right] \tag{42}$$

which proves uniform convergence for $|w| < 1$.

The second part consists in proving the identity

$$(-1)^l \binom{-\frac{1}{2} + j_2 - q_2}{l} = \sum_{k=0}^{l} \frac{(-1)^k}{k!} \frac{\Gamma(-2j(k)-k)}{\Gamma(-2j(k))} P_k^{j_1 j_2}(q_1, q_2) \times \frac{(\frac{1}{2} + j_1 - j_2 - j(k))_{l-k} (\frac{1}{2} - j(k) + q_1 + q_2)_{l-k}}{(l-k)!\,(1-2j(k))_{l-k}}. \tag{43}$$

The rhs can be expanded into a sum over

$$(\tfrac{1}{2} - j_1 + q_1)_n (\tfrac{1}{2} - j_2 + q_2)_{l-n}.$$

Comparing coefficients of both sides in (43), we need to prove only

$$\delta_{n0} = \sum_{k=0}^{l} (-1)^k \binom{l}{k} \times \frac{\Gamma(-2j(k)-k)(1-2j_2+k)_{l-k}}{\Gamma(-2j(k))(1-2j(k))_{l-k}} \times \sum_{m=0}^{k} (-1)^m \binom{k}{m} \binom{l-k}{n-m} (2j_1 - k)_{k-m} (2j_2 - k)_m. \tag{44}$$

We multiply with $(-z)^n$ and sum over n. The resulting hypergeometric polynomial

$${}_2F_1(-k, 2j_2 - k; 1 - 2j_1; z),$$

is rewritten in terms of

$${}_2F_1(-k, 2j_2 - k; 1 + 2j(k); 1 - z),$$

by means of one of Kummer's relations. The resulting

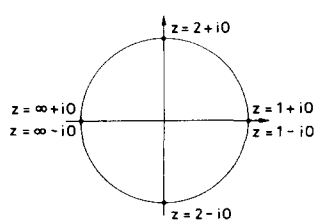

FIG. 2. The image of the cut z plane on the unit circle in the w plane.

function is expanded in powers of $(1-z)$, and, comparing coefficients, we obtain

$$\delta_{n0}=\sum_{k=0}^{n}(-1)^k\binom{n}{k}[(2j(k)-2l+3k)_k(1+2j(k)+4k-2l)_{n-k}]^{-1}. \tag{45}$$

One way to prove this elementary identity is by establishing periodicity under the substitution

$$2j(k)\to 2j(k)+1.$$

This completes the proof of the second part.

Our first derivation[4] of our main theorem by means of contour deformations in the completeness integral for the kernels K_1, K_3, K_1^d, K_3^d is much more imaginative and certainly easier to generalize to other groups. Nevertheless, completing it to a proof necessitates estimates of residual integrals that look rather hard. For this reason we prefer to establish the theorem by frontal attack.

As a mere corollary of Lemmas 1, 2, 3 and Theorem 1 we have

Theorem 2: The expansion (36) is termwise infinitesimally covariant in the sense

$$\begin{aligned}&(A^{\chi_1}+A^{\chi_2})\sum_{k=0}^{\infty}\gamma_k^{j_1j_2}\\&\times\int_0^{2\pi}Q(\chi_1,\varphi_1;\chi_2,\varphi_2|\chi(k),\varphi)F(\chi(k),\varphi)\,d\varphi\\&=\sum_{t=0}^{\infty}\gamma_k^{j_1j_2}(A^{\chi_1}+A^{\chi_2})\\&\times\int_0^{2\pi}Q(\chi_1,\varphi_1;\chi_2,\varphi_2|\chi(k),\varphi)F(\chi(k),\varphi)\,d\varphi,\end{aligned} \tag{46}$$

on $\mathbb{R}_2'$. For the sake of later application and generalization we need a weaker version of this Theorem 2. Write, using notations (39), (41),

$$\begin{aligned}&Q(\chi_1,\varphi_1;\chi_2,\varphi_2|\chi(k),\varphi)\\&=(2\pi)^{-2}\exp[i(\tfrac{1}{2}-j_2+\tau_2)S(\varphi_1-\varphi_2)-i\tau_2(\varphi_1-\varphi_2)\\&\times\sum_{n=0}^{\infty}w^{n+k}a_n\Big(k,-i\frac{\partial}{\partial\varphi_1}\Big)\delta_\tau(\varphi_1-\varphi),\quad |w|<1,\end{aligned} \tag{47}$$

and cut the sum at $n=N-k$:

$$\begin{aligned}&Q(\chi_1,\varphi_1;\chi_2,\varphi_2|\chi(k),\varphi)=Q_N(\chi_1,\varphi_1;\chi_2,\varphi_2|\chi(k),\varphi)\\&\qquad+O(w^{N+1}),\\&Q_N(\chi_1,\varphi_1;\chi_2,\varphi_2|\chi(k),\varphi)=0\ \text{ for } N<k.\end{aligned} \tag{48}$$

Comparison of (47) with (28) yields

$$\begin{aligned}a_n(k,q)=&\binom{-2k}{n}2^{2k}\\&\times{}_4F_3(\tfrac{1}{2}+j_1-j_2-j(k),\ \tfrac{1}{2}-j(k)+q,\ n+2k,\ -n;\\&1-2j(k),\ k,\ k+\tfrac{1}{2};1),\end{aligned} \tag{49}$$

as a polynomial in q of degree n.

We formulate then

Theorem 2′: The asymptotic expansion

$$\begin{aligned}g_1(\varphi_1)g_2(\varphi_2)=&\sum_{k=0}^{\infty}\gamma_k^{j_1j_2}\int_0^{2\pi}Q_N(\chi_1,\varphi_1;\chi_2,\varphi_2|\chi(k),\varphi)\\&\times F(\chi(k),\varphi)\,d\varphi+O(w)^{N+1}\end{aligned} \tag{50}$$

is asymptotically infinitesimally covariant in the sense

$$\begin{aligned}&(A^{\chi_1}+A^{\chi_2})\Big(g_1(\varphi_1)g_2(\varphi_2)\\&-\sum_{k=0}^{\infty}\gamma_t^{j_1j_2}\int_0^{2\pi}Q_N(\chi_1,\varphi_1;\chi_2,\varphi_2|\chi(k),\varphi)F(\chi(k),\varphi)\,d\varphi\Big)\\&=O(w^{N+1}),\end{aligned} \tag{51}$$

on $\mathbb{R}_2'$.

Proof: We need only to establish

$$\begin{aligned}&(A^{\chi_1}+A^{\chi_2})\int_0^{2\pi}Q_N(\chi_1,\varphi_1;\chi_2,\varphi_2|\chi(k),\varphi)\exp(iq\varphi)\,d\varphi\\&\quad-\int_0^{2\pi}Q_N(\chi_1,\varphi_1;\chi_2,\varphi_2|\chi(k),\varphi)A^{\chi(k)}\exp(iq\varphi)\,d\varphi\\&=O(w^{N+1})\end{aligned} \tag{52}$$

since (51) is obtained by a finite sum of terms (52). On the other hand (52) follows by elementary algebra.

3. THE RECOUPLING OF COVARIANT DIFFERENTIAL OPERATORS

We give three representations χ_1,χ_2,χ_3 and define

$$\begin{aligned}&\chi_{12}=(j_{12},\tau_{12}),\\&\tau_{12}\hat{=}\tau_1+\tau_2\ \mod 1,\\&j_{12}=-\tfrac{1}{2}+j_1+j_2-k_{12},\quad k_{12}=0,1,2,\cdots.\end{aligned} \tag{53}$$

χ_{12} is the result of coupling χ_1 and χ_2. We couple next χ_{12} with χ_3 to χ:

$$\begin{aligned}&\chi=(j,\tau),\\&\tau\hat{=}\tau_1+\tau_2+\tau_3\ \mod 1,\\&j=-\tfrac{1}{2}+j_{12}+j_3-k'_{12},\quad k'_{12}=0,1,2,\cdots,\\&\ =-1+j_1+j_2+j_3-k,\quad k_{12}+k'_{12}=k.\end{aligned} \tag{54}$$

This stepwise coupling of χ_1, χ_2, and χ_3 is achieved by

$$\begin{aligned}&F^{((\chi_1,\chi_2)\chi_{12},\chi_3)}(\chi,\varphi)\\&=\int P(\chi,\varphi;((\chi_1,\chi_2)\chi_{12},\chi_3)|\chi_1,\varphi_1;\chi_2,\varphi_2;\chi_3,\varphi_3)\\&\quad\times g_1(\varphi_1)g_2(\varphi_2)g_3(\varphi_3)\,d\varphi_1\,d\varphi_2\,d\varphi_3,\end{aligned}$$

$$\begin{aligned}&P(\chi,\varphi;((\chi_1,\chi_2)\chi_{12},\chi_3)|\chi_1,\varphi_1;\chi_2,\varphi_2;\chi_3,\varphi_3)\\&=\int_0^{2\pi}P(\chi,\varphi|\chi_{12},\varphi_{12};\chi_3,\varphi_3)P(\chi_{12},\varphi_{12}|\chi_1,\varphi_1;\chi_2,\varphi_2)\,d\varphi_{12}.\end{aligned} \tag{56}$$

The same representation χ results if we couple the three representations in a different order by means of

$$\begin{aligned}&P(\chi,\varphi;(\chi_1,(\chi_2,\chi_3)\chi_{23})|\chi_1,\varphi_1;\chi_2,\varphi_2;\chi_3,\varphi_3)\\&=\int_0^{2\pi}P(\chi,\varphi|\chi_1,\varphi_2;\chi_{23},\varphi_{23})P(\chi_{23},\varphi_{23}|\chi_2,\varphi_2;\chi_3,\varphi_3)\,d\varphi_{23},\end{aligned} \tag{57}$$

with

$$\chi_{23}=(j_{23},\tau_{23}),$$
$$\tau_{23}\hat{=}\tau_2+\tau_3 \mod 1, \tag{58}$$
$$j_{23}=-\tfrac{1}{2}+j_2+j_3-k_{23},\quad k_{23}=0,1,2,\cdots$$

and

$$\chi=(j,\tau),$$
$$\tau\hat{=}\tau_1+\tau_2+\tau_3 \mod 1, \tag{59}$$
$$j=-1+j_1+j_2+j_3-k,\quad k=k_{23}+k'_{23}.$$

In the first coupling scheme the representation χ can be arrived at in $k+1$ different fashions depending on the values of $k_{12}=0,1,2,\ldots,k$. Thus the representation χ appears $(k+1)$-fold degenerate in the product of χ_1, χ_2, and χ_3. The same representation χ results in the second coupling scheme. Therefore, we expect a relation

$$P(\chi,\varphi;((\chi_1,\chi_2)\chi_{12},\chi_3)\,|\,\chi_1,\varphi_1;\chi_2,\varphi_2;\chi_3,\varphi_3)$$
$$=\sum_{k_{23}=0}^{k}((j_1,(j_2,j_3)j_{23})j\,|\,((j_1,j_2)j_{12},j_3)j)$$
$$\times P(\chi,\varphi;(\chi_1,(\chi_2,\chi_3)\chi_{23})\,|\,\chi_1,\varphi_1;\chi_2,\varphi_2;\chi_3,\varphi_3), \tag{60}$$

whose coefficients are called "6j symbols."

In fact, such a finite matrix of 6j symbols is typical for the asymptotic reduction. In the Hilbert space reduction of a triple product of principal series representations a principal series representation appears infinitely often; in fact, the degeneracy parameter corresponding to k_{12} or k_{23} has a continuous spectrum. By an analytic continuation in j certain quantities for representations of SU(1,1) can be transformed into analogous quantities for representations of SU(2).[8] The spin $\mathcal{J}$ of SU(2) corresponds indeed to the analytically continued parameter $-\frac{1}{2}+j$. Thus there ought to be a connection between our 6j symbols and the 6J symbols of SU(2).[9] This is suggested by the following fact. We have for SU(2)

$$J_{12}=J_1+J_2-k_{12},\quad k_{12}=0,1,2,\ldots,\min(2J_1,2J_2),$$
$$J=J_{12}+J_3-k'_{12},\quad k'_{12}=0,1,2,\ldots,\max(2J_3,2J_{12}). \tag{61}$$

These relations differ from (53), (54) solely by the upper bound imposed on k_{12}, k'_{12}. We emphasize in particular that the expected relation between the 6j and the 6J symbols connects representations of *any* series (not only the discrete series) of $SU(1,1)^{uc}$ with representations of SU(2).

In order to investigate the 6j symbols, we apply (60) to elements of the canonical basis $\exp(iq_i\varphi_i)$, $i=1,2,3$ and then obtain

$$P^{j_1j_2}_{k_{12}}(q_1,q_2)\,P^{j_{12},j_3}_{k-k_{12}}(q_1+q_2,q_3)$$
$$=\sum_{k_{23}=0}^{k}((j_1,(j_2,j_3)j_{23})j\,|\,((j_1,j_2)j_{12},j_3)j)$$
$$\times P^{j_2j_3}_{k_{23}}(q_2,q_3)\,P^{j_1j_{23}}_{k-k_{23}}(q_1,q_2+q_3). \tag{62}$$

To stress on the matrix character of the 6j symbol, we use also the shorthand

$$A^{j_1j_2j_3j}_{k_{23},k_{12}}=((j_1,(j_2,j_3)j_{23})j\,|\,((j_1,j_2)j_{12},j_3)j),$$
$$B^{j_1j_2j_3j}_{k_{12},k_{23}}=((j_1,j_2)j_{12},j_3)j\,|\,(j_1,(j_2,j_3)j_{23})j). \tag{63}$$

Obviously the B matrix is defined by the inverse relation (62) so that

$$\sum_{k_{12}}A^{j_1j_2j_3j}_{k_{23}k_{12}}B^{j_1j_2j_3j}_{k_{12}k'_{23}}=\delta_{k_{23}k'_{23}}, \tag{64}$$
$$\sum_{k_{23}}B^{j_1j_2j_3j}_{k_{12}k_{23}}A^{j_1j_2j_3j}_{k_{23}k'_{12}}=\delta_{k_{12}k'_{12}}. \tag{65}$$

Now we make use of the obvious relation

$$P^{j_1j_2}_{k}(q_1,q_2)=(-1)^kP^{j_2j_1}_{k}(q_2,q_1). \tag{66}$$

This leads to

$$B^{j_1j_2j_3j}_{k_{12}k_{23}}=A^{j_3j_2j_1j}_{k_{12}k_{23}}, \tag{67}$$

and

$$\sum_{k_{23},k_{31}}(-1)^{k-k_{12}-k_{23}-k_{31}}A^{j_3j_1j_2j}_{k_{12}k_{31}}A^{j_2j_3j_1j}_{k_{31}k_{23}}A^{j_1j_2j_3j}_{k_{23}k_{12}},$$
$$=\delta_{k_{12}k'_{12}}. \tag{68}$$

Compared with the great many of relations for the 6J symbols of SU(2), the results (67), (68) look rather poor.

Both sides of (62) can be expanded into a series of

$$(\tfrac{1}{2}-j_1-q_1)_a(\tfrac{1}{2}-j_2-q_2)_b(\tfrac{1}{2}-j_3-q_3)_c, \tag{69}$$

with

$$a,b,c\geq 0,\quad a+b+c=k. \tag{70}$$

Comparison of the coefficients of (69) yields a set of equations

$$\sum_{n=\max(0,\,b-k_{12})}^{\min(b,\,k-k_{12}-c)}(-1)^{b+c+n}\binom{k_{12}}{b-n}\binom{k-k_{12}}{c}\binom{k-k_{12}-c}{n}$$
$$\times(2j_1-k_{12})_{b-n}(2j_2-k_{12})_{k_{12}-b+n}(2j_3-k+k_{12})_{k-k_{12}-c}$$
$$\times(2j_{12}-k+k_{12})_c$$
$$=\sum_{k_{23}=0}^{k-a}\sum_{n'=\max(0,\,c-k_{23})}^{\min(c,\,k-k_{23}-a)}A^{j_1j_2j_3j}_{k_{23}k_{12}}(-1)^{k-k_{23}+n'-a-c}$$
$$\times\binom{k_{23}}{c-n'}\binom{k-k_{23}}{a}\binom{k-k_{23}-a}{n'}$$
$$\times(2j_1-k+k_{23})_{k-k_{23}-a}(2j_2-k_{23})_{c-n'}(2j_3-k_{23})_{k_{23}-c+n'}$$
$$\times(2j_{23}-k+k_{23})_a. \tag{71}$$

These are $\frac{1}{2}(k+1)(k+2)$ equations for fixed k_{12} that overdetermine the $k+1$ unknowns $A^{j_1j_2j_3j}_{k_{23}k_{12}}$. Nevertheless, they must be compatible and we can select $k+1$ appropriate equations that we can solve. We choose $b=0$ and set $c=0,1,2,\cdots$. The resulting simplified equations can be solved recursively, and by induction we can prove that the solution is

$$A^{j_1 j_2 j_3 j}_{k_{23} k_{12}} = 2j_{23} \frac{\Gamma(2j_2 - k_{23})}{\Gamma(2j_2 - k_{12})} \times \sum_{n=\max(0, k_{12}+k_{23}-k)}^{k_{23}} \binom{k-k_{23}+n}{n} \binom{k-k_{12}}{k_{23}-n} \times (1-2j_1+k-k_{23})_n (2j_3-k+k_{12})_{k-k_{12}-k_{23}+n} \times (2j_{12}-k+k_{12})_{k_{23}-n} [(2j_{23}-k+k_{23})_{k-k_{23}+n+1}]^{-1}, \quad (72)$$

This expression is a finite ${}_4F_3$ series. In the cases $k_{23} = 0$ or $k_{12} = k$ it reduces to a single term. For $k_{23} = k$ or $k_{12} = 0$ it simplifies to a Saalschützian ${}_3F_2$ series and thus can be summed. We find

$$A^{j_1 j_2 j_3 j}_{k_{23} 0} = 2j_{23} \binom{k}{k_{23}} \frac{(2j_3 - k)_{k-k_{23}} (1 + 2j + k - k_{23})_{k_{23}}}{(2j_{23} - k + k_{23})_{k+1}}, \quad (73)$$

$$A^{j_1 j_2 j_3 j}_{k k_{12}} = \frac{(1-2j_1)_{k_{12}} (1+2j)_{k-k_{12}}}{(1+2j_{23})_k}. \quad (74)$$

4. THE ASYMPTOTIC REDUCTION OF A TRIPLE PRODUCT

From the results of the previous sections it is immediately clear that we may write

$$g_1(\varphi_1) g_2(\varphi_2) g_3(\varphi_3) = \sum_{k_{12}=0}^{\infty} \gamma^{j_1 j_2}_{k_{12}} \int_0^{2\pi} d\varphi_{12} Q(\chi_1, \varphi_1; \chi_2, \varphi_2 | \chi_{12}, \varphi_{12}) \times F(\chi_{12}, \varphi_{12}) g_3(\varphi_3)$$

$$= \sum_{k_{12}=0}^{\infty} \gamma^{j_1 j_2}_{k_{12}} \int_0^{2\pi} d\varphi_{12} Q(\chi_1, \varphi_1; \chi_2, \varphi_2 | \chi_{12}, \varphi_{12}) \quad (75)$$

$$\times \left\{ \sum_{k'_{12}=0}^{\infty} \gamma^{j_{12} j_3}_{k'_{12}} \int_0^{2\pi} d\varphi Q(\chi_{12}, \varphi_{12}; \chi_3, \varphi_3 | \chi, \varphi) \times F^{((\chi_1, \chi_2)\chi_{12}, \chi_3)}(\chi, \varphi) \right\}$$

for $g_i \in \mathcal{D}_{\tau_i}$ where the interior sum has to be supplemented at the points of divergence by continuity. Reordering the double sum and in particular introduction of the kernel

$$\int_0^{2\pi} d\varphi_{12} Q(\chi_1, \varphi_1; \chi_2, \varphi_2 | \chi_{12}, \varphi_{12}) Q(\chi_{12}, \varphi_{12}; \chi_3, \varphi_3 | \chi, \varphi), \quad (76)$$

corresponding to the coupling of the covariant differential operators does not make sense: Such an expression (76) is not infinitesimally covariant in the sense of Lemma 3 any more due to the integration over φ_{12}, and the range of the first kernel does not lie in the domain of the second kernel.

However, we can reorder (75) into an asymptotic series. Let us pick out one term with fixed k_{12} and k'_{12} ($k_{12} + k'_{12} = k$ as before). If we then replace the exterior Q by Q_N, this Q_N can be applied to any function that is C^∞ in a neighborhood of $\varphi_{12} = \varphi_1$. Moreover, it concentrates the function to which it is applied at $\varphi_{12} = \varphi_1$. Thus

$$Q_N^{(3)}(\chi_1, \varphi_1; \chi_2, \varphi_2; \chi_3, \varphi_3)((\chi_1, \chi_2)\chi_{12}, \chi_3)\chi, \varphi) = \int_0^{2\pi} d\varphi_{12} Q_N(\chi_1, \varphi_1; \chi_2, \varphi_2 | \chi_{12}, \varphi_{12}) Q(\chi_{12}, \varphi_{12}; \chi_3, \varphi_3 | \chi, \varphi) \quad (77)$$

makes sense provided that

$$|\varphi_1 - \varphi_2 - 2n\pi| \leq \pi - \epsilon, \quad |\varphi_1 - \varphi_3 - 2n'\pi| \leq \pi - \epsilon, \quad (78)$$

Denote

$$w_{12} = -i \tan \frac{\varphi_1 - \varphi_2}{4}, \quad w_{13} = -i \tan \frac{\varphi_1 - \varphi_3}{4}. \quad (79)$$

Whereas the factor Q_N in the integral is $O(w_{12}{}^N)$, it contains a derivation of maximal order $N - k_{12}$ that lowers the minimal power of the second factor Q. Namely, a term $w_{12}^{\nu+k_{12}}$, $0 \leq \nu \leq N - k_{12}$, contains a derivative of maximal degree ν, and thus lowers the order of Q to $O(w_{13}^{\max(k'_{12}-\nu, 0)})$. Thus if w_{12} and w_{13} are of the same order, say $O(w)$, then each term ν gives

$$\begin{cases} O(w^k) & \text{if } \nu \leq k'_{12}, \\ O(w^{k+\nu-k'_{12}}) & \text{if } \nu > k'_{12}. \end{cases} \quad (80)$$

The second case occurs only if $N > k$. The whole expression (77) is always $O(w^k)$.

Finally we study the covariance of $Q_N^{(3)}$ [(77)]. Provided (78) holds, we obtain from (52) and Lemma 3

$$(A^{\chi_1} + A^{\chi_2} + A^{\chi_3}) \int_0^{2\pi} d\varphi_{12} Q_N(\chi_1, \varphi_1; \chi_2, \varphi_2 | \chi_{12}, \varphi_{12}) \times Q(\chi_{12}, \varphi_{12}; \chi_3, \varphi_3 | \chi, \varphi)$$

$$= \int_0^{2\pi} d\varphi_{12} Q_N(\chi_1, \varphi_1; \chi_2, \varphi_2 | \chi_{12}, \varphi_{12}) \times Q(\chi_{12}, \varphi_{12}; \chi_3, \varphi_3 | \chi, \varphi) A^{\chi} + \text{remainder}. \quad (81)$$

The remainder is $O(w^{N+1})$. The whole expression $Q_N^{(3)}$ is $O(w^k)$. Thus $Q_N^{(3)}$ is asymptotically infinitesimally covariant in the sense of (51).

We can then reorder (75) into an asymptotic expansion with increasing N

$$g_1(\varphi_1) g_2(\varphi_2) g_3(\varphi_3) = \sum_{k=0}^{\infty} \sum_{k_{12}=0}^{k} \gamma^{j_1 j_2}_{k_{12}} \gamma^{j_{12} j_3}_{k-k_{12}} \int_0^{2\pi} d\varphi Q_N^{(3)}(\chi_1, \varphi_1; \chi_2, \varphi_2; \chi_3, \varphi_3 | ((\chi_1, \chi_2)\chi_{12}, \chi_3)\chi, \varphi) \times F^{((\chi_1, \chi_2)\chi_{12}, \chi_3)}(\chi, \varphi) + O(w^{N+1}), \quad (82)$$

for $\varphi_1, \varphi_2, \varphi_3$ in the domain (78), so that asymptotical infinitesimal covariance in the sense of Theorem 2′ is guaranteed.

5. REMARKS ON OPERATOR PRODUCT EXPANSIONS

Explicit calculations for quantum field theoretic models have been performed only for the Thirring model so far.[4] The variables φ are related with the coordinates x of two-dimensional Minkowski space by

$$x^0 \pm x^3 = \tan(\varphi_\pm/2), \quad (83)$$

where the (+) and (−) variables transform independently by a group $SU(1,1)^{uc}_{\pm}$. Instead of functions $g_i(\varphi_i) \in \mathcal{D}_{\tau_i}$ one has to expand products of field operators

$$f_1(\varphi_{1+}, \varphi_{1-}) f_2(\varphi_{2+}, \varphi_{2-}).$$

Such products are not C^∞ in either variable on a dense

domain in Fock space but singular at $\varphi_{1+}=\varphi_{2+}$, $\varphi_{1-}=\varphi_{2-}$. However, in the case of the Thirring model one can use the conformal covariant normal product of the field operators[3] instead, which is C^∞. If this extraction of a singular covariant factor were not possible, one could still modify the formalism developed in the preceding sections to include such singularities. Whereas the expansion (7) can be interpreted as a "covariantly reordered" Taylor expansion around $\varphi_1-\varphi_2=2n\pi$, such modified expansion looks more complicated and can only be derived if the Fourier transform (in the distribution sense) of the operator product can be realized as a meromorphic (analytic) functional.

Expansions of products of two operators are called "global" because of their convergence almost everywhere in Minkowski space. But this globality is lost in simultaneous expansions of three operators as we have learned in the present investigation. Nevertheless, interpreting these simultaneous expansions as asymptotic [such as in (82)] and expressing $F^{((x_1,x_2)x_{12},x_3)}(\chi,\varphi)$ by $F^{(x_1,(x_2,x_3)x_{23})}(\chi,\varphi)$ through the $6j$ symbols amounts to a reordering of the expansion in powers of w_{12}, w_{13} into an expansion in powers of w_{12}, w_{23}. In a similar fashion the asymptotic reductions of a four-point function in the s channel and the t channel are related to a $9j$ symbol.

6. CONCLUSIONS

The asymptotic reduction method is applied to products of representations of the group $SU(1,1)^{uc}$. Taking the second kind kernels as in Ref. 4, we prove termwise covariance of the asymptotic expansion (Lemmas 1, 3 and Theorem 2) and its convergence in a certain domain of regularity (Main Theorem). The representations appearing in the asymptotic expansions are obtained by a new kind of vector coupling coefficients [the $P_k^{j_1 j_2}(q_1,q_2)$]. Products of three representations allow us to introduce recoupling coefficients that are explicitly derived (Sec. 3). These recoupling coefficients are discrete matrices as in the case of compact Lie groups but contrary to the case of L^2 analysis on noncompact Lie groups. Their relation with the recoupling coefficients of SU(2) is still unknown. The very weak covariance properties of simultaneous asymptotic expansions for products of three representations are studied in Sec. 4.

[1]W. Rühl and B.C. Yunn, J. Math. Phys. **17**, 1521 (1976).
[2]M.A. Naimark, Am. Math. Soc. Transl. Ser. 2, **36**, 101 (1964).
[3]M. Toller, Nuovo Cimento **53**, 671 (1968); **54**, 295 (1968); A. Sciarrino and M. Toller, J. Math. Phys. **8**, 1252 (1967).
[4]W. Rühl and B.C. Yunn, Comm. Math. Phys. **48**, 215 (1976).
[5]W. Rühl and B.C. Yunn, Fortschr. Phys. **25**, 83 (1977).
[6]L.J. Slater, *Generalized Hypergeometric Functions* (Cambridge U.P., Cambridge, 1966).
[7]G.N. Watson, Trans. Cambr. Philos. Soc. **22**, 277 (1918); quoted in A. Erdélyi, W. Magnus, F. Oberhettinger, and F.G. Tricomi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. 1, p. 77.
[8]W.J. Holman and L. Biederharn, Ann. Phys. (N.Y.) **39**, 1 (1966); **47**, 205 (1968).
[9]A.R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, N.J., 1958), Chap. 6.

Local supersymmetry in (2+1) dimensions. II. An action for a spinning membrane

P. S. Howe and R. W. Tucker

Department of Physics, University of Lancaster, Lancaster, England
(Received 5 July 1977)

We present a locally supersymmetric action for a spinning membrane. This is obtained by supersymmetrizing the induced volume element action which is reformulated in terms of a three-dimensional field theory. We also discuss more complex actions which are possible due to the nontriviality of (super) gravity in three dimensions.

I. INTRODUCTION

In a separate paper, we have presented the theory of supergravity in three dimensions in some detail.[1] We remarked in that paper that owing to the nontrivial nature of free (super) gravity there exist several possibilities for the action describing two-dimensional extended systems (membranes). In this paper we discuss some of these possibilities and, in particular, write down a supersymmetrized version of the volume action for the relativistic membrane.[2] We believe that such a theory should serve as the simplest model of a spinning membrane and it provides a natural generalization to the much studied models of the relativistic string.[3] Closed membranes (bags) containing surface fermionic degrees of freedom have been studied before[4] but not in the manner described below. By constructing our action to be locally supersymmetric we endow our theory with a graded gauge structure which we believe is necessary for a consistent model. It would be premature at this stage to identify the quantized version with the hadron spectrum particularly in view of the intricasies involved in quantizing the spinless membrane.[2] However, we hope to return to an analysis of some of the simpler excitation modes of the system supplemented with the fermionic degrees of freedom to be described. It would, for example, be of considerable interest to charge our membrane and re-investigate the extended Dirac-electron[5] with intrinsic spin.

The organization of the paper is as follows. In Sec. II we reformulate the membrane action as a three-dimensional field theory; in Sec. III we supersymmetrize the action of the previous section. Finally in Sec. IV we discuss various other geometric actions for membranes both in the spinless and spinning cases.

II. THE SPINLESS MEMBRANE

The conventional action for a spinless membrane is given by the volume it generates as it moves through space–time[2,5]

$$S=-\int \sqrt{-g}\, d^3x, \tag{2.1}$$

where $g_{\mu\nu}$ is the metric induced on the submanifold generated by the membrane from the embedding flat-space metric

$$g_{\mu\nu}=\partial_\mu y^\alpha \partial_\nu y^\beta \eta_{\alpha\beta} \tag{2.2}$$

and

$$g=\det(g_{\mu\nu}).$$

The coordinates $y^\alpha(x^\mu)$ locate the membrane in the embedding space–time $[\alpha=0,1,2,3, \eta_{\alpha\beta}=(-,+++)]$ as a function of the internal coordinates $x^\mu(\mu=0,1,2)$ of the internal three-dimensional space.

As in the case of the spinless string we can rewrite (2.1) in terms of a three-dimensional scalar field theory providing we regard the set of scalar fields $\phi^\alpha \equiv y^\alpha$ and $g_{\mu\nu}$ as independent variables. Variation of S with respect to ϕ^α then provides the equation of motion whilst the metric variation gives the constraints. The action to use for this purpose is

$$S=-\tfrac{1}{2}\int d^3x \sqrt{-g}\{\partial_\mu\phi\partial_\nu\phi g^{\mu\nu}-1\}, \tag{2.3}$$

where we suppress (3 + 1) Lorentz contraction on the α indices. Observe that the only modification necessary in comparison with the string case is the inclusion of a "cosmological" term $\frac{1}{2}\sqrt{-g}$. The equations resulting from varying ϕ^α and $g_{\mu\nu}$ are

$$\Box\phi^\alpha=0 \tag{2.4}$$

and

$$t_{\mu\nu}=0. \tag{2.5}$$

Here, $\Box=(1/\sqrt{-g})\partial_\mu(g^{\mu\nu}\sqrt{-g}\,\partial_\nu)$ is the three-dimensional Laplace–Beltrami operator and $t_{\mu\nu}$ the energy momentum tensor for the system:

$$\begin{aligned} t_{\mu\nu} &\equiv -\frac{2}{\sqrt{-g}}\frac{\delta S}{\delta g^{\mu\nu}} \\ &= \partial_\mu\phi\partial_\nu\phi-\tfrac{1}{2}g_{\mu\nu}\{g^{\rho\sigma}\partial_\rho\phi\partial_\sigma\phi-1\}. \end{aligned} \tag{2.6}$$

To prove that (2.1) and (2.3) give identical theories, one can pass to the Hamiltonian formalism and identify the constraints.[1] Alternatively one can solve for $g_{\mu\nu}$ from (2.5)

$$g_{\mu\nu}=\lambda\partial_\mu\phi\partial_\nu\phi, \tag{2.7}$$

where $\lambda(x)$ is a scale factor. Substituting (2.7) into (2.3) then yields (2.1) up to a factor which may be absorbed by a reparametrization.

The action (2.3) is the one which, guided by our experience with string theory, we believe will furnish us with a model of a spinning membrane upon supersymmetrization. In terms of the differential forms introduced in Ref. 1, Eq. (2.3) reads

 0022-2488/78/1905-0981$1.00

$$S = \int \Lambda,$$

where

$$\Lambda = \tfrac{1}{2}\{d\phi \wedge *d\phi - \epsilon\} \tag{2.8}$$

is a Lagrangian three form.

III. AN ACTION FOR A SPINNING MEMBRANE

In order to supersymmetrize (2.8) let us first consider global supersymmetry for a massless scalar supermultiplet in flat 3-space. The Lagrangian 3-form is

$$\Lambda = \tfrac{1}{2} d\phi \wedge *d\phi + \tfrac{1}{2} i\bar{\psi} * \gamma \wedge d\psi, \tag{3.1}$$

where ψ is an SL(2, R) spinor-valued odd Grassman 0-form and $\gamma = \gamma_a e^a$.[1] On varying ϕ and ψ we obtain (up to an exact form)

$$\delta\Lambda = -\delta\phi(d * d\phi) + i\delta\bar{\psi}(*\gamma \wedge d\psi) \tag{3.2}$$

and hence the equations of motion

$$d * d\phi = 0, \quad (*\gamma) \wedge d\psi = 0. \tag{3.3}$$

The action corresponding to (3.1) is invariant under the global supersymmetry transformations

$$\delta\phi = i\bar{\alpha}\psi, \quad \delta\psi = (\boldsymbol{\gamma} \rfloor d\phi)\,\alpha, \tag{3.4}$$

where $\boldsymbol{\gamma} = \gamma^a \partial_a$, $e^a(\partial_b) = \delta^a_b$ and α is a constant SL(2, R) odd Grassman spinor. Explicitly, one finds

$$\begin{aligned}\delta\Lambda &= -d(i\bar{\alpha}\psi * d\phi + i\bar{\alpha}\gamma\psi \wedge d\phi) \\ &= i\bar{\alpha}\, dJ,\end{aligned} \tag{3.5}$$

where

$$J = -\psi * d\phi - \gamma\psi \wedge d\phi. \tag{3.6}$$

J is therefore a spinor-valued 2-form current and is conserved ($dJ = 0$) when the equations of motion are satisfied.

Our next step is to make (3.1) locally supersymmetric (i.e., generate an invariant with α an arbitrary x-dependent spinor valued 0-form). We first ensure reparameterization and SO(2, 1) invariance by replacing $d\psi$ with $D\psi = d\psi + \tfrac{1}{2}\gamma_a \omega^a \psi$. We then couple the spinor one form χ to the supercurrent J and, as in the case of the spinning string, include a quadratic χ term. Our final Lagrangian is

$$\begin{aligned}\Lambda = {} & \tfrac{1}{2} d\phi \wedge *d\phi + \tfrac{1}{2} i\bar{\psi} * \gamma \wedge D\psi - \tfrac{1}{2} i\bar{\chi} \wedge (\psi * d\phi + \gamma\psi \wedge d\phi) \\ & + \tfrac{1}{16}\bar{\psi}\psi\bar{\chi} \wedge (*\chi + \gamma \wedge \chi).\end{aligned} \tag{3.7}$$

The connection 1-form ω^a appears in the $D\psi$ term and since we are not including the free supergravity part in our action here we need to specify its structure. We choose

$$\omega^a = \hat{\omega}^a + \lambda^a, \tag{3.8}$$

where $\hat{\omega}^a$ is the Christoffel connection[1] and λ^a is related to the torsion two form T^a by

$$T^a = \epsilon_b{}^2 \wedge \lambda^b, \tag{3.9}$$

where

$$T^a = (i/4)\,\bar{\chi} \wedge \gamma^a \chi. \tag{3.10}$$

We observe that the only difference in structure between (3.7) and the corresponding action for the spinning string is the presence of the connection. This modification is, however, enough to complicate the proof of supersymmetry as we now demonstrate. In varying Λ we change not only ϕ and ψ but also the frame e^a and the field χ. The supersymmetry transformations are

$$\begin{aligned}&\delta e^a = i\bar{\alpha}\gamma^a \chi, \quad \delta\chi = 2D\alpha, \quad \delta\phi = i\bar{\alpha}\psi, \\ &\delta\psi = \{\boldsymbol{\gamma} \rfloor [d\phi - (i/2)\,\bar{\chi}\psi]\}\,\alpha.\end{aligned} \tag{3.11}$$

The proof that (3.7) is supersymmetric is somewhat tedious. One finds that, after performing the variations (3.11), the terms in $\delta\Lambda$ involving no χ's sum to give an exact form plus a torsion dependent part via the definition $T^a = De^a$. Then, utilizing the explicit form (3.10) for T^a one can compensate these pieces precisely by the terms involving two χ's in $\delta\Lambda$. A similar situation obtains for the one- and three-χ terms. The final result is

$$\begin{aligned}\delta\Lambda = {} & d\{\tfrac{1}{2} i\bar{\alpha}\psi * d\phi - \tfrac{1}{2} i\bar{\alpha}\gamma\psi \wedge d\phi - \tfrac{1}{8}\bar{\psi}\psi\bar{\alpha}[*\chi + \gamma \wedge \chi]\} \\ & + \tfrac{1}{8}\bar{\psi}\psi\bar{\alpha}\gamma \wedge D\chi.\end{aligned} \tag{3.12}$$

Consequently we conclude that (3.7) is supersymmetric provided

$$D\chi = 0. \tag{3.13}$$

The last term in (3.12) is a direct consequence of the variation of the connection in (3.7). The condition (3.13) does not need to be interpreted as a gauge restriction. Indeed, if we demand that it be supercovariant we have

$$\delta(D\chi) = D^2\alpha = \tfrac{1}{2} F^a \wedge \gamma_a \alpha \tag{3.14}$$

from Eq. (2.20) of Ref. 1. Hence α is unrestricted if

$$F^a = 0. \tag{3.15}$$

It appears, therefore, that the action corresponding to (3.7) is locally supersymmetric if the background supergravity satisfies (3.13), (3.15), and (3.10). (It may be possible to formulate the theory with the aid of auxiliary variables in a space without "background" constraints. A similar proposal for four-dimensional supergravity has recently been studied in Ref. 6.) Precisely these equations were derived in Ref. 1 to describe free supergravity.

To obtain the action describing a spinning membrane, however, we need to supersymmetrize (2.5). Thus in addition to (3.7) we are required to supersymmetrize the "cosmological term" ϵ. Just as with the matter free action discussed in Ref. 1 where the "super cosmological term" includes a mass-type term for χ, in this case we find that it is necessary to include a mass-type term for ψ. The Lagrangian three form for the spinning membrane is therefore taken to be:

$$\begin{aligned}\Lambda = {} & \tfrac{1}{2} d\phi \wedge *d\phi + \tfrac{1}{2} i\bar{\psi} * \gamma \wedge D\psi - \tfrac{1}{4} i\bar{\psi}\psi\epsilon - \tfrac{1}{2} i\bar{\chi} \wedge [\psi * d\phi + \gamma\psi \wedge d\phi] \\ & + \tfrac{1}{16}\bar{\psi}\psi\bar{\chi} \wedge [*\chi + \gamma \wedge \chi] - \tfrac{1}{2}\epsilon + \tfrac{1}{8} i\bar{\chi} \wedge \gamma \wedge \chi.\end{aligned} \tag{3.16}$$

The supersymmetry transformations (3.11) remain unchanged with the exception of the χ transformation which becomes

$$\delta\chi = 2D\alpha + \gamma\alpha. \tag{3.17}$$

We also choose the connection as in the previous case and (3.17) implies that there is a corresponding alteration in $\delta\omega^a$. The calculation of $\delta\Lambda$ is slightly modified now and we find

$$\delta\Lambda = d\{-\tfrac{1}{2}i\bar{\alpha}\gamma\psi\wedge d\phi + \tfrac{1}{2}i\bar{\alpha}\psi * d\phi - \tfrac{1}{8}\bar{\psi}\psi\bar{\alpha}[*\chi + \gamma\wedge\chi]\} + [\tfrac{1}{8}\bar{\psi}\psi\bar{\alpha}\gamma - \tfrac{1}{2}i\bar{\alpha}\gamma]\wedge(D\chi + \tfrac{1}{2}\gamma\wedge\chi). \quad (3.18)$$

We conclude, therefore, that (3.16) is invariant provided that

$$D\chi + \tfrac{1}{2}\gamma\wedge\chi = 0. \quad (3.19)$$

As before, the requirement that (3.19) be a supercovariant statement imposes the condition

$$F^a = \epsilon^a - T^a. \quad (3.20)$$

Thus, we find once more that the action corresponding to (3.16) is only supersymmetric if the background supergravity satisfies (3.10), (3.19) and (3.20). These equations are just the equations of motion that we derived for free supergravity with a cosmological term in Ref. 1.

The equations of motion for the spinning membrane obtained by varying (3.16) with respect to ϕ, ψ, χ, and e^a are

$$d * d\phi - \tfrac{1}{2}id[\bar{\psi} * \chi + \bar{\psi}\gamma\wedge\chi] = 0, \quad (3.21)$$

$$*\gamma\wedge D\psi + D(*\gamma)\,\psi - \tfrac{1}{2}\psi\epsilon - \tfrac{1}{2}\chi\wedge *d\phi + \tfrac{1}{2}\gamma\wedge\chi\wedge d\phi + \tfrac{1}{8}i\bar{\chi}\wedge[*\chi + \gamma\wedge\chi]\psi = 0, \quad (3.22)$$

$$\psi * d\phi + \gamma\psi\wedge d\phi + \tfrac{1}{4}i\bar{\psi}\psi * \chi + \tfrac{3}{8}i\bar{\psi}\psi\gamma\wedge\chi - \tfrac{1}{2}\gamma\wedge\chi = 0, \quad (3.23)$$

and

$$t_a = 0, \quad (3.24)$$

respectively. In (3.24), t_a is the total energy-momentum 2-form for the system. If we write $\Lambda = \mathcal{L}\epsilon$, then $\mathcal{L}$ is the conventional Lagrangian and

$$t_a = \epsilon_b t^b_\alpha, \quad (3.25)$$

where

$$t^b_a = \frac{\partial\mathcal{L}}{\partial(\partial_b\phi)}\,\partial_a\phi + \frac{\partial\mathcal{L}}{\partial(\partial_b\psi)}\,\partial_a\psi - \delta^b_a\mathcal{L}. \quad (3.26)$$

The equations (3.21)—(3.24) appear as a complicated nonlinear set and unlike the spinning string there does not seem to be an obvious gauge that will linearize any of them. We may interpret (3.21) and (3.22) as equations of motion for ϕ^α and ψ^α and we expect (3.23) and (3.24) to provide us with the constraints of the theory. Work on the clarification of these equations is currently in progress.

IV. OTHER THREE-DIMENSIONAL SYSTEMS

As we have remarked in the introduction and discussed in Ref. 1, the actions for matter-free (super) gravity in three dimensions are nontrivial unlike their counterparts in two dimension where at most they may modify the boundary conditions. In this section we enumerate a number of other possible geometric actions for both spinless and spinning membranes. We consider first a spinless self-gravitating membrane, described by the action

$$S = \int e^a\wedge F_a + \tfrac{1}{2}[d\phi\wedge * d\phi - \epsilon]. \quad (4.1)$$

Equation (4.1) can be interpreted in two ways: one may choose the frames and connection to be those induced from the embedding space [as in (2.2)] or one could choose to vary e^a and ω^a independently. In the first case (4.1) may be written

$$S = -\int d^3x\sqrt{-g}\,\{1 + \tfrac{1}{2}R\}, \quad (4.2)$$

where R is the curvature scalar calculated from the Christoffel connection of the induced metric (2.2). In the second case, the Euler—Lagrange equations for the connection tell us that there is no torsion but the frame equations of motion become second-order differential equations and consequently acquire a dynamical status.

Another possibility is of course to use the "free gravity" part of (4.1) alone, i.e.,

$$S = \int e^a\wedge F_a. \quad (4.3)$$

In this case, we have no alternative but to regard e^a and ω^a as those induced from the embedding space. [As we have seen in Ref. 1 the solution that renders (4.3) extremal for independent e^a and ω^a variations is trivial.]

We have, therefore, three actions in addition to (2.3) that may describe a spinless membrane-type system. To endow such systems with intrinsic spin it seems likely that the corresponding supersymmetrized version of (4.1) and (4.3) should be considered. However, whilst it is clear how one interpretes the induced metric the corresponding interpretation of the χ variable is not so obvious (It would of course be extremely elegant if one were able to settle this question by tackling the entire problem in superspace.[7]) This is also an interesting question which one could ask about the spinning string. Hence we shall simply contemplate here the supersymmetrized version of (4.1) (both with and without a cosmological term) where the gravity and matter variables are treated independently. The appropriate actions are

$$S_1 = \int \tfrac{1}{2}e^a\wedge F_a + \tfrac{1}{4}i\bar{\chi}\wedge D\chi + \tfrac{1}{2}d\phi\wedge * d\phi + \tfrac{1}{2}i\bar{\psi} * \gamma\wedge D\psi - \tfrac{1}{2}i\bar{\chi}\wedge(\psi * d\phi + \gamma\psi\wedge d\phi) + \tfrac{1}{16}\bar{\psi}\psi\bar{\chi}\wedge(*\chi + \gamma\wedge\chi) \quad (4.4)$$

and

$$S_2 = S_1 + \int \tfrac{1}{8}i\bar{\chi}\wedge\gamma\wedge\chi - \tfrac{1}{2}\epsilon - \tfrac{1}{4}i\bar{\psi}\psi\epsilon, \quad (4.5)$$

where S_2 incorporate the super "cosmological term."

We mention that, although in Sec. III we were forced to impose conditions (3.13), (3.15) or (3.19), (3.20) on the gravity variables to ensure supersymmetry in this case there is no such need. (We point out that for both these actions the connection equations give a contribution to the torsion from the ψ variable. This in turn implies that the variation of the connection involves the "matter" fields.) With the extra variables (4.4) is invariant under the transformations (3.11) and (4.5) is invariant under the transformations (3.11) amended by (3.17).

To summarize, we believe that the Lagrangian (3.16) adequately describes an extended two-dimensional

system with intrinsic spin and may help to supply an answer to the question left unsolved by Dirac.[5] In addition we have discussed a variety of more complex geometric actions for both spinless and spinning systems, although in some of these cases there are problems with interpretation at present.

[1]P.S. Howe and R.W. Tucker, J. Math. Phys. **19**, 869 (1978).
[2]P.A. Collins and R.W. Tucker, Nucl. Phys. B **112**, 150 (1976).
[3]J. Scherk, Rev. Mod. Phys. 47, 123 (1975); *Dual Theory*, edited by M. Jacob (North-Holland, Amsterdam, 1974).
[4]W. Bardeen, M. Chanowitz, S. Drell, M. Weinstein, and T.M. Yan, Phys. Rev. D **11**, 1094 (1975); R.C. Giles, *ibid.* 13 1670 (1976); R.C. Giles and S-H.H. Tye, *ibid.* 13, 1690 (1976).
[5]P.A.M. Dirac, Proc. Roy. Soc. A **268**, 57 (1962).
[6]P. Breitenlohner, Phys. Lett. B **67**, 49 (1977).
[7]B. Zumino, in *Gauge Theories and Modern Field Theory*, edited by R. Arnowitt and P. Nath (M.I.T. Press, Cambridge, Mass., 1976), p. 255.

Nontranslationally covariant currents and symmetries of the *S* matrix

W. D. Garber and H. Reeh

Institut für Theoretische Physik, Universität Göttingen, Bunsenstr. 9, D-3400 Göttingen, Germany
(Received 12 July 1977)

Generators of symmetries constructed from nontranslationally covariant currents are defined on scattering states and commute with the S matrix.

1. INTRODUCTION

As in Ref. 1, we consider the general case of a symmetry transformation generated by a translationally noncovariant current. There it was shown that the corresponding generator Q can be extended as a quadratic form to two forms on the asymptotic in and out states, resp. These forms belong to operators Q^{in} and Q^{out} which have a number of restrictive properties (Theorem 8.5 of Ref. 1). In the following we show that Q itself can be extended as an operator to smooth asymptotic multiparticle states (with disjoint support in velocity space), and there commutes with the S matrix.

We work within the Wightman framework of quantum field theory. For the reader's convenience, we sketch the assumptions of Ref. 1. Since in the present note we will consider powers of Q, some of the domain assumptions are slightly more restrictive:

The field theory is assumed to be given by a finite set of fields $\{\phi_i(x)\}$ with common dense domain D, stable and Lorentz invariant. $\mathcal{P}_{SL}$ denotes the polynomial algebra of the $\phi_i(f)$, $f\in\mathcal{D}(\mathbb{R}^4)$ (test functions of compact support), Ω denotes the vacuum vector, and $U(a)$ the unitary representation of the translations. The noncovariant current density $j^\mu(y,x)$, $\mu=0,1,2$ is assumed to have the following properties:

(i) $U(a)\,j^\mu(y,x)\,U^{-1}(a)=j^\mu(y+a,x)$;

(ii) $j^\mu(x,x)$ is an operator-valued tempered distribution such that $\int j^\mu(x,x)f(x)\,dx$ maps D into D, for all f in $\mathcal{D}(\mathbb{R}^4)$;

(iii) $j^\mu(x,x)$ is Hermitian;

(iv) $j^\mu(y,x)$, in its dependence on y, is local and relatively local with respect to $\mathcal{P}_{SL}$ for all x;

(v) $\partial_\mu j^\mu(y,y)=(\partial/\partial y^\mu)j^\mu(y,y)=0$.

(ii), (iii), and (v) immediately generalize, by (i), to $j^\mu(y,x)$.

We now enlarge $\mathcal{P}_{SL}$ to $\mathcal{P}'_{SL}$ by adding

$$\{\int j^\mu(y+a,y)f(y)\,d^4y \mid f\in\mathcal{D}(\mathbb{R}^4),\ a\in\mathbb{R}^4\}$$

to the generating elements of the strictly local polynomial algebra (this is convenient for defining powers of Q). The generator $Q(a)$ is defined by

$$\begin{aligned} &Q(a)\Omega := 0,\\ &Q(a)A\Omega := \lim_{r\to\infty}[Q_r(a),A]\Omega,\quad A\in\mathcal{P}'_{SL} \end{aligned} \tag{1.1}$$

with

$$Q_r(a) := \int j^0(y+a,y)\vartheta_r(\mathbf{y})\,\eta(y^0)\,d^4y,$$

$$\vartheta_r(\mathbf{y}) := \vartheta\left(\left|\frac{\mathbf{y}}{r}\right|\right),\quad \vartheta\in\mathcal{D}(\mathbb{R}^1),\quad \vartheta(s):=\begin{cases}1, & 0\le s\le 1,\\ 0, & s\ge 2,\end{cases}$$

$$\eta\in\mathcal{D}(\mathbb{R}^1),\quad \int\eta(y^0)\,dy^0=1.$$

$Q(a)$ is well defined on $\mathcal{D}_{Q(a)} := \mathcal{P}'_{SL}\Omega$ and maps it into itself. [Note that $Q(a+b)=U(b)\,Q(a)\,U^{-1}(b)$.] Hence powers of Q are also defined on $\mathcal{P}'_{SL}\Omega$.

We now assume invariance of the vacuum under the symmetry

$$\lim_{r\to\infty}(\Omega\,|\,[Q_r(a),A]\Omega)=0\quad\text{for all } A\in\mathcal{P}'_{SL}. \tag{1.2}$$

(In Ref. 1, it was sufficient to assume this for $A\in\mathcal{P}_{SL}$ only.) (1.2) implies that all powers of Q are symmetric and hence closable.

2. SOME TECHNICAL PROPERTIES OF Q

The statements of Sec. 3 in Ref. 1. can be generalized as follows:

2.1 *Lemma*: Let T be a closed operator with domain $\mathcal{D}_T$, $\varphi(x)$ a function from $\mathbb{R}^n$ to $\mathcal{D}_T$ such that $\varphi(x)$ and $T\varphi(x)$ are weakly continuous, and $\|\varphi(x)\|$, $\|T\varphi(x)\|$ are bounded by polynomials in $|x|^2=\sum_\nu|x_\nu|^2$. Then

$$\int\varphi(x)\,f(x)\,d^nx\in\mathcal{D}_T$$

and

$$T\int\varphi(x)\,f(x)\,d^nx=\int T\varphi(x)\,f(x)\,d^nx \tag{2.1}$$

for $f\in\gamma(\mathbb{R}^n)$, as weak integrals [$\gamma(\mathbb{R}^n)$ are the Schwartz test functions of rapid decrease].

Proof: Both integrals exist by assumption and can be approximated by finite sums. The statement follows by (III, 5.12) of Ref. 2. ∎

Lemma 2.1 can be used to enlarge the domain of powers of Q. We denote by $\mathcal{P}'_1$ the polynomial algebra spanned by $\{\int A(x)f(x)\,d^4x \mid A\in\mathcal{P}'_{SL},\ f\in\gamma(\mathbb{R}^4)\}$.

2.2. *Lemma*: (*i*) $\mathcal{P}'_1\Omega$ is in the domain of the closure of the product $Q(a_1)\cdots Q(a_n)$;

(ii) $\mathcal{P}'_1\Omega$ is in the domain of $Q(a)^*\overline{Q(a)}$ ($\bar{}$ denotes closure, * the adjoint).

Proof: (i) By definition, for $A\in\mathcal{P}'_{SL}$

$$\begin{aligned} &Q(a_1)\cdots Q(a_n)A_1(x_1)\cdots A_m(x_m)\\ &\quad=\lim_{r_1\to\infty}\cdots\lim_{r_n\to\infty}[Q_{r_1}(a_1),[\cdots[Q_{r_n}(a_n),A_1(x_1)]\cdots\\ &\quad\quad\times A_m(x_m)]\cdots]\Omega. \end{aligned} \tag{2.2}$$

 0022-2488/78/1905-0985$1.00

By locality, the limits are reached for finite $r_\nu \geq r_0 = C(A_1, \cdots, A_m; a_1 \cdots a_n) + \sum_{\mu=1}^{m} |x_m|$. Putting $r = r_0$ and using the temperedness of the Wightman fields and of the current, we see that the norm of (2.2) is bounded by a polynomial in $|x_1|, \ldots, |x_m|$ and thus bounded on any compact set. The scalar product of (2.2) with any vector in $\mathcal{D}'_{SL}$ is continuous. Hence (2.2) is weakly continuous and polynomially bounded as is $A_1(x_1) \cdots A_m(x_m)\Omega$. Lemma 2.1. implies (i).

(ii) Stability of $\mathcal{D}'_{SL}\Omega = \mathfrak{H}_Q$ under Q and hermiticity of Q imply $(\bar{Q})^* \bar{Q} \supset Q^2$. [Remember $(\bar{Q})^* = Q^*$.] Hence $Q^* \bar{Q} \supset Q^2$. ∎

In the next section, we extend Q to Haag–Ruelle scattering states.[3] For their construction, we assume a mass gap and consider time-dependent states

$$B_{\nu_1}^{f_1,t} \cdots B_{\nu_n}^{f_n,t}\Omega =: \psi(t) \tag{2.3}$$

with

$$B_\nu^{f,t} = \int B_\nu(t, \mathbf{a}) f(t, \mathbf{a})\, d^3\mathbf{a}$$

and with $B_\nu \in \mathcal{D}'_1$ chosen such that $B_\nu\Omega$ is in the subspace of one-particle states of type ν and mass m_ν. f is a smooth positive frequency solution of the Klein–Gordon equation for mass m_ν. The scattering states are limits of (2.3) for $t \to \pm\infty$.

2.3. *Lemma*: (i) $\bar{Q}$ and $Q^*\bar{Q}$ are defined on $\psi(t)$.

(ii) $\|\bar{Q}\psi(t)\|$, $\|Q^*\bar{Q}\psi(t)\|$ are polynomially bounded in t.

Proof: Consider

$$B_\nu(a) = \int A_\nu(x) h_\nu(x-a)\, d^4x, \quad A_\nu \in \mathcal{D}'_{SL}, \quad h_\nu \in \gamma(\mathbb{R}^\mu).$$

Then, by Lemmas 2.2 and 2.1,

$$\begin{aligned}\overline{Q^m} B_{\nu_1}(a_1) \cdots B_{\nu_n}(a_n)\Omega &= \int Q^m A_{\nu_1}(x_1) \cdots A_{\nu_n}(x_n)\Omega h_{\nu_1}(x_1 - a_1) \\ &\quad \cdots h_{\nu_n}(x_n - a_n)\, d^4x_1 \cdots d^4x_n\end{aligned} \tag{2.4}$$

is polynomially bounded in a since $h_{\nu_i} \in \gamma(\mathbb{R}^4)$ and $\|Q^m A_{\nu_1}(x_1) \cdots A_{\nu_n}(x_n)\Omega\|$ is polynomially bounded in x as shown in the proof of Lemma 2.2. Weak continuity of (2.4) follows again as in Lemma 2.2. Of course $\|B_{\nu_1}(a_1) \cdots B_{\nu_n}(a_n)\Omega\|$ is polynomially bounded and weakly continuous, too. By Lemma 2.1, $\psi(t) \in \mathfrak{H}_{\bar{Q}}$ proving (i), and

$$\begin{aligned}\overline{Q^m} B_{\nu_1}^{f_1,t} \cdots B_{\nu_n}^{f_n,t}\Omega &= \int \bar{Q} B_{\nu_1}(t, \mathbf{a}_1) \cdots B_{\nu_n}(t, \mathbf{a}_n)\Omega f_1(t, \mathbf{a}_1) \\ &\quad \cdots f_n(t, \mathbf{a}_n)\, d^3\mathbf{a}_1 \cdots d^3\mathbf{a}_n.\end{aligned} \tag{2.5}$$

For smooth solutions f of a Klein–Gordon equation,

$$\int |\mathbf{a}|^N |f(t, \mathbf{a})|\, d^3\mathbf{a} \leq C(1 + |t|)^{N+3/2}$$

(see, e.g., Lemma 6.5 in Ref. 1). Combined with (2.5), this proves (ii). ∎

3. Q ON SCATTERING STATES

In the Haag–Ruelle scattering theory it is shown that states of the form (2.3) converge strongly for $t \to \mp\infty$,

$$\text{s-}\lim_{t \to \mp\infty} B_{\nu_1}^{f_1,t} \cdots B_{\nu_n}^{f_n,t}\Omega =: \varphi^{\text{in}}_{\text{out}}{}_{\nu_1 \cdots \nu_n}$$

to asymptotic n-particle states with wavefunctions determined by $f_1 \cdots f_n$. The convergence is faster than any inverse power of t if the $f_1 \cdots f_n$ have pairwise disjoint support in velocity space.[3] The sets of these limits, $\tilde{D}^{\text{in}}$, $\tilde{D}^{\text{out}}$, are dense in the asymptotic Hilbert spaces.

3.1. *Theorem*: $\tilde{D}^{\text{in}}_{\text{out}} \subset \mathfrak{H}_{\bar{Q}}$.

Proof: We adopt the proof given in Ref. 3 for enlarging the domain of Wightman fields to $\tilde{D}^{\text{in}}_{\text{out}}$. Consider

$$\psi(t) := B_{\nu_1}^{f_1,t} \cdots B_{\nu_n}^{f_n,t}\Omega$$

and $d\psi(t)/dt$ which is a linear combination of expressions of the same type. $\psi(t)$ and $d\psi(t)/dt$ are in $\mathfrak{H}_{\bar{Q}}$ and $\mathfrak{H}_{Q^*\bar{Q}}$ by Lemma 2.3,

$$\begin{aligned}\left\|\bar{Q}\frac{d\psi}{dt}\right\|^2 &= \left(\frac{d\psi}{dt} \,\middle|\, (\bar{Q})^* \bar{Q}\frac{d\psi}{dt}\right) \\ &\leq \left\|\frac{d\psi(t)}{dt}\right\| \left\|Q^*\bar{Q}\frac{d\psi(t)}{dt}\right\|\end{aligned}$$

The first factor decreases faster than any inverse power of t for states with pairwise disjoint support in velocity space as a result of scattering theory. Since the second factor is polynomially bounded by Lemma 2.3, $\|\bar{Q}(d\psi/dt)\|$ is integrable. Hence $\bar{Q}(t)$ is a Cauchy sequence for $t \to +\infty$ as well as for $t \to -\infty$. Since $\bar{Q}$ is closed, the statement follows. ∎

Now we can show

3.2 *Theorem*: $\bar{Q}S = S\bar{Q}$ on $\tilde{D}^{\text{out}} \cup \tilde{D}^{\text{in}}$.

Proof: In Ref. 1 Q was extended as a form to two forms on $\tilde{D}^{\text{in}} \times \tilde{D}^{\text{in}}$ and $\tilde{D}^{\text{out}} \times \tilde{D}^{\text{out}}$. These forms are given by operators Q^{in}, Q^{out} leaving $\tilde{D}^{\text{in}}$, $\tilde{D}^{\text{out}}$ invariant. They obey

$$Q^{\text{in}} S = S Q^{\text{out}}$$

on $\tilde{D}^{\text{out}}$. [See (7.2) of Ref. 1.] From (3.1) it follows that Q^{in}, Q^{out} coincide with $\bar{Q}$ on $\tilde{D}^{\text{in}}$, $\tilde{D}^{\text{out}}$. Hence $\bar{Q}$ leaves $\tilde{D}^{\text{in}}$, $\tilde{D}^{\text{out}}$ invariant and

$$\bar{Q}S = S\bar{Q}$$

on $\tilde{D}^{\text{out}}$. The statement for $\tilde{D}^{\text{in}}$ follows by starting from $S^{-1}Q^{\text{in}} = Q^{\text{out}}S^{-1}$. ∎

We remark that the results of Ref. 1. for $Q^{\text{in}}_{\text{out}}$ hold for $\bar{Q}$, too.

[1] W.D. Garber and H. Reeh, "Non-translationally covariant currents and associated symmetry generators," J. Math. Phys. (in press).

[2] T. Kato, *Perturbation Theory for Linear Operators* (Springer, Berlin, 1966).

[3] K. Hepp, "On the connection between Wightman and LSZ quantum field theory," in *Axiomatic Field Theory, Brandeis University Summer Institute* 1965, Vol. I, edited by M. Chretien and S. Deser (Gordon and Breach, New York, 1966).

Electromagnetic solutions of Brans–Dicke theory of gravitation from Einstein theory

V. B. Johri and G. K. Goswami

Department of Mathematics, University of Gorakhpur, Gorakhpur-273001, India
(Received 18 July 1977)

A class of static and nonstatic solutions of the Brans–Dicke theory of gravitation is obtained in the presence of an electromagnetic field. The metric coefficients and fields (both scalar and electromagnetic) are supposed to be functions of any three independent variables. The major result of the paper may be stated as follows: "Corresponding to any diagonalizable solution of Einstein's vacuum field equations in which fields and metric coefficients are functions of not more than three variables, we can generate a solution of the coupled Brans–Dicke Maxwell field equations with nonzero electromagnetic field."

1. INTRODUCTION

This work is a continuation of our previous study[1] of the BD theory[2] of gravitation in which we have obtained a class of static and nonstatic solution of the BD vacuum field equations by transforming them into the Einstein vacuum field equations. Here we have considered the energy-momentum tensor due to source-free electromagnetic field, and have obtained a class of static and nonstatic solution of the coupled BD– Maxwell field equations by transforming them into Einstein vacuum field equations. The metric coefficients and fields (both electromagnetic and scalar) are taken to be functions of not more than three variables. Such type of work has been done by B. Kent Harrison[3] in which he has obtained solutions of the Einstein–Maxwell field equations by transforming them into the Einstein vacuum field equations. Recently Tiwari and Nayak[4] have also obtained a class of solutions of the BD electrostatic field equations by reducing them into the Einstein electrostatic field equations but here we have considered both static and stationary BD–Maxwell fields.

In Sec. 2, we have set up the BD–Maxwell field equations in a suitable form by assuming a functional relationship amongst g_{33}, scalar field ϕ and potential C; then we have established the main result of the paper. In Sec. 3, BD solutions are obtained corresponding to some well-known solution of Einstein theory. The last section contains some concluding remarks.

2. DERIVATION OF BD-MAXWELL FIELDS FROM EINSTEIN VACUUM FIELD

The BD field equations for source free electromagnetic field are

$$R_{ij} = -\frac{8\pi}{\phi}T_{ij} - \frac{\omega}{\phi^2}\phi_{,i}\phi_{,j} - \frac{\phi_{;ij}}{\phi}, \tag{1}$$

$$\phi^{k}_{;k} = 0 \quad (\omega \neq -\tfrac{3}{2}), \tag{2}$$

$$[ijkl]\frac{\partial}{\partial x^l}F_{jk} = 0, \tag{3}$$

and

$$\frac{\partial}{\partial x^k}[\sqrt{-g}\,g^{ij}g^{kl}F_{jl}] = 0 \tag{4}$$

with

$$T_{ij} = (F_{il}F^{l}_{j} - \tfrac{1}{4}g_{ij}F_{lm}F^{lm}) \tag{5}$$

and

$[ijkl] = (+1, -1)$ for (even, odd) permutation of i, j, k, l,

$= 0$ if any two of i, j, k, l are equal.

Where R_{ij} is the Ricci tensor, T_{ij} is the energy momentum tensor due to source-free electromagnetic field, F_{ij} is the electromagnetic field tensor, ϕ is the scalar field, and ω is the coupling constant.

If we put $\phi = e^{\lambda}$, (6)

then Eq. (1) is transformed to

$$R_{ij} = -(8\pi/e^{\lambda})T_{ij} - (\omega+1)\lambda_{,i}\lambda_{,j} - \lambda_{i;j}. \tag{7}$$

Now we consider a general line element

$$ds^2 = \exp(2V)(dx^3)^2 - \exp[-2(V+EU)][e_\alpha \Gamma_{\alpha\alpha}(dx^\alpha)^2], \tag{8}$$

where α, β, γ take the value 0, 1, 2. $\Gamma_{\alpha\beta} = 0$ for $\alpha \neq \beta$, $e_0 = -1, e_1 = e_2 = 1$ and U, V, and $\Gamma_{\alpha\alpha}$ are functions of only three variables x^0, x^1, and x^2, and E is an arbitrary constant.

The metric coefficients being independent of coordinate x^3, Eqs. (3) and (4) can be easily satisfied with the help of potentials as shown by Harrison. We define A, B, and C by

$$F^{\alpha\beta} = \sqrt{-g}\,\epsilon^{\alpha\beta\gamma}A_{,\gamma}, \tag{9}$$

$$F_{4\alpha} = B_{,\alpha}, \tag{10}$$

and C is connected to A and B by

$$A = C\cos D, \quad B = C\sin D, \tag{11}$$

D being a constant and $\epsilon^{\alpha\beta\gamma}$ being an alternating three-index symbol.

The transformation (11) may be looked upon as the duality rotation of Misner and Wheeler (1957). If the fields were independent of x^0 instead of x^3, B would be the electric potential and A would be the magnetic potential. We may also take the electromagnetic field to be purely electric or magnetic by choosing B or A to be zero, respectively.

Computing the components of energy-momentum tensor T_{ij} and Ricci tensor R_{ij} with the help of metric (8)

$$T_{\alpha\beta} = \exp(-2V)(A_{,\alpha}B_{,\beta} + A_{,\beta}B_{,\alpha}) = \exp(-2V)C_{,\alpha}C_{,\beta}, \quad (\alpha \neq \beta) \tag{12}$$

 0022-2488/78/1905-0987$1.00

$$T_{\alpha\alpha} = \exp(-2V)[(A^2_{,\alpha} + B^2_{,\alpha}) - \tfrac{1}{2}e_\alpha e_\beta \Gamma_{\alpha\alpha}\Gamma^{\beta\beta}(A^2_{,\beta} + B^2_{,\beta})]$$
$$= \exp(-2V)[C^2_{,\alpha} - \tfrac{1}{2}e_\alpha e_\beta \Gamma_{\alpha\alpha}\Gamma^{\beta\beta}C^2_\beta], \quad (13)$$

$$T_{33} = \exp[2(V+EU)]e_\alpha \Gamma^{\alpha\alpha}(A^2_{,\alpha} + B^2_{,\alpha})$$
$$= \exp[2(V+EU)]e_\alpha \Gamma^{\alpha\alpha}C^2_{,\alpha}, \quad (14)$$

$$R_{\alpha\beta} = P_{\alpha\beta} - EU_{\alpha;\beta} + 2V_{,\alpha}V_{,\beta} - E^2U_{,\alpha}U_{,\beta}, \quad (15)$$
$(\alpha \neq \beta)$

$$R_{\alpha\alpha} = P_{\alpha\alpha} - e_\alpha \Gamma_{\alpha\alpha}(V^\beta_{;\beta} + EU^\beta_{;\beta}) - EU_{\alpha;\alpha} + 2V^2_{,\alpha}$$
$$- E^2U^2_{,\alpha} + E^2e_\alpha \Gamma_{\alpha\alpha}\nabla_1(U) + Ee_\alpha \Gamma_{\alpha\alpha}\nabla_1(U, V), \quad (16)$$

$$R_{33} = \exp(4V + 2EU)[V^\alpha_{;\alpha} - E\nabla_1(U, V)], \quad (17)$$

where

$$\nabla_1(U, V) = e_\alpha \Gamma^{\alpha\alpha}U_{,\alpha}V_{,\alpha}, \quad (18)$$

$$\nabla_1(U) = e_\alpha \Gamma^{\alpha\alpha}U^2_{,\alpha}. \quad (19)$$

$P_{\alpha\beta}$ is the Ricci tensor formed with respect to the three metric $\Gamma_{\alpha\beta}$ and covariant deviatives are also taken with respect to the three metric $\Gamma_{\alpha\beta}$.

Therefore the field equations (1) to (4) become

$$P_{\alpha\beta} - EU_{\alpha;\beta} + 2V_{,\alpha}V_{,\beta} - E^2U_{,\alpha}U_{,\beta}$$
$$= -(8\pi/e^{2v}e^\lambda)C_{,\alpha}C_{,\beta} - (\omega+1)\lambda_{,\alpha}\lambda_{,\beta} - \lambda_{\alpha;\beta}$$
$$- E(\lambda_{,\alpha}U_{,\beta} + \lambda_{,\beta}U_{,\alpha}) - (\lambda_{,\alpha}V_{,\beta} + \lambda_{,\beta}V_{,\alpha}), \quad (20)$$

$$P_{\alpha\alpha} - e_\alpha \Gamma_{\alpha\alpha}(V^\beta_{;\beta} + EU^\beta_{;\beta}) - EU_{\alpha;\alpha} + 2V^2_{,\alpha} - E^2U^2_{,\alpha}$$
$$+ E^2e_\alpha \Gamma_{\alpha\alpha}\nabla_1(U) + Ee_\alpha \Gamma_{\alpha\alpha}\nabla_1(U, V)$$
$$= -\frac{8\pi}{e^{2V}e^\lambda}(C^2_{,\alpha} - \tfrac{1}{2}e_\alpha e_\beta \Gamma_{\alpha\alpha}\Gamma^{\beta\beta}C^2_{,\beta}) - (\omega+1)\lambda^2_{,\alpha}$$
$$- \lambda_{\alpha;\alpha} - E(2\lambda_{,\alpha}U_{,\alpha} - e_\alpha \Gamma_{\alpha\alpha}e_\beta \Gamma^{\beta\beta}U_{,\beta}\lambda_{,\beta})$$
$$- (2\lambda_{,\alpha}V_{,\alpha} - e_\alpha e_\beta \Gamma_{\alpha\alpha}\Gamma^{\beta\beta}V_{,\alpha}\lambda_{,\alpha}), \quad (21)$$

$$V^\alpha_{;\alpha} - E\nabla_1(U, V) = -(4\pi/e^{2V}e^\lambda)\nabla_1(C) - \nabla_1(V, \lambda), \quad (22)$$

with

$$\lambda^\alpha_{;\alpha} = -\nabla_1(\lambda) + E\nabla_1(U, \lambda) \quad (23)$$

and

$$C^\alpha_{;\alpha} = E\nabla_1(U, C) + 2\nabla_1(V, C). \quad (24)$$

Now we assume the functional relationship amongst $C, \phi, U,$ and V as

$$V = V(\phi, C), \quad U = U(\phi)$$

then it can be proved with the help of Eqs. (22), (23), and (24) that

$$e^{2V} = -\phi^{-1}(4\pi C^2 + GC + H) \quad (25)$$

$$\lambda = EU, \quad (26)$$

where G and H are constants.

Therefore adjusting constants we may write

$$\exp(2V) = -\exp(-\lambda)4\pi(C+L)^2 = -4\pi\exp(2\mu - \lambda)$$
$$= -4\pi\exp(2\mu - EU), \quad (27)$$

where

$$C + L = e^\mu \quad (28)$$

and L is a constant.

Now applying transformations (26) and (27) in Eqs. (22), (23), and (24), we get

$$\mu^\alpha_{;\alpha} - EU^\alpha_{;\alpha} = \nabla_1(\mu) \quad (29)$$

$$\mu^\alpha_{;\alpha} = \nabla_1(\mu) \quad (30)$$

which combine together to give

$$U^\alpha_{;\alpha} = 0. \quad (31)$$

Equations (20) and (21) reduce to

$$P_{\alpha\beta} + 2F^2U_{,\alpha}U_\beta = 0 \quad (\alpha \neq \beta) \quad (32)$$
$$P_{\alpha\alpha} + 2F^2U^2_{,\alpha} = 0,$$

where

$$F^2 = (E^2/4)(3 + 2\omega). \quad (33)$$

Putting

$$W = FU, \quad (34)$$

Eqs. (31) and (32) take the form

$$W^\alpha_{;\alpha} = 0 \quad (35)$$
$$P_{\alpha\alpha} + 2W^2_{,\alpha} = 0$$
$$P_{\alpha\beta} + 2W_{,\alpha}W_{,\beta} = 0 \quad (36)$$

which are Einstein vacuum field equations for the metric

$$ds^2 = \exp(2W)(dx^3)^2 + \exp(-2W)(e_\alpha \Gamma_{\alpha\alpha}(dx^\alpha)^2). \quad (37)$$

Thus we have established the following result:

Corresponding to any diagonalizable solution of the Einstein vacuum field equations in which fields and metric coefficients are functions of not more than three variables, we can generate a solution of the coupled BD—Maxwell field equations with nonzero electromagnetic field.

Mathematically, suppose the metric

$$ds^2 = \exp(2W)(dx^3)^2 + \exp(-2W)[e_\alpha \Gamma_{\alpha\alpha}(dx^\alpha)^2]$$

with $\Gamma_{\alpha\alpha}$ and W as functions of $x^0, x^1,$ and x^2 satisfies Einstein's vacuum field equations; then the metric

$$ds^2 = \exp(2\mu - EW/F)(dx^3)^2$$
$$+ \exp(-2\mu - EW/F)[e_\alpha \Gamma_{\alpha\alpha}(dx^\alpha)^2] \quad (38)$$

will satisfy the BD—Maxwell field equations with scalar field given by

$$\phi = \exp(EW/F), \quad E, F \text{ arbitrary constants} \quad (39)$$

and μ related to potential C through Eq. (28). The potential C may be obtained from Harrison's result

$$C = K[\exp(2W) - 1]/[\exp(2W) + 1], \quad (40)$$

k being constant.

The metric (37) shows that the solution obtained here is conformal to the metric

$$ds^2 = \exp(2\mu)(dx^3)^2 + \exp(-2\mu)[e_\alpha \Gamma_{\alpha\alpha}(dx^\alpha)^2]. \quad (41)$$

Through the conformal transformation

$$g_{\mu\nu(BD)} = \phi^{-1}g_{\mu\nu}. \quad (42)$$

3. SOME PARTICULAR SOLUTIONS CORRESPONDING TO WELL-KNOWN VACUUM SOLUTIONS OF EINSTEIN'S THEORY

In this section we have applied the result obtained in the previous section to some well known static and stationary vacuum solutions of Einstein's theory viz. the static plane symmetric solution of Taub,[5] the conformastat solution of Das,[6] and the nonstatic solution of Mishra and RadhaKrishna[7] of type II according to Pirani's criterion.[8]

A. Static solutions

I. Static plane symmetric solution

The static plane symmetric solution due to Taub is given by the metric

$$\begin{aligned} ds^2 &= -(k_1x+k_2)^{-1/2}(dx^2-dt^2)-(k_1x+k_2)(dy^2+dz^2) \\ &= (k_1x+k_2)^{-1/2}\,dt^2-(k_1x+k_2)^{1/2}[(k_1x+k_2)^{-1}\,dx^2 \\ &\quad +(k_1x+k_2)^{1/2}(dy^2+dz^2)]. \end{aligned} \tag{43}$$

The corresponding BD—Maxwell solution will be

$$g_{00}=\left\{k\left[\frac{(k_1x+k_2)^{-1/2}-1}{(k_1x+k_2)^{-1/2}+1}\right]+L\right\}^2(k_1x+k_2)^{E/4F},$$

$$g_{11}=\left\{k\left[\frac{(k_1x+k_2)^{-1/2}-1}{(k_1x+k_2)^{-1/2}+1}\right]+L\right\}^{-2}(k_1x+k_2)^{E/4F-1},$$

$$g_{22}=g_{33}=\left\{k\left[\frac{(k_1x+k_2)^{-1/2}-1}{(k_1x+k_2)^{-1/2}+1}\right]+L\right\}^{-2}(k_1x+k_2)^{E/4F+1/2}, \tag{44}$$

with scalar field defined by

$$\phi=(k_1x+k_2)^{-E/4F} \tag{45}$$

$$\text{electric potential } B=\left\{k\left[\frac{(k_1x+k_2)^{-1/2}-1}{(k_1x+k_2)^{-1/2}+1}\right]\sin D\right\} \tag{46}$$

$$\text{magnetic potential } A=\left\{k\left[\frac{(k_1x+k_2)^{-1/2}-1}{(k_1x+k_2)^{-1/2}+1}\right]\cos D\right\}.$$

This solution is due to an infnite charged plane parallel to the (y,z) plane. Like metric (43) this solution also has a singularity at $x=-k_2/k_1$. Thus we see that there is a one-to-one correspondence between the singularities of the Einstein vacuum solution and the BD—Maxwell solutions in this case.

II. Conformastat solution

The empty space conformastat solution of Das is given by the metric

$$ds^2=(1-mx)^{-2}\,dt^2-(1-mx)^2[(1-mx)^2(dx^2+dy^2+dz^2)]. \tag{47}$$

This is due to the gravitational field of an infinite plane parallel to the (y,z) plane. This metric has a singularity at $x=1/m$ where $m=$ constant. The corresponding solution of BD—Maxwell field equations is given by

$$g_{00}=\left\{k\left[\frac{(1-mx)^{-2}-1}{(1-mx)^{-2}+1}\right]+L\right\}^2(1-mx)^{E/2F} \tag{48}$$

$$g_{11}=g_{22}=g_{33}=\left\{k\left[\frac{(1-mx)^{-2}-1}{(1-mx)^{-2}+1}\right]+L\right\}^{-2}(1-mx)^{E/2F-2},$$

with scalar field

$$\phi=(1-mx)^{-E/2F} \tag{49}$$

$$\text{magnetic potential } A=\left\{k\left[\frac{(1-mx)^{-2}-1}{(1-mx)^{-2}+1}\right]\cos D\right\} \tag{50}$$

and

$$\text{electric potential } B=\left\{k\left[\frac{(1-mx)^{-2}-1}{(1-mx)^{-2}+1}\right]\sin D\right\}.$$

This solution also has a singularity at $x=1/m$. Thus there is again one-to-one correspondence between the singularities of metric (47) and solution (48).

B. Stationary solutions

I. Solution of type II according to Pirani's criterion

Let us consider a stationary metric of type II according to Pirani's criterion

$$\begin{aligned} ds^2 &= \exp[(m^2r^2/4+mt)](dt^2-dr^2)-r^2\exp(mt)d\phi^2 \\ &\quad -\exp(-mt)\,dz^2 \\ &= -\exp(-mt)\,dz^2-\exp(mt)[\exp(m^2r^2/4)\,dr^2 \\ &\quad +r^2d\phi^2-\exp(m^2r^2/4)\,dt^2], \end{aligned} \tag{51}$$

where m is a constant.

This metric was obtained by Mishra and Radha-Krishna. The corresponding solution of the BD—Maxwell field equation will be

$$g_{33}=\left\{k\left[\frac{[\exp(-mt)-1]}{[\exp(-mt)+1]}\right]+L\right\}^2\exp(E/4Fmt)$$

$$g_{11}=-g_{00}=\left\{k\left[\frac{[\exp(-mt)-1]}{[\exp(-mt)+1]}\right]+L\right\}^{-2}\exp(E/4Fmt+m^2r^2/4),$$

$$g_{22}=\left\{k\left[\frac{[\exp(-mt)-1]}{[\exp(-mt)+1]}\right]+L\right\}^{-2}r^2\exp(E/4Fmt), \tag{52}$$

with ϕ given by

$$\phi=\exp(-E/4Fmt) \tag{53}$$

and

$$\text{potential } C=\left\{k\left[\frac{[\exp(-mt)-1]}{[\exp(-mt)+1]}\right]\right\}. \tag{54}$$

C. Conclusion

The immediate use of the result derived in this paper is in obtaining exact solutions of the BD— Maxwell field equations, which are otherwise quite intricate, from the known solutions of Einstein vacuum field equations.

Besides, these solutions furnish examples of singularities occurring in the BD theory. It is found that there is a one-to-one correspondence between the singularities of Einstein vacuum solutions and BD—Maxwell solutions in the cases we have examined.

In conclusion, we at least hope that our investigations will lead to deeper understanding of the Einstein and BD theory. We also hope that some physical insight can be gained from the solution obtained in this paper.

ACKNOWLEDGMENTS

The authors are grateful to Professor K.B. Lal for his kind encouragement and interest in their work. Thanks are due to Sri S.K. Srivastava for some useful discussions and suggestions in this paper. This work was done under the U.G.C. research scheme.

[1]G.K. Goswami, "Some static and nonstatic solutions of Brans—Dicke theory of gravitation," J. Math. Phys. (to be published).
[2]C. Brans and R.H. Dicke, Phys. Rev. 124, 925 (1961).
[3]B. Kent Harrison, Phys.Rev. 138, B488 (1965).
[4]R.N. Tiwari and B.K. Nayak, Phys. Rev. D 14, 2502 (1976).
[5]A.H. Taub, Ann. Math. 53, 472 (1951).
[6]A. Das, J. Math. Phys. 12, 1136 (1971).
[7]M. Mishra and L. Radhakrishna, Proc. Natl. Inst. Sci. (India) A 28, 632 (1962).
[8]F.A.E. Pirani, Phys. Rev. 105, 1089 (1957).

Some constraints on finite energy solutions in non-Abelian gauge theories

M. Magg

Institut für Theoretische Physik, Rheinisch-Westfälische Technische Hochschule, Aachen, Federal Republic of Germany
(Received 24 August 1977)

We exploit dilatational invariance and some inequalities for the stress tensor to derive constraints for finite energy solutions in sourceless non-Abelian gauge theories. The results extend known no-go theorems considerably and provide some hints on what nondissipative finite energy solutions could look like.

In classical sourceless electrodynamics the complete set of finite energy solutions is given by the vacuum and the square integrable wavepackets. For non-Abelian gauge theories our knowledge on classical finite energy solutions is much poorer.[1] This ignorance can be a serious drawback if one is looking for a treatment of the quantized theory which goes beyond ordinary perturbation expansion.

In this paper we use dilatational and conformal invariance of sourceless non-Abelian gauge theories, supplemented by some inequalities for the gauge invariant, symmetric stress tensor to work out general constraints on finite energy solutions in physical Minkowski space. Although these conditions are not strong enough to settle the question of genuine non-Abelian finite energy solutions, they extend known no-go theorems[3,4] and may provide useful hints where to search for them.

The theory we are delaing with in this paper is the sourceless Yang—Mills theory in $(1+3)$-dimensional Minkowski space with a given compact gauge group. The gauge invariant, symmetric stress tensor for such a theory can be written in the form

$$\theta^{\mu\nu}=\sum_a [-F^{a\mu}{}_{\kappa}F^{a\nu\kappa}+\tfrac{1}{4}g^{\mu\nu}F^{a\kappa\lambda}F^{a}{}_{\kappa\lambda}]. \tag{1}$$

Here $F^{a\mu\nu}$ denotes the Yang—Mills field strength tensor. The latin index a labels the members of the adjoint representation of the gauge group algebra.

The stress tensor is traceless, which simply reflects dilatational and conformal invariance of the theory,

$$\theta^{\mu}{}_{\mu}(\mathbf{x},t)=0. \tag{2}$$

From expression (1) one can read off some inequalities which are important to us:

$$\theta^{00}\geq 0, \tag{3a}$$

$$\theta^{00}\geq |n^i\theta^{0i}| \quad (|\mathbf{n}|=1), \tag{3b}$$

$$\theta^{00}\geq |n^i\theta^{ij}n^j|. \tag{3c}$$

Vanishing energy density θ^{00} means vanishing field-strength tensor and this implies that the potentials vanish in some gauge.[3]

For any solution of the equations of motion the 4-momentum is locally conserved,

$$\partial_t\theta^{0\nu}(\mathbf{x},t)+\frac{\partial}{\partial x^i}\theta^{i\nu}(\mathbf{x},t)=0. \tag{4}$$

We restrict ourselves to finite energy solutions,

$$\int_{\mathbb{R}^3}dV\,\theta^{00}(\mathbf{x},t)<\infty \quad (t\in\mathbb{R}). \tag{5}$$

Without any further assumption on surface terms, Eqs. (3), (4), and (5) together imply the time independence of the total energy (5). The relations (2)—(5) form the starting point for our considerations. Contraction of the continuity equation (4) with the tensor $g_{\nu i}x^i$ and integration over a finite time interval (t_1,t_2) and a compact space region G leads to the virial formula,

$$\begin{aligned}&\int_G dV\int_{t_1}^{t_2}dt\,\theta^{00}(\mathbf{x},t)\\&\quad=\int_G dV\,x^j[\theta^{0j}(\mathbf{x},t_2)-\theta^{0j}(\mathbf{x},t_1)]\\&\qquad+\int_{\partial G}df^i\int_{t_1}^{t_2}dt\,x^j\theta^{ij}(\mathbf{x},t).\end{aligned} \tag{6}$$

We now evaluate Eq. (6) in the limit $(t_2-t_1)\to\infty$. The result is the following:

Theorem: For any finite energy solution of the equations of motion the energy density satisfies

$$\lim_{T\to\infty}\frac{1}{T}\int_{t_0}^{t_0+T}dt\,\theta^{00}(\mathbf{x},t)=0 \quad (\mathbf{x}\in\mathbb{R}^3). \tag{7}$$

This theorem has some immediate applications.

Corollary: For any finite energy solution of the equations of motion the energy density vanishes for large t in the sense

$$\lim_{t\to\infty}\theta^{00}(\mathbf{x},t)=0 \quad (\mathbf{x}\in\mathbb{R}^3).$$

Corollary: The vacuum is the unique finite energy solution with an almost periodic time dependence of the energy density,

The proof of the second corollary uses some elementary facts on almost periodic functions. For such solutions $\lim$ in Eq. (7) can be replaced by $\lim$. Because of positivity (3a) $\theta^{00}(\mathbf{x},t)$ has to vanish.[5] This means the solution was the vacuum. The last corollary covers the cases of static and ordinary periodic solutions.

Now we prove the theorem. Inequality (3b) provides us with a time independent bound for the volume integral on the right-hand side of Eq. (6),

$$\begin{aligned}&\left|\int_G dV\,x^j[\theta^{0j}(\mathbf{x},t_2)-\theta^{0j}(\mathbf{x},t_1)]\right|\\&\quad\leq 2\sup_{\mathbf{x}\in G}|\mathbf{x}|\int_{\mathbb{R}^3}dV\,\theta^{00}(\mathbf{x},t).\end{aligned}$$

The time average of Eq. (6) therefore becomes

 0022-2488/78/1905-0991$1.00

$$\lim_{T\to\infty}\left[\int_G dV\,\frac{1}{T}\int_{t_0}^{t_0+T} dt\,\theta^{00}(\mathbf{x},t) - \int_{\partial G} df^i\,\frac{1}{T}\int_{t_0}^{t_0+T} dt\,x^j\theta^{ij}(\mathbf{x},t)\right]=0. \tag{8}$$

It is not guaranteed that one can interchange the time limit with the space integration. Nevertheless we can conclude from "Fatou's lemma" that

$$\int_G dV \lim_{T\to\infty}\frac{1}{T}\int_{t_0}^{t_0+T} dt\,\theta^{00}(\mathbf{x},t) \le \lim_{T\to\infty}\int_{\partial G} df^i\,\frac{1}{T}\int_{t_0}^{t_0+T} dt\,x^j\theta^{ij}(\mathbf{x},t). \tag{9}$$

For the compact region G we now choose a ball with center at the origin and radius r. Then we use (3c) and continue inequality (9),

$$\lim_{T\to\infty} r^3\int_{|\mathbf{x}|=r} d\Omega\,\frac{1}{T}\int_{t_0}^{t_0+T} dt\,\frac{x^j}{r}\,\theta^{ij}(\mathbf{x},t)\,\frac{x^i}{r} \le r^3\lim_{T\to\infty}\int_{|\mathbf{x}|=r} d\Omega\,\frac{1}{T}\int_{t_0}^{t_0+T} dt\,\theta^{00}(\mathbf{x},t). \tag{10}$$

We apply "Fatou's lemma" once more in order to show that

$$r^2\lim_{T\to\infty}\int_{|\mathbf{x}|=r} d\Omega\,\frac{1}{T}\int_{t_0}^{t_0+T} dt\,\theta^{00}(\mathbf{x},t)$$

is an integrable function of r,

$$\int_0^\infty dr\,r^2\lim_{T\to\infty}\int_{|\mathbf{x}|=r} d\Omega\,\frac{1}{T}\int_{t_0}^{t_0+T} dt\,\theta^{00}(\mathbf{x},t) \le \lim_{T\to\infty}\frac{1}{T}\int_{t_0}^{t_0+T} dt\int_{\mathbb{R}^3} dV\,\theta^{00}(\mathbf{x},t) = \int_{\mathbb{R}^3} dV\,\theta^{00}(\mathbf{x},t).$$

Because $\lim_{r\to\infty} rf(r)$ vanishes for any integrable function $f(r)$ one gets the relation

$$\lim_{n\to\infty}\left[r_n^3\lim_{T\to\infty}\int_{|\mathbf{x}|=r_n} d\Omega\,\frac{1}{T}\int_{t_0}^{t_0+T} dt\,\theta^{00}(\mathbf{x},t)\right]=0 \tag{11}$$

for some sequence $r_n\to\infty$.
If one reads Eqs. (9), (10), and (11) in one line, one obtains

$$\int_{\mathbb{R}^3} dV \lim_{T\to\infty}\frac{1}{T}\int_{t_0}^{t_0+T} dt\,\theta^{00}(\mathbf{x},t)=0.$$

Because of (3a) this proves the theorem.

A complementary way to look at Eq. (6) is to hold the time interval (t_1,t_2) fixed and to extend the compact region to the whole space $\mathbb{R}^3$. The resulting formula expresses the time independence of the dilatational charge.

Theorem: For any finite energy solution of the equations of motion there exists a sequence $r_n>0$ with $r_n\xrightarrow[n\to\infty]{}\infty$ such that

$$\int_{\mathbb{R}^3} dV\,\theta^{00}(\mathbf{x},t) = \lim_{n\to\infty}\left\{\int_{|\mathbf{x}|\le r_n} dV\,\frac{x^j}{T}\left[\theta^{0j}(\mathbf{x},t_0+T)-\theta^{0j}(\mathbf{x},t_0)\right]\right\}. \tag{12}$$

Proof: Due to (3c) the surface term in Eq. (6) is bounded by

$$\left|\int_{|\mathbf{x}|=r} df^i\int_{t_1}^{t_2} dt\,x^j\theta^{ij}(\mathbf{x},t)\right| \le r^3\int_{|\mathbf{x}|=r} d\Omega\int_{t_1}^{t_2} dt\,\theta^{00}(\mathbf{x},t).$$

Because of

$$\int_0^\infty dr\,r^2\int_{|\mathbf{x}|=r} d\Omega\int_{t_1}^{t_2} dt\,\theta^{00}(\mathbf{x},t)<\infty,$$

an argument similar to that used in the previous proof shows the existence of the wanted sequence $r_n\to\infty$ with

$$\lim_{n\to\infty}\left[r_n^3\int_{|\mathbf{x}|=r_n} d\Omega\int_{t_1}^{t_2} dt\,x^i\theta^{ij}(\mathbf{x},t)\right]=0.$$

Corollary: For any finite energy solution of the equations of motions which satisfy in addition

$$\int_{\mathbb{R}^3} dV\,x^i\theta^{0i}(\mathbf{x},t)<\infty \quad (t\in\mathbb{R}),$$

one has the following relation for the total energy

$$\int_{\mathbb{R}^3} dV\,\theta^{00}(\mathbf{x},t)=\lim_{T\to\infty}\int_{\mathbb{R}^3} dV\,\frac{x^i}{T}\,\theta^{0i}(\mathbf{x},t_0+T). \tag{13}$$

This corollary can help to decide whether a given ansatz for a finite energy solution has a chance or not. For example we look for "lumplike" solution.[3] Let us assume that the lump has something like a center moving on a trajectory $\mathbf{y}(t)$ such that, relative to this center, the energy density dissipates only weakly. In more precise terms: let us assume that there is an integrable function $f(\mathbf{x})$ with a finite first moment which bounds the energy density in the following way,

$$\theta^{00}(\mathbf{x},t)\le f(\mathbf{x}-\mathbf{y}(t))(1+t) \quad (t>0). \tag{14}$$

With Lebesque's theorem on dominated convergence one can now calculate the limit in Eq. (13). The result is

$$\int_{\mathbb{R}^3} dV\,\theta^{00}(\mathbf{x},t)=\left[\int_{\mathbb{R}^3} dV\,\theta^{0i}(\mathbf{x},t)\right]\cdot\lim_{T\to\infty}\dot{y}^i(T) \tag{15}$$

Because of inequality (3b) the total 4-momentum can be timelike or lightlike. For timelike 4-momentum the solution has a rest frame and in this frame it is obvious that only the vacuum satisfies Eq. (15). The interesting case is lightlike 4-momentum (which is impossible for Abelian finite energy solutions). The center of the lump then moves asymptotically with the speed of light in the direction of the total momentum.

We therefore reach the following alternative for finite energy solutions of the sourceless Yang–Mills equations. Either the energy density dissipates more than weakly [in the sense of Eq. (14)] or the total 4-momentum is lightlike. Unfortunately our definition of "weakly dissipative" does not exactly fit with what Coleman calls a dissipative solution.[3]

A hint in the same direction comes from our first corollary. Assume $\underline{\lim}_{T\to\infty}[\sup_{\mathbf{x}\in\mathbb{R}^3}\theta^{00}(\mathbf{x},t)]$ is not zero. Then there is a constant $c>0$ and for every time t (t sufficiently large) there is at least one point $\mathbf{x}_0(t)$ such that $\theta^{00}(\mathbf{x}_0(t),t)>c$. The corollary tells us it is impossible to find a Lorentz frame where this point is at rest.

If one is willing to make stronger assumptions on the energy density, one can derive identities which involve second moments, e.g.,

$$\int_{\mathbb{R}^3} dV\,\theta^{00}(\mathbf{x},t) = \lim_{T\to\infty}\int_{\mathbb{R}^3} dV\,\frac{|\mathbf{x}|^2}{T^2}\,\theta^{00}(\mathbf{x},t_0+T).$$

We think that the essential points are already contained in the equations we have used.

ACKNOWLEDGMENTS

I want to thank Professor H. A. Kastrup, M. Kiera, and M. Rinke for several valuable discussions.

Note added in proof: After completion of this paper we received three preprints[6–8] containing similar results. In the first of these papers the possibility of lightlike lumps is eliminated.

[1]After completion of this paper we received a preprint by M. Lüscher which contains a nontrivial class of such solutions.[2]
[2]M. Lüscher, "SO(4)-Symmetric Solutions of Minkowskian Yang–Mills Field Equations," DESY preprint 77/32 (1977).
[3]S. Coleman, Lectures delivered at the 1975 International School of Subnuclear Physics "Ettore Majorana."
[4]S. Deser, Phys. Lett. B **64**, 463 (1976); H. Pagels, Phys. Lett. B **68**, 466 (1977).
[5]N. Dunford and J. T. Schwartz, *Linear Operators, Part I, General Theory* (Interscience, New York, 1958).
[6]S. Coleman, Commun. Math. Phys. **55**, 113 (1977).
[7]S. Coleman and L. Smarr, Commun. Math. Phys. **56**, 1 (1977).
[8]R. Weder, "Absence of Classical Lumps," Harvard Preprint (1977).

On the inverse problem of transport theory with azimuthal dependence

N. J. McCormick and J. A. R. Veeder

Department of Nuclear Engineering, University of Washington, Seattle, Washington 98195
(Received 2 September 1977)

The infinite medium inverse problem with an azimuthally dependent plane source leads to integral moments of the intensity over all space and angle. A new relationship has been derived between the moments and the coefficients of the expansion of powers of ν in terms of the $g_k^m(\nu)$ polynomials which arise in transport problems without azimuthal symmetry. This relationship has been used to obtain an improved method for determining the moments.

I. INTRODUCTION

The study of plane-symmetric one-speed neutron transport, with the anisotropic scattering kernel expressed in terms of the first $(N+1)$ Legendre polynomials of the scattering angle, involves a decomposition of the azimuthally dependent equations into a set of $(N+1)$ azimuthally independent equations. For the mth azimuthal Fourier component of the finite series solution for the particle field strength, a set of orthogonal $g_k^m(\nu)$ polynomials arise. For a historical perspective it is worth noting that these $g_k^m(\nu)$ polynomials were introduced by Chandrasekhar[1] in his treatment of the same transport equation in the theory of radiative energy transfer. Furthermore, these polynomials are those required in the solution of the transport equation by the spherical harmonics technique.[2]

For an inverse problem the neutron angular flux or the angular distribution of radiation in the body and on the boundaries may be assumed to be completely known, and from this the scattering properties of the medium are desired.[3] In the simplest inverse transport problem, corresponding to an infinite medium containing a localized azimuthally symmetric plane source (i.e., the Green's function problem), a method equivalent to the "method of moments" has been utilized to extract the scattering coefficients in terms of spatial and angular moments of the angular flux throughout the infinite medium.[4,5] Such a procedure involves use of a recursive set of moment equations of increasing complexity; for example, for the nth scattering coefficient it is necessary to solve a determinant of order $2n+n(n-1)/2$ for $n \geq 1$.[6] Solutions of the azimuthally-independent inverse problem also have been worked out for the energy-dependent[7] and time-dependent cases.[8]

The inverse problem with an azimuthally asymmetric source has also been solved, where it has been shown that a single moment of the azimuth-dependent Green's function can be related to a single scattering coefficient.[6] Both these moments and those moments for the azimuth-independent problems are special cases of a generalized family of moments which may be related to the anisotropic response of a detector in an anisotropically scattering medium, as will be shown.

The purpose of this work is to provide a relatively simple technique for determining these generalized moments and to illustrate their use for calculating even powers of the distance of travel of particles from the source. As a by-product of the analysis a new relationship between the moments and the coefficients of the expansion of powers of ν in terms of the $g_k^m(\nu)$ polynomials is derived.

II. THE INVERSE PROBLEM WITH AZIMUTHAL DEPENDENCE

For a plane source in an infinite medium, the radiation intensity (or neutron angular flux) $I(\tau, \mu, \phi)$ depends upon one coordinate (τ), on the cosine of the polar angle with respect to the positive τ axis (μ) and on the azimuth (ϕ). In the absence of all but localized sources, the equation of transfer may be written as[1]

$$\left(\mu \frac{\partial}{\partial \tau} + 1\right) I(\tau, \mu, \phi) = \frac{1}{4\pi} \int_{-1}^{1} d\mu' \int_{0}^{2\pi} d\phi' \, p(\cos\delta) I(\tau, \mu', \phi'), \quad \tau \neq 0, \tag{1}$$

where anisotropic scattering of finite order N is admitted.

$$p(\cos\delta) = \sum_{n=0}^{N} \bar{\omega}_n P_n(\cos\delta), \tag{2}$$

and where some absorption is assumed $(0 < \bar{\omega}_0 < 1)$. The prescription for the infinite-medium Green's function is completed with the conditions that $I(\tau, \mu, \phi)$ stays bounded as $\tau \to \pm\infty$ and that

$$I(0^+, \mu, \phi) - I(0^-, \mu, \phi) = \mu_o^{-1} \delta(\mu - \mu_o)\delta(\phi), \quad -1 \leq \mu \leq 1. \tag{3}$$

By an established procedure[1,9] the ϕ dependence in Eq. (1) can be eliminated by a finite Fourier expansion

$$I(\tau, \mu, \phi) = \sum_{m=0}^{N} (2 - \delta_{m0}) I^m(\tau, \mu)(1 - \mu^2)^{m/2} \cos m\phi + I_u(\tau, \mu, \phi), \tag{4}$$

where $I_u(\tau, \mu, \phi)$ is a portion of the uncollided distribution,

$$I_u(\tau, \mu, \phi) = \mu_o^{-1} \delta(\mu - \mu_o) \exp(-\tau/\mu_o) \times \left[\delta(\phi) - \frac{1}{2\pi} \sum_{m=0}^{N} (2 - \delta_{m0}) \cos m\phi\right]. \tag{5}$$

The resulting $(N+1)$ independent transport equations are

 0022-2488/78/1905-0994$1.00

$$\left(\mu\frac{\partial}{\partial\tau}+1\right)I^m(\tau,\mu)=\tfrac{1}{2}\int_{-1}^{1}dm(\mu')p^m(\mu,\mu')I^m(\tau,\mu'). \tag{6}$$

Here

$$p^m(\mu,\mu')=\sum_{k=m}^{N}\bar{\omega}_k\frac{(k-m)!}{(k+m)!}p_k^m(\mu)p_k^m(\mu'), \tag{7}$$

$$p_k^m(\mu)=\frac{d^m}{d\mu^m}P_k(\mu)=(1-\mu^2)^{-m/2}P_k^m(\mu), \tag{8}$$

and, for brevity,

$$dm(\mu)\equiv(1-\mu^2)^m\,d\mu \tag{9}$$

For a monodirectional plane source in an infinite medium, the function which must be considered is

$$K_{l,n}^m=2\pi\int_{-\infty}^{\infty}d\tau\,\tau^n\int_{-1}^{1}dm(\mu)p_l^m(\mu)I^m(\tau,\mu),\quad m\le N,$$
$$=0,\quad m>N. \tag{10}$$

Symmetry considerations[4,8] reveal that $K_{l,n}^m=0$ for $(n+l+m)$ odd and for $n<l-m$.

From Eq. (6) we derive the identity

$$2\pi(2l+1)\int_{-\infty}^{\infty}d\tau\,\tau^n\frac{d}{d\tau}\int_{-1}^{1}dm(\mu)I^m(\tau,\mu)\mu p_l^m(\mu)$$
$$+h_lK_{l,n}^m=0,\quad l\ge m, \tag{11}$$

where

$$h_l=2l+1-\bar{\omega}_l. \tag{12}$$

Use of the recursion relation for the modified associated Legendre polynomial, followed by an integration by parts, gives

$$(l-m+1)K_{l+1,n-1}^m+(l+m)K_{l-1,n-1}^m=\frac{h_l}{n}K_{l,n}^m,\quad l\ge m. \tag{13}$$

For $m=0$ Eq. (13) reduces to the recursion equation of McCormick and Kuščer[6] once we correct their result for a typographical error.

From Eq. (11) and the appropriate source condition, we find the starting conditions for the sets of equations are

$$K_{m,0}^m=\frac{(1-\mu_0^2)^{m/2}}{h_m}\prod_{n=0}^{m}(2n+1). \tag{14}$$

Equation (14) relates a single moment of the azimuth-dependent Green's function to a single h value, and has been derived previously.[6] Since Eq. (14) forms a closed set of equations from which the scattering coefficients of the medium can be determined in terms of the K moments, it represents a solution to the inverse problem. Alternatively, Eqs. (13) and (14) may be used to obtain the scattering coefficients in terms of a different set of moments.

If the angle of incident radiation from the plane source is normal to the plane so that $\mu_0=1$, then all $K_{l,n}^m$ values for $m\neq0$ will vanish as a consequence of the azimuthal symmetry.

III. CALCULATION OF THE $K_{l,n}^m$

In developing a scheme to facilitate the calculation of the $K_{l,n}^m$ it is useful to look at an array ordered by those l,n, and m for which $K_{l,n}^m$ exist and do not vanish. Remembering that $K_{l,n}^m$ vanishes for $n<l-m$, for $l<m$, and for $(n+l-m)$ odd, we construct Table I which is valid for $m\le N$.

For a particular m, the table shows that the nonvanishing $K_{l,n}^m$ are located in the lower right diagonal portion of the array. The elements of this lower diagonal portion are confined by an uppermost boundary of elements defined by the general term $K_{m+p,p}^m$, for all $p\ge0$ and $m\le N$. These "boundary" or "upper diagonal" elements follow immediately from recursion relation (13) since in this case the first term of that recursion relation vanishes, i.e., $K_{m+p+1,p-1}^m=0$. Thus

$$K_{p+m,p}^m=[p(p+2m)/h_{p+m}]K_{p+m-1,p-1}^m, \tag{15}$$

from which it follows that $K_{p+m,p}^m$ for $m\le N$ is

TABLE I. Table of m values for nonvanishing $K_{l,n}^m$ and $m\le N$.

l									
7	7*	6*	5*	4*	3*	2*	1*	0*	
			7	6	5,7	4,6	3,5,7	2,4,6	
6	6*	5*	4*	3*	2*	1*	0*		
			6	5	4,6	3,5	2,4,6	1,3,5	
5	5*	4*	3*	2*	1*	0*			
			5	4	3,5	2,4	1,3,5	0,2,4	
4	4*	3*	2*	1*	0*				
			4	3	2,4	1,3	0,2,4	1,3	
3	3*	2*	1*	0*[a]					
			3	2	1,3	0,2	1,3	0,2	
2	2*	1*	0*						
			2	1	0,2	1	0,2	1	
1	1*	0*	1	0	1	0	1	0	
$l=0$	0*		0		0		0		
	$n=0$	1	2	3	4	5	6	7	n

[a] This means that $K_{3,3}^m$ vanishes for all $m\neq0,2$ and the asterisk on $m=0$ indicates that the element should be calculated by use of Eq. (16).

$$K^m_{m+p,p} = K^m_{m,0} \prod_{n=1}^{p} \frac{n(n+2m)}{h_{n+m}}, \tag{16}$$

where $K^m_{m,0}$ is given by Eq. (14).

The calculation of the remaining nonvanishing $K^m_{l,n}$ in Table I would be cumbersome with the use of Eqs. (13) and (14).[6] Hence it is desirable to develop an improved procedure. To do this, it is necessary to introduce the set of functions which satisfy the recursion relation[1,9]

$$h_k \nu g^m_k(\nu) = (k+m) g^m_{k-1}(\nu) + (k-m+1) g^m_{k+1}(\nu), \quad k \geq m, \tag{17}$$

where the starting equation is[10]

$$g^m_m(\nu) = p^m_m(\nu) = \prod_{n=0}^{m-1} (2n+1). \tag{18}$$

The $g^m_l(\nu)$ are polynomials of order $(l-m)$, alternatively even and odd, and hence may be used in an expansion such as

$$\nu^n = \sum_{l=m}^{m+n} A^m_{l,n} g^m_l(\nu). \tag{19}$$

A convenient means for calculating the $g^m_l(\nu)$ is the determinant[11]

$$g^m_k(\nu) = \frac{g^m_m(\nu)}{(k-m)!} \times \begin{vmatrix} h_m\nu & 1 & 0 & 0 & \cdots & 0 \\ 2m+1 & h_{m+1}\nu & 2 & 0 & & \vdots \\ 0 & 2m+2 & h_{m+2}\nu & 3 & & \vdots \\ 0 & 0 & 2m+3 & h_{m+3}\nu & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & 0 \\ \vdots & & & h_{k-2}\nu & k+m-2 & k-m-1 \\ 0 & & \cdots & 0 & k+m-1 & h_{k-1}\nu \end{vmatrix} \tag{20}$$

which was derived from Eq. (17) by an inductive proof and which generalizes a result of İnönü[12] to the case for $m \neq 0$. By a straightforward expansion of Eq. (20), an alternative expression is

$$g^m_k(\nu) = \sum_{l=0}^{k-m} G^m_{l,k} \nu^l, \tag{21}$$

where $G^m_{l,k} = 0$ if $(k+l-m)$ is odd. Here

$$G^m_{k-m-2s,k} = (-1)^s G^m_{k-m,k} S^m_{k-m-2s,k}, \tag{22}$$

where we define the factors

$$S^m_{k-m-2s,k} = \sum_{j_1=0}^{k-m-2s} \sum_{j_2=j_1+2}^{k-m-2s+2} \cdots \sum_{j_s=j_{s-1}+2}^{k-m-2} w_{j_1} w_{j_2} \cdots w_{j_s}, \quad s \geq 1,$$
$$= 1, \quad s = 0, \tag{23}$$

$$G^m_{k-m,k} = \left[\prod_{n=0}^{m-1} (2n+1)\right] \left[\prod_{i=m}^{k-1} h_i\right] \left[(k-m)!\right]^{-1} \tag{24}$$

and where $G^0_{0,0} = 1$ in order to satisfy Eq. (18). The coefficient of the lowest power of ν in Eq. (21), for example, is given by

$$G^m_{1,k} = (-1)^{(k-m-1)/2} G^m_{k-m,k} \sum_{j_1=0}^{1} \sum_{j_2=j_1+2}^{3} \times \cdots \sum_{\substack{j_p=j_{p-1}+2 \\ p=(k-m-1)/2}}^{k-m-2} w_{j_1} w_{j_2} \cdots w_{j_p}, \tag{25}$$

if $(k-m)$ is odd, and

$$G^m_{0,k} = (-1)^{(k-m)/2} G^m_{k-m,k} w_0 w_2 w_4 \cdots w_{k-m-2}, \tag{26}$$

if $(k-m)$ is even. In Eqs. (23), (25), and (26) the term w_j is defined as

$$w_j = (j+1)(2m+j+1)/(h_{j+m} h_{j+m+1}). \tag{27}$$

Equations (22)–(27) reduce to those given by Inönü[12] and earlier by Mika[13] for the case $m = 0$.

In a manner similar to the proof of İnönü,[12] it may be shown that the g polynomials satisfy the orthogonality relations

$$\int_\sigma \frac{\nu}{N^m(\nu)} g^m_k(\nu) g^m_n(\nu) \, d\nu = \frac{2(k+m)!}{h_k (k-m)!} \delta_{nk}, \tag{28}$$

where $N^m(\nu)$ denotes the normalization functions defined in Ref. 9. Here the integral over the eigenvalues spectrum σ is actually a summation in the Stieltjes sense over $-1 \leq \nu \leq 1$ and the set of discrete eigenvalues. From Eqs. (19) and (28) it follows that

$$\int_\sigma \frac{\nu^{n+1}}{N^m(\nu)} g^m_l(\nu) \, d\nu = \frac{2A^m_{l,n}(l+m)!}{h_l (l-m)!}. \tag{29}$$

Equation (29) may be used to show that the $A^m_{l,n}$ and the $K^m_{l,n}$ are related by

$$K^m_{l,n} = A^m_{l,n} n! (l+m)! (1-\mu_o^2)^{m/2} \times \prod_{p=0}^{m} (2p+1)^2 \left[h_l (l-m)! \, (2m+1)!\right]^{-1}. \tag{30}$$

Equation (30) is verified by using Eq. (17) in Eq. (29) and then using Eq. (30) to recover Eq. (13), and by then using Eqs. (18) and (28) to check that Eq. (29) for $l = m$ and $n = 0$ reproduces Eq. (14).

To determine the $A^m_{l,n}$ needed to obtain $K^m_{l,n}$ from Eq. (30), we use Eq. (21) to rewrite Eq. (19) as

$$\nu^n = \sum_{j=m}^{m+n} A^m_{j,n} \sum_{l=0}^{j-m} G^m_{l,j} \nu^l. \tag{31}$$

Interchanging the orders of summation gives

$$\nu^n = \sum_{l=0}^{n} \nu^l \sum_{j=l}^{n} A^m_{j+m,n} G^m_{l,j+m}, \tag{32}$$

from which we obtain a set of $(n+1)$ equations for $A^m_{l,n}$ in terms of $G^m_{l,n}$,

$$A^m_{n+m,n} G^m_{n,n+m} = 1, \tag{33}$$

and, for $l = 0$ to $l = n-1$,

$$\sum_{j=l}^{n} A^m_{j+m,n} G^m_{l,j+m} = 0. \tag{34}$$

From Eqs. (33) and (34) it can be shown by inductive logic that the coefficient $A^m_{l,n}$ can be expressed in the following determinant form,[14] where use has been made of Eqs. (22) and (23):

$$A^m_{t-2j,n} = \frac{1}{G^m_{n-2j,t-2j}} \times \begin{vmatrix} S^m_{n-2,t} & 1 & 0 & \cdots & 0 \\ S^m_{n-4,t} & S^m_{n-4,t-2} & 1 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & & \ddots & 1 \\ S^m_{n-2j,t} & S^m_{n-2j,t-2} & \cdots & & S^m_{n-2j,t-2(j-1)} \end{vmatrix} \tag{35}$$

Here $j > 1$ while l has been defined as $l = n + m$ to simplify the notation.

The calculation of $K^m_{l,n}$ thus requires determination of $A^m_{l,n}$ from a determinant of order $(m+n-l)/2$, followed by use of Eq. (30). To demonstrate the facility with which $K^m_{l,n}$ can be found using this technique, we display the result

$$K^0_{0,8} = \frac{40\,320}{h_0^2 h_1^2}\left[\frac{1}{h_0^3 h_1^2} + \frac{12}{h_0^2 h_1^2 h_2} + \frac{48}{h_0 h_1^2 h_2^2} + \frac{64}{h_1^2 h_2^3} + \frac{72}{h_0 h_1 h_2^2 h_3} + \frac{288}{h_1 h_2^3 h_3} + \frac{324}{h_2^3 h_3^2} + \frac{576}{h_2^2 h_3^2 h_4}\right], \tag{36}$$

which follows with a fourth-order determinant from Eq. (35) plus use of Eqs. (23), (24), and (27); a determinant of 14th order would have been required had the procedure using Eqs. (13) and (14) been used.[6] Equation (36) also can be obtained from a result of Siewert *et al.*[7]

IV. POSSIBLE INTERPRETATIONS FOR $K^m_{l,n}$

A family of moments has been defined and determined which encompasses earlier results as special cases. These moments are suggestive of applications involving a general spherical harmonics expansion. The question remains as to how these additional moments might be utilized.

A possible use of the generalized moments $K^m_{l,n}$ is as a representation of higher-moments of the even powers of the distance of travel of particles from the source. That is, if we define

$$K^m_{m,n}/K^m_{m,0} = \langle \tau^n \rangle_m, \tag{37}$$

then $\langle \tau^n \rangle_m$ is the nth order distance of travel for particles for the mth azimuthal component. For example,

$$\begin{aligned}
&\langle \tau^2 \rangle_0 = 2/h_0 h_1,\\
&\langle \tau^4 \rangle_0 = 24(1/h_0^2 h_1^2 + 4/h_0 h_1^2 h_2),\\
&\langle \tau^2 \rangle_1 = 6/h_1 h_2,\\
&\langle \tau^4 \rangle_1 = 72(3/h_1^2 h_2^2 + 8/h_1 h_2^2 h_3),\\
&\langle \tau^2 \rangle_2 = 10/h_2 h_3,\\
&\langle \tau^4 \rangle_2 = 120(5/h_2^2 h_3^2 + 12/h_2 h_3^2 h_4).
\end{aligned} \tag{38}$$

Equation (38) demonstrates that the nth order distance of travel tends to decrease as m increases, as may be verified for various special scattering laws.

The additional moments also may be used to incorporate the effects of anisotropy of a detector response when determining the scattering properties of a medium from experimental measurements with the detector. From a set of measurements along the τ axis, we can construct the moments

$$M_n = \int_{-\infty}^{\infty} \tau^n d\tau \int_0^{2\pi} d\phi \int_{-1}^{1} D(\mu, \phi) I(\tau, \mu, \phi)\, d\mu. \tag{39}$$

For convenience we postulate that the detector response function can be expanded in spherical harmonics as

$$D(\mu, \phi) = \sum_{l=0}^{L} \sum_{m=0}^{l} D^m_l P^m_l(\mu) \cos m\phi \tag{40}$$

about the same reference azimuthal angle $\phi = 0$ defined by the Green's function $I(\tau, \mu, \phi)$. Here D^m_l are the $[(L+1)(L+2)/2]$ coefficients which are assumed known. If $D(\mu, \phi)$ does not rapidly change with variations in μ and ϕ, L will be small (i.e., ≤ 2).

When $I(\tau, \mu, \phi)$ in Eq. (39) is replaced by the expansion of Eq. (4), and after use of Eq. (40), it follows that

$$M_n = \sum_{l=0}^{L} \sum_{m=0}^{[l,N]} D^m_l K^m_{l,n}, \tag{41}$$

where $[a, b]$ means minimum value of the elements a and b. Of course, the constraints on nonvanishing K-moments that $n \geq (l-m)$ and $(n+l-m)$ be even are still applicable.

For each n there is a single equation involving at most $(N+1)$ unknown h_l's. To solve for these unknowns, we must produce the same number of independent equations as we have unknowns. The proper set of M_n measurements depends upon the D^m_l for the detector. In the simplest case, when $L > N$, then taking the set of equations with $n = 0$ to N suffices provided $D^0_l = 0$ for all $l < N$. Other situations may necessitate a more complicated unfolding algorithm.

The reverse use of Eq. (41) may also be envisioned, where now we wish to characterize the anisotropy of a detector response from a knowledge of the scattering properties of the medium. That is, the $[(L+1)(L+2)/2]$ values of D^m_l are unknown while the $K^m_{l,n}$ values are given. To solve for the D^m_l when $L < N$, the best procedure is to make measurements for a single μ_o and to then group the results according to whether n is even or odd. In this way we obtain two uncoupled sets of equations,

$$\mathbf{K}_e \mathbf{D}_e = \mathbf{M}_e \tag{42}$$

and

$$\mathbf{K}_o \mathbf{D}_o = \mathbf{M}_o. \tag{43}$$

Here $\mathbf{K}_e$ has matrix elements $K^m_{l,n}$ with n even, $\mathbf{M}_e$ has elements M_n with n even, and $\mathbf{D}_e$ has elements D^m_l with even $(l+m)$. The subscript o is for the odd elements. Thus the D values are obtained as solutions of the equations

$$\mathbf{D}_e = \mathbf{K}_e^{-1} \mathbf{M}_e, \tag{44}$$

$$\mathbf{D}_o = \mathbf{K}_o^{-1} \mathbf{M}_o \tag{45}$$

unless difficulties arise because of an ill-conditioned $\mathbf{K}_e$ or $\mathbf{K}_o$.

To illustrate the calculational procedure, we take the elementary case of $L = 1$ and $N > 1$, where

$$\begin{aligned}
\mathbf{K}_e &= \begin{bmatrix} K^0_{0,0} & K^1_{1,0} \\ K^0_{0,2} & K^1_{1,2} \end{bmatrix}\\
&= \begin{bmatrix} 2/h_0 & 3(1-\mu_0^2)^{1/2}/h_1 \\ 2/h_0^2 h_1 & 18(1-\mu_0^2)^{1/2}/h_1^2 h_2 \end{bmatrix}
\end{aligned} \tag{46}$$

and

$$\mathbf{D}_e = \begin{bmatrix} D^0_0 \\ D^1_1 \end{bmatrix}, \qquad \mathbf{M}_e = \begin{bmatrix} M_0 \\ M_2 \end{bmatrix}, \tag{47}$$

while

$$\mathbf{K}_o = K^0_{1,1} = 1/h_0 h_1 \tag{48}$$

and

$$\mathbf{D}_o = D^0_1, \quad \mathbf{M}_o = M_1. \tag{49}$$

In the event that $L > N$, then the procedure in Eqs. (44) and (45) will not lead to a determination of all the coefficients, but only to those D^m_l for which $m < N$. For example, for $L = 1$ and $N = 0$, Eqs. (48) and (49) are still valid; but now D^1_1 cannot be determined, so Eqs. (46) and (47) become

$$\mathbf{K}_e = K^0_{0,0} = 2/h_0 \tag{50}$$

and

$$\mathbf{D}_e = D^0_0, \quad \mathbf{M}_e = M_0. \tag{51}$$

ACKNOWLEDGMENTS

Several comments from Dr. A. G. Gibbs and Dr. C. E. Siewert were helpful.

[1]S. Chandrasekhar, *Radiative Transfer* (Oxford U. P., London and New York, 1950; Dover, New York, 1960).
[2]B. Davison, *Neutron Transport Theory* (Oxford U. P., London, 1957).
[3]S. Pahor, Phys. Rev. **175**, 218 (1968).
[4]K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley, Reading, Mass., 1967), pp. 99ff.
[5]K. M. Case, Phys. Fluids **16**, 1607 (1973).
[6]N. J. McCormick and I. Kuščer, J. Math. Phys. **15**, 926 (1974).
[7]C. E. Siewert, M. N. Özişik, and Y. Yener, Nucl. Sci. Eng. **63**, 95 (1977).
[8]E. Canfield, Nucl. Sci. Eng. **53**, 137 (1974).
[9]N. J. McCormick and I. Kuščer, J. Math. Phys. **7**, 2036 (1966).
[10]In this and subsequent equations substitute the value 1 for the symbol $\prod_{m=j}^{k}$ whenever $k < j$.
[11]I. Kuščer and N. J. McCormick, in *Proceedings of the UCLA International Conference on Radiation and Remote Probing of the Atmosphere*, edited by J. G. Kuriyan (Western Periodicals, North Hollywood, Calif., 1974), p. 196.
[12]E. İnönü, J. Math. Phys. **11**, 568 (1970).
[13]J. R. Mika, Nucl. Sci. Eng. **11**, 415 (1961).
[14]J. A. R. Veeder, MS Thesis (Dept. of Nucl. Eng., U. of Washington, Seattle, 1977).

Asymptotic behavior of group integrals in the limit of infinite rank[a)]

Don Weingarten

Physics Department, Indiana University, Bloomington, Indiana 47401
(Received 20 June 1977)

We show that in the limit $N\to\infty$ integrals with respect to Haar measure of products of the elements of a matrix in SO(N) approach corresponding moments of a set of independent Gaussian random variables. Similar asymptotic forms are obtained for SU(N) and Sp(N). An application of these results to Wilson's formulation of lattice gauge theory is briefly considered.

Let I_p be defined by

$$I_p=\int d\mu\, U_{i_1j_1}\cdots U_{i_pj_p}, \tag{1}$$

where (U_{ij}) is a matrix in SO(N), $d\mu$ is Haar measure on SO(N), and for convenience we have suppressed the dependence of I_p on the indices $i_1,j_1,\ldots,i_p,j_p$. For arbitrary N and p, I_p is a complicated function of $i_1,j_1,\ldots,i_p,j_p$. In the present article, however, we will show that the asymptotic behavior of I_p as $N\to\infty$ with $i_1,j_1,\ldots,i_p,j_p$ fixed is rather simple. For even $p=2q$, I_{2q} is given by

$$I_{2q}=N^{-q}\sum\delta_{i_{k_1}i_{l_1}}\delta_{j_{k_1}j_{l_1}}\cdots\delta_{i_{k_q}i_{l_q}}\delta_{j_{k_q}j_{l_q}} + O(N^{-q-1}), \tag{2}$$

where $\delta_{i_ki_l}$ and $\delta_{j_kj_l}$ are Kronecker deltas and the summation in (2) is carried out over all distinct partitions of the integers $1,\ldots,2q$ into pairs $(k_1,l_1),\ldots,(k_q,l_q)$. For odd p, I_p is identically 0 once $N>p$. In other words, our result is that as $N\to\infty$ with q fixed the moments with respect to Haar measure of the set of random variables $\{\sqrt{N}U_{ij}\}$, $1\le i,\ j\le q$, approach corresponding moments of a set of independent Gaussian random variables $\{V_{ij}\}$ with $\langle V_{ij}\rangle=0$, $\langle(V_{ij})^2\rangle=1$, $1\le i,\ j\le q$. Asymptotic forms similar to (2) will also be given for integrals over SU(N) and Sp(N).

An application of our results to the g^{-2} expansion of the Green's function of Wilson's lattice gauge theory[1] will be briefly discussed toward the end of this paper.

Consider the dependence of I_p on the indices $i_1j_1,\ldots,i_pj_p$ as $N\to\infty$. The invariance of I_p with respect to translations in SO(N) implies that once $N>p$, I_p can be expressed as a linear combination of products of Kronecker deltas of the form $\delta_{i_ki_l}$ or $\delta_{j_mj_n}$; the alternating index $\mathcal{E}_{i_{k_1}\cdots i_{k_N}}$ or $\mathcal{E}_{j_{l_1}\cdots j_{l_N}}$ has N arguments and therefore cannot occur if $N>p$. To find the asymptotic behavior of I_p as $N\to\infty$, we will expand I_p as a linear combination of products of deltas, then find the asymptotic behavior of the coefficient of each product which occurs in this expansion. Notice that since the alternating index cannot appear once $N>p$, I_p must vanish if $N>p$ and p is odd (as we already mentioned). Throughout the following discussion we will assume p is even and $N>p$.

Now any product of deltas which can contribute to I_p must contain each of the indices $i_1,j_1,\ldots,i_p,j_p$ exactly once. This makes it possible to construct a convenient diagramatic representation of each product. Choose p points on a surface and number them from 1 to p. Let $\delta_{i_ki_l}$ be represented by a solid line from k to l and let $\delta_{j_mj_n}$ be represented by a dashed line from m to n. Each product yields a set of closed loops with each loop joining an even number of points. For example, a term which can contribute to I_6 is $\delta_{i_1i_2}\delta_{j_2j_3}\delta_{i_3i_4}\delta_{j_4j_1}\delta_{i_5i_6}\delta_{j_6j_5}$. The corresponding diagram is shown in Fig. 1. For simplicity we will not distinguish between diagrams and the products of deltas which they represent, both of which will be called numbered loop diagrams.

But I_p is unaltered by any exchange of index pairs of the form

$$(i_k,j_k)\longleftrightarrow(i_l,j_l). \tag{3}$$

Moreover, any set of distinct numbered loop diagrams is linearly independent as a set of functions on the variables $i_1,j_1,\ldots,i_p,j_p$. Thus when I_p is expanded as a linear combination of numbered diagrams, any pair of diagrams which can be converted into each other by permutations of form (3) must occur with the same coefficient. The sum of all distinct diagrams which can be gotten from a particular diagram by permutation can be represented by the original diagram with numbers removed. Diagrams of this sort and the corresponding symmetrized combinations of deltas will be called (unnumbered) loop diagrams.

A loop diagram is uniquely specified by a sequence of nonnegative integers $m_1,\ldots,m_k,\cdots$ in which m_k gives the number of loops passing through $2k$ points. Since any diagram which contributes to I_p must include a total of p points we have $\sum km_k=p/2$, and the largest number of points which can appear in a single loop is p. The sequence of integers specifying a diagram can therefore be terminated with $m_{p/2}$. The $p/2$-component

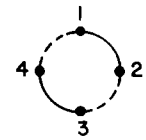

FIG. 1. A numbered loop diagram which can contribute to I_6.

[a)]Work supported in part by the United States Energy Research and Development Administration.

 0022-2488/78/1905-0999$1.00

vector $(m_1, \ldots, m_{p/2})$ will also be written m, and the loop diagram specified by m will be written $L_p(m)$ with dependence on $i_1, j_1, \ldots, i_p, j_p$ suppressed.

We can express I_p as a sum

$$I_p = \sum_{m \in S_p} f_p(m) L_p(m) \tag{4}$$

for a certain unique set of coefficients $f_p(m)$. The set S_p in (4) consists of all sequences of nonnegative integers $(m_1, \ldots, m_{p/2})$ such that $\sum k m_k = p/2$. For any value of p and for all but at most a finite set of $N > p$, the coefficient $f_p(m)$ for each $m \in S_p$ can be determined by using the equation

$$\sum_{j_q, j_{q-1}} \delta_{j_q j_{q-1}} I_q = \delta_{i_q i_{q-1}} I_{q-2} \tag{5}$$

for each even q, $2 \leq q \leq p$, where I_0 is defined to be 1. Eq. (5) follows from the normalization of Haar measure and the orthogonality relations for the matrix (U_{ij}) in (1). To prove that (5) determines each $f_p(m)$ the first step is to convert it to a set of equations directly on the collection of $f_q(m)$, $2 \leq q \leq p$.

Consider the effect of

$$\sum_{j_q j_{q-1}} \delta_{j_q j_{q-1}}$$

on a loop diagram $L_q(m)$ which contributes to I_q. The result will again consist of a linear combination of products of deltas. Each product will include one factor of the form $\delta_{i_q i_k}$ and one of the form $\delta_{i_{q-1} i_l}$ but $\delta_{j_q j_m}$ and $\delta_{j_{q-1} j_n}$ will not appear. If we represent one of these products by a numbered diagram using the rules introduced before, we obtain a set of closed loops plus one chain joining point q to point $q-1$. Since

$$\sum_{j_q j_{q-1}} \delta_{j_q j_{q-1}} L_q(m)$$

is unaffected by any permutation of form (3) which does not act on (i_q, j_q) or (i_{q-1}, j_{q-1}), any pair of numbered diagrams which can be converted into each other by a permutation of this sort must contribute to

$$\sum_{j_q j_{q-1}} \delta_{j_q j_{q-1}} L_q(m)$$

with the same coefficient. The sum of all distinct diagrams which can be gotten from a particular diagram by allowed permutations can be represented by the original diagram with numbers removed. Unnumbered symmetrized diagrams consisting of a chain plus some set of loops will be called chain diagrams. Each is uniquely determined by $q/2$ integers $m_1 \cdots m_k \cdots m_{q/2}$, where m_k, $2 \leq k \leq q/2 - 1$, gives the number of loops through $2k$ points, and $m_{q/2}$ gives half the number of points included in the remaining chain. The vector $(m_1, \ldots, m_{q/2})$ will be written m as before and the chain diagram determined by m will be written $C_q(m)$.

The preceding definitions yield the expansion:

$$\begin{aligned}\sum_{j_q j_{q-1}} \delta_{j_q j_{q-1}} L_q(m) = {} & \sum_{1 \leq k \leq q/2} N C_q(m - e_k + k e_{q/2}) \\ & + \sum_{1 \leq l < k \leq q/2} 2l(m_l + 1) C_q[m - e_k + e_l \\ & + (k - l) e_{q/2}] \\ & + \sum_{2 \leq k \leq q/2} (k-1) C_q(m - e_k + k e_{q/2}) \\ & + \sum_{1 \leq k, l \leq q/2 - 1} C_q[m - e_k - e_l \\ & + (k + l) e_{q/2}],\end{aligned} \tag{6}$$

where the kth component of e_k is 1 and all other components are 0 and $C_q(m)$ is defined to be 0 for $m \notin S_q$. If $L_q(m)$ is expanded as a sum of numbered loop diagrams, the first term on the right side of (6) is generated by diagrams in which points q and $q-1$ appear in the same loop joined by a dashed line; the second and third terms are generated by diagrams in which q and $q-1$ appear in the same loop but are not joined by a dashed line; and the fourth term comes from diagrams with q and $q-1$ in different loops.

The right-hand side of Eq. (5) can also be expanded as a sum of chain diagrams. Using (4) we obtain

$$\delta_{i_q i_{q-1}} I_{q-2} = \sum_{\substack{m \\ m_{q/2} = 1}} f_{q-2}(m) C_q(m), \tag{7}$$

where $f_{q-2}(m)$ for the $q/2$-component vector m is $f_{q-2}(m')$ with m' given by the first $q/2 - 1$ components of m. But the set of $C_q(m)$ for all distinct m is linearly independent as a collection of functions on the variables $i_1, j_1, \ldots, i_{q-2}, j_{q-2}, i_{q-1}, i_q$. Therefore, the coefficient of each $C_q(m)$ on the right side of (5) must equal its coefficient on the left side. Consider first the coefficient of $C_q(m - e_1 + e_{q/2})$ with $m_1 \geq 1$. Combining (5)—(7) yields

$$\begin{aligned}N f_q(m) + \sum_{2 \leq k \leq q/2} & (2k-2)(m_{k-1} - \delta_{k2}) f_q(m - e_1 - e_{k-1} + e_k) \\ & = f_{q-2}(m - e_1),\end{aligned} \tag{8}$$

where $f_q(m)$ is defined to be 0 for $m \notin S_q$. Now consider the coefficient of $C_q(m - e_l + l e_q)$, where $m_1 = m_2 = \cdots m_{l-1} = 0$, $m_l \geq 1$. Equations (5)—(7) give

$$\begin{aligned}(N + l - 1) & f_q(m) \\ & + \sum_{1 \leq k \leq q/2 - l} 2k(m_k - \delta_{kl}) f_q(m - e_l + e_{k+l} - e_k) \\ & + \sum_{1 \leq k \leq l-1} f_q(m - e_l + e_{l-k} + e_k) = 0.\end{aligned} \tag{9}$$

Equations (5)—(7) also imply other constraints in addition to (8) and (9), but as we will show, the set we have chosen is sufficient by itself to determine $f_p(m)$ for each $m \in S_p$. Moreover, the equations for $f_p(m)$ which we will not use must be consistent with (8) and (9) since we know from the derivations of (5)—(7) that at least one solution to the full set does exist.

If we introduce the rescaling

$$\bar{f}_q(m) = N^{q - \sum_k m_k} f_q(m), \tag{10}$$

Eqs. (8) and (9) become, respectively, (11) and (12)

$$\begin{aligned}\bar{f}_q(m) + N^{-2} \sum_{2 \leq k \leq q/2} & (2k-2)(m_{k-1} - \delta_{k2}) \\ & \times \bar{f}_q(m - e_1 - e_{k-1} + e_k) = \bar{f}_{q-2}(m - e_1)\end{aligned} \tag{11}$$

$$[1+N^{-1}(l-1)]\bar{f}_q(m)+N^{-2}\sum_{1\le k\le q/2-l}^{k}2k(m_k-\delta_{kl})$$

$$\times\bar{f}_q(m-e_l+e_{k+l}-e_k)$$

$$+\sum_{1\le k\le l-1}^{k}\bar{f}_q(m-e_l+e_{l-k}+e_k)=0. \tag{12}$$

Equations (11) and (12) hold for all even $q\ge 2$ if we adopt the convention $\bar{f}_0=1$. Thus to prove that (11) and (12) for each even q, $2\le q\le p$, together determine $\bar{f}_p(m)$ for every $m\in S_p$, it is sufficient to show that (11) and (12) can be solved for $\bar{f}_q(m)$ if $\bar{f}_{q-2}(m-e_1)$ is known. Now for each $m\in S_q$, Eqs. (11) and (12) can be rewritten in the form:

$$\sum_{m'\in S_q}[A(m,m')+N^{-1}B(m,m')$$

$$+N^{-2}C(m,m')]\bar{f}_q(m')$$

$$=\sum_{m''\in S_{q-2}}D(m,m'')\bar{f}_{q-2}(m''). \tag{13}$$

Equation (13) can be solved for $\bar{f}_q(m')$ if $\bar{f}_{q-2}(m'')$ is known and $\det(A+N^{-1}B+N^{-2}C)\neq 0$. $\det(A+N^{-1}B+N^{-2}C)$ is a polynomial in N^{-1} and can vanish for at most a finite set of values of N^{-2} if $\det A\neq 0$. On the other hand, $\det A\neq 0$ if (11) and (12) can be solved when N^{-1} is replaced by 0, (11) immediately determines all $\bar{f}_q(m)$ with $m_1\neq 0$. Equation (12), meanwhile, determines each $\bar{f}_q(m)$ with $m_1=m_2=\cdots=m_{l-1}=0$, $m_l\neq 0$ from the set of $\bar{f}_q(m)$ with at least one $m_k\neq 0$, $k<l$. By induction on l, (11) and (12) determine $\bar{f}_q(m')$ for all $m'\in S_q$ from $\bar{f}_{q-2}(m'')$. Thus $\det A\neq 0$. It follows that (11) and (12) for each even q, $2\le q\le p$, together determine $\bar{f}_p(m)$ for all $m\in S_p$ except possibly at a finite set of values of N.

In particular, (11) and (12) determine $\bar{f}_p(m)$ for all $m\in S_p$ if N is sufficiently large and imply the existence of the limit

$$\lim_{N\to\infty}\bar{f}_p(m)=F_p(m).$$

Equations (10)–(12) yield the asymptotic form

$$f_p(m)=N^{-p+\sum k m_k}F_p(m)+O(N^{-p+\sum k m_k-2}). \tag{14}$$

Equation (11) implies

$$F_p[(p/2,0,\ldots,0)]=1. \tag{15}$$

Finally (4), (14), and (15) give our main result, Eq. (2).

Equation (2) has an interesting corollary.[2] Consider the set of random variables $\{\sqrt{N}U_{ij}\}$ for $1\le i,\ j\le q$. Equations (1) and (2) imply that as $N\to\infty$, each moment of this set of variables approaches the corresponding moment of a set of q^2 independent Gaussian variables $\{V_{ij}\}$, with $\langle V_{ij}\rangle=0$, $\langle(V_{ij})^2\rangle=1$, $1\le i,\ j\le q$. Then by standard methods in probability theory[3] it can be shown that the joint cumulative distribution function of the set $\{\sqrt{N}U_{ij}\}$ approaches the distribution function of $\{V_{ij}\}$ pointwise everywhere.

Results similar to Eq. (2) can also be derived for integrals over SU(N) and Sp(N). For SU(N), let I'_{2q} be defined by

$$I'_{2q}=\int d\mu\, U_{i_1j_1}\cdots U_{i_qj_q}U^*_{i'_1j'_1}\cdots U^*_{i'_qj'_q}, \tag{16}$$

where (U_{ij}) is a matrix in SU(N) and $d\mu$ is Haar measure on SU(N). Integrals which do not include the same number q of U_{ij} and q' of U^*_{ij} vanish identically if $N>q,q'$. For I'_{2q}, if $N>q$ the invariance of $d\mu$ with respect to the action of elements of SU(N) implies I'_{2q} can be written as a linear combination of products of Kronecker deltas of the form $\delta_{i_ki'_l}$ and $\delta_{j_mj'_n}$. The behavior of the coefficients of these products as $N\to\infty$ can be determined by nearly the same method used for $f_q(m)$ in the case of SO(N). We obtain

$$I'_{2q}=N^{-q}\sum\delta_{i_1i'_{k_1}}\delta_{j_1j'_{k_1}}\cdots\delta_{i_qi'_{k_q}}\delta_{j_qj'_{k_q}}+O(N^{-q-1}),$$

where the sum is over all choices of $k_1\cdots k_q$ as a permutation of the integers $1,\ldots,q$.

The group Sp(N) can be represented as the subgroup of U($2N$) which fulfills

$$\sum_{ik}U_{ij}U_{kl}J_{ik}=J_{jl},$$

where J_{jl} is $-\delta_{jl-1}$ if j is even and δ_{jl+1} if j is odd. Define $U^1_{ij}=U_{ij}$, $U^2_{ij}=U^*_{ij}$, and let I''_p be

$$I''_p=\int d\mu\, U^{k_1}_{i_1j_1}\cdots U^{k_p}_{i_pj_p}, \tag{17}$$

where $d\mu$ is Haar measure on Sp(N). For $N>p$, I''_p vanishes if p is odd. Let M^{mn}_{kl} be J_{kl} if $m=n$ and δ_{kl} if $m\neq n$. Then the invariance of $d\mu$ with respect to Sp(N) implies, for $N>p$, I''_p can be written as a linear combination of products of $M^{k_li_m}_{i_li_m}$ and $M^{k_nk_r}_{j_nj_r}$. Calculating the behavior of the coefficients of these products, we find

$$I_{2q}=(2N)^{-q}\sum M^{k_1k_{m_1}}_{i_1i_{m_1}}M^{k_1k_{m_1}}_{j_1j_{m_1}}\cdots M^{k_1k_{m_q}}_{i_1i_{m_q}}M^{k_1k_{m_q}}_{j_1j_{m_q}}$$

$$+O[(2N)^{-q-1}]$$

as $N\to\infty$ with $i_1,j_1,k_1,\ldots,i_{2q},j_{2q},k_{2q}$ fixed. The sum in this relation is over all distinct partitions of $1,\ldots,2q$ into pairs $(l_1,m_1)\cdots(l_q,m_q)$.

In conclusion, it is perhaps worth mentioning that the method we have described for evaluating (1), (16), and (17) can probably be adapted to calculate the connected multistring vertices which appear in the g^{-2} expansion of the Green's functions of Wilson's formulation of gauge theory on a lattice.[1] The present results, for example, imply that for SO(N), SU(N), and Sp(N) as $N\to\infty$ the connected m-string vertices with $m>2$ fall faster than the 2-string vertex by at least one power of N.

ACKNOWLEDGMENTS

I would like to thank L. Pitt and particularly A. Lenard for helpful discussions.

[1] K. G. Wilson, Phys. Rev. D 10, 2445 (1975); K. G. Wilson, in *Gauge Theories and Modern Field Theory*, edited by R. Arnowitt and P. Nath (M. I. T. Press, Cambridge, Mass., 1975).

[2] I am grateful to A. Lenard for calling my attention to this result.

[3] A proof for the case of a single random variable is given by M. Loève, *Probability Theory* (Van Nostrand, New York, 1955), p. 185.

Borel summability and indeterminacy of the Stieltjes moment problem: Application to the anharmonic oscillators

S. Graffi[a), b)]

Dipartimento di Matematica e Informatica, Università di Ancona, Ancona, Italy

V. Grecchi[c)]

Istituto di Fisica, Università di Modena, Modena, Italy
and Istituto di Fisica, Università di Bologna, Bologna, Italy
(Received 20 July 1977)

An indeterminacy criterion is proven for the moment problem associated with the coefficients of a Borel summable power series of Stieltjes type which diverge faster than $(2n)!$. As an application we show that the Stieltjes type continued fraction corresponding to the Rayleigh–Schrödinger perturbation expansions for the energy eigenvalues of the anharmonic oscillators ($x^{2(m+1)}$ and in any finite number of dimensions) does not converge to the eigenvalues if $m > 2$. In particular, this implies the nonconvergence of the Padé approximants to the eigenvalues of $p^2 + x^2 + \lambda x^{2(m+1)}$ if $m > 2$.

I. INTRODUCTION

The Borel summability to the actual solution of the divergent perturbation expansions occurring in quantum theories, first proved for the anharmonic oscillators ($\lambda x^{2(m+1)}$ and in any finite number of degrees of freedom)[1] has been later extended to field theory [Simon[2] for the ground state eigenvalue of the spatially cutoff $(\phi^4)_2$; Eckmann, Magnen, Sénéor[3] for the Schwinger functions of the infinite volume $(\phi^4)_2$].

In all the above cases the proof rests on the verification of three conditions:

(i) a certain estimate on the behavior of the coefficients of the perturbation expansion,

(ii) analyticity of the solution in some sector $0 < |\lambda| < B$, $|\arg(\lambda)| < \pi/2 + \epsilon$, $B > 0$, $\epsilon > 0$,

(iii) a "geometric" bound on the remainder terms, valid uniformly in the above sector (for a more precise statement of these conditions, see below).

The aim of the present paper is to show that, when applied to a series of Stieltjes (see below) whose coefficients diverge faster than $\Gamma(2n + \epsilon n)$, $\epsilon > 0$, the above conditions yield an indeterminacy criterion for the Stieltjes moment problem associated with the coefficients. This result implies, as is known, the divergence of the Stieltjes type continued fraction associated with the power series. This means that the even and odd approximants sequences of the continued fraction [i.e., the (N/N) and $(N-1/N)$ Padé approximants sequences] converge to different functions, meromorphic in the whole complex plane except at the origin.

Applying in turn this result to the perturbation expansions of the eigenvalues of the one-dimensional anharmonic oscillators $p^2 + x^2 + \lambda x^{2(m+1)}$, which are known to be of Stieltjes type,[4] we can conclude that their associated sequences of Padé approximants do not converge to the eigenvalues for $m > 2$. This settles a question left open in the paper by Loeffel, Martin, Simon, and Wightman,[5] where the convergence of the Padé approximants was proven for $m \leq 2$. It also rigorously confirms the indications of a numerical analysis of the problem performed some years ago.[6]

As far as the multidimensional anharmonic oscillators are concerned it is not known whether or not the eigenvalues are Stieltjes functions, and in any case we still have the general result that the perturbation expansions of the eigenvalues are not Stieltjes summable for $m > 2$ [here $2(m+1)$ is the maximum degree of the interaction term].

The exposition proceeds as follows: in the next section we state our hypotheses and prove some preliminary technical lemmas; in Sec. III we state and prove the indeterminacy criterion, and in Sec. IV we present and discuss the applications to the quantum mechanical systems.

II. HYPOTHESES AND PROOF OF LEMMAS

Let $f(z)$ be a function of the complex variable $z = x + iy$, and $\sum_0^\infty c_n(-z)^n$ a power series such that

$$f(z) \sim \sum_0^\infty c_n(-z)^n . \tag{2.1}$$

The symbol $\sim$ means that the series $\sum_0^\infty c_n(-z)^n$ is the formal Taylor expansion of $f(z)$ around $z = 0$.

Throughout the rest of this paper, the function $f(z)$ and the power series $\sum_0^\infty c_n(-z)^n$ will be assumed to fulfill the following conditions:

(II.1) Let $m > 2$. For every $\epsilon > 0$ there exists $B > 0$ such that $f(z)$ is analytic in the domain $D: \{|z| < B, |\arg(z)| < (m+2)\pi/2 - \epsilon; |z| < \infty$ for $|\arg(z)| < \pi\}$ on the Riemann surface of $\log(z)$.

(II.2) There is β, $0 < \beta < +\infty$, such that $|f(z)| = O(|z|^\beta)$ for $|\arg(z)| < \pi$.

(II.3) For $|\arg(z)| < \pi$, $f(z)$ has the Herglotz property $\mathrm{Im} f(z)/\mathrm{Im}(z) > 0$.

a)Supported in part by C.N.R.
b)Present address: Istituto Matematico "G. Vitali," Università di Modena, Modena, Italy.
c)Supported in part by I.N.F.N., Sezione di Bologna.

 0022-2488/78/1905-1002$1.00

(II.4) Let $R_N(z)=f(z)-\sum_0^{N-1}c_n(-z)^n$ be the remainder after N terms of the expansion (2.1). Then there exist positive constants C,σ such that:

$$|R_N|(z)|<C\sigma^N\Gamma(mN+1)|z|^N,\quad N=1,2,\cdots \tag{2.2}$$

uniformly on compacts in D (m is as above).

Remarks:

(a) Condition (II.4) trivially yields

$$c_n=O(C\sigma^N\Gamma(mn+1)),\quad n=0,1,2,\cdots. \tag{2.3}$$

(b) Conditions (II.2) and (II.3), together with the analyticity on the whole first sheet $|\arg(z)|<\pi$ [Condition (II.1)] imply that the Stieltjes moment problem corresponding to the sequence $\{c_n\}_{n+[\alpha]+1}^{\infty}$ is solvable, i.e., there is at least one positive measure $d\mu$ on $[0,\infty)$ such that

$$c_{n+[\beta]+1}=\int_0^\infty x^n\,d\mu,\quad n=0,1,2,\cdots. \tag{2.4}$$

(c) A less restrictive version of Condition (II.1), i.e., analyticity of $f(z)$ in the sector $D_1:\{|z|<B,\ |\arg(z)|<m\pi/2+\epsilon;B>0,\epsilon>0\}$ of the Riemann surface of $\log(z)$ and Condition (II.4) [which implies (2.3)] represent the precise statement of conditions (i)–(iii) listed in Sec. I. They are sufficient to ensure the Borel summability of $\sum_0^\infty c_n(-z)^n$ to $f(z)$ in the whole sector $\{|z|<B,\ |\arg(z)|<\epsilon\}$ [see, e.g., (1)].

The meaning of such a statement, let us recall, is as follows. Let us define

$$F(z)=\sum_0^\infty c_n(-z)^n/\Gamma(mn+1). \tag{2.5}$$

This function, which by (2.3) is analytic in the circle $|z|<1/\sigma$, is called the mth Borel transform of $f(z)$, and has an analytic continuation into the whole sector $|\arg(z)|<\epsilon$, so that the representation

$$f(z)=1/m\int_0^\infty \exp(-a^{1/m})\,F(za)\,a^{-1+1/m}\,da \tag{2.6}$$

holds for $\{|z|<B,\ |\arg(z)|<\epsilon\}$, the convergence of the integral being uniform with respect to z.

As already emphasized, our aim is to prove the indeterminacy of the Stieltjes moment problem corresponding to the coefficients of the expansion in (2.1). We have thus to show that (II.1) and (II.4) imply the nonuniqueness of the measure $d\mu$ in (2.3) for $m>2$.

Let us begin by stating an elementary remark under the form of a lemma (without loss of generality we can take $[\beta]=-1$).

Lemma 2.1: Let the Stieltjes moment problem

$$c_n=\int_0^\infty x^n\,d\mu,\quad n=0,1,\cdots, \tag{2.7}$$

have at least one solution $d\Psi$.

Let $M>0$ be an arbitrary positive number. Then, if the Stieltjes moment problem

$$c_n'=\int_0^\infty x^n\,d\Psi(x+M),\quad n=0,1,\cdots, \tag{2.8}$$

is indeterminate (i.e., has infinite solutions), the same is true for (2.7).

Proof: Let $d\tau(x)\neq d\Psi(x+M)$ be a solution of (2.8). Then the measure $d\Psi_1$ with $d\Psi_1(x)=\theta(M-x)\,d\Psi(x)+\theta(x-M)\,d\tau(x-M)$ is a solution of (2.7) and $d\Psi_1\neq d\Psi$. The lemma is proved.

Now let $d\phi(x)=\lim_{\epsilon\to0}(1/\pi)\,\mathrm{Im}f(-x+i\epsilon)$, $0<x<+\infty$, be the discontinuity of $f(z)$ across the cut at $\arg(z)=\pi$. [Such a limit exists as a measure by a theorem of Herglotz, on account of (II.1) and (II.3), and let $d\tau(x)=d\phi(1/x)$.] In view of (II.1), the derivative $\phi'(z)$ exists and is analytic in the sectors $\{|z|<B,\ |\arg(z)|<m\pi/2-\epsilon\}$ so that $\tau'(z)$ will be analytic in the sector $\{|\arg(z-M)|<m\pi/2;\ 0<M<\infty\}$. Let $\zeta(x)=d\tau(x)/dx$, $x>M$. We have the following lemma.

Lemma 2.2: As $x\to+\infty$, the following asymptotic estimate holds,

$$\zeta(x)=O[\exp(-kx^{1/m})],\quad x\to\infty,\quad k>0. \tag{2.9}$$

Proof: Within our assumptions we can write, as is well known,

$$\begin{aligned}f(z)&=\int_0^\infty d\phi(t)/(z+t)=\int_0^\infty x\,d\tau(x)/(1+zx)\\&=\int_0^M x\,d\tau(x)/(1+zx)+\int_M^\infty x\,\zeta(x)/(1+zx),\\&\qquad |\arg(z)|<\pi,\end{aligned} \tag{2.10}$$

with

$$\begin{aligned}c_n&=(-1)^nf^{(n)}(0)/n!\\&=\int_0^M x^{n+1}\,d\tau(x)+\int_M^\infty x^{n+1}\zeta(x)\,dx\\&=O[\Gamma(mn+1)],\quad n=0,1,\cdots\end{aligned}$$

by (2.3).

Since of course there is an $A<\infty$ such that $\int_0^M x^n\,d\tau(x)<AM^n$, the assertion follows.

Next we prove that the asymptotic behavior (2.9) holds uniformly in the sector $\{|\arg(z-M)|<m\pi/2;\ 0<M<\infty\}$.

Lemma 2.3: As $z\to\infty$ within the sector $\{|\arg(z-M)|<m\pi/2;\ 0<M<\infty\}$, one has

$$\zeta(z)=O[\exp(-kz^{1/m})],\quad k>0. \tag{2.11}$$

Proof: Let us define

$$G(z)=\int_0^\infty \zeta(u+M)\,du/(1+zu). \tag{2.12}$$

By (2.9), the integral (2.12) defines $G(z)$ as an analytic function of z in the whole z plane cut from 0 to $-\infty$. Now it follows from (II.4) that $\phi'(u)$ vanishes for $u\to0$ more rapidly than any power of u, uniformly for $|\arg(u)|<m\pi/2$, and hence $\zeta(u)$ vanishes as $u\to\infty$ more rapidly than any power of $1/u$, uniformly in the same angular sector. Then by a well known result,[7] $G(z)$ will be regular on the whole sector $|\arg(z)|<(m+2)\pi/2$ on the Riemann surface of $\log(z)$, and within this sector we have

$$G^{(n)}(0)/n!=(-1)^n\int_0^\infty u^n\zeta(u+M)\,du=O(\Gamma(mn+1)) \tag{2.13}$$

by Lemma 2.2.

Now in view of the uniform decay property of $\zeta(u)$, we can replace $u\in\mathbb{R}$ in (2.12) by $\exp(i\alpha)u$, $u\in\mathbb{R}$, with $-m\pi/2<\alpha<m\pi/2$, and by (2.13) we get

$$G^{(n)}(0)/n! = (-1)^n \exp(i\alpha) \int_0^\infty (\exp(i\alpha)u)^n \zeta(\exp(i\alpha)u + M)\, du. \tag{2.14}$$

By Lemma 2.2, this implies that $\zeta(\exp(i\alpha)u + M)$ is $O(\exp(-ku^{1/m}))$, $k > 0$, as $u \to \infty$, and hence the result.

As we will see in Sec. IV, these preliminary results allow the application of the indeterminacy criterion stated in the next section to the case of the anharmonic oscillators.

III. THE INDETERMINACY CRITERION

Let us begin by stating the indeterminacy criterion for the Stieltjes moment problem whose proof is the object of this section.

Theorem 3.1: Let $\rho(x)$ be the derivative of the absolutely continuous part of a solution of the Stieltjes moment problem,

$$c_n = \int_0^\infty d\sigma(x), \quad n = 0, 1, \cdots \tag{3.1}$$

i.e., $\rho(x) = d\sigma_{ac}(x)/dx$.

Let ρ be an analytic function of $z = x + iy$ on the sector D: $\{z \mid |\arg(z)| \leq \pi/\alpha,\ \alpha > 2/m\}$ on the Riemann surface of $\log(z)$, and let $\rho(z) = O(\exp(-kz^{1/m}))$, $k > 0$, uniformly in any direction contained in D. Then the Stieltjes moment problem is indeterminate if $m > 2$ and $\alpha < 1$.

To prove Theorem 3.1, we will show that the stated hypotheses allow the application of a well known indeterminacy criterion for the Hamburger moment problem, with a suitable modification to account for the fact that (3.1) is a Stieltjes problem and not a Hamburger one.

Theorem A (See Akhiezer[8]): Let $\sigma'(u)$, the derivative of the absolutely continuous part of a measure $\sigma(u)$ on R, be a solution of the following Hamburger moment problem,

$$\mu_n = \int_{-\infty}^{+\infty} u^n d\sigma(u), \quad n = 0, 1, \cdots. \tag{3.2}$$

The problem (3.2) is indeterminate if

$$I = \int_{-\infty}^{+\infty} \log(\sigma'(u))/(1+u^2)\, du > -\infty. \tag{3.3}$$

The modification for the Stieltjes problem is easy, and is given by the following:

Lemma 3.1: The Stieltjes moment problem,

$$c_n = \int_0^\infty u^n d\Psi(u), \quad n = 0, 1, \cdots \tag{3.4}$$

is indeterminate if

$$J = \int_0^\infty [\log(\Psi'(u^2))/(1+u^2)]\, du > -\infty, \tag{3.5}$$

where $\Psi'(x)$ is the derivative of the absolutely continuous part of $\Psi(x)$.

Proof: As is known, a solution of a Stieltjes moment problem may be always obtained through the solution of the Hamburger moment problem $\mu_n = \int_{-\infty}^{+\infty} u^n d\sigma(u)$, $n = 0, 1, \ldots$, with $\sigma(u) = -\sigma(-u)$. This last condition yields $\mu_{2n} = 2\int_0^\infty u^{2n} d\sigma(u)$, $\mu_{2n+1} = 0$, so that the even moments μ_{2n} may be identified as the moments c_n of a Stieltjes problem with solution $\Psi(u) = 2\sigma(u^{1/2})$, $u > 0$. Hence the derivatives of the absolutely continuous part of the solutions are related by $\sigma'(u) = |u|\Psi'(u^2)$, $u \in \mathbb{R}$. Hence criterion (3.2) becomes:

$$I = \int_{-\infty}^{+\infty} [\log(u\Psi'(u^2))/(1+u^2)]\, du$$

$$= 2\int_0^\infty [\log(u\Psi'(u^2))/(1+u^2)]\, du > -\infty$$

which is of course equivalent to (3.5).

By Lemma 3.1, Theorem 3.1 will be proved once we show

$$\int_0^\infty [\log(\rho(u^2))/(1+u^2)]\, du > -\infty. \tag{3.6}$$

The proof of (3.6) relies upon the following representation theorem due to Nevanlinna:

Theorem B (Nevanlinna: See Boas[9]): Let $F(z)$ be regular and exponentially bounded in the half-plane $\mathrm{Im}(z) > -\epsilon$, $\epsilon > 0$; $F(z) \not\equiv 0$ and bounded on the real axis so that

$$\int_{-\infty}^{+\infty} \{\log_+|F(x)|/(1+x^2)\}\, dx < \infty. \tag{3.7}$$

[Here $\log_+(x) = \log(x)$, if $x \geq 1$; $\log_+(x) = 0$, if $x < 1$.] Let in addition $\{z_n\}_{n\neq 0}^\infty$ be the sequence of the zeros of $F(z)$ on the upper half plane.

Then the series $\sum_0^\infty \mathrm{Im}(1/z_n)$ is convergent, and the following representation holds,

$$\log|F(z)| = \log|B(z)| + y\pi^{-1}\int_{-\infty}^{\infty} \{\log|F(u)|/((u-x)^2+y^2)\}\, du + cy \tag{3.8}$$

($z = x + iy$, $y > 0$) where $-\infty < c < +\infty$, and the Blaschke product

$$\prod_0^\infty (1 - z/z_n)/(1 - z/\bar{z}_n) \tag{3.9}$$

converges uniformly on compacts in the half plane. In addition, if we define

$$\phi(z) = y\pi^{-1}\int_{-\infty}^{+\infty} \{\log|F(u)|/((u-x)^2+y^2)\}\, du + cy \tag{3.10}$$

we have

$$\lim_{z\to\infty} \phi(z)/|z| = c.\ \sin\theta, \quad 0 < \theta < \pi. \tag{3.11}$$

Next we prove the following.

Lemma 3.2: Let $F(z)$ be as in Theorem B. Then, if $0 < \alpha < 1$, one has

$$\int_{-\infty}^{+\infty} \{\log|F(iu^\alpha)|/(1+u^2)\}\, du > -\infty. \tag{3.12}$$

Proof: By the Nevanlinna representation (3.8), considered on the positive imaginary axis, we have

$$\log|F(iu^\alpha)| = \log|B(iu^\alpha)| + \phi(iu^\alpha),$$

where $\phi(z)$ is given by (3.10).

Now there is c', $-\infty < c' < +\infty$, such that

$$\int_0^\infty \{\phi(iu^\alpha)/(1+u^2)\}du$$
$$= \int_0^\infty \{\phi(iu^\alpha)/(u^\alpha+1)(u^\alpha+1)/(1+u^2)\}du$$
$$\geq c' \int_0^\infty \{u^{\alpha+1}/(1+u^2)\}du > -\infty \tag{3.13}$$

since $\phi(iu^\alpha)/(u^\alpha+1)$ is of course bounded for $u \geq 0$ and by (3.11) approaches the finite limit c. $\sin(\pi/2) = c > -\infty$ as $u \to +\infty$.

Next we consider the contribution of the Blaschke product defined by (3.9):

$$\int_0^\infty \{\log|B(iu^\alpha)|/(1+u^2)\}du$$
$$= \int_0^\infty \{\sum_0^\infty \log|(1-iu/z_n)/(1-iu/\bar{z}_n)|/(1+u^2)\}du$$
$$= -\tfrac{1}{2}\int_0^\infty \{\sum_0^\infty \Psi_n(u)/(1+u^2)\}du, \tag{3.14}$$

where $\Psi_n(u) = \log(1+f_n(u)) - \log(1-f_n(u)) \geq 0$, because $f_n(u) = 2a_n u^\alpha(1+u^{2\alpha}/|z_n|^2)^{-1} \geq 0$ for $u \geq 0$, since $a_n = -\operatorname{Im}(1/z_n) > 0$, $n = 0, 1, \cdots$. We have

$$\int_0^\infty \{\sum_0^\infty \Psi_n(u)/(1+u^2)\}du = \sum_0^\infty \int_0^\infty \{\Psi_n(u)/(1+u^2)\}du \tag{3.15}$$

by the uniform convergence of the Blaschke product and the monotone convergence theorem.

Now let $0 < \epsilon < 1$ and $a_n > \epsilon$; then, putting for $u > 0$:

$$w = a_n u^\alpha, \quad f(w) = 2w/(1+w^2),$$
$$\Psi(w) = \log(1+f(w)) - \log(1-f(w)),$$

one has

$$\int_0^\infty \{\Psi_n(u)/(1+u^2)\}du$$
$$\leq \alpha^{-1} a_n^{1/\alpha} \int_0^\infty \{\Psi(w)/(a_n^{2/\alpha}+w^{2/\alpha})\} w^{1/\alpha-1}dw \tag{3.16}$$

because $f_n(u) \leq f(w) \leq 1$, if $w \geq 0$, and $f(w) \neq 1$ for $w \neq 1$. Let us now define

$$F_1(x) = \int_0^\infty \{\Psi(w)/(x^{2/\alpha}+w^{2/\alpha})\} w^{1/\alpha-1}dw. \tag{3.17}$$

We have $F_1(a_n) \leq F_1(\epsilon)$ and hence by (3.16)

$$\int_0^\infty \{\Psi_n(u)/(1+u^2)\}du \leq \alpha^{-1} a_n^{1/\alpha} F_1(\epsilon) < \infty.$$

For $a_n < \epsilon$ we have

$$\int_0^\infty \{\Psi_n(u)/(1+u^2)\}du$$
$$\leq \alpha^{-1} a_n^{1/\alpha}[C(\epsilon)\int_0^\epsilon w^{1/\alpha}/(a_n^{2/\alpha}+w^{2/\alpha})\,dw + F_2(\epsilon)], \tag{3.18}$$

where $C(\epsilon) = \max_{0<w<\epsilon}(\Psi(w)/w)$ and $F_2(\epsilon) = \int_\epsilon^\infty \Psi(w) w^{-1-1/\alpha}dw$ are bounded functions on $(0, 1)$.

For the complete evaluation of the integral (3.18) we use the following inequality,

$$\int_0^\epsilon \{w^{1/\alpha}/(a_n^{2/\alpha}+w^{2/\alpha})\}dw$$
$$\leq a_n^{-2/\alpha}\int_0^{a_n}\{w^{1/\alpha}\}dw + \int_{a_n}^\epsilon \{w^{-1/\alpha}\}dw$$
$$= 2a_n^{1-1/\alpha}/(1-\alpha^2) - \alpha\epsilon^{1-1/\alpha}(1-\alpha). \tag{3.19}$$

Let $F(\epsilon) = \max(F_1(\epsilon), F_2(\epsilon))$. We can now state the result:

$$\int_0^\infty \{\log|B(iu^\alpha)|/(1+u^2)\}du$$
$$\geq \{-C(\epsilon)/(1-\alpha^2)\}\sum_0^\infty a_n - \{F(\epsilon)/(2\alpha)\}\sum a_n^{1/\alpha} > -\infty, \tag{3.20}$$

because $\sum_0^\infty a_n$ is a convergent series with positive terms and $1/\alpha > 1$. We are now in position to prove our criterion.

Proof of Theorem 3.1: Let $\rho(z)$ be as in Theorem 3.1. Set $\rho(z^2) = F(z')$, where $z' = iz^\alpha$, $2/m < \alpha < 1$, and $|\arg(z)| \leq \pi/(2\alpha)$. $F(z')$ will then fulfill all the assumptions of Theorem B. Therefore, we have

$$\int_0^\infty \{\log\rho(u^2)/(1+u^2)\}du = \int_0^\infty \{\log F(iu^\alpha)/(1+u^2)\}du > -\infty$$

by Lemma 3.2. Hence, in view of (3.6), Theorem 3.1 is proved.

Corollary 3.1: Let $f(z)$, $z = x+iy$, fulfill all the conditions (II.1)—(II.4), with $m > 2$. Then the Stieltjes moment problem (2.4) is indeterminate.

Proof: By Lemma 2.1, we can apply Theorem 3.1 to $\rho(x) = d\sigma(x)/dx$ and this proves the Corollary.

IV. APPLICATION TO THE ANHARMONIC OSCILLATORS

Consider first the one-dimensional anharmonic oscillators with interaction $x^{2(m+1)}$, i.e., the quantum mechanical systems whose Schrödinger operator, acting on $L^2(R)$, is given by

$$H_m = -\frac{d^2}{dx^2} + x^2 + \lambda x^{2(m+1)}, \quad m = 1, 2, \cdots. \tag{4.1}$$

(For an exhaustive mathematical treatment of such operators, the reader is referred to Simon.[10]) Let us denote by $E_m^n(\lambda)$, $n = 0, 1, \cdots$; $m = 1, 2, \cdots$, the eigenvalues of (4.1), and let us, from now on, drop the index n, because the present discussion does not depend on the particular eigenvalue. If we denote by

$$\sum_0^\infty A_n^m(-\lambda)^n \tag{4.2}$$

the Rayleigh—Schrödinger perturbation series for $E_m(\lambda)$, it is known that $E_m(\lambda)$ and $\sum_0^\infty A_n^m(-\lambda)^n$ verify all the conditions (II.1)—(II.4) [(II.1) is proved in (7) for the analyticity on the whole first sheet, and in (7) for the analyticity on $\{|z| < B; |\arg(z)| < (m+2)\pi/2 - \epsilon\}$; (II.2) and (II.3) are proved in (7), with $\beta = 1/3$; (II.4) is proved in (1)].

Therefore, by Corollary 3.1, we can apply directly the indeterminacy criterion of Sec. III, and conclude that the Stieltjes moment problem,

$$A_{n+1}^m = \int_0^\infty x^n d\mu^m(x), \quad n = 0, 1, 2, \cdots \tag{4.3}$$

is indeterminate for $m > 2$.

In this case it is well known that the corresponding Stieltjes type continued fraction diverges (see, e.g., Ref. 11), i.e., the even and odd approximants of the continued fraction corresponding to the series (4.2) [i.e., the (N/N) and $(N/N-1)$ Padé approximants se-

quences on the series (4.2)] converge to different functions, meromorphic in the whole complex λ plane except for a nonramified essential singularity at $\lambda = 0$. The same is true for any other $(N+j/N)$ Padé approximants sequence, $j = 1, 2, \cdots$.

Since by the already mentioned analyticity results of (4) and (7) [condition (II. 1)] the eigenvalues $E_n(\lambda)$ are not meromorphic in the whole λ plane except at the origin (as a matter of fact they have a discontinuity across the cut $\infty < \lambda \leq 0$ which is analytic for $|\lambda| < B$), the convergence of the Padé approximants to the eigenvalues cannot take place.

Hence on account of the convergence result for $m \leq 2$ proved by Loeffel, Martin, Simon and Wightman,[4] the following statement holds.

Theorem 4.1: The divergent Rayleigh–Schrödinger perturbation expansions of the eigenvalues of the anharmonic oscillators $p^2 + x^2 + \lambda x^{2(m+1)}$ are Stieltjes summable to the eigenvalues if, and only if, $m \leq 2$, i.e., the diagonal Padé approximants on the perturbation expansions converge to the eigenvalues if, and only if, $m \leq 2$.

Let us come now to the multidimensional case. If d is the space dimension, the Schrödinger operators, acting in $L^2(\mathbb{R}^d)$, are now given by

$$H_m = H_0 + \lambda V_m, \tag{4.4}$$

where

$$H_0 = \sum_{i=1}^{d} (-d^2/dx_i^2 + x_i^2), \tag{4.5}$$

$$V_m = \sum_{i_1 \cdots i_{2(m+1)}=1}^{d} x_{i_1} \cdots x_{i_{2(m+1)}}, \tag{4.6}$$

V_m being everywhere positive on the unit sphere of R^d. Denoting again by $E_m(\lambda)$ the eigenvalues of H_m and by $\sum_{n=0}^{\infty} A_n^m (-\lambda)^n$ their perturbation expansions, properties (II. 1), in the weaker version of analyticity in the domain D_1: $\{z \mid |z| < B, |\arg z| < (m+2)\pi/2 - \epsilon; \epsilon > 0, B > 0\}$ (see Remark b in Sec. II), (II. 2) and (II. 3) have been proved by Simon,[10] who in addition proved analyticity also near the whole positive real axis, and property (II. 4) has been proved in Ref. 1. Now two alternatives may occur: Either the energy values are analytic on the whole first sheet $|\arg(\lambda)| < \pi$, or there is some singularity. In the first case all the above considerations apply again; if there is some singularity [which by (II. 3) can be only a ramified essential singularity] the Stieltjes moment problem associated with the sequence $\{A_n^m\}_{n=1}^{\infty}$ has no solutions, since the cut plane analyticity is a necessary condition for its solvability. In this case therefore the Stieltjes type continued fraction corresponding to the perturbation expansion does not exist. Hence

Theorem 4.2: The divergent perturbation expansions of the eigenvalues of the multidimensional anharmonic oscillators (4.4) are not Stieltjes summable to the eigenvalues if $m > 2$.

Remark: In general, the nonsolvability of the Stieltjes moment problem does not prevent subsequences of diagonal Padé approximants from existing, since they may be uniquely defined as those rational functions $Q_{N+j}(\lambda)/P_N(\lambda)$ whose Taylor expansions around $\lambda = 0$ coincide with the given power series up to the $(2N+j+1)$th order. In this case they may still be interpreted as the approximants of some continued fraction, which simply is no more of Stieltjes type. Since there are no convergence (or divergence) statements under the present conditions, their convergence in some region of the complex λ plane cannot be *a priori* excluded. We feel that also in this occurrence there is convergence to the eigenvalues only for $m \leq 2$ as in the Stieltjes case, because all the asymptotic conditions are left unchanged.

[1] S. Graffi, V. Grecchi, and B. Simon, Phys. Lett. B 32, 631 (1970).
[2] B. Simon, Phys. Rev. Lett. 25, 1583 (1970).
[3] J. P. Eckmann, J. Magnen, and R. Sénéor, Commun. Math. Phys. 39, 251 (1975).
[4] J. J. Loeffel and A. Martin, Proceedings Program 25th Conference, Strasbourg, May, 1970.
[5] J. J. Loeffel, A. Martin, B. Simon, and A. S. Wightman, Phys. Lett. B 30, 656 (1969).
[6] S. Graffi, V. Grecchi, and G. Turchetti, Nuovo Cimento B 4, 313 (1971).
[7] See, e.g., G. Doetsch, *Theorie und Anwendungen der Laplace Transformation* (Springer-Verlag, Berlin, 1937), Chap. 14.1.
[8] N. I. Akhiezer, *The Classical Moment Problem* (Oliver and Boyd, Edinburgh, 1965).
[9] R. P. Boas, *Entire Functions* (Academic, New York, 1954).
[10] B. Simon, Ann. Phys. 58, 76 (1970).
[11] H. S. Wall, *Continued Fractions* (Van Nostrand, New York, 1947).

Analytic connection between configuration-interaction and coupled-cluster solutions[a)]

Tomislav P. Živković[b)] and Hendrik J. Monkhorst

Department of Physics, University of Utah, Salt Lake City, Utah 84112
(Received 17 May 1977)

The coupled-cluster (CC) equations in the work of Coester, Kümmel, Čižek, Paldus, and others are inhomogeneous, nonlinear and algebraic in the cluster operators to be determined. If taken to all orders, they are equivalent to complete configuration-interaction (CI) equations, except for states orthogonal to the reference state Φ. However, if taken only to nth order, they are not equivalent to the nth order CI equations, and due to their nonlinear form, the existence and the number of the solutions is not guaranteed. Also, the reality of the associated energy values is not certain since these values do not arise as eigenvalues of a Hermitian operator. We show that the equations can be cast in the form of perturbed CI equations, with the "perturbations" being non-Hermitian and nonlinear in the CI-like coefficients to be calculated. In the case of a finite number of single-particle states, we construct the solutions to the CC equations by analytic continuation from the CI solutions. Singularities peculiar to the method are identified and studied, and conditions for reality and the maximum multiplicity of solutions are given. In general, the energy will be real, and the number of solutions equals that of the associated CI problem. Singularities or instabilities in the coupled-cluster equations can be traced to unphysical assumptions in the basis set Hamiltonian, or a poor description to highly excited states.

I. INTRODUCTION

In their landmark publications of 1957, Goldstone[1] and Hubbard[2] proved the existence of the now-famous linked-cluster theorem for interacting fermion systems. The theorem, in effect, states that the perturbation corrections to the wavefunction and total energy beyond the independent-particle approximation can be represented by linked Feynman graphs. Earlier Brueckner[3] had proved this to hold for a few orders in the interaction strength by explicit computation. Goldstone and Hubbard generalized this to all orders. The significance of the linked-cluster theorem stems, in part, from the proportionality of the energy corrections for a crystal, plasma or nuclear matter fermions at a given density with the number of particles. It also provides a nice bookkeeping device for the myriad of perturbation correction terms.

Both Goldstone's and Hubbard's proofs use the interaction representation, thereby introducing time dependence in an intrinsically time-independent physical problem. Through the "time" integrations (resulting from the adiabatic switching of the interaction), the energy denominators of perturbation theory appear, and a cancellation of unlinked terms in both the exact wavefunction and correlation energy results. Even for bound states the Feynman diagrammatic language of "forward" and "backward" scattering of particles is kept, although this notion has only physical realism in scattering phenomena.

Notwithstanding its rigor and elegance, there are a number of disturbing aspects about the way the linked-cluster theorem was introduced and, subsequently, has been interpreted and applied. The first objection concerns the introduction of time dependence in the derivation. It certainly has made understanding of the proof more opaque from a mathematical standpoint, and even Goldstone admits that in his paper. Later Brandow[4] made the same point. In fact, it even leaves many authors still worrying to this date about the different types of cancellations of terms that we seem to have to distinguish.[5]

The second objection is the emphasis on the order-by-order feature of the theorem and, most importantly, its implementation. Yet it is well known that for Coulombic forces in extended systems perturbation theory diverges in all orders except the first one. The standard cure is partial summations of the most divergent terms to all orders, thereby eliminating this unphysical singularity. It is claimed as one of the strengths of the linked-cluster theorem to find, through topological arguments, which diagrams should be summed to infinite order. At the same time, in most cases, one lacks physical insight into the meaning of the particular partial summation. An infinite number of infinite-valued diagrams is always ignored thus casting doubt on the choice of those that are kept.

The third and main objection is the diversion of the attention away from the physical basis of the linked-cluster theorem to the algebraic and particularly topological meaning of linkedness or connectedness. In a sense, it is a question of representation in which one prefers to describe correlation effects. Its choice is inconsequential only when calculations can be carried to completion in the particular representation. Unfortunately this is almost never the case, so we are left with the choice of the "most efficient" representation leading to the "most favorable" convergence. (Of course, we entirely ignore the practical problem of our limitation to handle only a finite number of excitations in the zeroth order spectrum. But this applies to any representation.)

In a many-fermion configuration space, the exact

[a)]Supported in part by the National Science Foundation (Grant GP-42908).
[b)]Permanent address: The Rudjer Boskovic Institute, 41001 Zagreb, Croatia, Yugoslavia.

 0022-2488/78/1905-1007$1.00

wavefunction is "most of the time" well described by an independent-particle form. The probability for two or more particles to come close is small, and diminishes rapidly when an increasing number of them are involved. This is even attenuated by the Pauli exclusion principle that causes all fermions to carry a Fermi hole, even in the independent-particle model. Therefore, the energetic effect of this clustering of fermions, described by so-called linked terms, should decrease rapidly as their size increases. There is a significant probability for many pairs of fermions to cluster (or correlate) with antiparallel spin with no correlation of their relative positions. This leads to so-called unlinked terms in the wavefunction, expressible as the product of linked terms. Their occurrence is overwhelmingly important in large systems, and their neglect causes serious errors.

The importance of the linked-cluster theorem from a wavefunction point of view is its recognition of the fast convergence of the energetic effect from above linked-cluster type terms beyond the independent-particle approximation. At the same time, it includes products of these linked terms to higher orders in a most convenient manner. Goldstone and Hubbard were aware of this fact, and in Hubbard's paper it is brought out to some extent. However, the overall emphasis remained on the connection with the quantum-electrodynamical methods of Feynman, leading to the order-by-order view of many-body effects and scattering pictures.

Almost simultaneously, Coester[6] proposed a radically different approach to the same linked-cluster theory that does not suffer from any of the above disadvantages. It is time independent; it is algebraically simple; and it makes the physically significant corrections to the independent-particle wavefunction very explicit. The crucial idea is to express the exact wavefunction in occupation number formalism as

$$\Psi = e^T\Phi = (1 + T + T^2/2! + T^3/3! + \cdots)\Phi, \tag{1}$$

where the reference state Φ is defined by

$$\Phi = \prod_{\alpha=1} b_\alpha |\text{vac}\rangle, \quad b_\alpha = a_\alpha^\dagger, \tag{2}$$

$|\text{vac}\rangle$ is the vacuum state and b_α are the creation operators for the fermions in the reference state. We denote with α the occupied states, and with r the unoccupied states. The *cluster operator* T is defined by

$$T = \sum_{s=1} T_s, \tag{3}$$

where

$$\begin{aligned} T_1 &= \sum_{r\alpha} t_\alpha^r b_r a_\alpha, \\ T_2 &= \sum_{(rr')(\alpha\alpha')} t_{\alpha\alpha'}^{rr'} b_r b_{r'} a_{\alpha'} a_\alpha, \end{aligned} \tag{4}$$

and the general s-particle cluster operator T_s is given by

$$T_s = \sum_{R\Delta} t_\Delta^R b_{r_1} \cdots b_{r_s} a_{\alpha_s} \cdots a_{\alpha_1}, \tag{5}$$

where R is the ordered set of s "particle" indices $\{r_1, r_2, \cdots\}$, and Δ is that for the s "hole" indices $\{\alpha_1, \alpha_2, \cdots\}$. Here and in the following we will use the notation

$$\Phi_\Delta^R(s) = b_{r_1} \cdots b_{r_s} a_{\alpha_s} \cdots a_{\alpha_1}\Phi. \tag{6}$$

We sometimes write Φ_Δ^R when no confusion will arise. T_s creates s-particle-hole pairs, which makes Eq. (1) a nonunitary transformation of Φ. Its exponential form keeps track of the counting over distinct corrections to Φ only, expressed as replacements of one-particle states by cluster functions and their products, and represented by T_s and products thereof. The power of this description results from two aspects. First, without loss of generality, we can assume that the T_s describe linked clusters, meaning that these operators cannot be written as the product of two or more operators. This assumption is consistent with the equations that determine the cluster operators.

The second advantage is that we can perform a simple algebraic trick. When the Schrödinger equation is in our cluster expansion form

$$H(e^T\Phi) = E(e^T\Phi) \tag{7}$$

we can obtain its solution from the equivalent expression[6]

$$(e^{-T}He^T)\Phi = E\Phi. \tag{8}$$

But now we can express the left-hand side of Eq. (8) as a *finite* commutation series,

$$\begin{aligned} e^{-T}He^T = H + [H, T] + \tfrac{1}{2}[[H, T], T] \\ + \cdots + (1/4!)[[[[H, T], T], T], T]. \end{aligned} \tag{9}$$

The series terminates after five terms since H contains at most two particle operators. Using simple commutator algebra[7] one can reduce those expressions even further. In order to obtain equations for E and T_s one premultiplies Eq. (8) with $\langle\Phi|$ and $\langle\Phi_\Delta^R|$ and gets:

$$\langle\Phi|e^{-T}He^T|\Phi\rangle = E, \quad T = \sum_{s=1}^{n} T_s, \tag{10}$$

$$\langle\Phi_\Delta^R(s)|e^{-T}He^T|\Phi\rangle = 0 \quad (s = 1, 2, \ldots, n). \tag{11}$$

Each wavefunction nonorthogonal to Φ can be expressed in the form (1). However, usually one includes only up to n-particle cluster operators which leads to an approximation of the exact wavefunction. The corresponding equations are of the order n, as indicated in Eqs. (10) and (11). Moreover, we assume that the number of single-particle states is finite. This implies that both the number of particles K and the number of holes N is finite. Solutions of Eqs. (10) and (11) are hence approximate solutions of the Schrödinger equation for K particles in a finite dimensional Hilbert space. Obviously $n \leq N$. However, the case $n = N$ is of no practical interest. As shown in the Appendix, provided $n = N$, Eqs. (10) and (11) are equivalent to the corresponding finite dimensional configuration-interaction (CI) equations for all states nonorthogonal to Φ. Also, as n increases, the number of excitations increases so rapidly that it soon becomes impractical to solve these equations. Hence we assume $n \ll N$.

In actual calculations, Hartree–Fock or Brueckner orbitals for α and r states were used and a very limited number of nonzero T_s were kept. Obviously as many Eqs. (11) are needed as nonzero t_Δ^R amplitudes. It was found that in all electron systems considered, the pair

approximation $T \approx T_2$ leads to about 99% of the attainable correlation energy.[8] This is a very reassuring result which, as indicated above, can be well understood on statistical grounds.

In a long series of publications since 1960, Coester, Kümmel *et al.*[9] have used this method (named by them the exp-S method) for nuclear ground-state energies and properties. Čížek and Paldus[8,10] were the first to apply the method to a few atoms and molecules in calculations of ground-state correlation energies. Because of their pair approximations, the latter authors adopted the terms coupled-pair many electric theory (CPMET) and extended CPMET. To stress the general cluster aspects of the method, and the significant coupling between the clusters which results in nonlinear terms in Eqs. (11), we propose the name coupled-cluster (CC) method.

As expected, it turns out that by iteration of the algebraic equations (10) and (11) one generates all Goldstone-type linked perturbation terms, provided linked clusters to all orders are included. The marvel of this method seems to be that it zeros in directly to the heart of the correlation corrections in a most compact, yet transparent manner. This applies to fermion systems with any kind of interaction.

Even since its inception, Coester and Kümmel's approach has been largely ignored by the many-body theoretical community for a number of reasons. From a formal standpoint, it seemed to address itself only to calculating the ground-state correlation effects. These were not considered as very interesting quantities compared to response phenomena of many-body systems, which quite naturally shifted the attention to Green's function methods. Diagrammatic perturbation methods are used to compute these functions in a more or less systematic way. In a recent paper,[11] we have shown that one can indeed use the CC method to compute properties, both time dependent and time independent, by casting it in Green's functionlike form. Yet we can preserve the algebraic and conceptual simplicity, and we get easily solvable equations.

A serious objection was also the relative messiness of the coupled algebraic equations (11). These are nonlinear, inhomogeneous, and, if not carefully handled, of considerable complexity. Moreover, they are not in some eigenvalue form, thus not guaranteeing the reality and a definite multiplicity of their solutions. Coester[12] made some attempts to answer these questions.

We have now eliminated this objection by casting the CC method in the form of a perturbed configuration-interaction (CI) method. The perturbation is a non-Hermitian, unbounded operator with polynomial dependence on the CI-like coefficients to be determined.[13] Algebraically it is somewhat like the Hartree–Fock method, except that there the Coulomb and exchange operator are Hermitian and bounded. With the help of the analytic continuation method we will show a one-to-one correspondence of the CC solutions with CI solutions, and reality of E values for real Hamiltonians. Moreover, many mathematical properties such as the appearance of singularities will be derived. As we will see, these are more of numerical than physical significance.

In the next section we will present the derivation of the CI-like representation of the CC method followed by a section with numerical examples to illustrate some analytic properties of the formulation. In Sec. IV a summary is given of the mathematical analysis of the general CC equations, the details of which are given in the Appendix. We close with a section in which we point out the significance of the findings for the CC equations, and for the problem of solutions to coupled, nonlinear algebraic equations in general. As a byproduct, we obtain the number of Hartree–Fock solutions in a particular atomic orbital basis.

II. GENERAL

The most general state Ψ in cluster expansion form nonorthogonal to Φ can be written as

$$\Psi = e^T \Phi = (1 + C)\Phi, \tag{12}$$

where T is given by Eqs. (3)–(5), and

$$C = \sum_s C_s, \tag{13}$$

$$C_s = \sum_{R\Delta} C_\Delta^R b_r b_{r'} \cdots a_{\alpha'} a_\alpha. \tag{14}$$

Here R and Δ are, as before, the ordered sets of s "particle" and "hole" indices. In order to cast the CC equations in a form more accessible to mathematical analysis, we make use of the formal identities

$$e^T = 1 + C, \tag{15}$$

$$T = \ln(1 + C), \tag{16}$$

expressing T_s in CI-like operators C_s and vice versa. According to Theorem 1 in the Appendix, the CC equations (11), including up to n-tuple excitations, can be written as

$$\langle \Phi_\Delta^R(s) | (H - E)\{1 + C_1 + \cdots + C_n + [C_{n+1}(n) + C_{n+2}(n)]\} | \Phi \rangle = 0, \quad s = 0, 1, 2, \ldots, n, \tag{17}$$

where

$$\begin{aligned} C_{n+1}(n) &= \sum_{k=2}^{n+1} \frac{(-)^k}{k} \sum_{(s_i)} \delta(s_1 + \cdots + s_k, n+1) \prod_{i=1}^{k} C_{s_i}, \\ C_{n+2}(n) &= \sum_{k=2}^{n+1} \frac{(-)^k}{k} \sum_{(s_i)} \delta(s_1 + \cdots + s_k, n+1) \\ &\quad \times C_1 \prod_{i=1}^{k} C_{s_i} - \sum_{k=3}^{n+2} \frac{(-)^k}{k} \sum_{(s_i)} \delta(s_1 + \cdots + s_k, n+2) \\ &\quad \times \sum_{i=1}^{k} C_{s_i}, \end{aligned} \tag{18}$$

and $\sum_{(s_i)}$ is the summation over all possible choices of $s_1, s_2, \cdots$ such that their sum is in accord with the δ function. Instead of the CC Eqs. (17) we observe the following somewhat more general equations,

$$[H + \lambda V(\Psi)]\Psi = E\Psi, \tag{19}$$

where

$$H_{R\Delta, R'\Delta'} = \langle \Phi_\Delta^R | H | \Phi_{\Delta'}^{R'} \rangle, \tag{20}$$

$$\begin{aligned} V_{R\Delta, R'\Delta'} &= V_{R\Delta} \delta_{R'\Delta', 00} \\ &= \langle \Phi_\Delta^R | H[C_{n+1}(n) + C_{n+2}(n)] | \Phi \rangle \delta_{R'\Delta', 00}, \end{aligned} \tag{21}$$

$$\Psi_{R\Delta} = \langle \Psi_\Delta^R | 1 + \cdots + C_n | \Phi \rangle. \tag{22}$$

It can easily be shown that in the case of $\lambda=1$ Eq. (19) reduces to Eq. (17). The above equation is, however, written in the matrix form, where H is a Hermitian matrix with matrix elements given by Eq. (20). On the other hand, V is basically a non-Hermitian matrix with matrix elements depending on the components $\Psi_{R\Delta}$ of the column vector Ψ. As pointed out in the Introduction, we assume that the number of single-particle states is finite.

Equation (19) where λ is a parameter will be called the characteristic equation (CE). For the case $\lambda=0$, this equation reduces to the usual CI equation of order n with the subsidiary condition $\Psi_{00}=\langle\Psi|\Phi\rangle=1$. In this case Eq. (19) is a Hermitian equation, eigenvalues E are real, and the number of independent solutions Ψ is less than or equal to the dimension of the space involved. Our idea is to use this well-known property of Eq. (19) in the case $\lambda=0$, and by analytic continuation to extend the corresponding solutions to the whole complex λ plane. The particular case in point is $\lambda=1$ where the solution of the nth order CC Eqs. (17) should come. Thus, by that process, from the existence and reality of the solutions of CI equations, we hope to infer the existence and reality of the solutions to the CC equations. If, however, this fails, we are at least in the position to trace down the reason why it is so.

III. EXAMPLE

In order to get a clearer insight into the structure of the CE and the different possibilities which can emerge, we shall give an example of a CE which is simple enough to be solved analytically, yet incorporates all the characteristic features.

We consider a system consisting of two particles which can be distributed among four different states. In this case, there are six two-particle states; one non-excited, four singly excited, and one double excited state (see Fig. 1). If only single excitations are explicitly taken into account, the CE for the above system reads:

$$\langle\Phi_\Delta^R(s)|(H-E)\{1+C_1+\lambda[C_2(1)+C_3(1)]\}|\Phi\rangle=0,$$
$$C_2(1)=\tfrac{1}{2}C_1^2,\quad C_3(1)=\tfrac{1}{6}C_1^3,\quad s=0,1, \tag{23}$$

which can be easily deduced from Eqs. (17) and (18). The single excitation operator C_1 can be written in the form

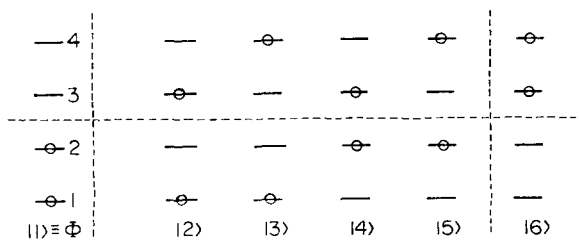

FIG. 1. Schematic representation of the ground state $|1\rangle\equiv\Phi$ and excited states $|2\rangle$ through $|6\rangle$ constructed from four one-particle states (denoted by 1 to 4), as used in examples of Sec. III. $|2\rangle$ to $|5\rangle$ are monoexcited states, and $|6\rangle$ is a doubly excited state.

$$C_1=x_2b_3a_2+x_3b_4a_2+x_4b_3a_1+x_5b_4a_1, \tag{24}$$

where

$$x_i=\langle i|C_1|\Phi\rangle.$$

This gives

$$C_2(1)|\Phi\rangle=(x_2x_5+x_3x_4)|6\rangle,\quad C_3(1)|\Phi\rangle=0. \tag{25}$$

We can now insert Eq. (25) into Eq. (23) and write the resulting equation in a matrix form,

$$\begin{pmatrix} h_{11} & \cdots & h_{15} \\ \vdots & & \vdots \\ h_{51} & & h_{55} \end{pmatrix}+\lambda(x_2x_5+x_3x_4)\begin{pmatrix} h_{16} & 0 & \cdots & 0 \\ \vdots & & & \vdots \\ h_{56} & 0 & \cdots & 0 \end{pmatrix}=E\begin{bmatrix} 1 \\ x_2 \\ \vdots \\ x_5 \end{bmatrix},$$
$$h_{ij}=\langle i|H|j\rangle,\quad i,j=1,\ldots,6. \tag{26}$$

Equation (26) incorporates some correlation effects since two particles are involved. On the other hand, only single excitations are explicitly taken into account, and hence the corresponding CC method is not equivalent to the CI method, which is known to have six linearly independent solutions with real eigenvalues. Thus Eq. (26) is the simplest nontrivial example of the CE. This model example will prove quite useful in the demonstration of the different properties of the CE.

So far, matrix elements h_{ij} are restricted only by the hermiticity condition $h_{ij}=h_{ji}^*$. We shall now solve Eq. (26) for some special choices of those elements.

Our first example is

$$\begin{pmatrix} u & 0 & v & v & 0 \\ 0 & 0 & 0 & 0 & 0 \\ v^* & 0 & 0 & 0 & 0 \\ v^* & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}+\lambda(x_2x_5+x_3x_4)\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ q & 0 & 0 & 0 & 0 \\ q & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}\begin{bmatrix} 1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}$$
$$=E\begin{bmatrix} 1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}. \tag{27}$$

We assume u, v, and q to be all nonzero and obviously $u^*=u$. Matrix equation (27) leads to a set of five equations in five unknowns:

$$\begin{aligned} &u+vx_3+vx_4=E, \\ &0=Ex_2, \\ &v^*+\lambda q(x_2x_5+x_3x_4)=Ex_3, \\ &v^*+\lambda q(x_2x_5+x_3x_4)=Ex_4, \\ &0=Ex_5, \end{aligned} \tag{27'}$$

If we put $E=0$ in Eq. (27') we obtain solutions satisfying

$$\begin{aligned} &x_3(u+vx_3)-vx_2x_5=vv^*/\lambda q, \\ &x_4=-(u/v+x_3),\quad E=0. \end{aligned} \tag{28}$$

For each $\lambda\neq0$ those solutions form a three-dimensional variety in the four-dimensional space spanned by x_2, x_3, x_4, and x_5. We can arbitrarily choose three eigenfunctions satisfying Eq. (28), and one possibility is

$$\Psi_1=\begin{bmatrix}1\\ x\\ 0\\ -u/v\\ x\end{bmatrix},\quad \Psi_2=\begin{bmatrix}1\\ 0\\ (-u+D)/2v\\ (-u-D)/2v\\ 0\end{bmatrix},\quad \Psi_3=\begin{bmatrix}1\\ 1/\lambda\\ (-u+D)/2v\\ (-u-D)/2v\\ 0\end{bmatrix}, \tag{29}$$

where

$$x=[(-v^*/\lambda q)]^{1/2},\quad D=[(u^2+4v^2v^*/\lambda)]^{1/2}.$$

We note that the space of the eigenfunctions (28) is not linear. If Ψ_1 and Ψ_2 are eigenvectors of Eq. (27) with the corresponding eigenvalue $E=0$, their linear combination $\Psi=\Psi_1+\beta\Psi_2$ such that $\langle\Psi|\Phi\rangle=1$ (i.e., $\alpha+\beta=1$) is generally not an eigenfunction to Eq. (27). This is the consequence of the nonlinear character of the equations (27), namely of the dependence of the matrix V on the eigenfunction Ψ. We can enlarge the notion of degenerate states to the general case that more than one eigenvector Ψ corresponds to a given eigenvalue E without those eigenvectors necessarily forming a linear space. In this sense, the solutions (28) are triply degenerate, forming a three-dimensional hypersurface in the four-dimensional space spanned by x_2 to x_5. As λ tends to zero, at least one of the components $x_i=\langle\Psi|i\rangle$ tends to infinity. Since $\langle\Psi|\Phi\rangle\equiv\langle\Psi|1\rangle=1$, this means that Ψ tends to be orthogonal to the nonexcited state Ψ. In particular if in our case we take a limit of Ψ_1, Ψ_2, and Ψ_3 as λ tends to zero we obtain, up to the normalization constant:

$$\Psi_1\to\begin{bmatrix}0\\1\\0\\0\\1\end{bmatrix},\quad \Psi_2\to\begin{bmatrix}0\\0\\1\\-1\\0\end{bmatrix},\quad \Psi_3\to\begin{bmatrix}0\\1\\0\\0\\0\end{bmatrix}. \tag{30}$$

Those three vectors are orthogonal to the vector Φ and though they are not eigenvectors of the CE, they are eigenvectors of the corresponding CI. This behavior of the above solutions of Eq. (27) reflects a general property of the solutions of CI. Since $\langle\Psi|\Phi\rangle=1$, no state orthogonal to Φ can be an eigenfunction of CE.

Besides degenerate eigenfunctions (28) corresponding to the eigenvalue $E=0$, Eq. (27) has in addition two nondegenerate eigenfunctions Ψ_4 and Ψ_5:

$$\Psi_4(\lambda)=\begin{bmatrix}1\\0\\x(\lambda)\\x(\lambda)\\0\end{bmatrix},\quad \Psi_5(\lambda)=\begin{bmatrix}1\\0\\x'(\lambda)\\x'(\lambda)\\0\end{bmatrix}, \tag{31}$$

$$E_4(\lambda)=\{u(v-\lambda q)+v([u^2+4v^*(2v-\lambda q)])^{1/2}\}/(2v-\lambda q),$$

$$E_5(\lambda)=\{u(v-\lambda q)-v([u^2+4v^*(2v-\lambda q)])^{1/2}\}/(2v-\lambda q),$$

where

$$x(\lambda)=\{-u+([u^2+4v^*(2v-\lambda q)])^{1/2}\}/(4v-2\lambda q),$$

$$x'(\lambda)=\{-u-([u^2+4v^*(2v-\lambda q)])^{1/2}\}/(4v-2\lambda q).$$

It should be observed that those two eigenfunctions represent one and the same analytic function in the complex λ plane, and that components of Ψ_4 and Ψ_5, as well as E_4 and E_5, lie on two different Riemann sheets of this function. There are two singular points of which one is a pole (for Ψ_5) and the other is a branch point,

$$\lambda_p=2v/q,\quad \lambda_b=(u^2+8vv^*)/4v^*q. \tag{32}$$

In the case of $\lambda=0$, which represents CI, vectors Ψ_4 and Ψ_5 reduce to

$$\Psi_4(0)=\begin{bmatrix}1\\0\\x\\x\\0\end{bmatrix},\quad \begin{aligned}x&=[-u+(u^2+8vv^*)^{1/2}]/4v,\\ E_4&=[u+(u^2+8vv^*)^{1/2}]/2,\end{aligned}$$

$$\Psi_5(0)=\begin{bmatrix}1\\0\\x'\\x'\\0\end{bmatrix},\quad \begin{aligned}x'&=[-u-(u^2+8vv^*)^{1/2}]/4v,\\ E_5&=[u-(u^2+8vv^*)^{1/2}]/2.\end{aligned} \tag{33}$$

If we start with the vector $\Psi_4(\lambda)$ for $\lambda=0$ and move along the line which encircles the branch point λ_b and then come back to the point $\lambda=0$, we arrive at the vector $\Psi_5(0)$ as indicated on Fig. 2. $\Psi_4(0)$ and $\Psi_5(0)$ are mutually orthogonal and their associated eigenvalues are real, which is not generally true for $\Psi_4(\lambda)$ and $\Psi_5(\lambda)$. Obviously, $\Psi_4(\lambda)$ and $\Psi_5(\lambda)$ cannot always remain orthogonal, since they are analytic continuations to each other, and in the branch point, λ_b, they coincide. Concerning reality of the corresponding eigenvalues, in the case of the real Hamiltonian (i.e., v and q real), energy as a function of (real) λ is real, as long as we do not pass a branch point. Once $\lambda>\lambda_b$ (for the case $\lambda_b>0$, and $\lambda<\lambda_b$ if $\lambda_b<0$), energy $E(\lambda)$ starts to be complex due to the negative number under the square root in Eq. (31). If, however, H is not real, energy $E(\lambda)$ is generally complex for each $\lambda\neq 0$. Hence the reality and nonreality of the eigenvalues of the CC equations depends on the basis in which matrix elements are written. If these matrix elements are complex, the eigenvalue is generally complex. If, however, these elements are real, the eigenvalue is real at least in some neighborhood of the point $\lambda=0$. This is to be compared with the CI method which always gives real eigenvalues for any choice of the basis vectors. We see that the most sensible basis choice for CC equations is a real basis where there is much more reason to believe that the corresponding eigenvalues will be real. As is well known, in the case of velocity-independent Hamil-

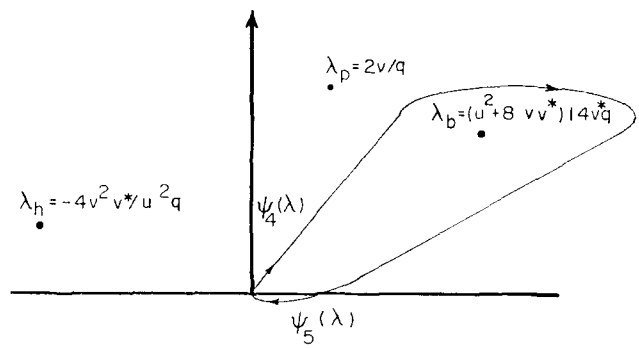

FIG. 2. Singular points in the complex λ plane and a path analytically connecting the solutions $\Psi_4(\lambda)$ and $\Psi_5(\lambda)$ of the examples of Sec. III. $\Psi_4(\lambda)$ and $\Psi_5(\lambda)$ belong to different Riemann sheets, and coincide in the branchpoint λ_b. λ_p and λ_h are a pole and hidden singularity, respectively.

tonians, such a basis always exists. It should be noted, however, that even in the case when H is complex, the eigenvalues are not completely without meaning: The imaginary part of the energy is a small quantity of first order in λ and as long as λ is small, the predominant contribution to $E(\lambda)$ is given by its real part, and the imaginary part can be considered as an error since, in any case, the CC method is only an approximation.

One can easily see that if H is real and if $\lambda_b < 0$ the CC equations will always have real eigenvalues. If, however, $\lambda_b > 0$, the reality will be insured by the requirement $\lambda_b \geq 1$, i.e.,

$$(u^2 + 8v^2)/4vq \geq 1. \tag{34}$$

This holds particularly true if q is small with respect to v or, from Eq. (27), if the mutual interaction of the states which are explicitly taken into account is big with respect to the interaction of those states with the states which are taken into account implicitly through the matrix V. That is, however, in accord with our intuitive feeling of the validity of the CC method.

Our next example is the CE,

$$\begin{pmatrix} 0 & 0 & u & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ u^* & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} + \lambda(x_2x_5 + x_3x_4) \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 2v & 0 & 0 & 0 & 0 \\ v & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{bmatrix} 1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}$$

$$= E \begin{bmatrix} 1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}, \tag{35}$$

where u and v are supposed to be different from zero. Equation (35) is equivalent to a set of equations:

$$\begin{aligned} &ux_3 = E, \\ &0 = Ex_2, \\ &u^* + 2\lambda v(x_2x_5 + x_3x_4) = Ex_3, \\ &\lambda v(x_2x_5 + x_3x_4) = Ex_4, \\ &0 = Ex_5. \end{aligned} \tag{35'}$$

Concerning (35') it can be shown that there are two possibilities: Either $\lambda \neq u/v$ or $\lambda = u/v$. If $\lambda \neq u/v$, there are two solutions to (35'): $\lambda \neq u/v$,

$$\Psi_1 = \begin{bmatrix} 1 \\ 0 \\ (u^*/u)^{1/2} \\ 0 \\ 0 \end{bmatrix}, \quad \Psi_2 = \begin{bmatrix} 1 \\ 0 \\ -(u^*/u)^{1/2} \\ 0 \\ 0 \end{bmatrix}, \quad \begin{aligned} E_1 &= (uu^*)^{1/2}, \\ E_2 &= -(uu^*)^{1/2}. \end{aligned} \tag{36}$$

If, however, $\lambda = u/v$, there is an infinite number of solutions to (35'),

$$\Psi(\lambda = u/v) = \begin{bmatrix} 1 \\ 0 \\ E/u \\ x \\ 0 \end{bmatrix}, \quad E = ux \pm (u^2x^2 + uu^*)^{1/2}. \tag{37}$$

Note that each energy E can be an eigenvalue, except for $E = 0$. Later we shall see that it is a general property of the CE that sometimes, for particular values of λ, an infinite number of eigenvalues are possible. We shall call such a point in a complex λ plane a resonance point.

The two examples above illustrate different types of anomalies which can be expected for the CE. We shall now examine the CE in more detail in order to show that the above two examples exhaust all but one of the anomalies which can be expected.

IV. GENERAL STRUCTURE OF CE

In the previous section we gave some examples of the CE in order to illustrate possible types of solutions. Now we are going to observe the CE from a general point of view.

Observe the CE

$$[\mathbf{H} + \lambda \mathbf{V}(x)]x = Ex, \quad x_1 = 1, \tag{38}$$

where $x = \{x_1, \ldots, x_m\}$ stands for a vector Ψ whose components $\Psi_{R\Delta}$ are renumerated in the order $1, 2, \ldots, m$ (see the Appendix). The two basic questions we want to answer are the existence of solution (E, x) and the reality of the eigenvalue E.

In order to reach these answers in the most complete sense, we must carefully analyze the mathematical structure of the CE. In particular, we have to identify singularities in the complex λ plane. In this section, we summarize the results of this analysis. Most of the details are given in the Appendix, particularly the proof of important theorems. We proceed by first discussing the special λ points for which the CE is singular in character, followed by rigorous answers to the above questions of existence, reality, and multiplicity of solutions (E, x).

(a) Resonance points

As illustrated by Example 2, the CE may have some points λ for which it has an infinite number of distinct eigenvalues. We called such a point a resonance point. It is shown in the Appendix that if λ_r is a resonance point, each E, except for a finite number of them, is an eigenvalue of the Eq. (38) where $\lambda = \lambda_r$. Thus, between resonance and nonresonance points there is a complete symmetry. In a nonresonance point there is a finite number of eigenvalues E; in a resonance point there is a finite number of E which are not eigenvalues.

Another question concerns the distribution of resonance points in a λ plane. This question is answered by Theorem 4 to the effect that there is either a finite number of resonance points, or there is a finite number that are not resonance points. If the first case occurs, we call the CE normal. However, the possibility that the CE is not normal is very unlikely. It would mean that for almost every λ Eq. (38) would have each E, except a finite number of them, as an eigenvalue. Physically it would imply that the CC equations, which correspond to the point $\lambda = 1$, are void of any significance, since even if $\lambda = 1$ were not a resonance point, each point infinitesimally close to it would be. However, we

were not able to rule out this possibility, although we believe that it actually does not take place. In any event, if the CC equations are to be given any physical significance, the corresponding CE should have, at most, a finite number of resonance points and hence it should be normal.

The solutions which are "specific" for a resonance point we call resonance solutions. Besides resonance solutions in a resonance point, we can also have nonresonance solutions—the distinction being that resonance solutions cannot be extended in the region outside the resonance point. Obviously, if each resonance solution could be extended continuously to some neighborhood $N(\lambda_r)$ of a point λ_r, then in each point of $N(\lambda_r)$ there would be an infinite number of distinct eigenvalues, and hence this point should be resonance as well. But this would imply that each λ except a finite number of them, is resonance, which would mean that the CE is not normal. Hence, in the case of a normal CE, only a finite number of the solutions in a resonance point λ_r can be extended continuously outside this point and those solutions we call nonresonance. Thus resonance solutions do exist only in a resonance point and cannot be extended outside it. They are isolated, and since the corresponding eigenvalue E can assume almost any value, be it real or complex, they are void of any physical meaning.

Note that the point $\lambda=0$, by the very definition, cannot be a resonance point, and hence in the case of normal CE there is a small neighborhood $N(0)$ of the point $\lambda=0$ where there is no resonance point. This is in accord with the intuitive idea that for small λ the term $\lambda V(x)$ should be considered as a perturbation, and hence should not destroy the "good" feature of the CE in the point $\lambda=0$. One must, however, be careful in such conclusions since, although $V(x)$ is continuous, it is not a bounded operator as shown in Lemma 7 of the Appendix. Hence there is no *a priori* reason why λV for infinitesimal λ should be infinitesimal.

(b) Singular points

CE (38), for a given (E,x) and a given λ, is well characterized by the Jacobian

$$D(\lambda,E,x)=\begin{vmatrix} 1 & (h_{12}+\lambda\partial^2 V_1(x)) & \cdots & (h_{1m}+\lambda\partial^m V_1(x)) \\ x_2 & (h_{22}-E+\lambda\partial^2 V_2(x)) & \cdots & \circ \\ & \cdot & & \cdot \\ & \circ & & \circ \\ x_m & (h_{m2}+\lambda\partial^2 V_m(x)) & \cdots & (h_{mm}-E+\lambda\partial^m V_m(x)) \end{vmatrix},$$

$$\partial^i=\frac{\partial}{\partial x_i}. \tag{39}$$

As shown in the Appendix (Theorem 5), if (E_0,x_0) is a solution of the CE in a point λ_0, and if a corresponding Jacobian $D(\lambda_0,E_0,x_0)$ is different from zero, then this solution can be analytically extended in the whole λ plane with the exception of a finite number of points. The solution $[E(\lambda),x(\lambda)]$, which is obtained by this process of analytic extension, we call normal. Our main concern, however, are those normal solutions which are defined in the point $\lambda=0$, which corresponds to the CI. Such solutions we call regular. They can bridge the gap between CI and CC, since they can be extended out of the point $\lambda=0$, to the whole λ plane, and provided they are defined in the point $\lambda=1$, they connect CI and CC solutions in a natural way. The normal solution $[E(\lambda),x(\lambda)]$, which is the extension of the solution (E_0,x_0) in the point $\lambda=\lambda_0$, is by the very definition nonresonant. Moreover, it is an algebraic function of λ, and as such it may have poles and branch points. As shown by Lemma 10, if $\lambda=\lambda_b$ is a branch point of a normal solution $[E(\lambda),x(\lambda)]$, the Jacobian $D[\lambda_b,E(\lambda_b),x(\lambda_b)]$ should be equal to zero. Also, according to Lemma 8, vanishing of the Jacobian can be due to a continuous degeneracy of the solution at this point (see Definition 3). If $\lambda=\lambda_d$ is such a point, we call it the point of accidental degeneracy. It is shown in the Appendix that vanishing of the Jacobian in a point $\lambda=\lambda'$ indicates instability of the normal solution $[E(\lambda'),x(\lambda')]$. Thus in a branch point and in the accidental degeneracy point the corresponding solution is unstable. It may also happen that the solution $[E(\lambda),x(\lambda)]$ is unstable in some point $\lambda=\lambda_h$ (i.e., the corresponding Jacobian is zero), but it is neither a branch nor an accidental degeneracy point. In this case, we say that $\lambda=\lambda_h$ is a hidden singularity of a normal solution $[E(\lambda),x(\lambda)]$.

It should be noted, however, that although $[E(\lambda),x(\lambda)]$ is not defined in a pole $\lambda=\lambda_p$, the corresponding Jacobian may well be defined in this point. $D(\lambda)\equiv D[\lambda,E(\lambda),x(\lambda)]$ is an analytic function in λ, and there is a possibility that there exists an analytic continuation of this function in a point $\lambda=\lambda_p$ which is a pole of a solution $[E(\lambda),x(\lambda)]$. It may happen that the Jacobian vanishes in this point; hence poles can also cause disappearance of the Jacobian.

In conclusion then, if $[E(\lambda),x(\lambda)]$ is a regular (normal) solution, it may have four types of singular points. First there are poles which are those points where $[E(\lambda),x(\lambda)]$ is not defined. Second there are branch points, third there are accidental degeneracies, and fourth there are hidden singularities.

(c) Examples of singular points

The above possibilities are illustrated in Sec. III. The regular eigenfunction $\Psi_5(\lambda)$ Eq. (31) has a pole in a point $\lambda_p=2v/q$ and a branch point in $\lambda_b=(u^2+8vv^*)/4v^*q$. In addition, there is a hidden singularity in a point $\lambda_h=-4v^2v^*/u^2q$. This last singular point can be found if we observe the corresponding Jacobian, which is

$$D(\lambda)=D[\lambda,E(\lambda),\Psi(\lambda)]$$

$$=\begin{vmatrix} 1 & 0 & v & v & 0 \\ x_2 & -E & 0 & 0 & 0 \\ x_3 & \lambda q x_5 & \lambda q x_4-E & \lambda q x_3 & \lambda q x_2 \\ x_4 & \lambda q x_5 & \lambda q x_4 & \lambda q x_3-E & \lambda q x_2 \\ x_5 & 0 & 0 & 0 & -E \end{vmatrix}. \tag{40}$$

Both eigenfunctions $\Psi(\lambda)$ in Eq. (31) satisfy $x_2=x_5=0$ and $x_3=x_4=x$. Hence

$$D(\lambda) = E^3[E + 2x(v - \lambda q)]. \tag{41}$$

Thus $D(\lambda) = 0$ if either $E = 0$ or $E + 2x(v - \lambda q) = 0$. The first possibility yields

$$E = 0 \Rightarrow \lambda_1 = \lambda_h = -4v^2v^*/u^2q, \quad \lambda_2 = \lambda_p = 2v/q. \tag{42}$$

The point λ_2 is a pole while the point λ_1, being neither a pole nor a branch point, is a hidden singularity. Another possibility yields

$$E + 2x(v - q) = 0 \Rightarrow \lambda = \lambda_b = (u^2 + 8vv^*)/4v^*q, \tag{43}$$

which is a branch point.

Concerning the eigenfunctions Ψ_1, Ψ_2, and Ψ_3, Eq. (29), one finds $D(\lambda) \equiv 0$ in accord with Lemma 8. Thus the vanishing of the Jacobian is due to a continuous degeneracy. However, no point λ is a point of accidental degeneracy, since the above eigenfunctions are degenerate identically, i. e., for each λ. Hence they are not normal eigenfunctions (see Theorem 5). In the same way, one finds $D(\lambda = u/v) = 0$ in the case of a resonance solution (37). This is in accord with Lemma 9.

(d) Physical meaning of singular points

The solutions close to a pole or a branch point can obviously have no physical meaning. Close to a pole, for example, at least one component of a vector $x(\lambda)$ tends to infinity. This, however, means that the operator V cannot be treated as a perturbation anymore. However, in the CC equation, V is introduced as an approximation which takes into account only some of the excited states. Hence, if a pole is close to a point $\lambda = 1$, the corresponding regular solution is not of much value. Formally, close to a pole a state $x(\lambda)$ tends to be orthogonal to a state Φ, and thus if $\langle x | \Phi \rangle$ is small for some solution (E, x) of the CC equations, we should regard this solution as poorly representing the real state. Hence the CC formalism is good only for a few low-lying states, expecially for a ground state which is likely to have significant overlap with a state Φ.

On the other hand, close to a branch point we have two different solutions with eigenfunctions whose overlap approaches to 1. However, at the point $\lambda = 0$, those two eigenfunctions are mutually orthogonal and hence distinct. Thus if operator V is to be considered as a perturbation, the solution close to a branch point can also have no physical meaning.

(e) Reality of solutions

The question of the reality of the solutions, especially, the reality of the eigenvalue E, is answered by Lemma 12. Provided Hamiltonian H is real, each regular solution $[E(\lambda), x(\lambda)]$ is real along the real λ axis, as long as we don't meet some branch point. It means that if there is no branch point between the points $\lambda = 0$ and $\lambda = 1$, the corresponding solution of CC equations ($\lambda = 1$) should be real. Note that the essential requirement is the reality of the Hamiltonian. As is well known, provided Hamiltonian H is velocity independent, there exists a real basis.

It is interesting to note that the reality and the nonreality of eigenvalues E depends to some extent on the basis in which matrix elements of the Hamiltonian H (and hence of the operator V) are written. Different orthonormal bases can be connected by unitary transformations and eigenvalues of CC equations are generally not invariant with respect to them. This can be illustrated using as an example the CE (27). The full Hamiltonian of the system described by this equation is a six-by-six matrix. Observe now the infinitesimal unitary transformation,

$$U = I + i\epsilon A, \quad a_{36} = a_{46} = a_{63} = a_{64} = 1 \quad \text{otherwise, } a_{ij} = 0, \quad i,j = 1,2,\ldots,6 \tag{44}$$

and assume that H is real,

$$h_{13} = h_{14} = v = v^*, \quad h_{36} = h_{46} = q = q^*. \tag{45}$$

To first order in ϵ the transformed CE reads

$$\begin{pmatrix} u & 0 & v & v & 0 \\ 0 & 0 & 0 & 0 & 0 \\ v & 0 & 0 & 0 & 0 \\ v & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} + \lambda(x_2x_5 + x_3x_4) \begin{pmatrix} -2i\epsilon v & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ q + i\epsilon P & 0 & 0 & 0 & 0 \\ q + i\epsilon P & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{bmatrix} 1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}$$

$$= E \begin{bmatrix} 1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}, \tag{46}$$

where

$$P = h_{66} = \langle 6 | H | 6 \rangle.$$

Due to (45) the eigenvalues of the nontransformed CE (27) are real for real λ values as long as $\lambda \leq \lambda_b$. This is obvious from Eqs. (31) and (32). However, the eigenvalues associated with the transformed CE (46) are complex, which results from the appearance of complex terms in the operator V.

It should be noted that the eigenvalues of CI equations are also generally not invariant with respect to unitary transformations. However, in the case of CI equations, eigenvalues remain always real, while in the case of CC equations they can, as shown above, assume complex values as well.

Unitary transformation (44) mixes singly excited states $|3\rangle$ and $|4\rangle$ with the doubly excited state $|6\rangle$. However, we can consider unitary transformations of one-particle type, which only mix occupied and unoccupied one-particle states among themselves. Such transformations express new creation (annihilation) operators as linear combinations of old creation (annihilation) operators, and they change the reference state Φ only up to the phase. Obviously CC equations (11), and hence the CE, are invariant with respect to those transformations. Usually, however, a velocity-independent Hamiltonian is written either in the real basis or in some basis which can be connected by the above type of unitary transformation with a real basis. We conclude that, provided Hamiltonian H is velocity independent, eigenvalues E are generally real. The reality of the eigenvalue can be violated only if the corresponding regular solution of the CE has a branch point of the real λ axis between real points $\lambda = 0$ and $\lambda = 1$.

(f) Modified characteristic equation

As mentioned above, regular solutions of the CE are our main concern, since only those solutions can connect CI and CC. However, a given CE can have very few or no regular solutions at all. In order that a solution $[E(\lambda),x(\lambda)]$ be regular, it is essential that the corresponding Jacobian is not identical to 0. This is a serious limitation, and we would like to extend the notion of a regular solution to as many solutions of the characteristic equation as possible.

This can be done by a slight modification of the characteristic equation. Instead of Eq. (38) we observe the equation

$$\{H+\epsilon A+\lambda[V(x)-\epsilon A]\}x=Ex, \quad x_1=1, \tag{47}$$

where operator A satisfies requirement (A45). We call this equation the modified characteristic equation (MCE). For $\epsilon=0$ it reduces to the usual CE, while for $\lambda=1$ it represents the CC equations. Formally, Eq. (47) can be considered as a characteristic equation of the CC method and a "modified" CI where Hamiltonian H is replaced with $H'=H+\epsilon A$, while at the same time operator V is replaced with $V'=V-\epsilon A$.

All the conclusions concerning the behavior and structure of the solutions of CE do apply to the solutions of MCE as well. As shown in the Appendix, for small enough ϵ, which we assume to be the case, MCE has exactly m regular solutions $[E(\lambda),x(\lambda)]$. Each of those solutions can have a finite number of singular points which are of four types: pole, branch point, accidental degeneracy, and hidden singularity. In a pole the solution $[E(\lambda),x(\lambda)]$ is not defined and at least one component of a vector $x(\lambda)$ blows up. Thus the pole corresponds to an eigenfunction which is formally orthogonal to Φ. In each neighborhood of a branch point there are at least two solutions which smoothly approach each other. In an accidental degeneracy point there are at least two solutions with the same energy. Finally, if the point λ is a hidden singularity point, none of the above holds, but the solution $[E(\lambda),x(\lambda)]$ is unstable and the corresponding Jacobian vanishes.

(g) Geometric interpretation of solutions

For each λ the CE is a set of m equations in m unknowns. Geometrically we can think of those equations as describing m hypersurfaces in an m-dimensional space X^m. Each solution is then a common intersection of those hypersurfaces. Such a picture can help us to understand the nature of resonance points and different types of singular points. In Fig. 3 are shown those different possibilities and their geometrical meaning. This figure deals with the simple case where we have only two hypersurfaces, $f_1(\lambda,E,x)=0$ and $f_2(\lambda,E,x)=0$. In Fig. 3 those hypersurfaces for a given λ value are shown. Figure 3(a) corresponds to a crossing of all hypersurfaces in a given point. In this point, $D\neq 0$ and hence it is a nonresonance solution. In Fig. 3(b) the two hypersurfaces do not cross; or, if one prefers, they meet in infinity. Hence there is no solution. However, the Jacobian still can be defined as the analytic continuation of the Jacobian in some neighboring λ point.

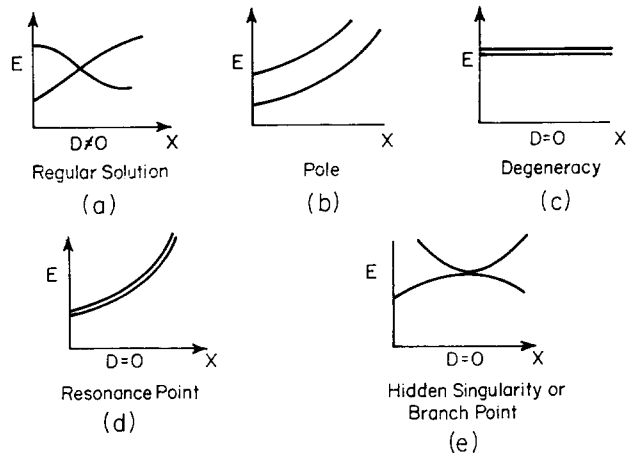

FIG. 3. Graphical depictions of possible solutions as intersections of m hypersurfaces in m-dimensional space, and their associated values of the Jacobian D.

In Figs. 3(c) and 3(d) both hypersurfaces do coincide which corresponds in a more general case to the dependence of hypersurfaces. In the former case, since E is always constant, we have degeneracy, while in the latter case it is a resonance point. Finally, Fig. 3(e) can describe either a hidden singularity or a branch point. In order to distinguish those two possibilities, we should draw at least a three-dimensional diagram including λ axes as well. In the first case (hidden singularity) the two hypersurfaces should only touch in some point. In a branch point, as we proceed along the λ axes starting from a point of touch, we should continuously arrive at two distinct crossing points corresponding to two solutions. It should be noted, however, that in reality two axes correspond to each axis: One for a real part and another for an imaginary part. Figure 3 is hence only approximate.

V. DISCUSSION

Exploiting the analytic connection of the CC with the CI method, we have come to the following main conclusions concerning the solutions of CC equations.

(1) Each MCE has as many regular solutions as the corresponding CI. Here, there are, in general, as many solutions to the CC equations as the number of solutions to the associated CI equations. The particular CI solution can be "lost" only if the point $\lambda=1$ is a pole or a branch point of a corresponding regular solution of MCE.

(2) Unless a branch point exists on the real axis in the complex λ plane between 0 and 1 the energy eigenvalue will be real for a real Hamiltonian. If the Hamiltonian matrix is complex, the energy might become complex.

(3) Solutions to the CC equations associated with highly excited states tend to become more and more orthogonal to Φ. They will also describe these states poorly. Therefore, it becomes more likely that those excited CC solutions will show singularities close to $\lambda=1$. Whether this causes the CC solution to be not defined ($\lambda=1$ is a pole) or has a complex energy value (branch point between 0 and 1 on a real axis) is dependent on the Hamiltonian matrix elements.

The reader will have noticed that little use was made of the actual algebraic form of Eqs. (38), peculiar for the CC equations. This is particularly the case when discussing the occurrence of resonance points, poles, and branch points. As evidenced by Lemmas 6 and 9 in the Appendix, it is important for proving the reality of the solution, but only to the extent that

$$\mathbf{V}(\mathbf{x}^*)=\mathbf{V}^*(\mathbf{x}).$$

This suggests that the developed procedure of analytically constructing the solutions to nonlinear, algebraic equations from a known problem (in this case the linear CI equations), is quite general. Thus, with some modification this method can be applied to Hartree–Fock (HF) equations as well. The modification is required since HF equations are not analytic in the unknowns. This follows from the fact that the HF potential depends on the products of the components of the eigenvectors Ψ_i with complex conjugates of those components. However, this difficulty can be overcome, and essentially the same method of the characteristic equation can be applied. One finds that, if HF equations describe a system consisting of N particles, and if there are M one-particle states (occupied plus unoccupied), then the characteristic equation corresponding to this HF has $\binom{M}{N}$ regular solutions. In order to obtain this result, it is necessary to redefine the characteristic equations and the definition of the regular solution in an appropriate way. This will be done elsewhere.[14]

APPENDIX

Lemma 1: Let T and C be cluster and configuration operators defined by

$$T=\sum_{s=1} T_s, \quad C=\sum_{s=1} C_s, \tag{A1}$$

where T_s and C_s,

$$\begin{aligned} T_s &= \sum_{\alpha_1\cdots\alpha_{s_1}r_1\cdots r_s} t^{r_1\cdots r_s}_{\alpha_1\cdots\alpha_s} b_{r_1}\cdots b_{r_s}a_{\alpha_s}\cdots a_{\alpha_1}, \\ C_s &= \sum_{\alpha_1\cdots\alpha_{s_1}r_1\cdots r_s} c^{r_1\cdots r_s}_{\alpha_1\cdots\alpha_s} b_{r_1}\cdots b_{r_s}a_{\alpha_s}\cdots a_{\alpha_1}, \end{aligned} \tag{A2}$$

are operators creating linear combinations of s-tuple p-h excitations. Further, let T and C be connected by the operator relation

$$e^T=1+C. \tag{A3}$$

The following two operator relations hold:

$$C_s=\sum_{k=1}^{s}\frac{1}{k!}\sum_{\{s_i\}}\delta(s_1+\cdots+s_k,s)\prod_{i=1}^{k}T_{s_i}, \tag{A4}$$

$$T_s=\sum_{k=1}^{s}\frac{(-)^{k-1}}{k}\sum_{\{s_i\}}\delta(s_1+\cdots+s_k,s)\prod_{i=1}^{k}C_{s_i}. \tag{A4'}$$

In Eq. (A4) and (A4') the sum "$\sum_{\{s_i\}}$" extends over all possible combinations of k numbers $s_1, s_2, \ldots, s_k$ such that $s_1+s_2+\cdots+s_k=s$.

Proof: Due to (A2) we have

$$[C_s, C_p]=[T_s, T_p]=0, \quad s,p=1,2,\ldots \tag{A5}$$

and hence

$$\begin{aligned} e^T &= \sum_{i=0} T^i/i! = \sum_{i=0}\Big(\sum_{p=1} T_p\Big)^i/i! \\ &= 1+T_1+(T_2+T_1^2/2)+(T_3+T_1T_2+T_1^3/6)+\cdots \\ &= 1+\sum_{s=1}\Big[\sum_{k=1}^{s}\frac{1}{k!}\sum_{\{s_i\}}\delta(s_1+\cdots+s_k,s)\prod_{i=1}^{k}T_{s_i}\Big], \end{aligned} \tag{A6}$$

where the last expression can be obtained by a simple combinatorial analysis. Comparison with (A3) proves (A4) now.

The inverse relation (A4') can be proven along the same lines using the operator relation

$$T=\ln(1+C)=\sum_{i=1}^{\infty}\frac{(-)^i}{i}C^i. \tag{A7}$$

(A7) is formally valid only if $|C|<1$. However, if the right-hand side in (A7) is applied to any state Ψ consisting of a finite number of excitations this series is finite and hence (A7) holds. Thus (A4') deduced from (A7) is valid for all finite-dimensional cases.

Theorem 1: Let

$$\begin{aligned} &\langle\Phi_\Delta^R(s)\,|\,e^{-T}(H-E)e^T\,|\,\Phi\rangle=0, \\ &T=T_1+\cdots+T_n, \quad s=0,1,\ldots,n, \end{aligned} \tag{A8}$$

be a set of CC equations including up to n excitations. This set of equations is equivalent to the CI-like equations,

$$\begin{aligned} &\langle\Phi_\Delta^R(s)\,|\,(H-E)\{1+C+[C_{n+1}(n)+C_{n+2}(n)]\}\,|\,\Phi\rangle=0, \\ &C=C_1+\cdots+C_n, \quad s=0,1,\ldots,n, \end{aligned} \tag{A9}$$

where

$$\begin{aligned} C_{n+1}(n) &= \sum_{k=2}^{n+1}\frac{(-)^k}{k}\sum_{\{s_i\}}\delta(s_1+\cdots+s_k,n+1)\prod_{i=1}^{k}C_{s_i}, \\ C_{n+2}(n) &= \sum_{k=2}^{n+1}\frac{(-)^k}{k}\sum_{\{s_i\}}\delta(s_1+\cdots+s_k,n+1)\,C_1\prod_{i=1}^{k}C_{s_i} \\ &\quad -\sum_{k=3}^{n+2}\frac{(-)^k}{k}\sum_{\{s_i\}}\delta(s_1+\cdots+s_k,n+2)\prod_{i=1}^{k}C_{s_i}. \end{aligned} \tag{A10}$$

Operators C_s and T_s for $s\le n$ are connected by the relations (A4) and (A4').

Proof: Let the space of states consist of, at most, N-tuple excitations. Observe (A3). Formally $T_s=C_s=0$ if $s>N$, while for the case $s\le N$ the connection between C_s and T_s is given by (A4) and (A4'). Now let T terminate after the nth term $T=T_1+T_2+\cdots+T_n$, and let us express C_s in terms of such T. In order to denote that C_s depends on n we will write $C_s(n)$. From (A4) it follows that $C_s(n)=C_s$ if $s\le n$. Let us now look at $C_{n+1}(n)$. This operator has the same dependence on T_s as the operator C_s, provided we put $T_{n+1}=0$ in (A4). Hence, from (A4),

$$C_{n+1}(n)=C_{n+1}-T_{n+1}. \tag{A11}$$

By the same argument, we obtain

$$C_{n+2}(n)=C_{n+2}-(T_{n+2}+T_1T_{n+1}). \tag{A12}$$

From (A4),

$$T_{n+1}=C_{n+1}+\sum_{k=2}^{n+1}\frac{(-)^{k-1}}{k}\sum_{\{s_i\}}\delta(s_1+\cdots+s_k,n+1)\prod_{i=1}^{k}C_{s_i},$$

$$T_{n+2}=C_{n+2}-C_1C_{n+1}+\sum_{k=3}^{n+2}\frac{(-)^{k-1}}{k}\sum_{\{s_i\}}\delta(s_1+\cdots+s_k,n+2)\times\prod_{i=1}^{k}C_{s_i}, \tag{A13}$$

and hence Eq. (A10) follows. This proves that (A8) is equivalent to

$$\langle\Phi_\Delta^R(s)\,|\,e^{-T}(H-E)\{1+C+[C_{n+1}(n)+C_{n+2}(n)]\}\Phi\rangle=0,\quad C=C_1+\cdots+C_n,\quad s=0,1,\ldots,n, \tag{A14}$$

where $C_{n+1}(n)$ and $C_{n+2}(n)$ are given by (A10). Operators C_{n+3}, C_{n+4}, etc., creating $(n+3)$-, $(n+4)$-tuple excitations do not enter into (A14), since Hamiltonian H destroys at most two excitations, operator $\exp(-T)$ can only create excitations and the state $\langle\Phi_\Delta^R(s)|$ contains at most n excitations.

By induction, we now conclude that operator $\exp(-T)$ can be omitted in (A14). First, observe the case $s=0$. Obviously,

$$\langle\Phi\,|\,e^{-T}(H-E)\{1+C+[C_{n+1}(n)+C_{n+2}(n)]\}\,|\,\Phi\rangle\equiv\langle\Phi\,|\,(H-E)\{1+C+[C_{n+1}(n)+C_{n+2}(n)]\}\,|\,\Phi\rangle=0, \tag{A15}$$

since $\exp(-T)=1-T+T^2/2!+\cdots$ can only create excitations and there is no excitation in $\langle\Phi|$. Second, we have

$$\langle\Phi_\Delta^R(1)\,|\,e^{-T}(H-E)\{1+C+[C_{n+1}(n)+C_{n+2}(n)]\}\,|\,\Phi\rangle\equiv\langle\Phi_\Delta^R(1)\,|\,(H-E)\{1+C+[C_{n+1}(n)+C_{n+2}(n)]\}\,|\,\Phi\rangle-\langle\Phi_\Delta^R(1)\,|\,T(H-E)\{1+C+[C_{n+1}(n)+C_{n+2}(n)]\}\,|\,\Phi\rangle=0. \tag{A16}$$

Operator T, when acting on any ket $|\Psi\rangle$, creates at least one more particle-hole pair in it. Therefore, according to (A15) the second term in (A16) is zero and hence

$$\langle\Phi_\Delta^R(1)\,|\,(H-E)\{1+C+[C_{n+1}(n)+C_{n+2}(n)]\}\,|\,\Phi\rangle=0. \tag{A17}$$

By induction it follows

$$\langle\Phi_\Delta^R(s)\,|\,(H-E)\{1+C+[C_{n+1}(n)+C_{n+2}(n)]\}\,|\,\Phi\rangle=0,\quad s=0,1,\ldots,n. \tag{A18}$$

This set of equations is equivalent to (A14) and hence to (A8). Theorem 1 is thus proved.

A kind of inverse theorem also holds:

Theorem 2: Let

$$\langle\Phi_\Delta^R(s)\,|\,(H-E)(1+C)\,|\,\Phi\rangle=0,\quad C=C_1+\cdots+C_n,\quad s=0,1,\ldots,n \tag{A19}$$

be a set of CI equations including up to n-tuple excitations. This set of equations is equivalent to the CC-like equations

$$\langle\Phi_\Delta^R(s)\,|\,e^{-T(n)}(H-E)e^{T(n)}\,|\,\Phi\rangle=0,\quad s=0,1,\ldots,n, \tag{A20}$$

where

$$T(n)=T_1+\cdots+T_n+T_{n+1}(n)+T_{n+2}(n),$$

$$T_{n+1}(n)=-\sum_{k=2}^{n}\frac{1}{k!}\sum_{\{s_i\}}\delta(s_1+\cdots+s_k,n+1)\prod_{i=1}^{k}T_{s_i},$$

$$T_{n+2}(n)=\sum_{k=2}^{n}\frac{1}{k!}\sum_{\{s_i\}}\delta(s_1+\cdots+s_k,n+1)T_1\prod_{i=1}^{k}T_{s_i}-\sum_{k=3}^{n}\frac{1}{k!}\sum_{\{s_i\}}\delta(s_1+\cdots+s_k,n+2)\prod_{i=1}^{k}T_{s_i}. \tag{A21}$$

Operators C_s and T_s for $s\leq n$ are connected by the relations (A4) and (A4′). The above theorem can be proven along the same lines as Theorem 1. Both theorems express a formal symmetry between CI and CC equations.

Observe the equations:

$$\langle\Phi_\Delta^R(s)\,|\,(H-E)\{1+C+\cdots+C_n+\lambda[C_{n+1}(n)+C_{n+2}(n)]\}\,|\,\Phi\rangle=0, \tag{A22}$$

$s=0,1,\ldots,n$.

These equations explicitly include up to n-tuple excitations while $(n+1)$ and $(n+2)$-tuple excitations are included through the operators $C_{n+1}(n)$ and $C_{n+2}(n)$, which depend on C_1 to C_n, according to (A10). For $\lambda=1$ this set of equations reduces to (A9), hence it is equivalent to CC equations including up to nth order clusters T_s. For $\lambda=0$, Eqs. (A24) reduce to a usual CI problem including up to n-tuple excitations.

Lemma 2: Equations (A22) can be written in the matrix form

$$[H+\lambda V(\Psi)]\Psi=E\Psi,\quad \Psi_{00}=1, \tag{A23}$$

where H and V are matrices with elements

$$H_{R\Delta,\,R'\Delta'}=\langle\Phi_\Delta^R\,|\,H\,|\,\Phi_{\Delta'}^{R'}\rangle,$$

$$V_{R\Delta,\,R'\Delta'}=V_{R\Delta}\delta_{R'\Delta',00}=\langle\Phi_\Delta^R\,|\,H[C_{n+1}(n)+C_{n+2}(n)]\,|\,\Phi\rangle\delta_{R'\Delta',00}, \tag{A24}$$

and Ψ is a column vector with components

$$\Psi_{R\Delta}(s)=\langle\Phi_\Delta^R(s)\,|\,C_s\,|\,\Phi\rangle=c_{\alpha_1\cdots\alpha_s}^{r_1\cdots r_s},\quad s=0,1,\ldots,n. \tag{A25}$$

Lemma 2 follows directly from (A22). The requirement $\Psi_{00}=1$ fixes the norm and phase of Ψ. Formally (A23) is an eigenvalue equation. Note that according to (A24), operator V is not Hermitian, and hence the reality of E is not guaranteed. We will call Eqs. (A22) and (A23) characteristic equations (CE). We shall analyze the solutions (eigenvectors and eigenvalues) of the CE as a function of λ. Besides CE (A23) we will sometimes consider the equation

$$\{H+\epsilon A+\lambda[V(\Psi)-\epsilon A]\}\Psi=E\Psi,\quad \Psi_{00}=1, \tag{A23′}$$

where A is a real Hermitian operator. Later on [see Eq. (A45)] we will impose some additional requirements on A. We will call Eq. (A23′) the modified characteristic equation (MCE). Obviously, for $\epsilon=0$ this equation reduces to CE. Note also that for $\lambda=1$ it reduces to CC equations. In what follows, we will refer mainly to CE. However, all the conclusions, unless otherwise specified, apply to MCE as well.

We now first examine some properties of the operator V and its matrix elements $V_{R\Delta}$.

Lemma 3: Each $V_{R\Delta}$ is either identically zero or it is a polynomial in the components of the vector Ψ.

(a) All matrix elements $V_{R\Delta}(s)$, corresponding to the s-tuple excitations where $s < n-1$ are identically zero.

(b) If there is no $(n+1)$-tuple excitation $|\Phi^{R'}_{\Delta'}(n+1)\rangle$ such that

$$\langle \Phi^{R}_{\Delta}(n-1)\,|\,H\,|\,\Phi^{R'}_{\Delta'}(n+1)\rangle \neq 0, \tag{A26}$$

the matrix element $V_{R\Delta}(n-1)$ corresponding to $(n-1)$-tuple excitations $|\Phi^{R}_{\Delta}(n-1)\rangle$ is identically zero. Otherwise $V_{R\Delta}(n-1)$ is polynomial in the components of Ψ of degree $(n+1)$ and with smallest power 2.

(c) If there is no $(n+1)$-tuple excitations $|\Phi^{R'}_{\Delta'}(n+1)\rangle$ such that

$$\langle \Phi^{R}_{\Delta}(n)\,|\,H\,|\,\Phi^{R'}_{\Delta'}(n+1)\rangle \neq 0 \tag{A27}$$

and at the same time no $(n+2)$-tuple excitations $|\Phi^{R'}_{\Delta'}(n+2)\rangle$ such that

$$\langle \Phi^{R}_{\Delta}(n)\,|\,H\,|\,\Phi^{R'}_{\Delta'}(n+2)\rangle \neq 0 \tag{A28}$$

matrix element $V_{R\Delta}(n)$ corresponding to n-tuple excitations $|\Phi^{R}_{\Delta}(n)\rangle$ is identically zero. If only (A26) holds, $V_{R\Delta}$ is a polynomial in the components of Ψ of the degree $(n+1)$ with smallest power 2 [same case as (b)]. If (A29) holds $V_{R\Delta}(n)$ is polynomial in components of Ψ of degree $(n+2)$ with smallest power 2.

Summarizing, Lemma 3 states that unless $V_{R\Delta}$ is identically zero, it is a polynomial of Ψ components of degree at most $(n+2)$ and at least 2.

Lemma 4: Polynomial $V_{R\Delta}$ contains each component of Ψ with power of at most 1.

Lemma 5: If the Hamiltonian H is real, matrix elements $V_{R\Delta}(\Psi)$ satisfy

$$[V_{R\Delta}(\Psi)]^{*} = V_{R\Delta}(\Psi^{*}) \tag{A29}$$

Lemma 6: If the Hamiltonian H is real, and if (E_0, Ψ_0) is a solution of CE for $\lambda = \lambda_0$, then (E_0^{*}, Ψ_0^{*}) is a solution of the same CE in the point $\lambda = \lambda_0^{*}$.

All these Lemmas follow from (A10) and (A24). For example,

$$\begin{aligned} V_{R\Delta}(n-1) &= \langle \Phi^{R}_{\Delta}(n-1)\,|\,H[C_{n+1}(n) + C_{n+2}(n)]\,|\,\Phi\rangle \\ &= \sum_{k=2}^{n+1} \frac{(-)^k}{k} \sum_{\{s_i\}} \delta(s_1 + \cdots + s_k, n+1) \\ &\quad \times \sum_{R'\Delta'} \langle \Phi^{R}_{\Delta}(n-1)\,|\,H\,|\,\Phi^{R'}_{\Delta'}(n+1)\rangle \\ &\quad \times \langle \Phi^{R'}_{\Delta'}(n+1) \prod_{i=1}^{k} C_{s_i}\,|\,\Phi\rangle . \end{aligned} \tag{A30}$$

If in the above expressions there is at least one vector $|\Phi^{R'}_{\Delta'}(n+1)\rangle$ such that relation (A26) holds, the term in (A30) corresponding to $k = n+1$ gives rise to the highest power $(n+1)$ in the components of Ψ. The lowest power is 2, corresponding to the term with $k=2$ in (A30). In the same way, matrix elements $V_{R\Delta}(n)$ can be analyzed. Lemma 4 follows from the fact that $a_i^2 = 0$ for each Fermion annihilation operator a_i^2. Since operators C_s can only create particle-hole pairs and never destroy them, once such pairs are created the repeated application of the same $p-h$ creation operators on the new state gives zero. Lemma 5 can be deduced straightforwardly from the explicit expressions for $V_{R\Delta}$, such as (A30), and from the assumed reality of the matrix elements of **H**. Lemma 6 follows from Lemma 5. The following lemma is also of some interest.

Lemma 7: Operator V is either identically zero or it is nonbounded.

Proof: Observe first that according to (A23) operator V is defined only for Ψ such that $\Psi_{00} = 1$. From Eq. (A24) it follows that

$$\Psi'_{R\Delta} = (V\Psi)_{R\Delta} = V_{R\Delta}(\Psi). \tag{A31}$$

Assume now that V is not identically zero. According to Lemma 3, there is $(R'\Delta')$ such that $V_{R'\Delta'}$ is a polynomial in the components of Ψ of the degree at least 2. Hence

$$\frac{|V\Psi|^2}{|\Psi|^2} = \frac{\sum_{R\Delta} |V_{R\Delta}(\Psi)|^2}{\sum_{R\Delta} |\Psi_{R\Delta}|^2} \geq \frac{|V_{R'\Delta'}(\Psi)|^2}{\sum_{R\Delta} |\Psi_{R\Delta}|^2}. \tag{A32}$$

However, $|V_{R'\Delta'}(\Psi)|^2$ is a polynomial of at least fourth degree, while $\sum_{R\Delta} |\Psi_{R\Delta}|^2$ is a polynomial of a degree 2, and hence the right-hand side in Eq. (A32) is not bounded. It follows that V is not bounded either.

Observe that although V is not bounded it is continuous since its matrix elements are polynomial (continuous) functions of the components of a vector Ψ.

We would now like to examine the properties of CE as a function of the parameter λ. Examples given in Sec. III suggest the following definitions:

Definition 1: Point λ_r in a complex λ plane is a resonance point of a CE if for $\lambda = \lambda_r$ the CE has an infinite number of eigenvalues.

Definition 2: Point λ_n in a complex λ plane is a nonresonance point of a CE if, for $\lambda = \lambda_n$ the CE has a finite number of eigenvalues.

The emphasis in Definitions 1 and 2 is on the number of eigenvalues and not on the number of eigenfunctions, which can be infinite due to degeneracy. It is clear that $\lambda = 0$ is a nonresonance point.

We now prove the following theorem.

Theorem 3: Let λ_r be a resonance point of a CE. Each value E, except a finite number of them, is an eigenvalue of the CE.

Proof: In order to prove this theorem, it is more convenient to renumerate the components $\Psi_{R\Delta}$ of a vector Ψ in such a way that we first take component $\Psi_{00} = 1$, then in some definite order all components corresponding to single excitations, then all components associated with double excitations, etc., up to n-tuple excitations. We denote the renumerated components by x_i, $i = 1, \ldots, m$. With this enumeration understood, we write, instead of Ψ, the vector x. Thus (A23) reads

$$[H + \lambda V(x)] = Ex, \quad x_1 = 1, \tag{A33}$$

where H and V are $m \times m$ square matrices with elements given by (A24), using the correspondence $\Psi_{R\Delta} \leftrightarrow x_i$. The matrix equation (A33) is a system of m equations in m unknowns:

$$
\begin{aligned}
&f_1 \equiv h_{11} + h_{12}x_2 + \cdots + h_{1m}x_m + \lambda V_1(x) - E = 0,\\
&f_2 \equiv h_{21} + h_{22}x_2 + \cdots + h_{2m}x_m + \lambda V_2(x) - Ex = 0,\\
&\quad\cdot\\
&\quad\cdot\\
&\quad\cdot\\
&f_m \equiv h_{m1} + h_{m2}x_2 + \cdots + h_{mm}x_m + \lambda V_m(x) - Ex_m = 0,
\end{aligned}
\tag{A34}
$$

where the unknowns are $E, x_2, \ldots, x_m$. Each polynomial f_i can be uniquely factorized in a set of $k(i)$ irreducible polynomials $f_i^{j_i}$, $j_i = 1, \ldots, k(i)$. For example, the polynomial

$$f = x_1^3 - x_2^3 + x_1x_2^2 - x_1^2x_2 + x_2 - x_1 = 0$$

can be factorized as

$$f \equiv (x_1^2 - x_2^2 - 1)(x_1 - x_2) = 0.$$

Hence the set (A34) is equivalent to $K \equiv \prod_i k(i)$ sets of equations

$$
\begin{aligned}
&f_1^{j_1} = 0, \quad j_1 = 1, 2, \ldots, k(1),\\
&f_2^{j_2} = 0, \quad j_2 = 1, 2, \ldots, k(2),\\
&f_m^{j_m} = 0, \quad j_m = 1, 2, \ldots, k(m).
\end{aligned}
\tag{A35}
$$

Each equation in (A35) represents a hypersurface in the m-dimensional space X^m spanned by $E, x_2, \ldots, x_m$. Geometrically, the solution of (A34) is the joint intersection of all hypersurfaces $f_1^{j_1} \cdots f_m^{j_m}$. Assume now that the point $\lambda = \lambda_r$ is a resonance point of (A34). In this case, there is at least one set (A35) which has an infinite number of intersections in X^m with different values of E. Due to the algebraic structure of the $f_i^{j_i}$ this can happen only if this set is dependent. But this means that it can be solved step by step to lead finally to an equation of the type

$$g[E, x_2(x_i), \ldots, x_{i-1}(x_i), x_i, x_{i+1}(x_i), \ldots, x_m(x_i)] = 0, \tag{A36}$$

which is algebraic in $E, x_2, \ldots, x_m$, where all $x_s(x_i)$ are algebraic in x_i. Therefore (A36) is algebraic in E and x_i, and, being satisfied for a denumerable number of distinct E values, it is satisfied for each E except, possibly, for a finite number of them. This proves Theorem 3.

Related to this Theorem is the following theorem.

Theorem 4: Each CE has either a finite number of resonance points, or each λ is a resonance point except for a finite number of them.

This theorem is proven along the same lines as the preceding one. Assume namely that a given CE has an infinite number of resonance points. From this infinite number, we can extract a denumerable subset of them. According to Theorem 3, for each resonance point the CE is satisfies for all E except for a finite number of them. It follows that for a denumerable number of resonance points there is a set consisting of all E except for, at most, a denumerable number of them. However, due to the algebraic character of the CE exclusion can apply only to a finite number of them. This means that the equations

$$f_i(\lambda, E, x) = 0 \tag{A37}$$

are satisfied for a denumerable number of λ values and for all E with the possible exception of a finite number of them. But then, since f_i are polynomials, (A37) should be satisfied for all λ with the exception of a finite number. This proves Theorem 4.

It is very likely that the number of resonance points can only be finite, meaning that the second possibility in Theorem 4 does not apply for realistic cases. However, we were unable to prove that.

Besides resonance points, the CE can also have singular points. In order to define these points, we first introduce the Jacobian of the CE which is defined for a given λ, E, and x_i:

$$D(\lambda, x, E) = \begin{vmatrix} \partial' f_1 & \cdots & \partial^m f_1 \\ \partial' f_m & \cdots & \partial^m f_m \end{vmatrix}, \tag{A38}$$

where

$$\partial' \equiv \frac{\partial}{\partial E}, \quad \partial^i = \frac{\partial}{\partial x_i}, \quad i = 2, \ldots, m.$$

In the case of CE we have

$$
D(\lambda, x, E) = -\begin{vmatrix}
1 & (h_{12} + \lambda\partial^2 V_1) & \cdots & (h_{1m} + \lambda\partial^m V_1)\\
x_2 & (h_{22} - E + \lambda\partial^2 V_2) & \cdots & (h_{2m} + \lambda\partial^m V_2)\\
\cdot & & & \\
\cdot & & & \\
\cdot & & & \\
x_m & (h_{m2} + \lambda\partial^2 V_m) & \cdots & (h_{mm} - E + \lambda\partial^m V_m)
\end{vmatrix}
\tag{A38'}
$$

and similarly for the MCE. Now the following theorem can be proven.

Theorem 5: Let the CE

$$[H + \lambda V(x)]x = Ex \tag{A39}$$

for $\lambda = \lambda_0$ have a solution (E_0, x_0). Further let the corresponding Jacobian be different from zero:

$$D(\lambda_0, E_0, x_0) \neq 0. \tag{A40}$$

Then there is $E(\lambda)$ and $x(\lambda)$ such that:

1. $E(\lambda)$ as well as each component of $x(\lambda)$ are analytic functions in the whole complex λ plane, with possible exceptions of, at most, a finite number of points.

2. $[E(\lambda), x(\lambda)]$ is a solution of a CE (A39) and for $\lambda = \lambda_0$ it concides with (E_0, x_0).

This solution $[E(\lambda), x(\lambda)]$ we shall call a *normal* solution. If $[E(\lambda), x(\lambda)]$ is defined in the point $\lambda = 0$ we shall call it *regular*. Also, eigenvalue $E(\lambda)$ and eigenfunction $x(\lambda)$ we shall call a normal (regular) eigenvalue and a normal (regular) eigenfunction, respectively. Note that by the very definition for each normal solution there is at least one point λ such that the corresponding Jacobian is different from zero.

Proof: The proof of the above Theorem 5 is based on the following Theorem 6.

Theorem 6: Let $g(x, y) = \{g_1, \ldots, g_n\}$ be continuous for vector $\mathbf{x}$ in a neighborhood of x_0 in R_m, and for vector $\mathbf{y}$ in a neighborhood y_0 in R_n, with $g(x_0, y_0) = 0$. Suppose g is continuously differentiable in y and that the determinant

$$J \equiv \left| \frac{\partial g_i(x_0, y_0)}{\partial y_j} \right| \neq 0. \tag{A41}$$

Then there is a neighborhood $N_1(x_0)$ in R_m and a neighborhood $N_2(y_0)$ in R_n such that for every x in N_1 there is a unique $y=\varphi(x)$ in N_2 for which $g(x,\varphi(x)]=0$. If $g(x,y)$ is k times continuously differentiable in x and y, then $\varphi(x)$ is k times continuously differentiable ($k\geq 1$).

The above theorem is proven in Ref. 15 for the case of real spaces R_m and R_n. However, it holds for the case of complex spaces as well, which can be proven along the same lines.

Putting $x\equiv\lambda$ and $y\equiv(E,x)$ the above theorem can immediately be applied to the CE. Since $V(x)$ is analytic in some neighborhood of a point $x=x_0$, $E(\lambda)$ and $x(\lambda)$ are analytic as well in some neighborhood $N(\lambda_0)$ of a point $\lambda=\lambda_0$ where a solution exists. Due to the algebraic character of the CE this solution shall be algebraic as well, and we can make an analytic continuation of the function $E(\lambda)$ and $x(\lambda)$ in the whole λ plane. Since these functions are algebraic in a small neighborhood $N(\lambda_0)$, they are algebraic in the whole λ plane as well. We have only to show that $[E(\lambda),x(\lambda)]$ is a solution of CE. In $N(\lambda_0)$ the solution $[E(\lambda),x(\lambda)]$ satisfies

$$f_i[\lambda,E(\lambda),x(\lambda)]=0. \tag{A42}$$

Since $E(\lambda)$ and $x(\lambda)$ are analytic in λ, and since f_i is analytic in $\lambda, E,$ and x, it is analytic in λ as well. On the other hand, f_i is identically zero in $N(\lambda_0)$ and, being analytic in λ, should therefore be identically zero for all λ. This proves that $[E(\lambda),x(\lambda)]$ satisfies CE for each λ where it is defined. However, $E(\lambda)$ and $x(\lambda)$ are algebraic functions of λ; hence they are defined for each λ with a possible exception of a finite number of poles and branch points.

It is interesting to see the meaning of the requirement (A40) in the above theorem. For that purpose, we use the following geometric interpretations:

Each equation $f_i(\lambda_0,E,x)=0$ for a given $\lambda=\lambda_0$ represents a finite number of hypersurfaces in the m-dimensional space X^m spanned by E and $x_2,\ldots,x_m$. The vector $\nabla f_i(\lambda_0,E,x)=\{\partial^1 f_i,\partial^2 f_i,\ldots,\partial^m f_i\}$ is orthogonal to the hypersurface in the point (E,x). The Jacobian (A38) consists of vectors $\nabla f_1,\nabla f_2,\ldots,\nabla f_m$ and hence the vanishing of this quantity implies a linear dependence of these vectors. $D\neq 0$ thus means that ∇f_1 to ∇f_m are not linearly dependent. If the point (E_0,x_0) satisfies $f_i(\lambda_0,E_0,x_0)=0$ it should lie on a common intersection of all hypersurfaces $f_i=0$. Moreover, if the corresponding ∇f_i are linearly independent, the vectors orthogonal to $f_i=0$ are linearly independent. Hence there is some neighborhood of point (E_0,x_0) in X^m where there is no other common intersection of hypersurfaces $f_i=0$. Thus the nonvanishing of the Jacobian [Eq. (A40)] implies that the solution (E_0,x_0) should be isolated in the space X^m.

Definition 3: Let (E_0,x_0) be a solution of the CE in the point $\lambda=\lambda_0$. Further, in each neighborhood of (E_0,x_0) let $(E_0,x_0')\neq(E_0,x_0)$ exist such that it is a solution of the CE for $\lambda=\lambda_0$ as well. We say that (E_0,x_0) is a continuously degenerate solution of the CE in the point $\lambda=\lambda_0$.

According to the definition, a continuously degenerate solution is not isolated. Also, Eq. (A36) implies that a resonance solution cannot be isolated. Hence, and in connection with the above conclusion, we did prove the following Lemmas.

Lemma 8: If (E_d,x_d) is a continuously degenerate solution of the CE for $\lambda=\lambda_d$, the corresponding Jacobian $D(\lambda_d,E_d,x_d)$ is equal to zero.

Lemma 9: If (E_r,x_r) is a resonance solution of the CE in a resonance point $\lambda=\lambda_r$, the corresponding Jacobian $D[\lambda_r,E(\lambda_r),x(\lambda_r)]$ is equal to zero.

Lemma 10: If $\lambda=\lambda_b$ is a branch point of a normal solution $[E(\lambda),x(\lambda)]$, the corresponding Jacobian $D(\lambda_b,E(\lambda_b),x(\lambda_b)]$ is equal to zero.

Namely if $D(\lambda_b)\neq 0$, by Theorem 5 the solution $[E(\lambda),x(\lambda)]$ should be analytic in the point $\lambda=\lambda_b$ which contradicts the assumption that this is a branch point.

The inverse is not true: Vanishing of the Jacobian does not necessarily mean that the point $\lambda=\lambda'$ is either a branch point, a resonance point, or a point of continuous degeneracy. The above geometrical interpretation implies that if $D(\lambda_0)=0$ the solution (E_0,x_0) should be unstable. Namely, if ∇f_i are linearly dependent, the point (E_0,x_0) is the common intersection of all $f_i=0$ to at least first order. Hence there are some vectors (E,x) in a space X^m which differ to the first order from (E_0,x_0), but also satisfy the CE to at least second order. Note that by the very definition, continuously degenerate solutions, solutions in a branch point, and resonance solutions are unstable as well.

Theorem 5 now provides a basis for the definition and classification of different types of singular points. Intuitively, each point where the normal solution $[E(\lambda),x(\lambda)]$ is not analytic, or is unstable due to any reason, we consider as singular. Thus singular points are singular points of a given normal solution $[E(\lambda),x(\lambda)]$, and each normal solution can have a different set of singular points. On the other hand, resonance points are characteristic of the CE as a whole. We can have the following singular points:

1. Point $\lambda=\lambda_p$ where solution $[E(\lambda),x(\lambda)]$ is not defined. Since $[E(\lambda),x(\lambda)]$ is algebraic in λ, such a point is a pole. As we approach $\lambda=\lambda_p$ at least one component of a vector $x(\lambda)$ tends to infinity. Hence

$$\lim_{\lambda\to\lambda_0}\frac{\langle x(\lambda)\,|\,\Phi\rangle}{[\langle x(\lambda)\,|\,x(\lambda)\rangle]^{1/2}}=0. \tag{A43}$$

This means that $x(\lambda)$ is more and more orthogonal to Φ. No vector orthogonal to Φ can be a solution to CE.

2. Point $\lambda=\lambda_b$ which is a branch point of the solution $[E(\lambda),x(\lambda)]$. Obviously, if λ_b is a branch point, $D(\lambda_b)$ should be equal to zero. If it were not, there would exist, according to Theorem 6, a unique analytic continuation of the solution $[E(\lambda_b),x(\lambda_b)]$ in some neighborhood of this point, which is contrary to the definition of a branch point. We call such a singularity a branch singularity.

3. Point $\lambda=\lambda_d$ where the normal solution $[E(\lambda),x(\lambda)]$ is continuously degenerate. According to Lemma 8, Jacobian $D(\lambda)$ vanishes in this point. We say that $\lambda=\lambda_d$ is a point of accidental degeneracy.

4. Point $\lambda=\lambda_h$ where the normal solution $[E(\lambda),x(\lambda)]$ is defined, the corresponding Jacobian is zero but the vanishing of the Jacobian is not due to any of the former reasons (branch point or accidental degeneracy). Such a point we call a point of a hidden singularity. As shown above, the solution $[E(\lambda),x(\lambda)]$ is unstable in this point.

In short, $\lambda=\lambda_0$ is a singular point of a normal solution $[E(\lambda),x(\lambda)]$ if either $[E(\lambda),x(\lambda)]$ is not defined in this point (pole) or, provided $[E(\lambda),x(\lambda)]$ is defined in $\lambda=\lambda_0$, the corresponding Jacobian vanishes (branch point, accidental degeneracy, and hidden singularity). Due to the algebraic character of the solution $[E(\lambda),x(\lambda)]$ and Jacobian $D(\lambda)$, there can be only a finite number of singular points. If it were not so, $D(\lambda)$ would be identically zero, contrary to the assumption that $[E(\lambda),x(\lambda)]$ is a normal solution.

The above definition of singular points applies only to the normal solution $[E(\lambda),\ x(\lambda)]$, i.e., such that there exist $\lambda=\lambda'$ with the property $D(\lambda')\neq 0$. However, we have no guarantee that in any point, including $\lambda=0$, there exists a solution with nonvanishing Jacobian. In order to overcome this limitation, instead of the CE (A23) we observe MCE (A23'). We chose a real Hermitian operator A in such a way that no eigenfunction of the equation

$$Ax=ax \tag{A44}$$

is orthogonal to Φ, and that all corresponding Jacobians are different from zero. In other words, A satisfies

$$Ax=ax \rightarrow \begin{cases} \langle x\,|\,\Phi\rangle \neq 0, \\ D(E,x)=-\begin{vmatrix} 1 & a_{12} & \cdots & a_{1m} \\ x_2 & a_{22}-E & \cdots & a_{2m} \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ x_m & a_{m2} & \cdots & a_{mm}-E \end{vmatrix} \neq 0, \end{cases} \tag{A45}$$

and

$$a_{ij}=a_{ij}^*=a_{ji}.$$

Obviously such an operator always exists which can be demonstrated by construction. We now prove Lemma 11.

Lemma 11: There exists $\epsilon_0\neq 0$ such that if $|\epsilon|\leq\epsilon_0$ and $\epsilon\neq 0$ no eigenfunction of the equation

$$(H+\epsilon A)x=Ex \tag{A46}$$

is orthogonal to Φ, and all corresponding Jacobians are different from zero.

Proof: If ϵ is big enough Eq. (A46) is close enough to Eq. (A44) and since A and H are bounded operators, each solution of Eq. (A46) is close enough to some solution of Eq. (A44), and vice versa. Hence for big enough ϵ, no solution of (A46) is orthogonal to Φ and all corresponding Jacobians are different from zero. However Eq. (A46) has the structure of a CE and, by Theorem 5, all those solutions are analytic in ϵ, with the possible exception of a finite number of ϵ points. It follows that overlaps $\langle x(\epsilon)\,|\,\Phi\rangle$ are analytic in ϵ, and hence can become zero only in a finite number of points. The same holds true for the corresponding Jacobians. Hence there is a small circle in the complex ϵ plane around a point $\epsilon=0$ where no solution of Eq. (A46) is orthogonal to Φ and where all the corresponding Jacobians are different from zero, with the possible exception of the point $\epsilon=0$. This proves Lemma 11.

We will consider the condition $\epsilon\neq 0$ and $|\epsilon|\leq\epsilon_0$ as a part of a definition of MCE. Also, without loss of generality, ϵ can be assumed real. With that in mind, and in combination with Theorem 5, the above Lemma proves the following Theorem 7.

Theorem 7: The modified characteristic equation has m regular solutions, i.e., the same number as the corresponding CI. This means that in each point $\lambda=\lambda'$ an MCE has m solutions which are analytically connected with m solutions of the CI equations (point $\lambda=0$), unless $\lambda=\lambda'$ happens to either be a pole or a branch point of some regular solution.

We can now state something about the reality of the solutions $[E(\lambda),x(\lambda)]$.

Lemma 12: Let the Hamiltonian H be real and let $[E(\lambda),x(\lambda)]$ be a regular solution of the CE. Assume furthermore that $[E(\lambda),x(\lambda)]$ has no branch point on the real axis between the real points $\lambda=0$ and $\lambda=\lambda_0$. Provided λ_0 is not a pole, the solution $[E(\lambda_0),x(\lambda_0)]$ exists and it is real.

Proof: Since $[E(\lambda),x(\lambda)]$ is regular, it is defined in $\lambda=0$. Since H is real, $[E(0),x(0)]$ will be real. We can now reach $\lambda=\lambda_0$ starting from $\lambda=0$ along some path P in the complex λ plane. P can always be chosen such that it avoids all possible singular points, and that the area enclosed by P and the real axis does not contain branch points [see Fig. 4]. Along P, we have the solution $[E(\lambda),x(\lambda)]$ which smoothly changes from $[E(0),x(0)]$ to $[E(\lambda_0),x(\lambda_0)]$. Now we can reach λ_0 as well along path P^* which is a complex conjugate of P. From Lemma 6, and because there are no branch points between P and the real axis, there is no branch point between P^* and the real axis either. Therefore, there is no branch point in the whole area enclosed by P and P^*. Hence the analytic continuation of $[E(0),x(0)]$ along either path will yield the same solutions in the point λ_0. However, from Lemma 6 and the reality of $[E(0),x(0)]$, it follows that

$$[E(\lambda),x(\lambda)]=[E^*(\lambda^*),x^*(\lambda^*)]$$

and hence, for $\lambda=\lambda_0=\lambda_0^*$,

$$[E(\lambda_0),x(\lambda_0)]=[E^*(\lambda_0),x^*(\lambda_0)] \tag{A47}$$

expressing the reality of the solution.

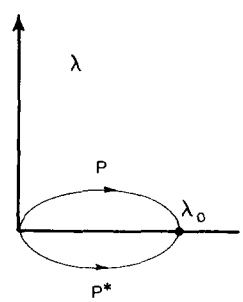

FIG. 4. Paths P and P^* in the complex λ plane connecting a configuration-interaction solution ($\lambda=0$) with that in point $\lambda=\lambda_0$ on the real axis. Solution $[E(\lambda_0),\ x(\lambda_0)]$ is real provided there is no branch-point on the real axis between 0 and λ_0.

Note in the above proof that the condition that there is no branch point between $\lambda = 0$ and $\lambda = \lambda_0$ on the real axis is essential since, if such a point would exist, the analytic continuations along P and P^* would give two different solutions—say, $[E_1(\lambda_0), x_1(\lambda_0)]$ and $[E_2(\lambda_0), x_2(\lambda_0)]$. All we could then infer is that $E_1(\lambda_0) = E_2^*(\lambda_0)$ and $x_1(\lambda_0) = x_2^*(\lambda_0)$. However, the absence of poles between 0 and λ_0 is not essential.

[1]J. Goldstone, Proc. R. Soc. London A **239**, 267 (1957).
[2]J. Hubbard, Proc. R. Soc. London A **240**, 539 (1957); **243**, 336 (1957).
[3]K.A. Brueckner, Phys. Rev. **100**, 36 (1955).
[4]B.H. Brandow, Rev. Mod. Phys. **39**, 771 (1967).
[5]P.W. Langhof and A.J. Hernandez, Int. J. Quant. Chem. **S10**, 337 (1976).
[6]F. Coester, Nucl. Phys. **7**, 421 (1958).
[7]H.J. Monkhorst, F.E. Harris, and D.L. Freeman (unpublished).
[8]J. Čižek, J. Chem. Phys. **45**, 4256 (1966); Adv. Chem. Phys. **14**, 35 (1969); see also Ref. 10.
[9]H. Kümmel, K.H. Lührmann, and J.G. Zabolitzky, "Many-Fermion Theory in exp-S Form," Phys. Rep. (in print) and references cited therein.
[10]J. Paldus, J. Čižek, and I. Shavitt, Phys. Rev. A **5**, 50 (1972) and references therein to their earlier work; J. Paldus in "Proceedings of Boulder Conference on Theoretical Chemistry," June 1972, Boulder, Colorado; J. Paldus and J. Čižek, in *Energy, Structure and Reactivity*, edited by D.W. Smith and W.B. Rae (Wiley, New York, 1973), p. 198.
[11]H.J. Monkhorst, Int. J. Quant. Chem. **S11**, 421 (1977).
[12]F. Coester, *Lectures in Theoretical Physics*, Vol. **11**, edited by K.T. Mahanthappa (Gordon & Breach, New York, 1969).
[13]T.P. Živković, Int. J. Quant. Chem. **S11**, 413 (1977).
[14]T.P. Živković (unpublished).
[15]T.L. Saaty and G. Bram, *Nonlinear Mathematics* (McGraw-Hill, New York, 1964), p. 39.

On the representation of linear operators in L^2 spaces by means of "generalized matrices"[a)]

R. Ascoli

Istituto Matematico del Politecnico di Torino, 10100 Torino, Italy

P. Gabbi and G. Palleschi

Istituto di Matematica dell'Università di Parma, Parma, Italy
(Received 20 October 1976)

This paper concerns the representation of linear operators of L^2 spaces by means of "generalized matrices" as it is usual, following Dirac, in quantum mechanics and in electronics. The known possibility of representing (on a nuclear test-function space) the bounded operators by means of distribution kernels is shown to extend to all the closable operators whose domain contains the test-function space (hence to all the Hermitian operators whose domain contains the Schwartz space $\mathcal{D}$ of the infinitely differentiable functions with compact support). The representation of the adjoint operator is considered, and the possibility of representing the product of operators by means of a suitably defined "Volterra convolution" is studied. In particular it is shown that $*$-algebras of unbounded operators (which are, for instance, generated by the canonical coordinates and momenta and the total energy of most quantum mechanical systems of n particles) may be represented isomorphically by means of distribution kernels, so that the Dirac's rules on "generalized matrices" apply in the sense of distributions without further assumptions.

1. INTRODUCTION

In recent years some remarkable papers[1] have appeared concerning the problem of giving a precise mathematical meaning to the "Dirac formalism."[2] The present note is intended as a contribution which is independent from using or not using rigged or equipped Hilbert spaces and concerns the particular problem[3] of representing any (linear) operator A of an L^2 space by means of a "generalized matrix" K^A, according to the formula

$$(Af)(t) = \int K^A(t, \tau) f(\tau)\, d\tau, \tag{1}$$

as it is usual in quantum physics and in electronics.

In his classical book[4] Dirac summarizes as follows the results that hold for ordinary matrices[5]:

(i) any linear operator is represented by a matrix,

(ii) the unit operator is represented by the unit matrix,

(iii) a real linear operator is represented by a Hermitian matrix,

(v) the matrix representing the product of two linear operators is the product of the matrices representing the two factors.

Then he states: "We arrange our definitions concerning these generalized matrices so that rules (i)–(v), which we had above for the discrete case, hold also for the continuous case."

It is known that it is not possible to represent every linear operator A of an $L^2(\Omega)$ space (Ω being an open set of R^n) by means of a kernel K^A which is a function on the square $\Omega\times\Omega$. However the situation becomes more favorable if the kernel K^A is allowed to be a "distribution kernel," that is a distribution on the square $\Omega\times\Omega$ [an element of $\mathcal{D}'(\Omega\times\Omega)$]. For instance, the identity operator I is represented by a Dirac δ, in accord with Dirac's statement (ii), the derivation operators by derivatives of the δ, and so on. Of course, it must be pointed out that such representations hold in general only for functions which belong to the space of the test functions[6] and the integral of the formula (1) has to be thought of in the sense of distributions.

The mathematical tool which seems to be the most suitable for this problem is the "kernels theorem" of distribution theory and it is indeed remarkable that Schwartz, presenting this highly nontrivial result at the International Congress of Mathematicians of 1950 at Cambridge,[7] has motivated his work by referring to the Dirac representation of linear operators. It is actually an immediate consequence of this theorem that any bounded linear operator of $L^2(\Omega)$ may be represented in the form (1) by means of a distribution kernel.

The main result of the present note (Theorem 1 below) is that such a representation is also possible for any closable linear operator of $L^2(\Omega)$ whose domain contains $\mathcal{D}(\Omega)$ (the space of the infinitely differentiable functions with compact support). So the representation holds in particular for any Hermitian operator whose domain contains $\mathcal{D}(\Omega)$. The result is an easy consequence of another profound proposition, i.e., the closed graph theorem for linear mappings from a barrelled space into a Frêchet space, the barrelled space being here $\mathcal{D}(\Omega)$.

So Theorem 1, as well as the remaining statements of this paper, remains valid when everywhere replacing the Schwartz space $\mathcal{D}(\Omega)$ with the Schwartz space $\mathcal{S}(R^n)$ of the rapidly decreasing infinitely differentiable functions.

Theorem 1 may be considered as characterizing a

[a)] This work has been performed in part within the Gruppo Nazionale per la Fisica Matematica del C.N.R.

 0022-2488/78/1905-1023$1.00

range where the above statement (i) of Dirac holds, when interpreting his "generalized matrices" as distribution kernels and it is perhaps difficult to find in the physical or electronical applications any particular linear operator which does not satisfy the conditions that are required by Theorem 1, provided Ω be suitably restricted (see the example below).

The consideration of statements (iii) and (v) of Dirac leads to Proposition 1 and Theorem 2 below. The case of Hermitian conjugation is trivial: Whenever both A and A^* allow the "generalized" representation (1), the corresponding distribution kernels K^A and K^{A^*} are Hermitian conjugates of each other.

Concerning the product AB of two linear operators A and B, such that A, B, and AB allow the "generalized" representation (1), it is shown here that whenever B maps a dense vector subspace V of $\mathcal{D}(\Omega)$ into $\mathcal{D}(\Omega)$ [case (a) of Theorem 2], then the distribution kernel K^{AB} is given by an easily defined[8] "Volterra convolution" $K^A \circ K^B$ of the kernels K^A and K^B, which may well be symbolically written

$$K^{AB}(t, s) = \int K^A(t, \tau) K^B(\tau, s)\, d\tau, \tag{2}$$

corresponding to the product of "generalized matrices" considered by Dirac. {In more general cases the definition of an appropriate "Volterra convolution" of distribution kernels is more complex. Here one such case is also treated [case (b) of Theorem 2]: The above considered subspace V of $\mathcal{D}(\Omega)$ is only required to be dense in $L^2(\Omega)$, provided the domain of $(AB|_{\mathcal{D}(\Omega)})^*$ contains $\mathcal{D}(\Omega)$.}

In the last statement (Theorem 3 below) the previous theorems are applied to the particularly favorable case of the $*$-algebra $C_{\mathcal{D}(\Omega)}$ of the linear operators that map $\mathcal{D}(\Omega)$ into $\mathcal{D}(\Omega)$ and have an adjoint which maps $\mathcal{D}(\Omega)$ into $\mathcal{D}(\Omega)$. Then all the requirements considered above are satisfied and the representation of the operators by means of distribution kernels provides an isomorphism of $*$-algebras (the product of the distribution kernels consisting in the Volterra convolution); that is Dirac's rules quoted above apply in the sense of distributions without further assumptions.

Using an easy proposition of Roberts[9] this last situation is seen to occur for instance for the $*$-algebra generated by the operators P_i, Q_i, and H that describe the canonical coordinates and momenta and the total energy in the wave mechanical description of a system of n particles with a potential energy W which is C^∞ on some open set whose complement has zero Lebesgue measure, provided this set is chosen as[10] Ω. So, in this example, all Dirac's rules considered above apply to the representation of the operators on $\mathcal{D}(\Omega)$.

Of course, in this example, the representation of many of the operators holds not only on the subspace $\mathcal{D}(\Omega)$ but even on some larger subspace of $L^2(R^n)$, such as $\mathcal{D}(R^n)$ or $\mathcal{S}(R^n)$: According to Theorem 1, this situation occurs for the operators whose domain contains such a larger subspace. When furthermore the product of two such operators is also defined on such a subspace, then even the rule concerning the representation of the product may be applied to the larger subspace, at least when the requirements of Theorem 2 are satisfied, for instance when also the adjoint of the product is defined on the larger subspace.

Starting from a one-dimensional one-particle particularization of the preceding example, Appendix A illustrates the various situations considered throughout the paper. Appendix B refers to some domain questions which arise when describing operators by means of distribution kernels.

2. CONTINUITY PROPERTIES OF CLOSABLE LINEAR OPERATORS OF THE HILBERT SPACE

We will use the following proposition, which is a consequence of the closed graph theorem.

Lemma 1: Let A be any closable densely defined linear operator of a Hilbert space. Let D be a dense vector subspace of the domain D_A of A and let a topology $\mathcal{T}$ be given on D which is finer than the topology induced by H and makes D a barrelled space,[11] and let us call $D^{\mathcal{T}}$ the space D endowed with the topology $\mathcal{T}$.

Then:

(i) the restriction $A|_D$ of A to D is a continuous mapping from $D^{\mathcal{T}}$ into[12] H;

(ii) the transposed mapping[13] $(A|_D)^{\mathcal{T}t}$ of this mapping $A|_D$, when identifying H' with H and embedding it canonically into $(D^{\mathcal{T}})'$, is a continuous mapping from H into $(D^{\mathcal{T}})'$ (endowed with the weak or the strong topology), which extends A^*.

We note that, due to the fact that $A \subset A^{**}$, statement (ii) is particularly interesting when both D_A and D_{A^*} include D: Then $(A^*|_D)^{\mathcal{T}t}: H \to (D^{\mathcal{T}})'$ extends continuously $A: D_A \to (D^{\mathcal{T}})'$.

Proof: Since A is assumed to be a closable linear operator of H, then $A|_D$ also has these properties. Hence its graph is closed in the product space $D \times H$ and *a fortiori* it is closed in the product space $D^{\mathcal{T}} \times H$. This means that $A|_D$ is a closed linear mapping from $D^{\mathcal{T}}$ into H. Then, as a consequence of the assumption on $D^{\mathcal{T}}$, the closed graph theorem for linear operators from a barrelled space into a Fréchet space[14] may be applied and it follows that $A|_D$ is a continuous linear mapping from $D^{\mathcal{T}}$ into H. So statement (i) is proved. Concerning statement (ii), the continuity of the mapping $A|_D$ from $D^{\mathcal{T}}$ into H is known to imply the continuity (for both the weak and the strong topologies)[15] of its transposed $(A|_D)^{\mathcal{T}t}$ from H into $(D^{\mathcal{T}})'$, which obviously extends A^*.

3. REPRESENTATION OF LINEAR OPERATORS OF L^2 BY MEANS OF DISTRIBUTION KERNELS

The purpose of the next two definitions is to give a precise meaning to the statement that an operator A on $L^2(\Omega)$ is represented by a distribution kernel.

Definition 1: Let Ω be an open set of R^n. A distribution $K \in \mathcal{D}'(\Omega \times \Omega)$ on the open set $\Omega \times \Omega$ of R^n is usually

called a distribution kernel on Ω and it is written

$$K:\mathcal{D}(\Omega\times\Omega)\to C:\varphi\to\langle K_{t,\tau},\varphi(t,\tau)\rangle.$$

The Hermitian conjugate K^* of K is defined by $K^*_{t,\tau}=\overline{K}_{\tau,t}$ (where the bar means the complex conjugation); K is Hermitian if $K_{t,\tau}=\overline{K}_{\tau,t}$.

Moreover, for any test function $\varphi\in\mathcal{D}(\Omega)$, the distribution $\langle K_{\cdot,\tau},\varphi(\tau)\rangle\in\mathcal{D}'(\Omega)$ is defined by

$$\langle K_{\cdot,\tau},\varphi(\tau)\rangle:\mathcal{D}(\Omega)\to C:\psi\to\langle K_{t,\tau},\varphi(\tau)\psi(t)\rangle.$$

In this manner K determines the continuous linear mapping from $\mathcal{D}(\Omega)$ into $\mathcal{D}'(\Omega)$ defined by

$$\varphi\to\langle K_{\cdot,\tau},\varphi(\tau)\rangle \tag{3}$$

and the linear operator A^K in the Hilbert space $L^2(\Omega)$, defined on the vector subspace $D_{A^K}=\{\varphi\in\mathcal{D}(\Omega)\mid \langle K_{\cdot,\tau},\varphi(\tau)\rangle\in L^2(\Omega)\}$ by (3) [here, of course, $L^2(\Omega)$ is embedded into $\mathcal{D}'(\Omega)$ by identifying each $f\in L^2(\Omega)$ to the mapping[16]

$$\varphi\to\langle f,\varphi\rangle=(f\mid\bar{\varphi})=\int_\Omega f(t)\varphi(t)\,dt\Big]\ . \tag{4}$$

We note that for any $\varphi\in D_{A^K}$ the equalities

$$A^K\varphi=\langle K_{\cdot,\tau},\varphi(\tau)\rangle \tag{5}$$

and

$$\forall\psi\in\mathcal{D}(\Omega),\quad (A^K\varphi\mid\psi)=\langle K_{t,\tau},\varphi(\tau)\bar{\psi}(t)\rangle \tag{6}$$

are equivalent as a consequence of (4); actually (5) states the equality of the distributions and (6) states the equality of their values.

This last remark justifies the equivalence of the two conditions in the next definition.

Definition 2: We say that any linear operator A in $L^2(\Omega)$ is represented [on $\mathcal{D}(\Omega)$] by the distribution kernel $K^A\in\mathcal{D}'(\Omega\times\Omega)$ if the domain of A contains $\mathcal{D}(\Omega)$ and for any $\varphi\in\mathcal{D}(\Omega)$

$$A\varphi=\langle K^A_{\cdot,\tau},\varphi(\tau)\rangle,$$

that is,

$$(A\varphi)(t)=\langle K^A_{t,\tau},\varphi(\tau)\rangle\ \text{ a.e. on }\Omega$$

or equivalently

$$\forall\psi\in\mathcal{D}(\Omega),\quad (A\varphi\mid\psi)=\langle K^A_{t,\tau},\varphi(\tau)\bar{\psi}(t)\rangle.$$

Any distribution kernel K^A representing some linear operator A, is uniquely determined by the operator A. In fact, if K^A_1 also represents A, then for any $\varphi\in\mathcal{D}(\Omega)$, $\langle(K^A-K^A_1)_{\cdot,\tau},\varphi(\tau)\rangle=0$, hence $K^A=K^A_1$.

The next statement follows immediately.

Proposition 1: Let A be any linear operator of $L^2(\Omega)$. If A and A^* are represented by the distribution kernels K^A and K^{A^*} respectively, then K^A and K^{A^*} are Hermitian conjugates of each other, i.e., $K^{A^*}=(K^A)^*$.

We can now formulate the main statement.

Theorem 1: Let A be any closable linear operator of $L^2(\Omega)$ whose domain contains $\mathcal{D}(\Omega)$. Then A is represented by a distribution kernel $K^A\in\mathcal{D}'(\Omega\times\Omega)$.

Proof: $\mathcal{D}(\Omega)$ is known to be a barrelled space. Then Lemma 1 applies [with $H=L^2(\Omega)$ and $D=\mathcal{D}(\Omega)$] to the linear operator A: It follows that $A|_{\mathcal{D}(\Omega)}$ is continuous from $\mathcal{D}(\Omega)$ into $\mathcal{D}'(\Omega)$. Now we can apply the Schwartz kernels theorem.[17] According to this theorem the canonical homomorphism of $\mathcal{D}'(\Omega\times\Omega)$ into $L(\mathcal{D}(\Omega),\mathcal{D}'(\Omega))$ [the space of all continuous linear mappings from $\mathcal{D}(\Omega)$ into $\mathcal{D}'(\Omega)$] which to any $K\in\mathcal{D}'(\Omega\times\Omega)$ assigns the mapping

$$\mathcal{D}(\Omega)\to\mathcal{D}'(\Omega):\varphi\to\langle K_{\cdot,\tau},\varphi(\tau)\rangle$$

introduced after definition 1, is an isomorphism. That is, for any continuous linear mapping $B:\mathcal{D}(\Omega)\to\mathcal{D}'(\Omega):\varphi\to B\varphi$ there exists one distribution kernel $K\in\mathcal{D}'(\Omega\times\Omega)$ such that for any $\varphi\in\mathcal{D}(\Omega)$, $B\varphi=\langle K_{\cdot,\tau},\varphi(\tau)\rangle$. We apply this last statement to $B=A|_{\mathcal{D}(\Omega)}$. Then according to definition 2, operator A is represented by a distribution kernel K^A and the theorem is proved.

It is important to note that this theorem, as well as all the statements of Secs. 3 and 4, hold unchanged, when one replaces everywhere the Schwartz space $\mathcal{D}(\Omega)$ with any other nuclear barrelled subspace of $L^2(\Omega)$ satisfying the Schwartz kernels theorem, and in particular, if $\Omega=R^n$, the Schwartz space $\mathcal{S}(R^n)$ of the rapidly decreasing infinitely differentiable functions on R^n.

Corollary 1: Any symmetric operator A of $L^2(\Omega)$ whose domain contains $\mathcal{D}(\Omega)$ is represented by a Hermitian distribution kernel $K^A=(K^A)^*$.

In fact $A\subset A^{**}$ and $\mathcal{D}(\Omega)\subset D_A$ imply $\mathcal{D}(\Omega)\subset D_{A^*}$: So according to Theorem 1 both A and A^* are represented by distribution kernels, which coincide because $A|_{\mathcal{D}(\Omega)}=A^*|_{\mathcal{D}(\Omega)}$ implies $K^A=K^{A^*}$. It follows from Proposition 1 that $K^A=(K^A)^*$, that is, K^A is Hermitian.

4. ON THE REPRESENTATION OF THE PRODUCT OF OPERATORS IN L^2

This section concerns the representation of the product AB of two linear operators A and B of $L^2(\Omega)$ where A, B, and AB satisfy the conditions of Theorem 1, and consequently are closable operators whose domains contain $\mathcal{D}(\Omega)$.

The mutually corresponding conditions (a) of definition 3 and Theorem 2 refer to the simpler situation in which the vector subspace $V=\mathcal{D}(\Omega)\cap B^{-1}\mathcal{D}(\Omega)$ of $\mathcal{D}(\Omega)$, that B maps into $\mathcal{D}(\Omega)$, is dense in $\mathcal{D}(\Omega)$, whereas the mutually corresponding conditions (b) refer to a more complex situation in which V is only required to be dense in $L^2(\Omega)$ (examples are given in Appendix A).

The remark after Theorem 1, concerning the validity of all the statements when $\mathcal{D}(\Omega)$ is replaced by $\mathcal{S}(R^n)$, holds again.

Definition 3: Let H,K be two distribution kernels belonging to $\mathcal{D}'(\Omega\times\Omega)$. Then we define their "Volterra convolution" $H\circ K\in\mathcal{D}'(\Omega\times\Omega)$ by starting from the formula

$$\langle(H\circ K)_{\cdot,s},\varphi(s)\rangle=\langle H_{\cdot,\tau},\langle K_{\tau,s},\varphi(s)\rangle\rangle, \tag{7}$$

whose right-hand side, $P(\varphi)$, has a meaning on the vector subspace

$$V=\{\varphi\in\mathcal{D}(\Omega)\mid\langle K_{\cdot,s},\varphi(s)\rangle\in\mathcal{D}(\Omega)\}$$

of $\mathcal{D}(\Omega)$, in the following two (partially overlapping)

situations:

(a) $V=\mathcal{D}(\Omega)$ [more generally V needs only to be dense in $\mathcal{D}(\Omega)$] and the above defined mapping P from V into $\mathcal{D}'(\Omega)$ is continuous when V carries the topology of $\mathcal{D}(\Omega)$: then $H\circ K$ is defined uniquely by formula (7);

(b) V is dense in $L^2(\Omega)$ and the mapping P is continuous when V carries the topology of $L^2(\Omega)$: then $H\circ K$ is defined uniquely by extending this continuous mapping to a mapping $\hat{P}$ from $\mathcal{D}(\Omega)$ into $\mathcal{D}'(\Omega)$ and identifying the left-hand side of (7) to $\hat{P}(\varphi)$.

The existence of one distribution kernel satisfying (7) follows from the Schwartz kernels theorem, because the assumptions allow in any case to extend P to a continuous linear mapping from $\mathcal{D}(\Omega)$ into $\mathcal{D}'(\Omega)$.

Theorem 2: Let A, B two linear operators of $L^2(\Omega)$. If A, B, AB are represented respectively by the distribution kernels K^A, K^B, K^{AB}, then we have for any $\varphi\in V=\mathcal{D}(\Omega)\cap B^{-1}\mathcal{D}(\Omega)$,

$$\langle K^{AB}_{\circ,s},\varphi(s)\rangle=\langle K^{A}_{\circ,\tau},\langle K^{B}_{\tau,s},\varphi(s)\rangle\rangle. \tag{8}$$

Moreover, with reference to definition 3, $K^{AB}=K^A\circ K^B$ if at least one of the following conditions is satisfied:

(a) $V=\mathcal{D}(\Omega)$ [more generally V is dense in $\mathcal{D}(\Omega)$]

(b) V is dense in $L^2(\Omega)$ and the domain of $(AB|_{\mathcal{D}(\Omega)})^*$ contains $\mathcal{D}(\Omega)$.

Proof: The first statement of the theorem comes out from repeated application of definition 2: In fact, for any $\varphi\in V$

$$\langle K^{AB}_{\circ,s},\varphi(s)\rangle=AB\varphi=\langle K^{A}_{\circ,\tau},(B\varphi)(\tau)\rangle$$
$$=\langle K^{A}_{\circ,\tau},\langle K^{B}_{\tau,s},\varphi(s)\rangle\rangle.$$

Let us prove the remaining statements. According to definition 3 we have to consider the mapping

$$P=AB|_V: V\to\mathcal{D}'(\Omega):\varphi$$
$$\to\langle K^{A}_{\circ,\tau},\langle K^{B}_{\tau,s},\varphi(s)\rangle\rangle=AB\varphi.$$

In case (a) the assumption that AB is represented by a distribution kernel K^{AB} implies the mapping P to be continuous when V carries the topology induced by $\mathcal{D}(\Omega)$. Hence condition (a) of definition 3 is satisfied.

In case (b), the assumption that the domain of $(AB|_{\mathcal{D}(\Omega)})^*$ contains $\mathcal{D}(\Omega)$ implies, according to the remark following Lemma 1, the linear operator $AB|_{\mathcal{D}(\Omega)}$ to be continuous from $\mathcal{D}(\Omega)$, endowed with the topology of $L^2(\Omega)$, into $\mathcal{D}'(\Omega)$, so that the mapping P is continuous when V carries the topology of $L^2(\Omega)$. Hence condition (b) of definition 3 is satisfied.

So, in both cases (a) and (b), definition 3 characterizes through continuous extension the Volterra convolution $K^A\circ K^B$. By comparing formula (7) of definition 3 with the first statement of this theorem, we obtain

$$\forall\,\varphi\in V,\quad\langle(K^A\circ K^B)_{\circ,s},\varphi(s)\rangle=\langle K^{AB}_{\circ,s},\varphi(s)\rangle.$$

Taking into account, in both cases (a) and (b), the density properties that have been assumed for V, together with the continuity properties of the right-hand side that have been remarked upon, this identity extends to $\mathcal{D}(\Omega)$, so that $K^A\circ K^B=K^{AB}$. So the theorem is proved.

The next proposition concerns a situation, which occurs frequently in the applications, where all the requirements of Theorems 1 and 2 and Proposition 1 are satisfied.

Theorem 3: Let $C_{\mathcal{D}(\Omega)}$ be the $*$-algebra which consists of the linear operators of $L^2(\Omega)$ with domain $\mathcal{D}(\Omega)$ that map $\mathcal{D}(\Omega)$ into $\mathcal{D}(\Omega)$ and whose adjoints map $\mathcal{D}(\Omega)$ into $\mathcal{D}(\Omega)$ [the $*$-operation transforming any operator into the restriction to $\mathcal{D}(\Omega)$ of its adjoint].[18]

Let $K_{\mathcal{D}(\Omega)}$ be the $*$-algebra which consists of the distribution kernels $K\in\mathcal{D}'(\Omega\times\Omega)$ such that for any $\varphi\in\mathcal{D}(\Omega)$ $\langle K_{\cdot,\tau},\varphi(\tau)\rangle\in\mathcal{D}(\Omega)$ and $\langle K^*_{\cdot,\tau},\varphi(\tau)\rangle\in\mathcal{D}(\Omega)$, the product consisting in the Volterra convolution [as defined in Definition 3(a)].

Then the mapping

$$j: C_{\mathcal{D}(\Omega)}\to K_{\mathcal{D}(\Omega)}: A\to K^A$$

(where K^A is defined by Theorem 1) is an isomorphism of $*$-algebras.

Proof: Every element of $C_{\mathcal{D}(\Omega)}$ is represented by a distribution kernel, because it satisfies the requirements of Theorem 1; such a distribution kernel belongs to $K_{\mathcal{D}(\Omega)}$, as it follows from Proposition 1, taking into account the definitions of $C_{\mathcal{D}(\Omega)}$ and $K_{\mathcal{D}(\Omega)}$: so $A\to K^A$ maps $C_{\mathcal{D}(\Omega)}$ into $K_{\mathcal{D}(\Omega)}$. Conversely, according to the remark following definition 1, any element of $K_{\mathcal{D}(\Omega)}$ describes an operator of $L^2(\Omega)$ which belongs to $C_{\mathcal{D}(\Omega)}$ as it follows, from Proposition 1 and the definitions of $C_{\mathcal{D}(\Omega)}$ and $K_{\mathcal{D}(\Omega)}$. Hence the mapping j is bijective. At last using Proposition 1 and Theorem 2, it is immediately verified that the mapping j is a homomorphism.

ACKNOWLEDGMENTS

We are particularly indebted to Dr. M. V. Valdes[19] and also to Dr. L. Indorato for the preceding collaboration with one of us (R.A.) in an unpublished related work on the distribution-kernels associated with the bounded operators of L^2 spaces.

APPENDIX A: SOME ILLUSTRATING EXAMPLES

As a particularization of the situation considered by Roberts and quoted at the end of the Introduction let us consider a one-dimensional one-particle system, whose potential energy W is given by the Heaviside step function $u(t)$ [$u(t)=0$ for $t<0$ and $u(t)=1$ for $t>0$, t being the *space* variable]. Then all the operators of the $*$-algebra generated by Q (the multiplication by the variable t), $P=-i\hbar D=-i\hbar(d/dt)$ and W, may be represented on $\mathcal{D}(\Omega)$, with $\Omega=R\setminus\{0\}$ and the Dirac rules hold for them.

Such a $*$-algebra is actually a $*$-subalgebra of the $*$-algebra $C_{\mathcal{D}(\Omega)}$ considered in Theorem 3.

As an illustration of the last remark in the Introduction we observe that in this example many operators, like Q, D, and W, are represented also on the wider space $\mathcal{S}(R^n)$. For instance the distribution kernels of D and W are $K^D_{t,\tau}=\delta'(t-\tau)$ and $K^W_{t,\tau}=u(t)\delta(t-\tau)$ [where $u(t)\delta(t-\tau)$ is defined by $\langle u(t)\delta(t-\tau),\varphi(t,\tau)\rangle=\langle u(t),\langle\delta(w),\varphi(t,t-w)\rangle\rangle$].

Furthermore the operator WD is defined on $\mathcal{S}(R)$ and

it is represented on $\mathcal{S}(R)$ by the Volterra composition product, $K^W \circ K^D$, as defined according to case (a) of definition 3 and Theorem 2: An easy calculation, based on formula (8) of Theorem 2 leads to $K^{WD}_{t,\tau} = (K^W \circ K^D)_{t,\tau} = u(t)\delta'(t-\tau)$ [to be defined by analogy with $u(t)\delta(t-\tau)$].

On the contrary, the domain of the operator DW doesn't contain $\mathcal{S}(R)$: Of course DW is represented on $\mathcal{D}(\Omega)$ by $K^{DW}_{t,\tau} = (K^D \circ K^W)_{t,\tau} = u(t)\delta'(t-\tau)$ (in fact $DW|_{\mathcal{D}(\Omega)} = WD|_{\mathcal{D}(\Omega)}$.

Finally the product QW provides an example where condition (b) but not condition (a) of definition 3 and Theorem 2 is satisfied. In fact QW, Q, and W are represented on $\mathcal{S}(R)$, according to Theorem 1, by the distribution kernels $K^Q_{t,\tau} = t\delta(t-\tau)$, $K^W_{\tau,s} = u(\tau)\delta(\tau - s)$, and $K^{QW}_{t,s} = tu(t)\delta(t-s)$, respectively. However the vector space V where the right-hand side of formula (7) of definition 3 has a meaning is $V = \{\varphi \in \mathcal{S}(R) \mid \forall n \in N, \varphi^{(n)}(0) = 0\}$ which is dense in $L^2(\Omega)$ but not in $\mathcal{S}(R)$, so that definition 3(a) cannot be applied. On the other hand, QW is Hermitian on $\mathcal{S}(R)$, hence $QW \subset (QW)^* \subset (QW|_{\mathcal{S}(R)})^*$, so that the domain of $(QW|_{\mathcal{S}(R)})^*$ contains $\mathcal{S}(R)$: So condition (b) of Theorem 2 is satisfied and we may still write, according to part (b) of definition 3, $K^{QW} = K^Q \circ K^W$. Here, of course, the explicit calculation of $K^Q \circ K^W$, using formula (7) of definition 3 also requires some continuity considerations.

APPENDIX B: RECOVERING THE OPERATORS FROM THE REPRESENTATION

As the representation (1) holds only on the test function space, the distribution kernel K^A does not in general characterize the operator A. However in important cases some "auxiliary" information enables us to recover in a trivial way a linear operator A of an Hilbert space H from its restriction $A|_D$ to a dense vector subspace D. Apart from the case of bounded A, situations of this kind occur, when A is known to be the closure of $A|_D$ or to be essentially self-adjoint on D. Perhaps the most usual case occurs when A is known to be Hermitian and moreover its domain D_A is given (as it happens in many applications, where the "right" domain is chosen according to physical requirements). Then $A = A^*|_{D_A} = A|_D{}^*|_{D_A}$.

More generally any closable operator A of a Hilbert space H may trivially be recovered from its restriction $A|_D$ to a dense vector subspace D whenever its domain D_A and also some vector subspace D^* of the domain D_{A^*} of its adjoint A^* are given,

$$A = A^{**}|_{D_A} = A^*|_{D^*}{}^*|_{D_A} = A|_D{}^*|_{D^*}{}^*|_{D_A}.$$

This situation occurs frequently when both D_A and D_{A^*} include the same D. If furthermore a topology $\overline{\tau}$ on D is given so as to satisfy the requirements of Lemma 1, then, according to the remark following the lemma, A (and analogously A^*) may also be recovered through continuous extension of $A|_D$, considered as an operator from H into $(D^{\overline{\tau}})'$.

So, with reference to Theorem 1, whenever both D_A and D_{A^*} include $\mathcal{D}(\Omega)$, the distribution kernel K^A characterizes the operator A, provided also its domain D_A is given, and A may be recovered, either through two iterated restrictions to $\mathcal{D}(\Omega)$ and adjunctions, or through continuous extension of $A|_{\mathcal{D}(\Omega)}$, considered as an operator from H into $\mathcal{D}'(\Omega)$. Moreover in such a case both A and A^* satisfy the requirements of Theorem 1 and A^* is represented by $(K^A)^*$.

[1] See Marlow, J. Math. Phys. 6, 919 (1965); J.E. Roberts, J. Math. Phys. 6, 1097 (1966); J.P. Antoine, J. Math. Phys. 1, 53, 2276 (1969); A. Böhm, Lect. Theor. Phys. 9A, 255 (1967); J.M. Jauch, in *Aspects in Quantum Theory*, edited by A. Salam and E.P. Wigner (Cambridge U.P., Cambridge, 1972), p. 137; E. Prugovecki, J. Math. Phys. 10, 1410 (1973); O. Melsheimer, J. Math. Phys. 7, 902, 917 (1974). References to preceding (more mathematical) works are found in these papers.

[2] See P.A.M. Dirac, Proc. Roy. Soc. London Ser. A 113, 621 (1926–7); *The Principles of Quantum Mechanics* (Clarendon, Oxford, 1958), IV ed.

[3] The main statement of the present paper may especially be compared with Sec. IV D, case 3 of Antoine's work (see Ref. 1): in his paper the continuity (from his space Φ into H) of the operator, which is considered, is a hypothesis, whereas here it is a thesis for all Hermitian (more generally closeable) operators with domain Φ. The present paper may as well be considered as a development of Sec. 12 of Jauch's paper (see Ref. 1).

[4] See Ref. 2, Sec. 14.

[5] We are not concerned here with point (iv), which refers to diagonal representations.

[6] Concerning this restriction cf. Appendix B.

[7] See L. Schwartz, Proc. Int. Congr. Math. I, 220 (1950).

[8] See Eq. (8) of Theorem 2. In the communication of L. Schwartz quoted above,[7] the Volterra convolution is introduced under conditions which are in general different from those relevant here.

[9] See Sec. 4 of the reference of Roberts in Ref. 1.

[10] It may easily be seen that the $*$-algebra operations may be defined unambiguously on the set of operators considered, just because $\mathcal{D}(\Omega)$ provides a common invariant domain for the operators and their adjoints. Concerning these kinds of $*$-algebras of unbounded operators see R. Ascoli, A. Restivo, and G. Epifanio, Commun. Math. Phys. 18, 291 (1970) and also Riv. Mat. Univ. Parma 3, 21 (1974).

[11] Concerning topological vector spaces see, for instance, N. Bourbaki, *Elêments de mathêmatique: Espaces Vectorielles Topologiques* (Hermann, Paris, 1966), Chaps. I, II and (Hermann, Paris, 1953), Chaps. III, IV, and V. Here see in particular Chap. III, Sec. 8, n. 8.

[12] This statement holds more generally if H is any Fréchet space.

[13] We use the notation $(A|_D)^{\overline{\tau}t}$ to recall that the domain D of $A|_D$ in endowed with the topology $\overline{\tau}$: If instead D is endowed with the scalar product topology, then the "transposition" of $A|_D$ leads to the adjoint $(A|_D)^*$, which in general restricts $(A|_D)^{\overline{\tau}t}$ to some proper vector subspace of H, in accord with the fact that the very definition of $(A|_D)^*$ amounts to "restricting" the range of $(A|_D)^{\overline{\tau}t}$ to H.

[14] See T. Husain, *The Open Mapping Theorem in Topological Vector Spaces* (Oxford Mathematical Monographs, London, 1965), Chap. IV, Sec. 1, Proposition 3, and Sec. 5, Theorem 8.

[15] See Ref. 11, Chap. IV, Sec. 4, n. 2 Cor. Proposition 6.

[16] Note that, according to this identification, for any densely defined linear operator A of $L^2(\Omega)$, A^t and A^* are linked by $A^* f = \overline{A^t f}$ for any $f \in D_{A^*}$.

[17] See F. Treves, *Topological Vector Spaces, Distributions and Kernels* (Academic, New York, 1967), Chap. 51, Theorem 7.

[18] See references in Ref. 10.

[19] Communication at the 53° Congresso Nazionale della Società Italiana di Fisica Bologna 1967; see R. Ascoli and M.V. Valdes, Boll. S.I.F. 55, 91 (1967).

Entropy of an n-system from its correlation with a k-reservoir

Elihu Lubkin

Physics Department, The University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201
(Received 11 October 1977)

Let a random pure state vector be chosen in nk-dimensional Hilbert space, and consider an n-dimensional subsystem's density matrix P. P will usually be close to the totally unpolarized mixed state if k is large. Specifically, the rms deviation of a probability from the mean value $1/n$ is $[(1-1/n^2)/(kn+1)]^{1/2}$. "Random" refers to unitarily invariant Haar measure.

1. MOTIVATION

With Lucretius,[1] I find disorder in an initial state of the universe repugnant. Furthermore, a pure state remains pure. Entropy is somehow to be developed without fundamental entropy.

Indeed, my favorite key to understanding quantum mechanics is that a subsystem cannot be isolated by tracing from an enveloping pure state without generating impurity[2]: The probabilities associated with measurement develop because the observer must implicitly trace himself away from the observed system. So I even agree with Lucretius in tying a notion of ordered initial state together with a notion of quantum-mechanical indeterminacy. I feel that he was very smart to have understood this connection without density matrices.

With this wild beginning, I hope to have convinced the reader that whereas there may be nothing practical about the subsequent calculation, I find it meaningful in a happily nebulous, philosophic sense.

2. THE n, k ENTROPY PROBLEM

Let a "system of interest" have an n-complex-dimensional state space, within a larger nk-complex-dimensional tensor-product context; the "reservoir" to be ignored is k-dimensional. Let the over-all nk-state be pure, a vector $|x\rangle$, and trace out the k-space labels from the projection operator $|x\rangle\langle x|$ to get the n-space density matrix P which is appropriate for neglecting k-space effects. If $|x\rangle$ is chosen "at random," what distribution of probability n-tuples $(p_1,\ldots,p_n)$, the eigenvalues of P, follows? The problem becomes completely defined by interpreting "random" nk-state as a weighting of nk-state unit vectors by a measure invariant under U_{nk}, the nk by nk unitary transformations. By considering the unit complex nk-sphere as a homogeneous space of U_{nk}, then again as U_{nk} itself but organized into cosets, the measure is seen to be Haar measure on U_{nk}, thus unique.

If an orthonormal basis is chosen, and the $2nk$ real and imaginary parts of the components are taken as $2nk$ real Cartesian coordinates in $2nk$-real-dimensional Euclidean space, then the unit vectors $|x\rangle$ form the origin-centered real unit sphere, S_{2nk-1}. Since the usual geometric hyperarea of this sphere is invariant under all O_{2nk} origin-centered real orthogonal transformations, it is in particular invariant under those which happen to represent U_{nk} action. Thus, the unique unitary-invariant measure desired coincides with the usual geometric hyper-area on the unit sphere S_{2nk-1}, and that is how I will handle it.

I have found the distribution of $(p_1,\ldots,p_n)$ n-tuples too hard to present explicitly by quoting a weighting over n-tuples of positive numbers; yet I have determined $\sigma_{\rm rms}=\langle(1/n)\sum_{i=1}^{n}(p_i-1/n)^2\rangle^{1/2}$, the rms deviation of a typical probability from the over-all mean. The answer is

$$\sigma_{\rm rms}=\left(\frac{1-1/n^2}{kn+1}\right)^{1/2}. \tag{1}$$

Neither have I been able to determine the distribution of the entropy

$$S=-\sum p_i\ln p_i$$

or its mean, precisely; but (1) can be converted into a reasonable $\langle S\rangle$ estimate as follows. From Taylor expansion of each $\ln p_i$ about $p_i=1/n$, with

$$p_i=\frac{1}{n}-\frac{q_i}{n}, \tag{2a}$$

one gets the formal expansion

$$S=\ln n-\frac{1}{n}\left(\frac{1}{1\cdot 2}\sum_{i=1}^{n}q_i^2+\frac{1}{2\cdot 3}\sum_{i=1}^{n}q_i^3+\cdots\right), \tag{2b}$$

convergent if $|q_i|<1$, i.e., if

$$0<p_i<2/n. \tag{2c}$$

Since $\sigma_{\rm rms}$ is small for large k, most of the measure will indeed lie with $p_i<2/n$. This makes it plausible that if

$$S\approx\ln n-\text{def}, \tag{3}$$

most of the defect or information def is represented by (2b)'s leading defect term $(1/2n)\sum q_i^2=\frac{1}{2}n^2\sigma^2$, which with (1) gives

$$\langle\text{def}\rangle\approx\frac{1}{2}\frac{n^2-1}{kn+1}. \tag{4}$$

Thus, the mean entropy $\ln n-\langle\text{def}\rangle$ is not much less than the maximum entropy, $\ln n$, if $k\gtrsim n$, say. As the reservoir dimension $k\to\infty$, $\langle\text{def}\rangle\to 0$.

Hence neglecting the correlations of a small system with a large surround [$n\ll k$, or even only $k=O(n)$] already loses enough information to justify maximizing

 0022-2488/78/1905-1028$1.00

entropy of the small system, even if there is no further device for randomization, at least, if the correlations are likely to have become rich enough to justify the U_{nk} ergodicity assumption.

3. THE CALCULATION

Let the nk complex components of $|x\rangle$ be X_{iA}, where a lower-case index runs from 1 to n, a capitalized index, from 1 to k. $|x\rangle\langle x|$'s matrix elements are $X_{iA}X^*_{jB}$. The k-space trace of this is $P_{ij}=\sum_A X_{iA}X^*_{jA}=(XX^\dagger)_{ij}$, or $P=XX^\dagger$, in a notation which takes X to be an n by k rectangular complex matrix. We wish to determine aspects of the set of eigenvalues $p_1,\ldots,p_n$ of P. Of course, each $p_i\geq 0$, and $\sum_{i=1}^n p_i=1$; the latter, for example, is equivalent to $\sum_{iA}|X_{iA}|^2=1$. If

$$p(\lambda)=\det(\lambda-P)=\prod_{i=1}^{n}(\lambda-p_i)$$

$$=\lambda^n-\sigma_1\lambda^{n-1}+\sigma_2\lambda^{n-2}-\cdots, \tag{5}$$

the σ_i are the well-known elementary symmetric functions of the $p_1,\ldots,p_n$,

$$\sigma_m=\sum_{a<b\Rightarrow i_a<i_b}\ \prod_{a=1}^{m} p_{i_a}$$

being the sum of all m-fold products of distinct p's. Hence

$$\sigma_1=\sum p_i=1, \tag{6}$$

and

$$\sigma_1^2=(p_1+p_2+\cdots)^2$$

$$=p_1^2+p_2^2+\cdots+2(p_1p_2+p_1p_3+\cdots)$$

$$=\sum p_i^2+2\sigma_2,$$

or

$$\sum p_i^2=\sigma_1^2-2\sigma_2=1-2\sigma_2. \tag{7}$$

The mean square deviation of the p_i about their mean $1/n$ is

$$\sigma^2=\frac{1}{n}\sum\left(p_i-\frac{1}{n}\right)^2=\frac{1}{n}\left(\sum p_i^2-\frac{1}{n}\sum p_i+n\frac{1}{n^2}\right)$$

$$=\frac{1}{n}\left(\sum p_i^2-\frac{1}{n}\right);$$

$$\sigma^2=\frac{1}{n}\left(1-\frac{1}{n}-2\sigma_2\right) \tag{8}$$

shows that a determination of the distribution of σ_2 alone can decide whether the deviation of a typical p_i from $1/n$ is small. I will calculate $\langle\sigma_2\rangle$, where $\langle\ \rangle$ indicates expectation value over the orthogonally invariant measure on S_{2nk-1}; the "rms" deviation which corresponds to this is $\sigma_{\rm rms}\equiv\langle\sigma^2\rangle^{1/2}$, not $\langle(\sigma^2)^{1/2}\rangle$; this clarifies the meaning of Eq. (1).

From expansion of the determinant (5), σ_2 is seen to be the sum of the diagonal order-2 minors of P,

$$\sigma_2=\sum_{r<s}P_{rr}P_{ss}-P_{rs}P_{sr}.$$

Since a zero $r=s$ term is harmless,

$$\sigma_2=\tfrac{1}{2}\sum_{r,s}P_{rr}P_{ss}-P_{rs}P_{sr}=\tfrac{1}{2}\left[\left(\sum_r P_{rr}\right)^2-\sum_{r,s}P_{rs}P_{sr}\right]$$

$$=\tfrac{1}{2}[(\mathrm{Tr}P)^2-\mathrm{Tr}(P^2)]=\tfrac{1}{2}(\sigma_1^2-\mathrm{Tr}P^2);$$

$$\sigma_2=\tfrac{1}{2}(1-\mathrm{Tr}P^2). \tag{9}$$

Thus, we wish the expectation value of

$$\mathrm{Tr}P^2=\mathrm{Tr}XX^\dagger XX^\dagger=\sum_{ijAB}X_{iA}X^*_{jA}X_{jB}X^*_{iB}.$$

If a contraction on an n-fold index is represented by a dash, on a k-fold index by a double dash, then we wish to mean the loop

$$\begin{matrix}X=X^*\\ |\quad |\\ X^*=X.\end{matrix} \tag{10}$$

Let

$$X_{iA}=u_{iA}+\sqrt{-1}\,v_{iA}$$

introduce real-component notation. The single complex loop (10) gives rise to 16 real-component loop terms, eight of which bear the $\sqrt{-1}$ factor and of course cancel, leaving

$$\begin{matrix}u=u & u=u & u=v & u=v & v=u & v=u & v=v & v=v\\ |\quad | & |\quad | & |\quad | & |\quad | & |\quad | & |\quad | & |\quad | & |\quad |\\ u=u & v=v & u=v & v=u & u=v & v=u & u=u & v=v\\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8\end{matrix}$$

Term 2 = term 7, term 3 = term 6. In the mean, the symbols u and v can be interchanged, each being nk of the $2nk$ real variables which sum-square to 1; this shows that, in the mean, term 1 = term 8, term 4 = term 5. Hence,

$$\tfrac{1}{2}\langle\mathrm{Tr}P^2\rangle=\left\langle\begin{matrix}u=u\\ |\quad |\\ u=u\end{matrix}+\begin{matrix}u=u\\ |\quad |\\ v=v\end{matrix}+\begin{matrix}u=v\\ |\quad |\\ u=v\end{matrix}-\begin{matrix}u=v\\ |\quad |\\ v=u\end{matrix}\right\rangle. \tag{11}$$

Each term is the mean over the unit sphere of a sum of products of four Cartesian coordinates, hence a sum of means $\langle x_Ix_Jx_Kx_L\rangle$, where now I,J,K,L are each $2nk$-valued. Equal contributions from one coordinate's positive and negative ranges cause cancellation, if any I,J,K,L is distinct from the other three. Nonzero values thus require two equal pairs, or else the completely equal fourth-power case $I=J=K=L$. Both alternatives are to be considered for the *uuuu* term, but of course the *uuvv* terms must simply have the two u's coincide and the two v's coincide, as each u is a distinct coordinate from each v. Thus,

$$\left\langle\begin{matrix}u=u\\ |\quad |\\ v=v\end{matrix}\right\rangle=\sum_{ijAB}\langle u_{iA}u_{jA}v_{jB}v_{iB}\rangle\delta_{ij}$$

$$=\sum_{iAB}\langle u_{iA}^2v_{iB}^2\rangle=nk^2\langle x^2y^2\rangle, \tag{12a}$$

where x,y represent any two distinct Cartesian coor-

dinates in $2nk$ space. Similarly,

$$\left\langle \begin{matrix} u=v \\ \| \quad \| \\ v=u \end{matrix} \right\rangle = \sum_{ijAB} \langle u_{iA} v_{jA} u_{jB} v_{iB} \rangle \delta_{ij} \delta_{AB}$$

$$= \sum_{iA} \langle u_{iA}^2 v_{iA}^2 \rangle = nk \langle x^2 y^2 \rangle. \tag{12b}$$

Interchange of n and k in the $nk^2\langle x^2y^2\rangle$ result, or similar work, gives

$$\left\langle \begin{matrix} u=v \\ \| \quad \| \\ u=v \end{matrix} \right\rangle = n^2 k \langle x^2 y^2 \rangle. \tag{12c}$$

Finally,

$$\left\langle \begin{matrix} u=u \\ \| \quad \| \\ u=u \end{matrix} \right\rangle = \sum_{ijAB} \langle u_{iA} u_{jA} u_{jB} u_{iB} \rangle = \text{coincide} + 12 \text{ pair}$$

$$+ 13 \text{ pair} + 14 \text{ pair},$$

where "pair" is meant to exclude fourfold coincidence and, explicitly,

$$\text{coincide} = \sum_{iA} \langle u_{iA}^4 \rangle = nk \langle x^4 \rangle,$$

$$12 \text{ pair} = \sum_{i,A,B,A\neq B} u_{iA}^2 u_{iB}^2 = nk(k-1)\langle x^2 y^2 \rangle,$$

$$13 \text{ pair} = 0,$$

$$14 \text{ pair} = \sum_{i,j,A,i\neq j} u_{iA}^2 u_{jA}^2 = kn(n-1)\langle x^2 y^2 \rangle$$

thus,

$$\left\langle \begin{matrix} u=u \\ \| \quad \| \\ u=u \end{matrix} \right\rangle = nk\langle x^4 \rangle + nk(k-1+n-1)\langle x^2 y^2 \rangle. \tag{12d}$$

Substitution of (12a)—(12d) into (11) yields

$$\tfrac{1}{2}\langle \mathrm{Tr} P^2 \rangle = nk\langle x^4 \rangle + nk(2n+2k-3)\langle x^2 y^2 \rangle. \tag{13}$$

We will need $\langle x^4 \rangle$, $\langle x^2 y^2 \rangle$, and in order to delay normalization to the end, $\langle 1 \rangle$. These are integrals over the unit sphere $S_{2nk-1} = S_{N-1}$, where $N = 2nk$ is even. The required element of hyperarea is the factor of $r^{N-1}dr$ in the polar-coordinate hypervolume element for Euclidean N-space, where polar coordinates $r, \theta, \phi, \cdots$ are defined in terms of the Cartesian x_I by

$$x = x_1 = r\cos\theta$$

$$y = x_2 = r\sin\theta\cos\phi,$$

$$x_3 = r\sin\theta\sin\phi\cos\psi,$$

$$x_4 = r\sin\theta\sin\phi\sin\psi\cos\chi,$$

$$\cdots$$

All angles except the last run from 0 to π; the last, from 0 to 2π. The requisite factor of $r^{N-1}dr$ is the product $d\theta\, d\phi\, d\psi\, d\chi \cdots$ times the determinant of

$$\begin{matrix}
\cos\theta & \sin\theta\cos\phi & \sin\theta\sin\phi\cos\psi & \sin\theta\sin\phi\sin\psi\cos\chi\cdots \\
-\sin\theta & \cos\theta\cos\phi & \cos\theta\sin\phi\cos\psi & \cos\theta\sin\phi\sin\psi\cos\chi\cdots \\
0 & -\sin\theta\sin\phi & \sin\theta\cos\phi\cos\psi & \sin\theta\cos\phi\sin\psi\cos\chi\cdots \\
0 & 0 & -\sin\theta\sin\phi\sin\psi & \sin\theta\sin\phi\cos\psi\cos\chi\cdots \\
0 & 0 & 0 & -\sin\theta\sin\phi\sin\psi\sin\chi\cdots \\
0 & 0 & 0 & 0 \quad \cdots \\
\cdots & & & \cdots
\end{matrix}$$

which yields up its pattern upon expansion in the first column and factorization of $\cos\theta$ and $\sin\theta$ factors; then similarly for ϕ; the result is of the form

$$\sin^{N-2}\theta \sin^{N-3}\phi F(\psi, \chi, \cdots)\, d\theta\, d\phi\, d\psi\, d\chi \cdots.$$

The arbitrary normalization now allows dropping the work in angles other than θ (because of $x = \cos\theta$) and ϕ (because of $y = \sin\theta\cos\phi$), to give

$$\langle 1 \rangle = \int \sin^{N-2}\theta \sin^{N-3}\phi\, d\theta\, d\phi,$$

$$\langle x^4 \rangle = \int \cos^4\theta \sin^{N-2}\theta \sin^{N-3}\phi\, d\theta\, d\phi, \tag{14}$$

$$\langle x^2 y^2 \rangle = \int \cos^2\theta \sin^2\theta \cos^2\phi \sin^{N-2}\theta \sin^{N-3}\phi\, d\theta\, d\phi.$$

Let

$$\int_0^\pi \sin^m\theta\, d\theta = S_m \begin{cases} = \pi(m-1)!!/m!!, & m \text{ even} \\ = 2(m-1)!!/m!!, & m \text{ odd} \end{cases}. \tag{15}^3$$

Then

$$\langle 1 \rangle = S_{N-2}S_{N-3}, \quad \langle x^4 \rangle = (S_{N-2} - 2S_N + S_{N+2})S_{N-3},$$

$$\langle x^2 y^2 \rangle = (S_N - S_{N+2})(S_{N-3} - S_{N-1}).$$

Upon proper normalization to $\langle 1 \rangle = 1$, these become

$$\langle x^4 \rangle = 1 - 2\frac{S_N}{S_{N-2}} + \frac{S_{N+2}}{S_{N-2}},$$

$$\langle x^2 y^2 \rangle = \left(\frac{S_N}{S_{N-2}} - \frac{S_{N+2}}{S_{N-2}}\right)\left(1 - \frac{S_{N-1}}{S_{N-3}}\right).$$

The double factorials mostly cancel to give

$$\langle x^4 \rangle = 3/N(N+2), \quad \langle x^2 y^2 \rangle = 1/N(N+2).$$

Then (13) simplifies ($N = 2nk$) to

$$\sum_{i=1}^{n} p_i^2 = \langle \mathrm{Tr} P^2 \rangle = \frac{n+k}{nk+1},$$

and (9), (8) yield

$$\sigma_2 = \frac{1}{2}\left(1 - \frac{n+k}{nk+1}\right),$$

and

$$\sigma_{\mathrm{rms}} = \left(\frac{1-1/n^2}{kn+1}\right)^{1/2}$$

as stated in Eq. (1).

Of course, the argument as given needs $N \geq 4$. Nevertheless, $N = 2nk = 2$ corresponds only to the trivial case $n = k = 1$, when because $n = 1$ there is only one probability; hence $\sigma_{\mathrm{rms}} = 0$, which is indeed the value given by (1). Indeed, $n = 1$ alone must make σ_{rms} zero, whatever the value of k; the formula (1) satisfies this check.

Higher moments and other symmetric polynomials of higher degree in the p_i's will depend on more coefficients σ_m of the polynomial (5), which will require more and longer trace loops, and trigonometric integrals involving more than two angles. If this comment is taken as "clear," the p_i distribution is in fact "given" here.

4. DISCUSSION

When all the p_i are $1/n$, the rms width is of

course $\sigma = 0$. For any p_i, σ can be enlarged by decreasing a probability and simultaneously increasing the largest probability if that is not already 1; hence σ is maximum for the extreme case $p_1 = 1$, $p_2 = p_3 = \cdots = 0$, in which case (8) immediately yields $\sigma = \sigma_{\max}(n) = n^{-1}(n-1)^{1/2} \approx n^{-1/2}$ for n large.[4] That $\sigma_{\rm rms} \ll 1$ for $n \gg 1$ is thus without significance. However, when $n \gg 1$, $\sigma_{\rm rms} \approx n^{-1/2}k^{-1/2}$; hence $\sigma_{\rm rms} \ll \sigma_{\max}(n)$ provided that $k^{-1/2} \ll 1$. Thus, $\sigma_{\rm rms}(n,k)/\sigma_{\max}(n) \to 0$ in the limit $k \to \infty$, but, interestingly, is already small for $k = O(n)$. A neglected reservoir need not be enormous to produce near-maximum entropy.

Two oddities will not have escaped the reader. One is the finite dimensionality of the state spaces. The other is the freedom of the over-all state vector to roam over its whole nk-dimensional Hilbert space, without worrying about conservation principles. Perhaps these two faults or limitations can to some degree destroy each other, if we first limit the discussion to a space of ergodicity, where special boundary conditions (a perfect box, say), and conservation principles combine to produce the effective overall finite dimensionality.

Note added in manuscript: Charles Goebel and Nicholas Papastamatiou have led me to the following computation of the normalized $\langle x^4 \rangle$ and $\langle x^2 y^2 \rangle$, which avoids explicit trigonometric integrals over the volume element.

If x is any vector, then $x_I x_J x_K x_L$ is a fourth rank symmetric tensor. Similarly if the x are a set of vectors, then $\sum_A x_{IA} x_{JA} x_{KA} x_{LA}$ is a fourth rank symmetric tensor. $\langle x_I x_J x_K x_L \rangle$ averaged over the $(N-1)$-sphere is a limit of such a sum or linear combination, with the extra property that the set of vectors as a whole is O_N-invariant. Hence, $\langle x_I x_J x_K x_L \rangle$ is an O_N-invariant symmetric fourth-rank tensor, t_{IJKL}. Hence, if a, b, c, d are vectors, $\sum_{IJKL} t_{IJKL} a_I b_J c_K d_L$ is a symmetric scalar function of a, b, c, d only. The most general symmetric scalar function linear in the vector arguments is a multiple of $(a \cdot b)(c \cdot d) + (a \cdot c)(b \cdot d) + (a \cdot d)(b \cdot c)$: this by specializing the well-known theorem in H. Weyl's *Classical Groups* (Princeton U. P., Princeton, New Jersey, 1946) about rational invariants of vectors under O_N, to the linear symmetric case. Hence $t_{IJKL} = A(\delta_{IJ}\delta_{KL} + \delta_{IK}\delta_{JL} + \delta_{IL}\delta_{JK}) = \langle x_I x_J x_K x_L \rangle$. In order to determine A, contract on K, L, using $\sum_{K=1}^{N} x_K^2 = 1$, to get $A(\delta_{IJ}N + \delta_{IJ} + \delta_{IJ}) = \langle x_I x_J \rangle$; i. e., $\langle x_I x_J \rangle = A(N+2)\,\delta_{IJ}$. Then contract on I, J to get $1 = AN(N+2)$. Hence $A = [1/N(N+2)]$, and $\langle x_I x_J x_K x_L \rangle = [1/N(N+2)] \times (\delta_{IJ}\delta_{KL} + \delta_{IK}\delta_{JL} + \delta_{IL}\delta_{JK})$. The nonzero cases, where all four indices coincide, or where the indices form two distinct equal pairs, give the desired information

$$\langle x^4 \rangle = 3/N(N+2), \quad \langle x^2 y^2 \rangle = 1/N(N+2).$$

Higher order moments will have similar symmetrical Kronecker δ expressions, with over-all coefficients which may be determined inductively by contraction.

Note added in proof: Estimate (4) of the entropy defect is a quadratic-term truncation of (2b). Hence the hypotheses $|q_i| \ll 1$, roughly equivalent to $k \gg n$, should be noted in connection with (4). The "rough equivalence" is in the sense that $(q_i)_{\rm rms} \equiv \langle \sum q_i^2 \rangle / n = n^2 \langle \sigma^2 \rangle = (n^2 - 1)/(kn + 1) \approx n/k$. Equation (4) should not be trusted for $k \approx n$.

[1] Lucretius, *On the Nature of the Universe*, translated by R. E. Latham (Penguin, Toronto, 1951), p. 66.

[2] L. D. Landau, Z. Physik 45, 430 (1927); or J. Von Neuman, *Mathematical Foundations of Quantum Mechanics*, translated by R. Beyer (Princeton U. P., Princeton, N.J., 1955), Sec. VI. 2.

[3] Adapted from I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic, New York, 1965), p. 369, or induction through integration by parts.

[4] This extreme, pure-case answer $\sigma = n^{-1}(n-1)^{1/2}$ coincides with the specialization to $k = 1$ of Eq. (1), because $k = 1$ forces purity by eliminating the reservoir. This is another check of (1).

Quantization of the Liouville mechanics for systems with singular Lagrangians

B. Boisseau and C. Barrabes

Département de Physique, Faculté des Sciences de Tours, Parc Grandmont, Tours 37, France
(Received 15 August 1977)

Quantum mechanics, in phase space formulation, is directly deduced from the Liouville mechanics by a correspondence principle. The latter is applied to systems with singular Lagrangians.

1. INTRODUCTION

The phase-space formulation of quantum mechanics, initiated by Wigner,[1] and later developed by Moyal,[2] is essentially characterized in the following manner:

The quantum state is represented by the phase-space distribution $R(p,q,t)$.

The time dependence of $R(p,q,t)$ is governed by the Wigner equation. The latter is analogous to the Liouville equation in statistical mechanics.

If $\hat{A}$ is the quantum operator corresponding to the physical quantity $A(p,q)$ in classical mechanics, one must have:

$$\langle \psi | \hat{A} | \psi \rangle = \int A(p,q) R(p,q,t)\, dp dq. \tag{1.1}$$

This more or less ends the similarity between quantum mechanics and statistical mechanics.

However, $R(p,q,t)$ is not nonnegative and cannot be interpreted as a probability density. Moreover, $R(p,q,t)$ is not unique; as a matter of fact the function $R(p,q,t)$ satisfying (1.1) is determined by the correspondence which maps the classical quantity A onto the quantum operator $\hat{A}$. Wigner's distribution function is thus associated with Weyl's rule, Blokhintsev's distribution function to the standard rule, etc. We refer the reader to Shewell,[3] Metha[4] and Cohen.[5]

The latter[5] succeeded in unifying some of these distributions by introducing an arbitrary function. More recently, Agarwal and Wolf[6,7] have given a general standard formalism.

Other authors[8,9] have tried to introduce a nonnegative distribution function.

In this paper, we shall consider the quasiprobability distribution $R(p,q,t)$ as only representing the quantum state of the system; its statistical aspect will be a heuristic way to establish a correspondence principle between the Liouville mechanics and quantum mechanics in phase-space formulation.

We shall choose the Blokhintsev representation,[10] and shall not longer impose the relation (1.1), this allows us to have a "fuzzy" correspondence principle. In usual mechanics (free from constraints) ambiguity shall vanish, and the correspondence will be perfectly determined. When constraints occur, certain compatibility conditions will allow us to reduce the "fuzziness."

The Hamiltonian quantization of systems with singular Lagrangians was first carried out by Dirac.[11-13] Difficulties occur in the choice of the ordering of the operator factors in order that the constraints satisfy the compatibility conditions.

Here it appears necessary to employ for such systems a "fuzzy" correspondence principle, this being solved in a pragmatic way depending on the particular system.

Faddeev[14] has solved the difficulty by using the Feynman path integral[15] quantization in a Hamiltonian form.[16,17]

After having introduced in Sec. 2 the algebra of quantum quantities used in phase-space, we give in Sec. 3 a quick translation from the Schrödinger mechanics into phase-space formulation, we then give a correspondence principle between the Liouville mechanics and quantum mechanics in phase-space formulation. This is then applied in Sec. 4 to systems with singular Lagrangians giving first class constraints following Dirac's terminology.

2. BLOKHINTSEV REPRESENTATION OF THE QUANTUM OPERATOR ALGEBRA

Let us consider a quantum system with a finite number N of degrees of freedom, where $q=(q_1, q_2 \cdots q_N)$ designates a coordinate system, and $p=(p_1,p_2 \cdots p_N)$ designates another system of variables, which we shall interpret as a momentum system. The set (p,q) describes describes the phase space.

Let us consider a quantum operator $\hat{A}$; its Blokhintsev transform[18] is a phase-space function defined as

$$\tilde{A}(p,q) = \beta(\hat{A}) = \int \langle q | \hat{A} | q' \rangle \exp[-i(p/\hbar)(q-q')]\, dq' \tag{2.1}$$

with $pq = p_1q_1 + p_2q_2 + \cdots + p_Nq_N$ and $dq = dq_1 dq_2 \cdots dq_N$

This mapping is linear:

$$\beta(\lambda\hat{A} + \mu\hat{B}) = \lambda\tilde{A}(p,q) + \mu\tilde{B}(p,q). \tag{2.2}$$

One shows easily that the transform of a product of operators can be written under the form

$$\beta(\hat{A}\cdot\hat{B}) = \int \tilde{B}(p,q'') \langle q | \hat{A} | q'' \rangle \exp[-i(p/\hbar)(q-q'')]\, dq''. \tag{2.3}$$

We develop the series

$$\tilde{B}(p,q'') = \tilde{B}(p,q) + \sum_{\alpha \geq 1} \frac{(q''-q)^\alpha}{\alpha!} \frac{\partial^\alpha \tilde{B}(p,q)}{\partial q^\alpha}, \tag{2.4}$$

where $\alpha \equiv \alpha_1 + \alpha_2 + \cdots + \alpha_N$, $\alpha! \equiv \alpha_1!\ \alpha_2! \cdots \alpha_N!$,

 0022-2488/78/1905-1032$1.00

$$(q''-q)^{\alpha} \equiv (q_1''-q_1)^{\alpha_1}(q_2''-q_2)^{\alpha_2}\cdots(q_N''-q_N)^{\alpha_N},$$
$$\partial^{\alpha}\tilde{B}(p,q)/\partial q^{\alpha} \equiv (\partial^{\alpha_1+\alpha_2+\cdots+\alpha_N}/\partial q_1^{\alpha_1}\partial q_2^{\alpha_2}\cdots\partial q_N^{\alpha_N})\,\tilde{B}(p,q).$$

Upon introducing the development (2.4) into Eq. (2.3), (2.3), we obtain

$$\beta(\hat{A}\cdot\hat{B}) = \tilde{B}(p,q)\tilde{A}(p,q) + \sum_{\alpha\geq 1}\frac{1}{\alpha!}\left(\frac{\hbar}{i}\right)^{\alpha}\frac{\partial^{\alpha}\tilde{B}(p,q)}{\partial q^{\alpha}}\,\frac{\partial^{\alpha}\tilde{A}(p,q)}{\partial p^{\alpha}}, \tag{2.5}$$

which allows use to define a noncommutative law of multiplication in phase space[19]:

$$\tilde{A}(p,q)*\tilde{B}(p,q) = \tilde{B}(p,q)\tilde{A}(p,q) + \sum_{\alpha\geq 1}\frac{1}{\alpha!}\left(\frac{\hbar}{i}\right) \times\frac{\partial^{\alpha}\tilde{B}(p,q)}{\partial q^{\alpha}}\,\frac{\partial^{\alpha}\tilde{A}(p,q)}{\partial p^{\alpha}}, \tag{2.6}$$

This operation is associative and distributive with regard to addition:

$$(\tilde{A}(p,q)*\tilde{B}(p,q))*\tilde{C}(p,q) = \tilde{A}(p,q)*(\tilde{B}(p,q)*\tilde{C}(p,q)), \tag{2.7}$$
$$\tilde{A}(p,q)*(\tilde{B}(p,q)+\tilde{C}(p,q)) = \tilde{A}(p,q)*\tilde{B}(p,q) + \tilde{A}(p,q)*\tilde{C}(p,q). \tag{2.8}$$

The relations (2.7) and (2.8) are derived from the corresponding properties of operators $\hat{A}$, $\hat{B}$, $\hat{C}$.

Let us remark that the product $\tilde{A}(p,q)*\tilde{B}(p,q)$ becomes the ordinary product when limit $\hbar \to 0$.

From the relations (2.5) and (2.6), we deduce the transform of a commutator:

$$\beta([\hat{A},\hat{B}]) = \tilde{A}(p,q)*\tilde{B}(p,q) - \tilde{B}(p,q)*\tilde{A}(p,q) = i\hbar\left[\sum_{\alpha\geq 1}\frac{1}{\alpha!}\left(\frac{\hbar}{i}\right)^{\alpha-1}\{\tilde{A}(p,q),\tilde{B}(p,q)\}_{\alpha}\right] \tag{2.9}$$

with the definition

$$\{\tilde{A}(p,q),\tilde{B}(p,q)\}_{\alpha} = \frac{\partial^{\alpha}\tilde{A}(p,q)}{\partial q^{\alpha}}\,\frac{\partial^{\alpha}\tilde{B}(p,q)}{\partial p^{\alpha}} - \frac{\partial^{\alpha}\tilde{B}(p,q)}{\partial q^{\alpha}} \times\frac{\partial^{\alpha}\tilde{A}(p,q)}{\partial p^{\alpha}}. \tag{2.10}$$

When $\alpha = 1$, the definition (2.10) is reduced to the classical Poisson brackets. This makes us introduce what we call quantum Poisson brackets:

$$\{\tilde{A}(p,q),\tilde{B}(p,q)\}_{\text{Q.P.}} = \sum_{\alpha\geq 1}\frac{1}{\alpha!}\left(\frac{\hbar}{i}\right)^{\alpha-1}\{\tilde{A}(p,q),\tilde{B}(p,q)\}_{\alpha} = \frac{1}{i\hbar}[\tilde{A}(p,q)*\tilde{B}(p,q) - \tilde{B}(p,q)*\tilde{A}(p,q)]. \tag{2.11}$$

In the limit $\hbar = 0$, (2.11) is reduced to classical Poisson brackets.

The quantum Poisson brackets satisfy the properties of the Lie algebra:

Antisymmetry:

$$\{\tilde{A},\tilde{B}\}_{\text{Q.P.}} = -\{\tilde{B},\tilde{A}\}_{\text{Q.P.}}, \tag{2.12}$$

Linearity:

$$\{\lambda\tilde{A}+\mu\tilde{B},\tilde{C}\}_{\text{Q.P.}} = \lambda\{\tilde{A},\tilde{C}\}_{\text{Q.P.}} + \mu\{\tilde{B},\tilde{C}\}_{\text{Q.P.}} \tag{2.13}$$

Jacobi identity:

$$\{\tilde{A},\{\tilde{B},\tilde{C}\}_{\text{Q.P.}}\}_{\text{Q.P.}} + \{\tilde{B},\{\tilde{C},\tilde{A}\}_{\text{Q.P.}}\}_{\text{Q.P.}} + \{\tilde{C},\{\tilde{B},\tilde{A}\}_{\text{Q.P.}}\}_{\text{Q.P.}} = 0. \tag{2.14}$$

One can add the product law:

$$\{\tilde{A}*\tilde{B},\tilde{C}\}_{\text{Q.P.}} = \tilde{A}*\{\tilde{B},\tilde{C}\}_{\text{Q.P.}} + \{\tilde{A},\tilde{C}\}_{\text{Q.P.}}*\tilde{B}. \tag{2.15}$$

These last four relations are deduced from the corresponding properties of operators.

We shall now specify the Blokhintsev transform of some functions of position $\hat{q}$ and momentum $\hat{p}$ operators:

$$\beta(\hat{q}) = q, \quad \beta(\hat{p}) = p. \tag{2.16}$$

From Eq. (2.16), using relations (2.2) and (2.5), one obtaines the following results:

$$\beta(V(\hat{q})) = V(q), \tag{2.17}$$
$$\beta(F(\hat{p})) = F(p), \tag{2.18}$$
$$\beta(V(\hat{q})\cdot F(\hat{p})) = V(q)\cdot F(p). \tag{2.19}$$

Let us consider an operator $F(\hat{p}_k, \hat{q}_1)$ constructed with commutative operators $\hat{p}_k$, $\hat{q}_1$, (a component of angular momentum for example); it is possible to put its factors in the same order as in Eq. (2.19). Consequently, one obtains

$$\beta(F(\hat{p}_k, \hat{q}_1)) = F(p_k, q_1). \tag{2.20}$$

According to Born and Jordan the operator $\hat{p}_k{}^n F(\hat{q}_k)$ can be put into the form[20]:

$$\hat{p}_k{}^n F(\hat{q}_k) = \sum_{m=0}^{n}\frac{n!}{m!(n-m)!}(-i\hbar)^m\frac{d^m F}{dq_k{}^m}\hat{p}^{n-m}$$

Its transform is then obtained by using (2.19):

$$\beta(\hat{p}_k{}^n F(\hat{q}_k)) = \sum_{m=0}^{n}\frac{n!}{m!(n-m)!}(-i\hbar)^m\frac{d^m F}{dq_k{}^m}p_k{}^{n-m}. \tag{2.21}$$

To put it more simply,

$$\beta(\hat{p}_k{}^n F(\hat{q}_k)) = p_k{}^n F(q_k) + O(\hbar) \tag{2.22}$$

where $O(\hbar)$ is a term of order $\hbar$. More generally, for any operator $\varphi(\hat{p},\hat{q})$,

$$\beta(\varphi(\hat{p},\hat{q})) = \tilde{\varphi}(p,q) = \varphi(p,q) + O(\hbar). \tag{2.23}$$

The density operator ρ, representing the state of the system, plays a special role, and shall be expressed as a complex function in phase space, usually called quasiprobability of the phase space:

$$R(p,q,t) = (2\pi\hbar)^{-N}\beta(\rho(t)). \tag{2.24}$$

We shall limit ourselves to pure systems:

$$\rho(t) = |\psi\rangle\langle\psi|. \tag{2.25}$$

The function R satisfies the following properties:

$$\int R(p,q)\,dp = \psi^*(q)\,\psi(q) > 0, \tag{2.26}$$
$$\int R(p,q)\,dq = \phi^*(p)\,\phi(p) > 0, \tag{2.27}$$

with $\phi(p) = (2\pi\hbar)^{-N/2}\int\psi(q)\exp(-ipq/\hbar)\,dq$. So,

$$\int R(p,q)\,dp\,dq = 1. \tag{2.28}$$

Moreover, the characterization $\rho^2 = \rho$ of a pure case gives

$$(2\pi\hbar)^N R*R = R. \tag{2.29}$$

We now have all the elements necessary to proceed to a quantization in phase space, which we shall carry out in the next section.

3. QUANTUM MECHANICS IN PHASE-SPACE FORMULATION, QUANTIZATION OF THE LIOUVILLE MECHANICS

A. Quantum mechanics in phase-space mechanics

Quantum mechanics is essentially characterized by the following datas:

(a) an equation of motion: the Schrödinger equation;

(b) an operation rule which allows to obtain the measurements: the eigenvalue equation;

(c) an operation rule which allows to get the classical mechanics quantities: the quantum average built upon probability amplitudes.

Let us translate (a), (b), (c) into the phase-space language.

(a) The equation of motion equivalent to the Schrödinger equation is obtained from the density equation

$$i\hbar \frac{d}{dt}\rho(t) = [\hat{H}, \rho(t)] \tag{3.1}$$

Taking the Blokhintsev transform of the two members, one obtains

$$\frac{\partial}{\partial t} R(p,q,t) = \{\tilde{H}(p,q), R(p,q,t)\}_{Q.P.} \tag{3.2}$$

In the ordinary case in which $\tilde{H}(p,q)$ is reduced to the classical Hamiltonian $H(p,q)$, Eq. (3.2) is the Wigner equation in the Blokhintsev representation.[21]

Omitting the terms of order $\hbar$ and replacing the quasiprobability $R(p,q,t)$ by the probability $f(p,q,t)$ in phase space, one obtains the Liouville equation

$$\frac{\partial}{\partial t} f(p,q,t) = \{H(p,q), f(p,q,t)\}. \tag{3.3}$$

(b) The eigenvalue equation, written as follows:

$$\hat{A}\,|\psi_a\rangle\langle\psi_a| = a\,|\psi_a\rangle\langle\psi_a| \tag{3.4}$$

translates into phase-space formulation:

$$\tilde{A}(p,q) * R_a(p,q,t) = aR_a(p,q,t) \tag{3.5}$$

(c) One has the following equalities:

$$\begin{aligned}
&\int \tilde{A}(p,q) * R(p,q)\,dp\,dq \\
&\quad = (2\pi\hbar)^{-N} \int \beta(\tilde{A} \circ \rho)\,dp\,dq \\
&\quad = \int \langle q|\hat{A}|\psi\rangle\langle\psi|q'\rangle\,(2\pi\hbar)^{-N} \\
&\qquad \times \exp[-ip/\hbar(q-q')]\,dp\,dq\,dq' \\
&\quad = \int \langle q|\hat{A}|\psi\rangle\langle\psi|q'\rangle\,\delta(q-q')\,dq dq' = \langle\psi|\hat{A}|\psi\rangle.
\end{aligned}$$

One then deduces expectation value in phase-space:

$$\langle\hat{A}\rangle = \langle\tilde{A}\rangle = \int \tilde{A}(p,q) * R(p,q,t)\,dp\,dq \tag{3.6}$$

Omitting the terms of order $\hbar$ and replacing the quasiprobability R by the probability f in phase space, one obtains the classical average in phase space,

$$\overline{A} = \int A(p,q) f(p,q,t)\,dp\,dq. \tag{3.7}$$

The transition amplitude between two states ψ and ψ_a is given by

$$\langle\psi_a|\psi\rangle = (2\pi\hbar)^N \int R_a(p,q) * R(p,q)\,dq\,dp. \tag{3.8}$$

This result is obtained in the same way as was Eq. (3.6).

Having obtained an equivalent formulation in phase space to the Schrödinger mechanics, we shall now examine how to arrive directly to this formulation by a correspondence principle using the Liouville mechanics.

B. Quantization of the Liouville mechanics

The Liouville mechanics of a system with N degrees of freedom consists of a space with $2N$ dimensions (p,q) called the phase space, and of a probability density function $f(p,q,t)$ whose time evolution is governed by the Liouville equation

$$\frac{\partial}{\partial t} f(p,q,t) = \{H(p,q), f(p,q,t)\}. \tag{3.9}$$

Let us remark, following Cabannes,[22] that this equation is equivalent to the Hamiltonian equations.

We define the correspondence principle:

$$f(p,q,t) \to R(p,q,t) \tag{3.10}$$

$$A(p,q) \to \tilde{A}(p,q) \tag{3.11}$$

$$\{\ \} \to \{\ \}_{Q.P.} \tag{3.12}$$

$$\text{product} \to * \tag{3.13}$$

As a matter of convention we shall always choose $\tilde{A} * R$ (R on the right-hand side).

The quantum quantities appear to be an extension of corresponding classical quantities and are defined on the same phase space (p,q). The Liouville mechanics is then used as an heuristic way of constructing the quantum mechanics in phase space.

Thus the correspondences (3.10), (3.11), (3.12) allow us to map the Liouville equation onto the Wigner equation:

$$\frac{\partial}{\partial t} f = \{H, f\} \to \frac{\partial}{\partial t} R = \{\tilde{H}, R\}_{Q.P.}. \tag{3.14}$$

In the same way the correspondences (3.10), (3.11), (3.13) induce the quantum average:

$$\begin{aligned}
&\overline{A} = \int A(p,q) f(p,q,t)\,dp dq \\
&\quad \to \langle\tilde{A}\rangle = \int \tilde{A}(p,q) * R(p,q,t)\,dp dq.
\end{aligned} \tag{3.15}$$

Finally one easily deduces the meaning of the eigenvalue equation

$$\tilde{A}(p,q) * R(p,q,t) = aR(p,q,t) \tag{3.16}$$

from a classical situation where the dynamical quantity $A(p,q)$ takes a value a independent of p,q, i.e., without dispersion.

We shall end this paragraph by remarking that the correspondence (3.11) is somewhat "fuzzy":

$$A(p,q) \to \tilde{A}(p,q) = A(p,q) + O(\hbar) \tag{3.17}$$

This is due to the possible variety of choices of the term $O(\hbar)$, as shown by Eq. (2.21) and (2.22). This difficulty is not inherent in phase space; it is produced by transition from a commutative to a noncommutative algebra. Numerous rules have been proposed[3]; their abundance show that none of them is entirely satisfactory.

But for the usual quantities in ordinary mechanics, we can apply Eqs. (2.17), (2.18), (2.20); consequently the correspondence is then perfectly defined since

$$\tilde{A}(p,q)=A(p,q), \tag{3.18}$$

that is, the case for example with the Hamiltonian $p^2/2m + V(q)$ of energy, momentum, angular momentum, etc.

The introduction of constraints is going to modify the former simplicity we shall examine this in the next paragraph.

4. QUANTIZATION OF MECHANICAL SYSTEMS WITH SINGULAR LAGRANGIANS

We follow here Faddeev.[14]

Consider a classical system with singular Lagrangian $L(q,\dot{q})$ in the sense that the equations:

$$p_i=\frac{\partial L(q,\dot{q})}{\partial \dot{q}_i} \tag{4.1}$$

cannot be solved for the $\dot{q}$. The canonical variables p,q do not vary through the phase space (p,q) of dimension $2N$ but have to verify a certain number of constraints

$$\varphi^a(p,q)=0, \quad a=1,\ldots,m \tag{4.2}$$

These constraints are independent; the equations (4.2) define a surface M of dimension $2N-m$ in phase space.

An arbitrary function $g(p,q)$ vanishing on M is a combination of constraints:

$$g(p,q)=\sum_a C_a(p,q)\,\varphi^a(p,q). \tag{4.3}$$

The equations of motion are obtained from the variation of the action

$$S(p,q,\lambda)=\int\Big(\sum_i p_i\dot{q}_i - H-\sum_a \lambda_a\varphi^a\Big)\,dt, \tag{4.4}$$

where the $\lambda_a(t)$ are Lagrangian multipliers:

$$\begin{aligned}\dot{q}_i&=\frac{\partial H}{\partial p_i}+\sum_a\lambda_a\frac{\partial\varphi^a}{\partial p_i},\\ \dot{p}_i&=-\frac{\partial H}{\partial q_i}-\sum_a\lambda_a\frac{\partial\varphi^a}{\partial q_i}\end{aligned} \tag{4.5}$$

Thus for any function $f(p,q,t)$ i.e.,

$$\begin{aligned}\dot{f}&=\frac{\partial f}{\partial t}+\frac{\partial f}{\partial q_i}\dot{q}_i+\frac{\partial f}{\partial p_i}\dot{p}_i,\\ \dot{f}&=\frac{\partial f}{\partial t}+\{f,H\}+\sum_a\lambda_a\{f,\varphi^a\}.\end{aligned} \tag{4.6}$$

The constraints should obviously satisfy

$$\dot{\varphi}^a=0 \quad \text{on } M \tag{4.7}$$

To avoid complications, we shall impose on the constraints additional conditions assuring (4.7):

$$\{\varphi^a,\varphi^b\}=\sum_c C_c^{ab}\varphi^c, \tag{4.8}$$

$$\{H,\varphi^a\}=\sum_b C_b^a\varphi^b. \tag{4.9}$$

The constraints satisfying (4.8) are first class ones, following Dirac's terminology.[11-13]

Now, if $f(p,q,t)$ represents the phase space density, Liouville theorem ($\dot{f}=0$), and Eq. (4.6) lead to Liouville's equation

$$\frac{\partial f}{\partial t}+\{f,H\}+\sum_a\lambda_a\{f,\varphi^a\}=0. \tag{4.10}$$

Let us examine the conditions imposed upon the density $f(p,q,t)$ by the constraints (4.2).

Let us call CM the complement of the surface M in the phase-space (p,q):

$$f(p,q,t)=0 \quad \text{on } CM, \tag{4.11}$$

$$\varphi^a(p,q)=0 \quad \text{on } M. \tag{4.12}$$

From (4.11) and (4.12) one deduces

$$\varphi^a(p,q)\cdot f(p,q,t)=0 \quad \forall\ p,q. \tag{4.13}$$

On the other hand, Eq. (4.11) is equivalent to saying that $f(p,q,t)$ is identically null "almost everywhere" on the phase space excepting on a surface M; one deduces that

$$\frac{\partial f}{\partial p_i}=0, \quad \frac{\partial f}{\partial q_i}=0 \quad \text{on } CM, \tag{4.14}$$

and hence

$$\{f,\varphi^a\}=0 \quad \text{on } CM, \tag{4.15}$$

$$\{f,H\}=0 \quad \text{on } CM. \tag{4.16}$$

The evolution of the density $f(p,q,t)$ should not depend on the arbitrary parameters $\lambda_a(t)$; we shall therefore impose

$$\{f,\varphi^a\}=0 \quad \text{on } M. \tag{4.17}$$

From (4.15) and (4.17) one deduces that

$$\{f,\varphi^a\}=0 \quad \forall\ p,q \tag{4.18}$$

The classical system is therefore described by Eqs. (4.10), (4.13), and (4.18); by using the correspondence principle, we obtain the quantum equations

$$\frac{\partial R}{\partial t}+\{R,H\}_{Q.P.}+\sum_a\lambda_a\{R,\tilde{\varphi}^a\}_{Q.P.}=0, \tag{4.19}$$

$$\tilde{\varphi}^a * R=0, \tag{4.20}$$

$$\{R,\tilde{\varphi}^a\}_{Q.P.}=0. \tag{4.21}$$

(We have supposed $\tilde{H}=H$.)

We shall remark that Eqs. (4.20) and (4.21) become necessary once the correspondence principle has been used.

Up till now we never talked about the hermiticity of operators and its translation into phase space.

Briefly, we shall only say that a quantity $\tilde{\varphi}(p,q)$ is Hermitian if the corresponding operator $\hat{\varphi}$ is Hermitian itself.

We shall suppose that the quantum constraints $\tilde{\varphi}^a(p,q)$ are Hermitian, this imposes a certain choice of $O(\hbar)$ and partly reduces the ambiguity of $\tilde{\varphi}^a(p,q)$. It is easy to see, since R is Hermitian, that Eqs. (4.21) are deduced from Eqs. (4.20), which in turn are equivalents to the conditions given by Dirac[11–13]:

$$\hat{\varphi}^a|\psi\rangle = 0. \tag{4.22}$$

Let us examine folowing Dirac the compatibility of Eqs. (4.20).

Multiplying (4.20) by $\tilde{\varphi}^b$,

$$\tilde{\varphi}^b * \tilde{\varphi}^a * R = 0. \tag{4.23}$$

In the same way,

$$\tilde{\varphi}^a * \tilde{\varphi}^b * R = 0. \tag{4.24}$$

We deduce that

$$\{\tilde{\varphi}^a, \tilde{\varphi}^b\}_{Q.P.} * R = 0. \tag{4.25}$$

We must not introduce new constraints; Eq. (4.25) must be included in Eq. (4.20). Therefore, let us require that

$$\{\tilde{\varphi}^a, \tilde{\varphi}^b\}_{Q.P.} = \sum_c \tilde{C}^{ab}_c * \tilde{\varphi}^c. \tag{4.26}$$

In the same way, the compatibility of Eqs. (4.19) with Eqs. (4.20) require

$$\{H, \tilde{\varphi}^a\}_{Q.P} = \sum_b \tilde{C}^a_b * \tilde{\varphi}^b. \tag{4.27}$$

Let us remark (coherence of quantization) that Eqs. (4.26) and (4.27) are quantum translations of the classical equations (4.8) and (4.9). We then have the advantage of a certain free choice of the term $O^a(\hbar)$ (belonging to $\tilde{\varphi}^a$) in such a way as to satisfy Eqs. (4.26) and (4.27). If one requires that the theory be equivalent to Dirac's,[11–13] the number and expression of the possible possible terms $O^a(\hbar)$ are well determined. This freedom corresponds to the choice of the order in which operators $\hat{p}$ and $\hat{q}$ are put in $\hat{\varphi}^a$.

More precisely, let us take a simple example devoide of any physical sense. Take a classical quantity

$$\varphi = q^2 p. \tag{4.28}$$

One can make correspond to it the following quantum operators $\hat{q}^2\hat{p}$, $\hat{q}\hat{p}\hat{q}$, $\hat{p}\hat{q}^2$ and all their linear combinations whose sum of coefficients equals 1. To these operators corresponds in phase space the quantities $q^2p + O(\hbar)$, where $O(\hbar)$ has respectively the values 0, $i\hbar q$, $-2i\hbar q$ and all their linear combinations whose sum of the coefficient equals 1. One deduces that the choice of $O(\hbar)$ has to be limited to

$$O(\hbar) = \lambda i\hbar q \quad (\lambda \text{ a complex number}) \tag{4.29}$$

If one requires $\tilde{\varphi}$ to be Hermitian,

$$\lambda = -1. \tag{4.30}$$

In this example it is no longer possible to impose other conditions.

More generally, the compatibility conditions on the constraints does not necessarily have a solution.

Nevertheless, the translation of quantum mechanics into phase-space formulation suggests new possibilities: If the problem has no solution when using the limited choice of possible forms of $O(\hbar)$, one could consider widening the choice to any function of p and q of the order of $\hbar$. It then may be necessary to add some criteria to eliminate all indeterminations.

The study of that hypothesis necessitates its application to particular systems. We shall not do it here. As a matter of fact the most interesting systems with singular Lagrangians are to be found in field theory; we have considered systems with only a finite number of degrees of freedom.

5. CONCLUSION

After setting up a "fuzzy" correspondence principle between the Liouville mechanics and quantum mechanics in phase-space formulation, we have now been able to treat in general the systems with singular Lagrangians. In the case of first class Hermitian constraints, we obtain results identical to Dirac's.[11–13] The process has several advantage from a theoretical point of view:

(a) It shows that, for such systems a clearly defined correspondence cannot be chosen first hand.

(b) The equations of constraints and their compatibilities are necessarily introduced as direct consequence of the correspondence principle.

(c) If Dirac's theory has no solution, it suggests a possible way of solving the problem.

[1]E. Wigner, Phys. Rev. 40, 749 (1932).
[2]J. E. Moyal, Proc. Cambridge Phil. Soc. 45, 99 (1949).
[3]J. R. Shewell, Am. J. Phys. 27, 16 (1959).
[4]C. L. Metha, J. Math. Phys. 5, 677 (1964).
[5]L. Cohen, J. Math. Phys. 7, 781 (1966).
[6]G. S. Agarwal and E. Wolf, Phys. Rev. D 2, 2161 (1970).
[7]G. S. Agarwal and E. Wolf, Phys. Rev. D 2, 2187 (1970).
[8]V. V. Kuryshkin, Ann. Inst. H. Poincaré 17, 81 (1972).
[9]S. Twareque Ali and E. Prugovecki, J. Math. Phys. 18, 219 (1977).
[10]D. I. Blokhintsev, *Mécanique quantique et application à l'étude de la structure de la matière* (Masson, Paris, 1967), Chap. 7, p. 163.
[11]P. A. M. Dirac, Can. J. Math. 2, 129 (1950).
[12]P. A. M. Dirac, Proc. Roy. Soc. A 246, 326 (1958).
[13]P. A. M. Dirac, Lectures on Quantum Mechanics (Belfer Graduate School of Sciences, Yeshiva University, New York, 1964).
[14]L. D. Faddeev, Teor. Mat. Fis. 1, 3 (1969).
[15]R. P. Feynman, Rev. Mod. Phys. 20, 367 (1948).
[16]R. P. Feynman, Phys. Rev. 84, 108 (1951).
[17]C. Garrod, Rev. Mod. Phys. 38, 483 (1966).
[18]D. I. Blokhintsev (Ref. 10) uses the transformation (2.1) divided by the factor $(2\pi\hbar)^N$.
[19]The operation defined by (2.6) has been studied by C. L. Metha (Ref. 4).
[20]M. Born and P. Jordan, Z. Physik 34, 873 (1925).
[21]For convenience, Eq. (3.2) shall called Wigner's equation.
[22]H. Cabannes, *Cours de mécanique générale* (Dunod, Paris, 1962), p. 288.

The zero-energy Coulomb problem

Augustine C. Chen

Physics Department, St. John's University, New York, New York 11439
(Received 25 October 1976)

A new operator denoted as a zero-energy Runge–Lenz vector is used to derive a differential equation for the zero-energy Schrödinger wavefunction in the **u** representation. An integral equation is also derived in terms of a generalized Fock 4-vector which, in the zero-energy limit, yields the correct integral equation for the zero-energy Green's function. The zero-energy aspect of the problem is further exemplified by extracting the zero-energy results directly from those for negative energies through the taking of the zero-energy limit, which is construed as making a group contraction from O(4) to E(3).

1. INTRODUCTION

The Kepler or Coulomb problem has been a subject of interest to physicists and mathematicians alike for a long time. The reason for this sustained interest is perhaps twofold. It is the only isolated two-body system occurring in nature and it represents the only case in which the potential is a solution of the Laplace equation with a point source.[1] Together with the harmonic oscillator and the Ising model, it is also one of the three exactly solvable problems in all of quantum mechanics.

This paper deals with the problem specificly when the total energy of the Kepler or Coulomb system is zero. This aspect of the problem does not seem to have been fully explored. While the zero-energy Kepler problem may have applications in astrophysics, the zero-energy Coulomb problem is of interest in atomic scatterings of three-body Coulomb systems. Indeed, the two-body Coulomb T-matrix has been widely used in the treatment of such systems by the Faddeev formalism.[2] In this paper, we take a closer look at the zero-energy Coulomb problem.

It has long been known that the nonrelativistic Coulomb problem has interesting symmetry properties. Fock[3] explained the degeneracy of the levels of the hydrogen atom in terms of the symmetry group in a four-dimensional space. Schwinger[4] constructed the Green's function for the problem by exploiting this O(4) rotational invariance. More recently, Rogers,[5] using quarternion algebra, gave a geometric interpretation of the classical transformations for the negative-energy Kepler problem. For positive energies, it has been pointed out[6] that the problem acquires the symmetry properties of the O(3, 1) group. Finally, for $E=0$, the invariance group becomes the three-dimensional Euclidean group E(3) isomorphic to the restricted Galilean group. While the Green's function[7] for positive energies can be obtained by analytic continuation from negative energies, a separate integral equation has been derived for the zero-energy Green's function[6] for its solution. Similarly, the zero-energy wavefunction has been obtained by solving an appropriate integral equation.[6]

The plan of this paper is as follows. In Sec. 2 we show that the zero-energy Runge–Lenz vector $\mathbf{A}_0$, a self-adjoint operator, and a generalized coordinate vector **u** satisfy the canonical commutation relations and are used, under the zero-energy constraint, to derive a differential equation for the zero-energy wavefunction. In Sec. 3 we derive a unified integral equation for the Green's function in terms of a generalized Fock four-vector, which, in the zero-energy limit, yields directly the integral equation for the zero-energy Green's function. In Sec. 4 we demonstrate how the zero-energy Green's function and wavefunction can be extracted directly from those for negative energies by taking the zero-energy limit, which is tantamount to making a group contraction from O(4) to E(3). Some concluding remarks are made in Sec. 5.

2. THE SCHRÖDINGER WAVEFUNCTION

The fact that the Coulomb system has not only the geometrical symmetry of the group O(3) but also the dynamical symmetry of the group O(4) referred to as the "hidden" symmetry means that there exists an additional constant of the motion besides the angular momentum **L**. Such a constant is known as the Runge–Lenz vector given in the Hermitian form by

$$\mathbf{A}=\mathbf{r}/r-(\mathbf{p}\times\mathbf{L}-\mathbf{L}\times\mathbf{p})/(2p_c), \tag{2.1}$$

where the constant p_c stands for Ze^2m. The operator **A** satisfies the following identity:

$$\mathbf{A}\cdot\mathbf{A}-1=2m(L^2+\hbar^2)H/p_c^2, \tag{2.2}$$

where

$$H=p^2/(2m)-Ze^2/r \tag{2.3}$$

is the Hamiltonian. It should be noted in this connection that the existence of **A** leads to the degeneracy of the energy levels of the hydrogen atom.

We now rewrite Eq. (2.1) in the form

$$\mathbf{A}=\mathbf{B}-m\mathbf{r}H/p_c \tag{2.4}$$

so that when the energy E becomes zero it leads immediately to the zero-energy limit

$$\mathbf{A}(E=0)=\mathbf{B}. \tag{2.5}$$

The operator **B** is not self-adjoint but has the following decomposition:

$$\mathbf{B}=\mathbf{A}_0+i\mathbf{C}, \tag{2.6}$$

where

$$\mathbf{A}_0=-\frac{\mathbf{r}p^2+p^2\mathbf{r}}{4p_c}+\frac{\mathbf{p}(\mathbf{p}\cdot\mathbf{r})+(\mathbf{r}\cdot\mathbf{p})\mathbf{p}}{2p_c}, \tag{2.7}$$

 0022-2488/78/1905-1037$1.00

and

$$\mathbf{C} = \hbar\mathbf{p}/(2p_c) \tag{2.8}$$

are noncommuting Hermitian operators. We shall refer to $\mathbf{A}_0$ as the zero-energy Runge–Lenz vector because it is the properly symmetrized form of its classical counterpart defined by

$$\mathbf{a}_0 = -\mathbf{r}p^2/(2p_c) + \mathbf{p}(\mathbf{p}\cdot\mathbf{r})/p_c. \tag{2.9}$$

The self-adjoint operator $\mathbf{A}_0$ and the generalized coordinates

$$\mathbf{u} = 2p_c\mathbf{p}/p^2 \tag{2.10}$$

satisfy the commutation relation for canonical operators. From Eqs. (2.7) and (2.8) and with the aid of the following commutation relations

$$[\mathbf{p}\cdot\mathbf{r}, p^2] = 2i\hbar p^2, \tag{2.11a}$$

$$[\mathbf{p}\circ\mathbf{r}, \mathbf{p}] = i\hbar\mathbf{p}, \tag{2.11b}$$

$$[\mathbf{r}, \mathbf{p}\cdot\mathbf{r}] = i\hbar\mathbf{r}, \tag{2.11c}$$

$$[\mathbf{r}, p^2] = 2i\hbar\mathbf{p}, \tag{2.11d}$$

it can be easily verified that

$$p_c^2 A_{0\alpha}A_{0\beta} = (\mathbf{p}\cdot\mathbf{r})^2 p_\alpha p_\beta - \tfrac{1}{2}(\mathbf{p}\cdot\mathbf{r})p^2(x_\alpha p_\beta + p_\alpha x_\beta) + \tfrac{1}{4}i\hbar p^2(x_\alpha p_\beta + p_\alpha x_\beta) + \tfrac{1}{4}p^4 x_\alpha x_\beta + \tfrac{1}{4}\hbar^2 p_\alpha p_\beta, \tag{2.12a}$$

$$u_\alpha A_{0\beta} = 2(\mathbf{p}\cdot\mathbf{r})p_\alpha p_\beta/p^2 + 3i\hbar p_\alpha p_\beta/p^2 - p_\alpha x_\beta, \tag{2.12b}$$

$$A_{0\beta}u_\alpha = 2(\mathbf{p}\circ\mathbf{r})p_\alpha p_\beta/p^2 + 3i\hbar p_\alpha p_\beta/p^2 - x_\beta p_\alpha. \tag{2.12c}$$

It then follows that

$$[A_{0\alpha}, A_{0\beta}] = 0, \tag{2.13a}$$

$$[u_\alpha, A_{0\beta}] = i\hbar\delta_{\alpha\beta}. \tag{2.13b}$$

Consequently, $\mathbf{A}_0$ admits the representation[8]

$$\mathbf{A}_0 = -i\hbar\nabla_u. \tag{2.14}$$

From Eqs. (2.6) and (2.8) we also find that the operator $\mathbf{B}$ can be represented by

$$\mathbf{B} = -i\hbar\nabla_u + i\hbar\mathbf{u}/u^2 = -i\hbar u\nabla_u u^{-1}. \tag{2.15}$$

It should be pointed out that, since $\mathbf{C}$ commutes with $\mathbf{u}$, the non-Hermitian operator $\mathbf{B}$ of Eq. (2.6) satisfies the same commutations relations (2.13a and b) with $\mathbf{u}$ as $\mathbf{A}_0$.

To derive the differential equation for the zero-energy Schrödinger wavefunction in the $\mathbf{u}$ representation, we proceed by first evaluating $\mathbf{B}\cdot\mathbf{B}$. With the use of commutation relations in Eq. (2.11) and the following additional ones:

$$[\mathbf{p}\cdot\mathbf{r}, r^{-1}] = i\hbar r^{-1}, \tag{2.16a}$$

$$[p^2, r] = -i\hbar(r^{-1}\mathbf{r}\cdot\mathbf{p} + \mathbf{p}\circ\mathbf{r}r^{-1}), \tag{2.16b}$$

we obtain

$$p_c^2\mathbf{A}_0\cdot\mathbf{A}_0 = \tfrac{1}{4}rp^2rp^2 + \tfrac{1}{2}\hbar^2p^2 - \tfrac{1}{2}i\hbar(\mathbf{p}\cdot\mathbf{r})p^2, \tag{2.17}$$

$$p_c^2\mathbf{C}\cdot\mathbf{C} = \tfrac{1}{4}\hbar^2p^2, \tag{2.18}$$

$$p_c^2\mathbf{A}_0\cdot\mathbf{C} = \tfrac{1}{4}\hbar(\mathbf{p}\cdot\mathbf{r})p^2, \tag{2.19}$$

$$p_c^2\mathbf{C}\cdot\mathbf{A}_0 = \tfrac{1}{4}\hbar(\mathbf{p}\cdot\mathbf{r})p^2 + \tfrac{1}{4}i\hbar^2p^2. \tag{2.20}$$

Upon combining these results, we get

$$\mathbf{B}\cdot\mathbf{B} = rp^2rp^2/(4p_c^2). \tag{2.21}$$

Since Eq. (2.3) gives

$$rp^2/(2p_c) = 1 + mrH/p_c, \tag{2.22}$$

Eq. (2.21) then becomes

$$\mathbf{B}\cdot\mathbf{B} = (1 + mrH/p_c)^2. \tag{2.23}$$

Since the zero-energy Schrödinger wavefunction satisfies $H\Phi_0 = 0$, it follows from Eq. (2.23) that

$$(\mathbf{B}\cdot\mathbf{B} - 1)\Phi_0 = 0. \tag{2.24}$$

We note that Eq. (2.24) can be obtained from Eqs. (2.2) and (2.5) with, however, less rigor. Upon utilizing Eq. (2.15), we obtain finally the desired differential equation

$$-\hbar^2[\nabla_u^2\Phi_0(\mathbf{u}) - 2(\mathbf{u}/u^2)\nabla_u\Phi_0(\mathbf{u})] - \Phi_0(\mathbf{u}) = 0. \tag{2.25}$$

Its solution is given by

$$\Phi_0(\mathbf{u}) = u\exp(i\boldsymbol{\lambda}\circ\mathbf{u}/\hbar), \tag{2.26}$$

with $|\boldsymbol{\lambda}| = 1$. We note that our solution is related to the solution $\Psi_0(\mathbf{u}) = \exp(i\boldsymbol{\lambda}\circ\mathbf{u}/\hbar)$ obtained from the integral equation by

$$\Phi_0(\mathbf{u}) = u\,\Psi_0(\mathbf{u}) \tag{2.27}$$

and , consequently, is related to the ordinary momentum space wavefunction by

$$\Phi_0(\mathbf{u}) = \Psi_0(\mathbf{p})/u^3, \tag{2.28}$$

where

$$\Psi_0(\mathbf{p}) = \exp[2ip_c\boldsymbol{\lambda}\circ\mathbf{p}/(\hbar p^2)]/p^4. \tag{2.29}$$

Furthermore, we have from Eqs. (2.14), (2.15), and (2.27)

$$\mathbf{B}\Phi_0(\mathbf{u}) = -i\hbar u\nabla_u\exp(i\boldsymbol{\lambda}\circ\mathbf{u}/\hbar) = -i\hbar u\mathbf{A}_0\exp(i\boldsymbol{\lambda}\circ\mathbf{u}/\hbar) = u\boldsymbol{\lambda}\exp(i\boldsymbol{\lambda}\cdot\mathbf{u}/\hbar). \tag{2.30}$$

Hence, $\boldsymbol{\lambda}$ is an eigenvalue of $\mathbf{A}_0$. We can accordingly set $\boldsymbol{\lambda} = \mathbf{a}_0$ as given by Eq. (2.9) with $|\mathbf{a}_0| = 1$. Therefore, in our approach, we have not only derived a differential equation for the zero-energy Schrödinger wavefunction but also unambiguously identified the unit vector $\boldsymbol{\lambda}$ which has been left undetermined by the other approach.

3. THE INTEGRAL EQUATION

The Coulomb's Green function in terms of the Fock coordinates

$$\boldsymbol{\xi} = 2p_0\mathbf{p}/(p^2 + p_0^2), \tag{3.1}$$

$$\xi_0 = (p^2 - p_0^2)/(p^2 + p_0^2), \tag{3.2}$$

is given by[4,6]

$$G(\xi, \xi') = \sum_{nlm}\frac{nY_{nlm}(\xi)\,Y^*_{nlm}(\xi')}{n-\nu}. \tag{3.3}$$

where $Y_{nlm}(\xi)$ are the four-dimensional spherical harmonics and $\nu = p_c/p_0$. It satisfies the integral equation

$$G(\xi, \xi') - \frac{\nu}{2\pi^2}\int\frac{d^3\xi''G(\xi', \xi'')}{\xi_0|\xi - \xi''|^2} = \delta(\xi - \xi'). \tag{3.4}$$

It also satisfies the identity

$$G(\xi,\xi') = -\,(16mp_0^3)^{-1}(p^2+p_0^2)^2\,G(\mathbf{p},\mathbf{p}';E)(p'^2+p_0^2)^2. \tag{3.5}$$

The Green's function $G(\xi,\xi')$ diverges at $E=-p_0^2/(2m)=0$. This is understandable since, as the energy, or p_0, approaches zero, the unit sphere defined by $\boldsymbol{\xi}^2+\xi_0^2=1$ collapses into a point given by $\boldsymbol{\xi}=0$, $\xi_0=1$, at which $G(\xi,\xi')$ becomes singular. The failure for $G(\xi,\xi')$ to yield a well-defined zero-energy limit is related to the fact that the integral equation (3.4) is incapable to switch its invariance properties through group contraction[9] converting O(4) to E(3). In order to obtain an integral equation that is susceptible to group contraction, we make a variable change by introducing the generalized Fock coordinates $v=(\mathbf{v},v_0)$ defined by

$$\mathbf{v}=\nu\boldsymbol{\xi}, \tag{3.6}$$

$$v_0=\nu\xi_0. \tag{3.7}$$

These coordinates define, in turn, a sphere in the four-dimensional space $\mathbf{v}^2+v_0^2=\nu^2$, the radius of which, ν, tends to infinity as the energy tends to zero. The use of these coordinates transforms Eq. (3.4) into

$$G(v,v')-\frac{\nu}{2\pi^2}\int\frac{d^3v''}{v_0}\frac{G(v'',v')}{|v-v''|^2}=\delta(v-v'), \tag{3.8}$$

where we have identified

$$G(v,v')=G(\xi,\xi')/\nu^3. \tag{3.9}$$

In the zero-energy limit, we have

$$\lim_{\nu\to\infty}\mathbf{v}=\mathbf{u},\quad \lim_{\nu\to\infty}v_0/\nu=1, \tag{3.10}$$

and Eq. (3.8) becomes the integral equation, invariant with respect to E(3), for the zero-energy Green's function[6]

$$G_0(\mathbf{u},\mathbf{u}')-\frac{1}{2\pi^2}\int\frac{d\mathbf{u}''G_0(\mathbf{u}'',\mathbf{u}')}{|\mathbf{u}-\mathbf{u}'|^2}=\delta(\mathbf{u}-\mathbf{u}'), \tag{3.11}$$

where we have set

$$G_0(\mathbf{u},\mathbf{u}')=\lim_{\nu\to\infty}G(v,v'). \tag{3.12}$$

From Eqs. (3.9) and (3.12), we obtain the desired relationship

$$G_0(\mathbf{u},\mathbf{u}')=\lim_{\nu\to\infty}G(\xi,\xi')/\nu^3 \tag{3.13}$$

for extracting $G_0(\mathbf{u},\mathbf{u}')$ directly from $G(\xi,\xi')$.

4. THE EXTRACTION OF G_0 AND Ψ_0

To proceed with the taking of the limit, we first note that the functions $Y_{nlm}(\xi)$ satisfy the sum rule

$$\sum_{lm}Y_{nlm}(\xi)\,Y^*_{nlm}(\xi')=\frac{n}{2\pi^2}\,\frac{\sin n\chi}{\sin\chi}\,, \tag{4.1}$$

where the angle χ is defined by

$$\cos\chi=1+2p_c^2|\mathbf{p}-\mathbf{p}'|^2/[\nu^2(p^2+p_0^2)(p'^2+p_0^2)]. \tag{4.2}$$

Utilizing the sum rule, we obtain from Eqs. (3.3) and (3.13)

$$G_0(\mathbf{u},\mathbf{u}')=\lim_{\nu\to\infty}\frac{1}{\nu^3}\sum_n\frac{n^2}{2\pi^2(n-\nu)}\,\frac{\sin n\chi}{\sin\chi}\,. \tag{4.3}$$

When ν becomes very large, the angle χ becomes very small. Expanding $\cos\chi$ to χ^2 and letting

$$t=|\mathbf{u}-\mathbf{u}'|=2p_c|\mathbf{p}-\mathbf{p}'|/(pp'), \tag{4.4}$$

we obtain from Eq. (4.2)

$$\nu\chi=t. \tag{4.5}$$

We then introduce the parameter

$$x=n/\nu. \tag{4.6}$$

In taking the limit, we set

$$\sum_{n=0}^{\infty}\to\nu\int_0^\infty dx. \tag{4.7}$$

Thus, we obtain the desired result

$$G_0(\mathbf{u},\mathbf{u}')=\frac{1}{2\pi^2}\int_0^\infty\frac{dx\,x^2\sin xt}{t(x-1)}\,, \tag{4.8}$$

which can be rewritten in the form[6]

$$G_0(\mathbf{u},\mathbf{u}')=\frac{1}{(2\pi)^3}\int\frac{d\mathbf{x}\,x\exp(i\mathbf{x}\cdot\mathbf{t})}{x-1}\,. \tag{4.9}$$

It should be noted that the singularity at $x=1$ physically corresponds to the bound states of the hydrogen atom.

It is of interest to note that by separating out the singular parts we can express $G_0(\mathbf{u},\mathbf{u}')$ of Eq. (4.9) as

$$G_0(\mathbf{u},\mathbf{u}')=\delta(\mathbf{u}-\mathbf{u}')+\frac{1}{2\pi^2}\,\frac{1}{t^2}-\frac{1}{2\pi^2 t}\times[\sin t\,\mathrm{ci}(-t)+\cos t\,\mathrm{si}(-t)], \tag{4.10}$$

where $\mathrm{ci}(-t)$ and $\mathrm{si}(-t)$ are, respectively, the cosine and sine integrals.

In a similar manner, the method can be applied to obtain the zero-energy wavefunction by evaluating the residue of $G(\nu,\nu')$ in the ν plane at the pole $\nu=n$ and then letting n tend to infinity. Utilizing Eqs. (3.3), (3.5), and (3.9), we obtain

$$\begin{aligned}\sum_{ln}\Psi_{\nu lm}(\mathbf{p})\Psi^*_{\nu lm}(\mathbf{p}') &= \mathrm{Res}[G(\mathbf{p},\mathbf{p}';E)\text{ at }\nu=n]\\ &=\frac{16mp_c^3}{(p^2+p_0^2)(p'^2+p_0^2)n^2}\sum_{lm}Y_{nlm}(\xi)Y^*_{nlm}(\xi')\\ &=\frac{8mp_c^3}{\pi^2(p^2+p_0^2)(p'^2+p_0^2)}\,\frac{\sin n\chi}{n\sin\chi}\,.\end{aligned} \tag{4.11}$$

Taking the limit $\nu=n\to\infty$ and defining the zero-energy wave function by

$$\Psi_{0lm}(\mathbf{p})\Psi^*_{0lm}(\mathbf{p}')=\lim_{\nu\to\infty}\Psi_{\nu lm}(\mathbf{p})\Psi^*_{\nu lm}(\mathbf{p}'), \tag{4.12}$$

we obtain from Eqs. (4.5) and (4.11)

$$\sum_{lm}\Psi_{0lm}(\mathbf{p})\Psi^*_{0lm}(\mathbf{p}')=\frac{8mp_c^3}{\pi^2p^4p'^4}\,\frac{\sin t}{t}=\frac{8mp_c^3}{\pi^2p^4p'^4}\,j_0(t), \tag{4.13}$$

where j_0 is the spherical Bessel function. By using the partial wave expansion and the addition theorem for spherical Bessel functions, we obtain the desired zero-energy wavefunction[6]

$$\Psi_{0lm}(\mathbf{p})=\frac{\text{const}}{p^4}\,j_l\!\left(\frac{2p_c}{p}\right)Y_{lm}\!\left(\frac{\mathbf{p}}{p}\right), \tag{4.14}$$

which is in agreement with the partial wave expansion of Eq. (2.29).

It should be noted that the wavefunction $\Psi_{\nu lm}(\mathbf{p})$ satisfy the normalization condition

$$\int \Psi_{\nu lm}(\mathbf{p})\Psi^*_{\nu' lm}(\mathbf{p})\, d\mathbf{p} = \delta[E(\nu) - E(\nu')] \tag{4.15}$$

on the energy scale, where

$$E(\nu) = -p_c^2/(2m\nu^2). \tag{4.16}$$

They can be shown to satisfy the relation

$$\Psi_{\nu lm}(\mathbf{p}) = [\nu(m\nu)^{1/2}/p_c]\Psi_{nlm}(\mathbf{p}), \tag{4.17}$$

where $\Psi_{nlm}(\mathbf{p})$ are the Coulomb wavefunctions in the momentum space and satisfy the normalization condition

$$\int \Psi_{nlm}(\mathbf{p})\Psi^*_{n'lm}(\mathbf{p})\, d\mathbf{p} = \delta_{nn'}. \tag{4.18}$$

Furthermore, it is interesting to observe that the residue of the Green's function in the ν plane yields wavefunctions normalized on the energy scale. The success of our approach is undoubtedly based upon the fact that the bound-state poles in the ν plane are equally spaced instead of an infinite accumulation of such poles at zero energy on the energy plane or the p_0 plane.

5. CONCLUDING REMARKS

In examining the zero-energy aspect of the Coulomb problem, we have uncovered a new self-adjoint operator, the zero-energy Runge–Lenz vector, which has enabled us to derive a differential equation for the zero-energy Schrödinger wavefunction. In addition, we have derived an integral equation in generalized Fock coordinates, which applies to all energies through analytic continuation and by group contraction. We have also extracted the zero-energy Green's function and wavefunction by taking the zero-energy limit in the ν plane, in which the bound state poles are equally spaced instead of accumulating to result in a condensation of levels. The usefulness of these results in Coulomb scattering problems and the existence of wavefunction in the **u** representation for nonzero energies should be explored.

[1] A. O. Barut, Phys. Today 26, 57 (August 1973).
[2] See for example: G. L. Nutt, J. Math. Phys. 9, 796 (1968); J. S. Ball, J. C. Y. Chen, and D. Y. Wong, Phys. Rev. 173, 202 (1968); C. S. Shastry, L. Kumar, and J. Callaway, Phys. Rev. A 2, 781 (1970); J. C. Y. Chen, A. C. Chen, and P. J. Kramer, *ibid.* 4, 1982 (1971).
[3] V. Fock, Z. Phys. 98, 145 (1935).
[4] J. Schwinger, J. Math. Phys. 5, 1606 (1964).
[5] H. H. Rogers, J. Math. Phys. 14, 1125 (1973).
[6] A. M. Perelomov and V. S. Popov, Sov. Phys.-JETP 23, 118 (1966).
[7] For a review of the Coulomb Green's function, see J. C. Y. Chen and A. C. Chen, Advan. At. Mol. Phys. 8, 71 (1972).
[8] Classically, the zero-energy Runge–Lenz vector $\mathbf{a}_0$ of Eq. (2.9) and the generalized coordinates **u** of Eq. (2.10) satisfy the Poisson bracket relations. In terms of **u** and $\mathbf{a}_0$, the Hamiltonian can be written as $H = 2p_c^2(1 - 1/a_0)/(mu^2)$, which is valid for all energies instead of just the zero energy as noted by A. Norcliffe, J. Phys. B 4, 143 (1971). Hamilton's equations of motion can be obtained accordingly and, for the zero-energy case, can be solved exactly. The operator representation can be obtained by simply invoking the correspondence principle.
[9] For a discussion on group contraction, see P. Chand, C. L. Mehta, N. Mukunda, and E. C. G. Sudarshan, J. Math. Phys. 8, 2048 (1967).

Four Euclidean conformal group in atomic calculations: Exact analytical expressions for the bound–bound two-photon transition matrix elements in the H atom

J. P. Gazeau

Laboratoire de Chimie Physique [a] *de l'Université Pierre et Marie Curie, 75231 Paris Cedex 05, France*
(Received 7 June 1977)

By combining Sturmian–Coulomb techniques with a local representation of the four Euclidean conformal group SU*(4) ≃Spin(1,5), a compact analytical form, suitable for any analytic continuation on the energy variable, is obtained for the following bound–bound two-photon transition matrix element in the H atom: $I_{NN'}(E) = \langle N|\mathbf{p}\cdot\epsilon_1 \exp(i\mathbf{k}_1\cdot\mathbf{r})G(E)\mathbf{p}\cdot\epsilon_2\exp(i\mathbf{k}_2\cdot\mathbf{r})|N'\rangle$, where $G(E)$ is the Coulomb Green's function.

INTRODUCTION

In this paper, yet another method is presented for the calculation of matrix elements of two photon transitions between two hydrogenic bound states. Although a large amount of literature exists on this subject already,[1–12] general compact expressions valid for arbitrary states, and suitable for analytic continuation on the energy variable, are still unknown.

We propose a group theoretical technique, solving this problem completely, by using the four Euclidean conformal group, isomorphic to Spin(1, 5), the universal covering group of $SO_+(1,5)$.[13] This approach describes the transformation induced by a general boost in an "energy–momentum" space and the Fock stereographic projection[14] in terms of a conformal transformation in the quaternion field $\mathbb{H}$. Thus, this method differs from the usual ones where the boost is described by exponentiation of the infinitesimal action of the Lie algebra so(2, 4) on the Hydrogenic states.[9,15]

The organization of this paper is as follows. In Sec. 1 we formulate the Fock treatment of the H atom by introducing a "coupling constant" operator which acts on a Hilbert space, denoted by $\mathcal{H}(p_0)$ [identical to $L^2_c(\mathrm{SU}(2))$]. Both its eigenvectors or "Sturmian functions" and its eigenvalues are dependent on the energy E which is a fixed parameter $p_0 \equiv (-2mE)^{1/2}$.[16–18]

In Sec. II, we introduce the group SU*(4) ≈ Spin(1, 5) and define its action on $\mathbb{H}$ as a conformal transformation. Then we consider a local representation of this group on $L^2_C(\mathrm{SU}(2))$ which is linear when it is restricted to Spin(1, 4), and define some matrix elements which are computed in Appendix A.

In Sec. III, the above elements are used to give exact analytic expressions for transition matrix elements in the H atom, and to recover the classical results for the elastic transitions in the dipole approximation and to extend it to higher orders. The formulas which are obtained are suitable for any analytic continuation.

1. A SURVEY OF THE STURMIAN PROBLEM AND THE FOCK METHOD

The Schrödinger equation for two charged particles without spin in terms of relative coordinates and momenta can be written as (in natural units)

$$(p_0^2 + p^2)\psi = c^{-1}V\psi, \tag{1.1}$$

where

$$p_0 = (-2mE)^{1/2}, \quad c^{-1} = 2m\alpha z, \quad V = \frac{1}{r}.$$

For fixed E or equivalently fixed p_0, Eq. (1.1) is also the eigenvalue equation for the so-called "Sturmian operator" or "coupling constant operator,"[16–18]

$$C \equiv (p_0^2 + p^2)^{-1}V, \tag{1.2}$$

$$C\psi = c\psi. \tag{1.3}$$

When $E < 0$, the spectrum of C is infinite and discrete, whereas it is continuous for $E > 0$, and the algebraic relation between the eigenvalues c of C and the parameter p_0 allows one to find the energy spectrum, but the basic difference between the Hamiltonian problem and the Sturmian problem must be emphasized. For instance, C is not Hermitian in $L^2_C(\mathbb{R}^3)$, the Hilbert space of complex square integrable functions on $\mathbb{R}^3$. It is possible to render it Hermitian in a pre-Hilbertian space in correspondence with the first by the p_0 dependent transformation: Suppose E negative, for all $\psi \in L^2_C(\mathbb{R}^3)$ such that $|(\psi, (p_0^2 + p^2)\psi)| < \infty$; we associate a weighted state

$$\psi' = (p_0^2 + p^2)^{1/2}\psi, \tag{1.4}$$

where the square root makes sense since $(p_0^2 + p^2)$ is diagonal in momentum space. Then

$$C' = (p_0^2 + p^2)^{1/2}C(p_0^2 + p^2)^{-1/2} = (p_0^2 + p^2)^{-1/2}V(p_0^2 + p^2)^{-1/2} \tag{1.5}$$

is clearly Hermitian on the space generated by the ψ'.

A similar treatment was used by Fock.[6,14,19,20] The Fock method consists of two operations.

The change of the integration variable in the scalar product of $L^2_C(\mathbb{R}^3)$ introduces the Hilbert space $L^2_C(S^3)$, where S^3 is the unit sphere of $\mathbb{R}^4$. The multiplication of the states by a weight renders C Hermitian in $L^2_C(S^3)$.

Explicitly, the Fock stereographic projection, denoted by $s(p_0)$, brings the unit sphere S^3 onto the compactified hyperplane $H(p_0)$ which is isomorphic to the momentum space,

[a] Laboratoire "Matière et Rayonnement" associé au C.N.R.S.

 0022-2488/78/1905-1041$1.00

$$x=(p_0,\mathbf{p})\in H(p_0)$$
$$\rightarrow \xi=s^{-1}(p_0)(x)\begin{cases}\boldsymbol{\xi}=\dfrac{2p_0\mathbf{p}}{p_0^2+p^2},\\ \xi_0=\dfrac{p_0^2-p^2}{p_0^2+p^2}.\end{cases}\tag{1.6}$$

The relation between the Euclidean measure $d^3\mathbf{p}$ and the O(4) invariant measure on S^3 is

$$d\mu(\xi)=\left(\frac{2p_0}{p_0^2+p^2}\right)^3 d^3\mathbf{p}$$
$$=\left(\frac{|\underline{1}+\xi|^2}{2p_0}\right)^3 d^3\mathbf{p},\tag{1.7}$$

where $\underline{1}\equiv(1,0)$.

The Fock correspondence $\mathcal{J}_{p_0}$ between the two Hilbert spaces $L^2_{\mathbb{C}}(\mathbb{R}^3)$ and $L^2_{\mathbb{C}}(S^3)$, the latter denoted by $\mathcal{H}(p_0)$, is

$$\psi\in L^2_{\mathbb{C}}(\mathbb{R}^3)\xrightarrow{\mathcal{J}_{p_0}}\phi$$
$$=\frac{1}{\sqrt{p_0}}\left(\frac{2p_0}{|\underline{1}+\xi|^2}\right)^2\psi\circ s(p_0),\tag{1.8}$$
$$\phi\in L^2_{\mathbb{C}}(S^3)\approx\mathcal{H}(p_0),$$

and reciprocally

$$\phi\in\mathcal{H}(p_0)\xrightarrow{\mathcal{J}_{p_0}^{-1}}\psi$$
$$=\sqrt{p_0}\left(\frac{2p_0}{p_0^2+p^2}\right)^2\phi\circ s^{-1}(p_0).\tag{1.8'}$$

Their respective scalar products are related by

$$(\psi_1,\psi_2)_{L^2_{\mathbb{C}}(\mathbb{R}^3)}=p_0\left(\phi_1,\left(\frac{|\underline{1}+\xi|^2}{2p_0}\right)\phi_2\right)_{\mathcal{H}(p_0)},\tag{1.9}$$

$$(\phi_1,\phi_2)_{\mathcal{H}(p_0)}=\frac{1}{p_0}\left(\psi_1,\left(\frac{2p_0}{p_0^2+p^2}\right)\psi_2\right)_{L^2_{\mathbb{C}}(\mathbb{R}^3)},\tag{1.9'}$$

and the eigenvalue equation (1.3) for the Sturmian operator is written as an integral equation,

$$\hat{C}\phi(\xi)=\frac{1}{2\pi^2}\int_{S^3}d\mu(\xi')|\xi-\xi'|^{-2}\phi(\xi')$$
$$=\hat{c}\phi(\xi),\tag{1.10}$$

where

$$\phi=\mathcal{J}_{p_0}\psi,\quad \hat{C}=\mathcal{J}_{p_0}C\mathcal{J}_{p_0}^{-1},\quad \hat{c}=2p_0c=\frac{p_0}{m\alpha z}.$$

$\hat{C}$ is clearly Hermitian and O(4) invariant. Its eigenvalues are

$$\hat{c}_n=\frac{1}{n},\quad n\in\mathbb{N}^*.\tag{1.11}$$

A natural system of eigenvectors is the set of the spherical harmonics Y_{nlm}[20] on S^3.

Returning to the Hamiltonian problem, by solving (1.11) in p_0 and carrying out the corresponding transformation $\mathcal{J}_{p_0}^{-1}$, the well-known eigenvalues and related eigenstates in the momentum space are obtained:

$$p_0=\frac{\lambda}{n}\equiv p_n\quad \lambda\equiv m\alpha z\quad\left(\text{i.e.},\ E_n=\frac{\lambda^2}{2mn^2}\right),\tag{1.12}$$

$$\psi_{nlm}=\mathcal{J}_{p_n}^{-1}Y_{nlm},\quad \xi_n=s^{-1}(p_n)((p_n,\mathbf{p})),$$
$$\psi_{nlm}(\mathbf{p})=\frac{4p_n^{5/2}}{(p_n^2+p^2)^2}Y_{nlm}(\xi_n).\tag{1.13}$$

2. SU*(4) APPROACH TO THE STURMIAN-COULOMB PROBLEM

The dynamical symmetry O(4,2) of the H atom has already been intensively used in atomic calculations, mainly by Barut and Kleinert[15] and by Fronsdal.[8] See also Refs. 9 and 10. The main difficulty in the generalization of these methods to many atomic calculations (transition matrix elements, etc.,$\cdots$) is the translation of the action of operators on the states space in terms of "abstract rotations" deduced by exponentiation of the action of a representation of the Lie algebra o(4,2) or of its enveloping algebra. Part of the physical meaning is lost when the matrix elements which describe the processes are re-expressed in terms of real or imaginary angles.

We exploit all the resources of the quaternionic calculus. This is an advantage in itself. Indeed, when the calculation of matrix elements describing the transition from one Coulomb state to another is required, the energy jumps and momentum transfers characterizing a "general physical boost" are described by a translation in an "energy–momentum" space. This, one naturally identifies as the quaternion field $\mathbf{H}$ when the "energy component" $p_0=(-2mE)^{1/2}$ is real. Now, the Fock projection transforms the translation group element to an $Sp(1,1)\approx Spin(1,4)$ element.[13] More generally if other effects are taken into account (e.g., intermediate summation in the form of a Green function, as arises in perturbation calculations) a Fock projection will give an element of the four Euclidean conformal group SU*(4)[13] which is isomorphic to Spin(1,5). Explicitly, in this way a convenient representation of Spin(1,5) is obtained in the form of a group of 2×2 quaternionic matrices, denoted by SU*(4), which acts on $\mathbf{H}$ as a homographic transformation. It is then a simple matter to multiply 2×2 matrices together as other processes follow. Moreover, the fact that their matrix elements, which are physically undimensioned quadrivectors, are very simply and strangely [see Eq. (2.8)] connected to the general boost parameters of the process under consideration merits deeper understanding. This will be gone into elsewhere.

Now the way in which the conformal group appears naturally in the Sturmian problem and Fock method when $p_0=(-2mE)^{1/2}$ is real, is explained.

Let us consider the quaternion field $\mathbf{H}$, the elements of which will be denoted by $x=(x_0,\mathbf{x})$, where x_0 is the scalar part and $\mathbf{x}$ the vector part[21]:

$$x=x_0\mathbf{1}+x_1\mathbf{i}+x_2\mathbf{j}+x_3\mathbf{k},$$
$$xx'=(x_0x_0'-\mathbf{x}\cdot\mathbf{x}_0',x_0\mathbf{x}'+x_0'\mathbf{x}+\mathbf{x}\times\mathbf{x}'),\tag{2.1}$$
$$\bar{x}\equiv(x_0,-\mathbf{x}).$$

Spin(1,5) is isomorphic to the group SU*(4) of 2×2 matrices with quaternionic entries verifying a scalar relation

$$\text{SU*(4)}$$
$$=\left\{g=\begin{pmatrix}a & b\\ c & d\end{pmatrix};a,b,c,d\in\mathbf{H};\ |c||d||ac^{-1}-bd^{-1}|=1\right\},\tag{2.2}$$

where $x \to |x|$ is the Euclidean norm in $\mathbf{H}$, also called the modulus of the quaternion x.

SU*(4) acts on $\mathbf{H}$ via conformal transformations:

$$x \in \mathbf{H}, \quad g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SU}^*(4), \qquad x \to g \circ x = (ax+b)(cx+d)^{-1}. \tag{2.3}$$

The following important property holds:

$$|cx+d|\,|g\cdot x - g\cdot y|\,|cy+d| = |x-y| \tag{2.4}$$

for all $x, y \in \mathbf{H}$, $g \in \mathrm{SU}^*(4)$.

The Fock stereographic projection $s(p_0)$ is a particular case of this action. It establishes a one to one correspondence between the subgroup [isomorphic to, and briefly denoted by SU(2)] of unit modulus quaternions and the hyperplane of the quaternions having the same scalar part p_0; let us put:

$$x = (p_0, \mathbf{p}),$$

$$s(p_0) = \frac{1}{\sqrt{2p_0}} \begin{pmatrix} 2p_0 & 0 \\ 1 & 1 \end{pmatrix}, \quad s^{-1}(p_0) = \frac{1}{\sqrt{2p_0}} \begin{pmatrix} 1 & 0 \\ -1 & 2p_0 \end{pmatrix}, \tag{2.5}$$

where

$$\lambda \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} \lambda a & \lambda b \\ \lambda c & \lambda d \end{pmatrix}, \quad \lambda \in \mathbf{R},$$

and 0 and 1 henceforward denote the zero and unit elements both in $\mathbf{H}$ and in $\mathbf{R} \subset \mathbf{H}$. (1.6) is now written

$$\xi = s^{-1}(p_0)\cdot x = x\bar{x}^{-1},$$

where $\bar{x}$ is the quaternionic conjugate of x,

$$\bar{x} = (p_0, -\mathbf{p}). \tag{2.6}$$

Now, we describe in terms of the action (2.3) the general physical "boost" which also includes the Galilean boost in momentum space

$$\mathbf{p} \to \mathbf{p}' = \mathbf{p} + \mathbf{k},$$

as a scalar boost

$$p_0 \to p_0' = p_0 + k_0,$$

$$x = (p_0, \mathbf{p}) \to x' = (p_0', \mathbf{p}') = (p_0 + k_0, \mathbf{p} + \mathbf{k}) = t_K \circ x, \tag{2.7}$$

where

$$t_K = \begin{pmatrix} 1 & K \\ 0 & 1 \end{pmatrix} \in T_4 \subset \mathrm{SU}^*(4), \qquad K \equiv (k_0, \mathbf{k}),$$

T_4: group of translations in $\mathbf{H}$.

The transformation induced on S^3 by the translations in $\mathbf{H}$ after the inverse stereographic projection is

$$\xi' = (s^{-1}(p_0')t_K s(p_0))\cdot \xi = h\cdot\xi, \tag{2.8}$$

where

$$\xi' = s^{-1}(p_0')\cdot x', \quad \xi = s^{-1}(p_0)\circ x,$$

$$h = \frac{1}{2(p_0 p_0')^{1/2}} \begin{pmatrix} K_+ & K_- \\ \bar{K}_- & \bar{K}_+ \end{pmatrix}, \quad K_\pm = (p_0' \pm p_0, \mathbf{k}).$$

h is an element of the subgroup $\mathrm{Sp}(1,1) \approx \mathrm{Spin}(1,4)$, which leaves both SU(2) and the unit ball invariant under the conformal transformation. Spin(1,4) has been intensively studied (see for instance Fronsdal[22]), particularly under its quaternionic representation by Takahaski[23] and Strömm.[24] In atomic computations we shall use the following (nonunitary) irreducible linear representation of this group[22]:

$$f \in L^2_{\mathbf{C}}(\mathrm{SU}(2)) \approx L^2_{\mathbf{C}}(S^3),$$

$$h^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \Leftrightarrow h = \begin{pmatrix} \bar{a} & -\bar{c} \\ -\bar{b} & \bar{d} \end{pmatrix},$$

$$T^\rho(h) f(\xi) = |c\xi + d|^{-2\rho} f(h^{-1}\cdot\xi)$$

with $\rho = 1$ or 2. (2.9)

We define the matrix elements of this representation with respect to the orthonormal basis $\{Y_N\}$, $N = nlm$,

$$T^\rho_{NN'}(h) = (T^\rho(h) Y_{N'}, Y_N)_{L^2_{\mathbf{C}}(\mathrm{SU}(2))}. \tag{2.10}$$

Since

$$d\mu(h^{-1}\cdot\xi) = d\mu(\xi)\,|c\xi + d|^{-6}, \tag{2.11}$$

these matrix elements verify

$$T^2_{NN'}(h) = (T^1_{N'N}(h^{-1}))^*. \tag{2.12}$$

One may show that (see Appendix B):

$$T^1_{NN'}(h) = \frac{n'}{n}(T^1_{N'N}(h^{-1}))^*. \tag{2.13}$$

It follows from (2.12) and (2.13) that

$$T^2_{NN'}(h) = \frac{n}{n'} T^1_{NN'}(h). \tag{2.14}$$

Now we extend the representation T^ρ to SU*(4) in the following way: Let us continue an element f of $L^2_{\mathbf{C}}(\mathrm{SU}(2))$ to a function defined inside (outside) $S^3 \approx \mathrm{SU}(2)$:

$$f \to F^\zeta_\gtrless,$$

$$F_<(x) = \frac{1}{2\pi^2} \int_{\mathrm{SU}(2)} d\mu(\xi') f(\xi') \frac{1-|x|^2}{|\xi' - x|^4}, \quad |x| < 1, \tag{2.15}$$

$$F_>(x) = |x|^{-2} F_<(x^{-1}), \quad |x| > 1,$$

$$F_<(x) = F_>(x) = f(x), \quad |x| = 1.$$

Putting $x = |x|\,\xi$, $F_<$ (resp. $F_>$) as a function of the variable ξ belongs to $L^2_{\mathbf{C}}(\mathrm{SU}(2))$.

We define the local representation of SU*(4) on $L^2_{\mathbf{C}}(\mathrm{SU}(2))$ in the following way,

$$f \in L^2_{\mathbf{C}}(\mathrm{SU}(2)), \quad g^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SU}^*(4),$$

$$T^\rho(g) f(\xi) = |c\xi + d|^{-2\rho} F^\zeta_\gtrless(g^{-1}\cdot\xi), \tag{2.16}$$

$\gtrless$ according as to whether $|g^{-1}\cdot\xi| \gtrless 1$.

The representation T^ρ is local in the sense that it is linear for all g in some neighborhood of e, unit element of SU*(4). Now, let us consider the integral

$$S^\rho_{NN'}(g) = \int_{\mathrm{SU}(2)} d\mu(\xi)\,|c\xi + d|^{-2\rho} \mathcal{Y}_{N'}(g^{-1}\cdot\xi') Y^*_N(\xi), \tag{2.17}$$

where $\mathcal{Y}_N(x)$ is the harmonic polynomial ("solid harmonic") deduced from the surface harmonic by the homogeneity formula

$$\mathcal{Y}_N(x) = |x|^{n-1} Y_N\left(\frac{x}{|x|}\right).$$

It is homogeneous of degree $n-1$.

If $|g^{-1}\circ\xi|\leq 1$ for all $\xi\in SU(2)$, $S^{\rho}_{NN'}(g)$ is the matrix element $T^{\rho}_{NN'}(g)$ of the operator $T^{\rho}(g)$. The same remains true if $|g^{-1}\circ\xi|\geq 1$ for all $\xi\in SU(2)$. We give the analytic expression of the integral (2.17) in Appendix A for $\rho=1$ and $|c|<|d|$.

3. CALCULATION OF THE BOUND-BOUND TWO PHOTON TRANSITION MATRIX ELEMENTS IN THE H ATOM

In a set of particular cases, exact analytic expressions have already been available for about ten years for the following matrix elements (between bound or continuum states),

$$I^2_{NN'}(E,\epsilon_1,\mathbf{k}_1,\epsilon_2,\mathbf{k}_2)\equiv I^2_{NN'}(E)=\langle N|\mathbf{p}\cdot\epsilon_1\exp(i\mathbf{k}_1\circ\mathbf{r})G(E)\mathbf{p}\cdot\epsilon_2\exp(i\mathbf{k}_2\circ\mathbf{r})|N'\rangle, \tag{3.1}$$

where $G(E)=(H-E)^{-1}$ is the Coulomb Green function.

ϵ_i (resp. $\mathbf{k}_i$) is the polarization vector (resp. momentum) of the ith photon.

The following various techniques were used:

(i) analytic methods in configuration space, by use of the Hostler integral representation[25] or the Sturmian Coulomb Green function: in the dipole approximation[1-3] and with the retardation effects[3];

(ii) analytic methods in momentum space, by use of the Schwinger integral representation of the Coulomb Green function[26]: in the dipole approximation[4] and with retardation effects[5];

(iii) analytic methods in L^2_C [SU(2)], by use of the Fock method and the Schwinger integral representation and the harmonic analysis on SU(2): in the dipole approximation[6] and with retardation effects[1,7];

(iv) algebraic techniques, by use of the dynamical group O(4, 2), with retardation effect[8-10];

(v) numerical techniques, by numerical integration of inhomogeneous differential equations.[11,12]

None of these methods is able to give a general compact analytical expression for the matrix elements $I^2_{NN'}(E)$ between any initial and final states.

It is evident that our expressions provide their own analytic continuation in contrast to the usual situation.

The detail of the method is given in Appendix C. We shall now examine the results.

Let us define the two elements of Sp(1,1),

$$h_1=\frac{1}{2\sqrt{p_0p_n}}\begin{pmatrix}K_{1+} & -\bar{K}_{1-}\\ -K_{1-} & \bar{K}_{1+}\end{pmatrix},\quad h_2=\frac{1}{2\sqrt{p_0p_{n'}}}\begin{pmatrix}K_{2+} & K_{2-}\\ \bar{K}_{2-} & \bar{K}_{2+}\end{pmatrix},\quad \text{and } \exp[(x/2)\upsilon]=\begin{pmatrix}e^{x/2} & 0\\ 0 & x^{-x/2}\end{pmatrix}, \tag{3.2}$$

where

$$p_n=(-2mE_n)^{1/2},\quad p_{n'}=(-2mE_{n'})^{1/2},\quad p_0=(-2mE)^{1/2},\quad K_{1\pm}=(p_0\pm p_{n'},\mathbf{k}_1),\quad K_{2\pm}=(p_0\pm p_{n'},\mathbf{k}_2). \tag{3.3}$$

Let us put $\nu=m\alpha z/p_0$.

Then, we obtain for the matrix element (3.1) the following integral representation

$$I^2_{NN'}(E)=-m\nu(nn')^{-1/2}\sum_{N_0,N_0'}C_{N_0^{\dagger}N'}(\epsilon_2)C^{*}_{NN_0}(\epsilon_1)n_0\int_0^{+\infty}dx\,e^{\nu x}T^1_{N_0N_0'}(g(x)), \tag{3.4}$$

with

$$SU^*(4)\ni g^{-1}(x)=(h_1\exp[(x/2)\upsilon]h_2)^{-1}=\frac{1}{4p_0\sqrt{p_np_{n'}}}\begin{pmatrix}\bar{K}_{2+}\bar{K}_{1+}e^{-x/2}-K_{2-}K_{1-}e^{x/2} & \bar{K}_{2+}\bar{K}_{1-}e^{-x/2}-K_{2-}K_{1+}e^{x/2}\\ -\bar{K}_{2-}\bar{K}_{1+}e^{-x/2}+K_{2+}K_{1-}e^{x/2} & -\bar{K}_{2-}\bar{K}_{1-}e^{-x/2}+K_{2+}K_{1+}e^{x/2}\end{pmatrix}. \tag{3.5}$$

The matrix elements $C_{NN'}(\epsilon)$ are given by (D2). We can show that

$$|-\bar{K}_{2-}\bar{K}_{1+}e^{-x/2}+K_{2+}K_{1-}e^{x/2}|<|-\bar{K}_{2-}\bar{K}_{1-}e^{-x/2}+K_{2+}K_{1+}e^{x/2}|, \tag{3.6}$$

which corresponds to the particular case given in Appendix A. Expression (3.4) can be reduced to a finite sum of integrals $\mathcal{G}^b_a$ defined by[1,7]

$$\mathcal{G}^b_a(y,y')=\int_0^{+\infty}dx\exp(-bx)|y'-e^{-x}y|^{-2a}=|y'|^{2a}b^{-1}F_1(b,a,a,b+1;\rho e^{i\omega},\rho e^{-i\omega}), \tag{3.7}$$

where

$$y,y'\in\mathbf{H},\quad |y|<|y'|,\quad \rho=|y||y'|^{-1},\quad \omega=(y,y')\hat{}.$$

F_1 is an Appell function.[27] Thus

$$n_0\int_0^{+\infty}dx\,e^{\nu x}T^1_{N_0N_0'}(g(x))=\sum_{q,q'}C^{qq'}_{N_0N_0'}(h_1,h_2)\mathcal{G}^{q-\nu}_{q'}(\bar{K}_{2-}\bar{K}_{1-},K_{2+}K_{1+}). \tag{3.8}$$

The expression of the coefficients $C^{qq'}_{N_0N_0'}(h_1,h_2)$ is given in Appendix E.

Let us consider explicitly two important particular cases. In the dipole approximation ($\mathbf{k}_1 = \mathbf{k}_2 = 0$) we have

$$g^{-1}(x) = \frac{(1+\nu/n)(1+\nu/n')e^{x/2}}{4(\nu^2/nn')^{1/2}} \begin{pmatrix} e^{-x} - r_n r_{n'} & r_n e^{-x} - r_{n'} \\ r_n - r_{n'} e^{-x} & 1 - r_n r_{n'} e^{-x} \end{pmatrix}, \quad \text{where } r_n = \frac{1-\nu/n}{1+\nu/n}, \quad r_{n'} = \frac{1-\nu/n'}{1+\nu/n'} \tag{3.9}$$

and (3.8) now becomes

$$\begin{aligned} n_0 \int_0^{+\infty} dx\, e^{\nu x} \mathcal{T}^1_{N_0 N'_0}(g(x)) = \delta_{l_0 l'_0} \delta_{m_0 m'_0} & [n_0 n'_0 (n_0 - l_0 - 1)!\,(n'_0 - l_0 - 1)!\,(n_0 + l_0)!\,(n'_0 + l_0)!\,]^{1/2} \sum_{q,\sigma} \left(\frac{(1-\nu^2/n^2)(1-\nu^2/n'^2)}{16\nu^2/nn'} \right)^{q-n_<} \\ & \times \frac{(-1)^\sigma r_<^{n_< + \sigma}\, {}_2F_1(-q, -\sigma; q + n_> - n_< - \sigma + 1; (r_>/r_<)^2)}{\sigma!\, q!\, (n_> - n_< + q - \sigma)!\, (n_< - l_0 - 1 - q)!)(n_< + l_0 - q)!} \\ & \times (n_> + q - \sigma - \nu)^{-1}\, {}_2F_1(n_0 + n'_0, n_> + q - \sigma - \nu; n_> + q - \sigma - \nu + 1; r_n r_{n'}), \end{aligned} \tag{3.10}$$

where

$$n_{\gtrless} = \sup_{\inf}(n_0, n'_0), \quad r_< = \begin{matrix} r_{n'} & \text{if } n_0 = n_< \\ r_n & \text{if } n'_0 = n_< \end{matrix}, \quad r_> = \begin{matrix} r_{n'} & \text{if } n_0 = n_> \\ r_n & \text{if } n'_0 = n_> \end{matrix}.$$

For $n=2$, $n'=1$, $l=l'=m=m'=0$, this expression is in agreement with Ref. 3. Particularly, for $N=N'$, and $\epsilon_1 = \epsilon_2 = \hat{z}$ (elastic scattering), we give a complete expression,

$$I^2_{NN}(E) = -\frac{m\nu}{4n} \sum_{l_0} \{A_{l l_0} S_{n-1, n-1}(l_0, \nu) + B_{l l_0} S_{n+1, n+1}(l_0, \nu) - 2C_{l l_0} S_{n-1, n+1}(l_0, \nu)\}, \tag{3.11}$$

where

(1) $S_{n_0 n'_0}(l_0, \nu)$

$$\begin{aligned} = & [n_0 n'_0 (n_0 - l_0 - 1)!\,(n'_0 - l_0 - 1)!\,(n_0 + l_0)!\,(n'_0 + l_0)!\,]^{1/2} \left(1 + \frac{\nu}{n}\right)^{-2(n_0 + n'_0)} \\ & \times (-1)^{n_0 - n'_0} \sum_{q=0}^{n_< - l_0 - 1} \frac{(1-\nu^2/n^2)^{2q + n_> - n_<} (4\nu/n)^{2n_< - 2q}}{(n_< - l_0 - 1 - q)!\,(n_< + l_0 - q)!\,(n_> - n_< + q)!\,q!} \\ & \times \frac{\Gamma(n_< - q - \nu)\Gamma(2q + n_> - n_< + 1)}{\Gamma(q + n_> + 1 - \nu)}\, {}_2F_1(n_0 + n'_0, n_< - q - \nu; n_> + q + 1 - \nu; r_{n'}^2), \end{aligned}$$

$$n_{\gtrless} = \sup_{\inf}(n_0, n'_0).$$

(2)
$$A_{l l_0} = \frac{(n+l)(n+l-1)(l+m)(l-m)}{n(n-1)(2l+1)(2l-1)} \delta_{l-1\, l_0} + \frac{(n-l-1)(n-l-2)(l+1+m)(l+1-m)}{n(n-1)(2l+1)(2l+3)} \delta_{l+1\, l_0},$$

$$B_{l l_0} = \frac{(n-l)(n-l+1)(l+m)(l-m)}{n(n+1)(2l+1)(2l-1)} \delta_{l-1\, l_0} + \frac{(n+l+1)(n+l+2)(l+1+m)(l+1-m)}{n(n+1)(2l+1)(2l+3)} \delta_{l+1\, l_0},$$

$$\begin{aligned} C_{l l_0} = & \left(\frac{(n+l)(n+l-1)(n-l)(n-l+1)}{n^2(n-1)(n+1)} \right)^{1/2} \frac{(l+m)(l-m)}{(2l+1)(2l-1)} \delta_{l-1\, l_0} \\ & + \left(\frac{(n-l-1)(n-l-2)(n+l+1)(n+l+2)}{n^2(n-1)(n+1)} \right)^{1/2} \frac{(l+1+m)(l+1-m)}{(2l+1)(2l+3)} \delta_{l+1\, l_0}. \end{aligned}$$

For $n=1$ and 2, $l=m=0$, this expression is in agreement with the results of the previous works.[1-4]

CONCLUSION

All these results can be analytically continued to positive energies $E = -p_0^2/2m > 0$ or equivalently to purely imaginary values of ν. Moreover, the method can be easily generalized to the higher order processes, where the following integrals must appear at the end of the calculus,

$$\int_0^{+\infty} dx_1 \cdots \int_0^{+\infty} dx_n \exp\left(\sum_{i=1}^n \nu_i x_i \right) \mathcal{T}^1_{N_0 N'_0} \left(\left(\prod_{i=1}^n h_i e^{(x_i/2)v} \right) h_{n+1} \right),$$

$h_i \in \mathrm{Sp}(1,1)$. (3.12)

Then one has to express (3.12) in terms of known special functions. Finally, similar group theoretical methods can be used to calculate the matrix elements $I_{NN'}(E)$ between arbitrary continuum states. These different cases will be dealt with in another work.

ACKNOWLEDGMENTS

The author expresses his gratitude to Professor Y. Heno, Professor M. Gavrila, Dr. S. Klarsfeld, and Dr. A. Maquet for useful discussions.

APPENDIX A

Our aim is to calculate the integral

$$S^1_{NN'}(g) = \int_{SU(2)} d\mu(\xi) |c\xi + d|^{-2} \mathcal{Y}_{N'}(g^{-1} \cdot \xi) Y^*_N(\xi), \tag{A1}$$

for

$$g^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SU}^*(4), \quad |c| < |d|.$$

It is easier to evaluate the integral

$$S^1_{N,N'}(g) \equiv \int_{SU(2)} d\mu(\xi) |c\xi + d|^{-2} D^{j'}_{m'_1 m'_2}(g^{-1} \cdot \xi) D^{j*}_{m_1 m_2}(\xi), \tag{A2}$$

$$\mathcal{N} \equiv j m_1 m_2,$$

where the $D^j_{m_1 m_2}(x)$ are the homogeneous harmonic polynomials on $\mathbb{H}$ deduced from the usual matrix elements of the unitary irreducible representations of SU(2) by the homogeneity formula[28]

$$D^j_{m_1 m_2}(x) = |x|^{2j} D^j_{m_1 m_2}\left(\frac{x}{|x|}\right).$$

In another paper[23] we have established three fundamental properties verified by these polynomials: A finite difference equation, an addition formula, and an expansion formula.

Putting

$$\sigma^j_m = [(j-m)!\,(j+m)!\,]^{-1/2},$$
$$\sigma^j_{m_1 m_2} = \sigma^j_{m_1}\sigma^j_{m_2}, \tag{A3}$$

we have successively:

a finite difference equation:

$$(\sigma^j_{m_1 m_2}) D^j_{m_1 m_2}(x) = \binom{2j}{2j'} \sum_{m'_1, m'_2} (\sigma^{j-j'}_{m_1 - m'_1 m_2 - m'_2}) D^{j-j'}_{m_1 - m'_1 m_2 - m'_2}(x) \times (\sigma^{j'}_{m'_1 m'_2}) D^{j'}_{m'_1 m'_2}(x); \tag{A4}$$

an addition theorem,

$$x, x' \in \mathbb{H},$$

$$\sigma^j_{m_1 m_2} D^j_{m_1 m_2}(x + x') = \sum_{j', m'_1, m'_2} \sigma^{j-j'}_{m_1 - m'_1 m_2 - m'_2} D^{j-j'}_{m_1 - m'_1 m_2 - m'_2}(x) \sigma^{j'}_{m'_1 m'_2} D^{j'}_{m'_1 m'_2}(x'); \tag{A5}$$

an expansion theorem:

$$x, x' \in \mathbb{H}, \quad |x| < |x'|:$$

$$(\sigma^j_{m_1 m_2})^{-1} |x + x'|^{-2} D^j_{m_1 m_2}((x + x')^{-1}) = \sum_{j', m'_1, m'_2} (-1)^{2j'} \sigma^{j'}_{m'_1 m'_2} D^{j'}_{m'_1 m'_2}(x) (\sigma^{j+j'}_{m_1 + m'_2 m_2 + m'_1})^{-1} \times |x'|^{-2} D^{j+j'}_{m_1 + m'_2 m_2 + m'_1}(x'^{-1}). \tag{A6}$$

Let us add the fundamental group representation property, $x, x' \in \mathbb{H}$:

$$D^j_{m_1 m_2}(xx') = \sum_{m'} D^j_{m_1 m'}(x) D^j_{m' m_2}(x'). \tag{A7}$$

Combining these four formulas, we easily obtain the "four Euclidean conformal transformation" formula, $|cx| < |d|$,

$$|cx + d|^{-2} D^{j'}_{m'_1 m'_2}((ax + b)(cx + d)^{-1}) = \sum_{\mu} D^{j'}_{m'_1 \mu}((ax + b)) |cx + d|^{-1} D^{j'}_{\mu m'_2}((cx + d)^{-1}) = \sum_{j, m_1, m_2} F(j m_1 m_2; j' m'_1 m'_2; g) \frac{\sigma^{j'}_{m_2} \sigma^{j}_{m_1}}{\sigma^{j'}_{m'_1} \sigma^{j}_{m_1}} D^j_{m_1 m_2}(x),$$

where

$$F(j m_1 m_2; j' m'_1 m'_2; g) = \sum_{\substack{j_i, m_{i1}, m_{i2} \\ 1 \le i \le 4}} \delta_{j', j_1 + j_2} \delta_{j_4, j' + j_3} \delta_{j, j_1 + j_3} \delta_{m'_1, m_2 + m_{21}} \times \delta_{m_{42}, m'_2 + m_{31}} \delta_{m_{41}, m_2 + m_{22}} \delta_{m_1, m_{12} + m_{32}} (-1)^{2j_3} (\sigma^{j_4}_{m_{41} m_{42}})^{-2} |d^2|^{-(2j_4 + 1)} \prod_{i=1}^{4} \sigma^{j_i}_{m_{i1} m_{i2}} D^{j_i}_{m_{i1} m_{i2}}(g_i) \tag{A8}$$

with $g_1 = a$, $g_2 = b$, $g_3 = c$, $g_4 = \bar{d}$.

By use of the unitary transformation for connecting a given $j = (n-1)/2$, the set of the $\mathcal{Y}_{nlm}$, and that of the $D^j_{m_1 m_2}$,

$$\mathcal{Y}_{nlm}(x) = \left(\frac{n}{2\pi^2}\right)^{1/2} i^l \sum_{m_1, m_2} (2l+1)^{1/2} (-1)^{j - m_2} \times \begin{pmatrix} j & j & l \\ m_1 & -m_2 & m \end{pmatrix} D^j_{m_1 m_2}(x), \tag{A9}$$

we finally obtain ($n = 2j+1$, $n' = 2j'+1$):

$$S^1_{NN'}(g) = i^{l'-l} [(2l+1)(2l'+1)]^{1/2} \left(\frac{n'}{n}\right)^{1/2} \times \sum_{\substack{m_1, m_2 \\ m'_1, m'_2}} (-1)^{j - m_2 + j' - m'_2 - l} \begin{pmatrix} j & j & l \\ m_1 & -m_2 & m \end{pmatrix} \begin{pmatrix} j' & j' & l' \\ m'_1 & -m'_2 & m' \end{pmatrix} \times \frac{\sigma^j_{m_2} \sigma^{j'}_{m'_2}}{\sigma^j_{m_1} \sigma^{j'}_{m'_1}} F(j m_1 m_2; j' m'_1 m'_2; g). \tag{A10}$$

For $g \in \mathrm{SL}(2, \mathbb{R}) \subset \mathrm{SU}^*(4)$, this expression is reduced to a hypergeometric polymial,

$$S^1_{NN'}(g) = \delta_{ll'} \delta_{mm'} \left(\frac{n'}{n}\right)^{1/2} \left(\frac{(n_> - l - 1)!\,(n_> + l)!}{(n_< - l - 1)!\,(n_< + l)!}\right)^{1/2} \times d^{-(n_> + l + 1)} a^{n_< - l - 1} \frac{(\gamma(b,c))^{n_> - n_<}}{(n_> - n_<)!} \times {}_2F_1\left(l + 1 - n_<, n_> + l + 1; n_> - n_< + 1; \frac{bc}{ad}\right) \tag{A11}$$

$$n_{\gtrless} = \sup_{\inf}(n, n'), \quad \gamma(b,c) = \begin{cases} b, & n_> = n', \\ -c & n_> = n. \end{cases}$$

In the latter case, another expression is deduced from the following formula,[27]

$${}_2F_1(\alpha, \beta; \gamma; z) = (1-z)^{-\alpha} {}_2F_1\left(\alpha, \gamma - \beta; \gamma; \frac{z}{z-1}\right)$$

$$S^1_{NN'}(g) = \delta_{ll'} \delta_{mm'} \left(\frac{n'(n_> - l - 1)!\,(n_> + l)!}{n(n_< - l - 1)!\,(n_< + l)!}\right)^{1/2} \times d^{-(n + n')} \frac{(\gamma(b,c))^{n_> - n_<}}{(n_> - n_<)!} \times {}_2F_1(l + 1 - n_<, -(n_< + l), n_> - n_< + 1; -bc). \tag{A12}$$

APPENDIX B

We state here the formula (2.13)

$$T^1_{NN'}(h) = \frac{n'}{n} T^1_{N'N}(h^{-1}), \quad \text{for all } h \in \mathrm{Sp}(1,1). \tag{2.13}$$

For all $h \in \mathrm{Sp}(1,1)$, we have the following factorization[23]

$$h = k \alpha(t) k', \tag{B1}$$

where $k, k' \in \mathrm{Spin}(4)$, maximal compact subgroup of $\mathrm{Sp}(1,1)$, $\alpha(t) \in A$, one parameter subgroup of the matrices:

$$\alpha(t) = \begin{pmatrix} \cosh t/2 & \sinh t/2 \\ \sinh t/2 & \cosh t/2 \end{pmatrix}, \quad \alpha^{-1}(t) = \alpha(-t).$$

Thus

$$T^1_{nlm, n'l'm'}(k \alpha(t) k') = \sum_{\lambda \mu} T^1_{nlm, n\lambda\mu}(k) T^1_{n\lambda\mu, n'\lambda\mu}(\alpha(t)) T^1_{n'\lambda\mu, n'l'm'}(k'). \tag{B2}$$

$\mathcal{T}^1$ is unitary when restricted to Spin(4). Thus

$$\mathcal{T}^1_{n\ell m,\, n\lambda\mu}(k) = \mathcal{T}^{1*}_{n\lambda\mu,\, n\ell m}(k^{-1}),$$
$$\mathcal{T}^1_{n'\lambda\mu,\, n'\ell'm'}(k') = \mathcal{T}^{1*}_{n'\ell'm',\, n'\lambda\mu}(k'^{-1}). \tag{B3}$$

On the other hand, we can state from (A12) that $\mathcal{T}^1$ satisfies

$$\mathcal{T}^1_{n\lambda\mu,\, n'\lambda\mu}(\alpha(t)) = \frac{n'}{n}\mathcal{T}^1_{n'\lambda\mu,\, n\lambda\mu}(\alpha(-t)), \quad \text{for all } \alpha(t)\in A. \tag{B4}$$

Let us also remark that this matrix element is real. Inserting (B3) and (B4) in (B2), we finally obtain:

$$\begin{aligned}
&\mathcal{T}^1_{n\ell m,\, n'\ell'm'}(k\alpha(t)k') \\
&= \frac{n'}{n}\sum_{\lambda,\mu}\mathcal{T}^{1*}_{n'\ell'm',\, n'\lambda\mu}(k'^{-1})\mathcal{T}^1_{n'\lambda\mu,\, n\lambda\mu}(\alpha^{-1}(t))\mathcal{T}^{1*}_{n\lambda\mu,\, n\ell m}(k^{-1}) \\
&= \frac{n'}{n}\mathcal{T}^{1*}_{n'\ell'm',\, n\ell m}(k'^{-1}\alpha^{-1}(t)k^{-1}) \\
&= \frac{n'}{n}\mathcal{T}^{1*}_{n'\ell'm',\, n\ell m}(h^{-1}).
\end{aligned} \tag{B5}$$

APPENDIX C

In this appendix, we shall examine the details of the calculation of the matrix elements $I_{NN'}(E)$.

In a first step, we suppose E negative and we re-express the calculus on $L^2_C(\mathrm{SU}(2))$ as in Refs. 7 and 26 by means of:

(i) the inverse stereographic projection (1.6) or (2.6),

$$\xi_n = s^{-1}(p_n)\cdot(p_n,\mathbf{p}), \quad \xi'_{n'} = s^{-1}(p_{n'})\cdot(p_{n'},\mathbf{p}'),$$
$$\xi = s^{-1}(p_0)\cdot(p_0,\mathbf{p}), \tag{C1}$$

where

$$p_n = (-2mE_n)^{1/2}, \quad p_{n'} = (-2mE_{n'})^{1/2}, \quad p_0 = (-2mE)^{1/2},$$

and consequently

$$\mathbf{p}\cdot\epsilon_1 = \left(\frac{p_n^2+p^2}{2p_n}\right)\xi_n\circ\epsilon_1,$$
$$\mathbf{p}'\cdot\epsilon_2 = \left(\frac{p_{n'}^2+p'^2}{2p_{n'}}\right)\xi'_{n'}\circ\epsilon_2; \tag{C2}$$

(ii) the Fock transformations (1.13):

$$Y_N\in H(p_n) \to \mathcal{J}_n^{-1}Y_N = \psi_N,$$
$$Y_{N'}\in H(p_{n'}) \to \mathcal{J}_{n'}^{-1}Y_{N'} = \psi_{N'}; \tag{C3}$$

(iii) the Schwinger–Sturmian expansion[26] of the Coulomb Green function, taking into account the retardation effect:

$$\begin{aligned}
&(\exp(i\mathbf{k}_1\cdot\mathbf{r})G(E)\exp(i\mathbf{k}\cdot\mathbf{r}))(\mathbf{p},\mathbf{p}') \\
&= G(E)(\mathbf{p}-\mathbf{k}_1,\mathbf{p}'+\mathbf{k}_2) \\
&= -\frac{m}{p_0}\left(\frac{p_0^2+|\mathbf{p}-\mathbf{k}_1|^2}{2p_0}\right)^{-1}\left(\frac{p_0^2+|\mathbf{p}'+\mathbf{k}_2|^2}{2p_0}\right)^{-2} \\
&\quad\times\sum_{N_1}\left(1-\frac{\nu}{n_1}\right)^{-1}Y_{N_1}(h_1^{-1}\cdot\xi_n)Y^*_{N_1}(h_2\cdot\xi_{n'}),
\end{aligned} \tag{C4}$$

where $\nu = m\alpha z/p_0$, and h_1 and h_2 are elements of Sp(1,1) induced by the general boosts in $\mathbf{H}$,

$$\bar{K}_1 = (p_0 - p_n, -\mathbf{k}_1), \quad K_2 = (p_0 - p_{n'}, \mathbf{k}_2),$$

and the inverse stereographic projections

$$h_1^{-1} = s^{-1}(p_0)t_{\bar{K}_1}s(p_n),$$
$$h_2 = s^{-1}(p_0)t_{K_2}s(p_{n'}).$$

By introducing

$$K_{1\pm} = (p_0 \pm p_n, \mathbf{k}_1),$$
$$K_{2\pm} = (p_0 \pm p_{n'}, \mathbf{k}_2), \tag{C5}$$

we obtain

$$h_1 = \frac{1}{2\sqrt{p_0p_n}}\begin{pmatrix} K_{1+} & -\bar{K}_{1-} \\ -K_{1-} & \bar{K}_{1+}\end{pmatrix},$$
$$h_2 = \frac{1}{2\sqrt{p_0p_{n'}}}\begin{pmatrix} K_{2+} & K_{2-} \\ \bar{K}_{2-} & \bar{K}_{2+}\end{pmatrix}; \tag{C6}$$

(iv) the relations (1.9) between the scalar products of the two Hilbert spaces $L^2_C(\mathbb{R}^3)$ and $L^2_C(\mathrm{SU}(2))$.

Finally, we have to compute

$$\begin{aligned}
I^2_{NN'}(E) = &-\frac{m}{p_0}(p_np_{n'})^{1/2}\sum_{N_1}\left(1-\frac{\nu}{n_1}\right)^{-1} \\
&\times(\xi'_{n'}\circ\epsilon_2Y_{N'},\, \mathcal{T}^2(h_2^{-1})Y_{N_1})_{H(p_{n'})} \\
&\times(\mathcal{T}^2(h_1)Y_{N_1},\, \xi_n\circ\epsilon_1Y_N)_{H(p_n)},
\end{aligned}$$

where $\mathcal{T}^2$ is defined by (2.9).

Reducing the product $\xi\cdot\epsilon Y_N(\xi)$ is easy. One may define an operator $C(\epsilon)$ which is the image of an element of the SU*(4) enveloping algebra under its representation in $\mathcal{L}(L^2_C(\mathrm{SU}(2)))$.

The matrix elements $C_{N'N}(\epsilon)$ are given explicitly in Appendix D:

$$\begin{aligned}
\xi\cdot\epsilon Y_N(\xi) &\equiv C(\epsilon)Y_N(\xi) \\
&= \sum_{N'}C_{N'N}(\epsilon)Y_{N'}(\xi).
\end{aligned} \tag{C8}$$

Thus, in terms of the matrix elements of the representation $\mathcal{T}^2$ of Sp(1,1) and of the operators $C(\epsilon_1)$ and $C(\epsilon_2)$, Eq. (C7) is written:

$$\begin{aligned}
I^2_{NN'}(E) = &-\frac{m}{p_0}(p_np_{n'})^{1/2}\sum_{N_0,N'_0}C_{N'_0N'}(\epsilon_2)C^*_{NN_0}(\epsilon_1) \\
&\times\left[\sum_{N_1}\left(1-\frac{\nu}{n_1}\right)^{-1}\mathcal{T}^{2*}_{N'_0N_1}(h_2^{-1})\mathcal{T}^2_{N_0N_1}(h_1)\right].
\end{aligned} \tag{C9}$$

By use of (2.12) and (2.14), the expansion between the brackets is equal to

$$\sum_{N_1}(n_1-\nu)^{-1}\mathcal{T}^1_{N_0N_1}(h_1)\mathcal{T}^1_{N_1N'_0}(h_2). \tag{C10}$$

Now, as[9,26]

$$(n_1-\nu)^{-1} = \int_0^{+\infty}dx\,\exp[(\nu-n_1)x], \tag{C11}$$

let us put

$$\upsilon \equiv \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}, \quad \exp[(x/2)\upsilon] = \begin{pmatrix}\exp(x/2) & 0\\ 0 & \exp(-x/2)\end{pmatrix}, \tag{C12}$$

and realize the following homogeneity property, deduced from the integral (2.17)

$$\exp(-n_1x)\mathcal{T}^1_{N_0N_1}(h_1) = S^1_{N_0N_1}(h_1\exp[(x/2)\upsilon]). \tag{C13}$$

Since $|(\exp[-(x/2)\upsilon]h)\cdot\xi| < 1$ for all $x > 0$, $\xi\in\mathrm{SU}(2)$, $h\in\mathrm{Sp}(1,1)$, $S^1_{N_0N_1}(h_1\exp[(x/2)\upsilon])$ is merely the matrix element of the local representation (2.16) of SU*(4), and we can apply the fundamental property of the group representation to sum the expansion (C10) and to obtain Eq. (3.4),

$$\sum_{N_1} T^1_{N_0N_1}(h_1 \exp[(x/2)\upsilon])T^1_{N_1N_0}(h_2) = T^1_{N_0N_0'}(h_1 \exp[(x/2)\upsilon]h_2). \tag{C14}$$

APPENDIX D

We give in this appendix the expression of the matrix elements $C_{N'N}(\epsilon)$ defined by (C8),

$$\xi\cdot\epsilon Y_N(\xi) \equiv C(\epsilon)Y_N(\xi) = \sum_{N'} C_{N'N}(\epsilon)Y_{N'}(\xi). \tag{C8}$$

For all $N=(n, l, m)$, let us define the numbers:

$$a(n,l) = \frac{1}{2}\left(\frac{(n+l)(n+l-1)}{n(n-1)}\right)^{1/2}, \quad a(1,0)=0,$$

$$b(l,m) = \left(\frac{(l+m)(l-m)}{(2l+1)(2l-1)}\right)^{1/2}, \quad b(0,0)=0, \tag{D1}$$

$$c(l,m) = \left(\frac{(l+m)(l+m-1)}{(2l+1)(2l-1)}\right)^{1/2}, \quad c(0,0)=0.$$

Then

$$C_{N'N}(\epsilon) = \epsilon_z\delta_{mm'}T^m_{nl,n'l'} + \tfrac{1}{2}(\epsilon_x + i\epsilon_y)\delta_{m-1,m'}U^m_{nl,n'l'} - \tfrac{1}{2}(\epsilon_x - i\epsilon_y)\delta_{m+1,m'}U^{-m}_{nl,n'l'}, \tag{D2}$$

where

$$T^m_{nl,n'l'} = \delta_{n-1,n'}(\delta_{l-1,l'}a(n,l)b(l,m) - \delta_{l+1,l'}a(n,-l-1) \times b(l+1,m)) - \delta_{n+1,n'}(\delta_{l-1,l'}a(n+1,-l)b(l,m) - \delta_{l+1,l'}a(n+1,l+1)b(l+1,m)), \tag{D3}$$

$$U^m_{nl,n'l'} = \delta_{n-1,n'}(\delta_{l-1,l'}a(n,l)c(l,m) + \delta_{l+1,l'}a(n,-l-1)c(l+1,-m)) - \delta_{n+1,n'}(\delta_{l-1,l'}a(n+1,-l)c(l,m) + \delta_{l+1,l'}a(n+1,l+1)c(l+1,-m)). \tag{D4}$$

APPENDIX E

In this appendix, we make explicit the coefficient $C^{qq'}_{N_0N_0'}(h_1,h_2)$ appearing in Eq. (3.8). If we consider the expression

$$T^1_{N_0N_0'}(g(x)) \equiv S^1_{N_0N_0'}(g(x))$$

which is given by Eq. (A10), we note that it is a question of expanding, in powers of e^{-x} and $|K_{2+}K_{1+} - e^{-x}\bar{K}_{2-}\bar{K}_{1-}|$, the function

$$F\left(\frac{n_0-1}{2}m_1m_2;\frac{n_0'-1}{2}m_1'm_2';g(x)\right).$$

Let us note that $g^{-1}(x)$, given by Eq. (3.5), is the sum of two matrices,

$$g^{-1}(x) = \frac{e^{x/2}}{4p_0\sqrt{p_np_{n'}}}\left[\begin{pmatrix} -K_{2-}K_{1-} & -K_{2-}K_{1+} \\ K_{2+}K_{1-} & K_{2+}K_{1+}\end{pmatrix} + e^{-x}\begin{pmatrix} \bar{K}_{2+}\bar{K}_{1+} & \bar{K}_{2+}\bar{K}_{1-} \\ -\bar{K}_{2-}\bar{K}_{1+} & -\bar{K}_{2-}\bar{K}_{1-}\end{pmatrix}\right]. \tag{E1}$$

Thus, by use of the homogeneity properties of the polynomials $D^j_{m_1m_2}$ and of the addition theorem (A5), we obtain:

$$j_0 = \frac{n_0-1}{2}, \quad j_0' = \frac{n_0'-1}{2},$$

$$C^{qq'}_{N_0N_0'}(h_1,h_2) = 16p_0^2p_np_{n'}[n_0n_0'(2l_0+1)(2l_0'+1)]^{1/2}i^{-(l_0+l_0')} \times \sum_{\substack{j_i,j_i',m_{i\frac{1}{2}},m_{i\frac{1}{2}}' \\ m_1,m_2,m_1',m_2'}} \delta_{2j_4+1,q'}\delta_{2\Sigma_ij_i'+1,q}\delta_{j_0',j_1'+j_2} \times \delta_{j_4,j_0'+j_3}\delta_{j_0,j_1'+j_3}\delta_{m_1',m_{11}+m_{21}}\delta_{m_{42},m_2'+m_{31}} \times \delta_{m_{41},m_2+m_{22}}\delta_{m_1,m_{12}+m_{32}}(-1)^{j_0+j_0'+2j_3-m_2-m_2'} \times \begin{pmatrix} j_0 & j_0 & l_0 \\ m_1 & -m_2 & m_0\end{pmatrix}\begin{pmatrix} j_0' & j_0' & l_0' \\ m_1' & -m_2' & m_0'\end{pmatrix} \times (\sigma^{j_4}_{m_{41}m_{42}})^{-2}\frac{\sigma^{j_0}_{m_2}\sigma^{j_0'}_{m_2'}}{\sigma^{j_0}_{m_1}\sigma^{j_0'}_{m_1'}}\prod_{i=1}^{4}\sigma^{j_i-j_i'}_{m_{i1}-m_{i1}m_{i2}-m_{i2}} \times \sigma^{j_i'}_{m_{i1}'m_{i2}'}D^{j_i-j_i'}_{m_{i1}-m_{i1}'m_{i2}-m_{i2}'}(A_i)D^{j_i'}_{m_{i1}'m_{i2}'}(B_i),$$

$$A_1 = -K_{2-}K_{1-}, \quad A_2 = -K_{2-}K_{1+},$$
$$A_3 = K_{2+}K_{1-}, \quad A_4 = \bar{K}_{1+}\bar{K}_{2+},$$
$$B_1 = \bar{K}_{2+}\bar{K}_{1+}, \quad B_2 = \bar{K}_{2+}\bar{K}_{1-},$$
$$B_3 = -\bar{K}_{2-}\bar{K}_{1+}, \quad B_4 = -K_{1-}K_{2-}. \tag{E2}$$

[1] A. Maquet, Phys. Rev. A **15**, 1088 (1977).
[2] Ya.I. Granovskii, Zh. Eksp. Teor. Fiz. **56**, 605 (1969) [Sov. Phys. JETP **29**, 333 (1969)].
[3] S. Klarsfeld, Lett. Nuovo Cimento, Ser. I, **1**, 682 (1969); **2**, 548 (1969); **3**, 395 (1970).
[4] M. Gavrila, Phys. Rev. **163**, 147 (1967).
[5] M. Gavrila and A. Costescu, Phys. Rev. A **2**, 1752 (1970); **4**, 1688 (1971).
[6] M. Lieber, Phys. Rev. **174**, 2037 (1968).
[7] J.P. Gazeau, C.R. Acad. Sci. Paris B **272**, 501 (1971); J.P. Gazeau and A. Maquet, C.R. Acad. Sci. Paris B **272**, 1097 (1971).
[8] C. Fronsdal, Phys. Rev. **179**, 1513 (1969).
[9] R.W. Huff, Phys. Rev. **186**, 1367 (1969).
[10] G. Bisiacchi and G. Calucci, Phys. Rev. **181**, 185 (1969).
[11] A. Dalgarno and J.T. Lewis, Proc. Phys. Soc. A **233**, 70 (1969).
[12] Y. Gontier and M. Trahin, Phys. Rev. **172**, 83 (1968); Phys. Rev. A **4**, 1896 (1971).
[13] S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962); G. Grensing, J. Math. Phys. **16**, 312 (1975).
[14] W. Fock, Z. Phys. **98**, 145 (1935).
[15] A.O. Barut and H. Kleinert, Phys. Rev. **156**, 1541 (1967); C. Fronsdal, Phys. Rev. **156**, 1665 (1967); A.O. Barut and H. Kleinert, Phys. Rev. **157**, 1180 (1967); A.O. Barut in *Lie Algebras, Applications and Computational Methods*, edited by B. Kolman (Soc. Ind. Appl. Math. Philadelphia, 1973), p. 79.
[16] M. Rotenberg, Ann. Phys. (N.Y.) **19**, 262 (1962).
[17] M. Rotenberg, *Advances in Atomic and Molecular Physics* edited by R.D. Bates and I. Estermann (Academic, New York, 1970), Vol. 6.
[18] A. Joseph, Int. J. Quant. Chem. **1**, 535 (1967).
[19] A. Decoster, thesis, Rapport C.E.A.-R-4112 (1970); Nuovo Cimento A **68**, 105 (1970).
[20] M. Bander and C. Itzyckson, Rev. Mod. Phys. **38**, 330 (1966); R.J. Finkelstein, J. Math. Phys. **8**, 443 (1967); R. Finkelstein and D. Levy, J. Math. Phys. **8**, 2147 (1967).
[21] N. Bourbaki, *Topologie générale* (Hermann, Paris, 1974), Chap. 8.
[22] C. Fronsdal, "Lectures on Bound Electron Problems," International Centre for Theoretical Physics, Miramare Trieste (1970).
[23] R. Takahashi, "Sur les représentations unitaires des groupes de Lorentz généralisés," Bull. Soc. Math. France **91**, 289 (1963).
[24] S. Ström, Ark. Fys. **33**, 465 (1967); Ann. Inst. H. Poincaré **13**, 77 (1970).
[25] L. Hostler, J. Math. Phys. **5**, 1235 (1964).
[26] J. Schwinger, J. Math. Phys. **5**, 1606 (1964).
[27] P. Appell and J. Kampe de Feriet, *Fonctions Hypergéométriques and Hypersphériques. Polynômes d'Hermite* (Gauthiers-Villars, Paris, 1926).
[28] J.D. Talman, *Special functions. A Group Theoretic Approach* (Based on lectures by Eugène P. Wigner) (Benjamin, New York, 1968).
[29] J.P. Gazeau, "Fifth International Colloquium in Group Theoretical Methods in Physics," Montreal, July 1976.

Invariance and conservation laws for Lagrangian systems with one degree of freedom

Willy Sarlet

Instituut voor Theoretische Mechanica, Rijksuniversiteit Gent, Krijgslaan 271 S9, B-9000 Gent, Belgium
(Received 4 March 1977)

For a general class of linear transformations of both the dependent and the independent variable in Lagrangian systems with one degree of freedom, we investigate the relationship between invariances for the Lagrangian, the equation of motion, and a constant of the motion. Unlike Denman's work on this subject, we try to show that for every invariance in the equation of motion a Lagrangian can be found with the related invariance. Furthermore, a constant of the motion is considered as being implied by the invariance of the equation of motion if it has the same invariance as the Lagrangian.

1. INTRODUCTION

In a series of papers[1–4] Denman investigated the problem of invariance and conservation laws in classical mechanics from a viewpoint differing from the usual one in that both concepts are not necessarily related to some property of the Lagrangian of the system. He shows by various, mostly one-dimensional examples how the equation of motion can have some invariance property and induce some constant of the motion without a corresponding symmetry in the Lagrangian. If at first sight this is surprising, it becomes evident as soon as one realizes that the Lagrangian governing a system of differential equations is far from unique. This fact was systematically studied by Currie and Saletan,[5] who introduced the notion of q-equivalent Lagrangians. Perhaps it is worth mentioning that these authors used the term "fouling" when talking about a transformation from the commonly used Lagrangian of the problem to an equivalent Lagrangian. So one could ask if one is not "fouling" the symmetry in considering Lagrangians which do not have the invariance property of the equations of motion.

To be more precise let us take the following example: the equation of motion of a linearly damped oscillator,

$$\ddot{q} + 2\gamma\dot{q} + \omega_0^2 q = 0, \tag{1}$$

is autonomous and therefore invariant for time translations. Denman[3] points out that (1) can be derived from the Lagrangian (introduced by Bateman[6]),

$$L = \tfrac{1}{2} e^{2\gamma t}(\dot{q}^2 - \omega_0^2 q^2), \tag{2}$$

which is obviously not time-translation invariant.

The equation of motion (1) induces a constant of the motion which obviously cannot be considered as corresponding to an invariance of the Lagrangian (2). However, there exists a time-independent Lagrangian governing (1), namely,[7]

$$L = \frac{\dot{q}+\gamma q}{\omega q}\arctan\left(\frac{\dot{q}+\gamma q}{\omega q}\right) - \tfrac{1}{2}\ln[\omega^2 q^2 + (\dot{q}+\gamma q)^2], \tag{3}$$

with $$\omega^2 = \omega_0^2 - \gamma^2. \tag{4}$$

Although this Lagrangian is less elegant than (2) it allows retaining the invariance throughout the whole description. It is this feature that we want to emphasize here. We first consider a general linear transformation of both the dependent variable q and the independent variable t in the form

$$Q = aq + bt + \lambda, \qquad T = cq + dt + \mu, \tag{5}$$

and investigate under what conditions an invariance of the Lagrangian for this transformation induces a corresponding invariance for the equation of motion. In this way we systematically come to ten types of basic transformations including all types of transformations discussed by Denman, together with some cases of periodicity in one of the variables. Then for each of these basic transformations we try to treat the inverse problem, and the essential difference between Denman's approach and ours can be explained as follows. Starting from an equation of motion which exhibits some transformation invariance Denman often mentions the partial differential equation which must be satisfied by the Lagrangian in order to produce such an invariant equation of motion. Then he looks for a particular solution of this equation with no invariance properties at all, while we just intend to show that this equation always has a particular solution with corresponding symmetry (here the words symmetry and invariance are used as synonyms). A second difference of conceptual nature is related to the notion of "constant of the motion implied by an invariance of the equation of motion". For Denman a constant of the motion is implied by the invariance of the equation of motion if this invariance suggests a substitution which reduces the second-order differential equation to a first-order one which then yields the first integral by a simple quadrature. Respecting the philosophy of keeping the symmetry throughout every aspect of the description of the problem we would like to call a constant of the motion being "implied by the invariance of the equation of motion" if it exhibits a similar invariance of the Lagrangian. With this concept we can associate a constant of the motion to invariances which do not imply a first integral in Denman's approach (like the case of time or coordinate inversion). For the damped oscillator (1), for instance, we simply require the constant of the motion to be time-independent as well. This gives

$$\phi = \tfrac{1}{2}\ln[\omega^2 q^2 + (\dot{q}+\gamma q)^2] - \frac{\gamma}{\omega}\arctan\frac{\dot{q}+\gamma q}{\omega q} = C, \tag{6}$$

which here coincides with Denman's "implied constant

 0022-2488/78/1905-1049$1.00

of the motion",[3] and is equivalent to the constant Hamiltonian also mentioned by Havas.[7]

2. GENERAL CONSIDERATIONS

Suppose that a given Lagrangian $L(q,\dot{q},t)$ is invariant for the transformation (5). By this we mean that

$$L(q,\dot{q},t)\equiv L(Q,Q',T), \tag{7}$$

where

$$Q'\equiv\frac{dQ}{dT}=\frac{a\dot{q}+b}{c\dot{q}+d}, \tag{8}$$

and

$$\Delta\equiv ad-bc\neq 0. \tag{9}$$

For the general considerations of this section the parameters a,b,c,d,λ, and μ may be prescribed constant values or may run through some arbitrary parameter family. Of course, if too many parameters may take an arbitrary number of values, only a trivial constant or even zero will satisfy an invariance condition like (7), but this is not important because we will select afterwards from (5) only the interesting cases.

If we denote the equation of motion resulting from L by

$$\ddot{q}+f(q,\dot{q},t)=0, \tag{10}$$

then f is determined by

$$\frac{\partial^2 L}{\partial q\partial\dot{q}}\dot{q}+\frac{\partial^2 L}{\partial\dot{q}\partial t}-\frac{\partial L}{\partial q}=\frac{\partial^2 L}{\partial\dot{q}^2}f(q,\dot{q},t). \tag{11}$$

A straightforward calculation shows that (10) will be invariant for the same transformation (5) if and only if

$$f(q,\dot{q},t)\equiv\Delta^{-1}(c\dot{q}+d)^3 f(Q,Q',T), \tag{12}$$

which by (11) yields

$$c\left[\frac{\partial^2 L}{\partial\dot{q}^2}\left(\frac{\partial L}{\partial t}+\dot{q}\frac{\partial L}{\partial q}\right)-2\frac{\partial L}{\partial\dot{q}}\left(\frac{\partial^2 L}{\partial q\partial\dot{q}}\dot{q}+\frac{\partial^2 L}{\partial\dot{q}\partial t}-\frac{\partial L}{\partial q}\right)\right]=0. \tag{13}$$

Apart from the unsignificant case of a Lagrangian canceling the term between square brackets, we see that the equation of motion inherits the symmetry of the Lagrangian for an arbitrary transformation of the form

$$Q=aq+bt+\lambda,\quad T=dt+\mu, \tag{14}$$

with $ad\neq 0$.[8]

The inherited invariance for the function f is given by

$$f(q,\dot{q},t)=(d^2/a)f(Q,Q',T). \tag{15}$$

As is known, in the case of one degree of freedom, every second-order differential equation linear in $\ddot{q}$ can be derived from Hamilton's variational principle. This is simply the interpretation of the fact that under suitable conditions (upon which we do not want to insist here), Eq. (11) with given f always yields solutions for L. The point that interests us now is whether there can be found a solution for L satisfying (7), when f has the invariance property (15). At the same time we look for a first integral of the system, i.e., a solution of the equation

$$\frac{\partial\phi}{\partial q}\dot{q}-f\frac{\partial\phi}{\partial\dot{q}}+\frac{\partial\phi}{\partial t}=0, \tag{16}$$

having the same symmetry:

$$\phi(q,\dot{q},t)\equiv\phi(Q,Q',T). \tag{17}$$

Now it is easy to prove that if $\phi(q,\dot{q},t)$ is a solution of (16), where f satisfies (15), then

$$\Phi(q,\dot{q},t)=\phi(Q,Q',T) \tag{18}$$

is another solution. Indeed,

$$\begin{aligned}\frac{\partial\Phi}{\partial q}\dot{q}-f\frac{\partial\Phi}{\partial\dot{q}}+\frac{\partial\Phi}{\partial t}&=a\frac{\partial\phi}{\partial Q}\dot{q}-\frac{a}{d}f\frac{\partial\phi}{\partial Q'}+b\frac{\partial\phi}{\partial Q}+d\frac{\partial\phi}{\partial T}\\&=d\left[\left(\frac{a\dot{q}+b}{d}\right)\frac{\partial\phi}{\partial Q}-\frac{a}{d^2}f\frac{\partial\phi}{\partial Q'}+\frac{\partial\phi}{\partial T}\right]\\&=d\left[\frac{\partial\phi}{\partial Q}Q'-f(Q,Q',T)\frac{\partial\phi}{\partial Q'}+\frac{\partial\phi}{\partial T}\right]\\&=0,\end{aligned}$$

in view of the assumption.

This is a property similar to the "related integral theorem" in the work of Katzin and Levine.[9] An analogous property can be shown to hold for the Lagrangian: if $L(q,\dot{q},t)$ is a solution of (11) with f satisfying (15), then

$$\Lambda(q,\dot{q},t)=L(Q,Q',T) \tag{19}$$

is another solution.

However, we want something more, namely a solution for ϕ and L, such that $\phi(q,\dot{q},t)\equiv\Phi(q,\dot{q},t)$ and $L(q,\dot{q},t)\equiv\Lambda(q,\dot{q},t)$.

Before going into the details of this question, we want to make the following remark. Suppose for a moment that we have solved the problem and that ϕ is a symmetric solution of (16). What are the consequences of the eventual existence of a second, independent solution ψ with the same property?

Let us put

$$\phi(q,\dot{q},t)=\alpha, \tag{20}$$

$$\psi(q,\dot{q},t)=\beta, \tag{21}$$

α and β being arbitrary constants.

These two first integrals completely determine the solutions of the equation of motion. So suppose that (20) can be solved for q, giving

$$q=q_1(\alpha,\dot{q},t)$$

and that

$$\chi(\alpha,\dot{q},t)\equiv\psi(q_1(\alpha,\dot{q},t),q,t)=\beta$$

can be solved for $\dot{q}$, giving

$$\dot{q}=\dot{q}_1(\alpha,\beta,t),$$

so that the solution finally has the form

$$q(\alpha,\beta,t)\equiv q_1(\alpha,\dot{q}_1(\alpha,\beta,t),t).$$

Then using the symmetry of ϕ and ψ, we first get

$$\alpha\equiv\phi(q_1(\alpha,\dot{q},t),\dot{q},t)\equiv\phi(aq_1(\alpha,\dot{q},t)+bt+\lambda,Q',T),$$

from which it follows that

$$q_1(\alpha,Q',T)\equiv aq_1(\alpha,\dot{q},t)+bt+\lambda. \tag{22}$$

Using this result, we obtain

$$\chi(\alpha, Q', T) \equiv \psi\,(q_1(\alpha, Q', T), Q', T)$$
$$\equiv \psi(aq_1(\alpha, \dot{q}, t) + bt + \lambda, Q', T)$$
$$\equiv \psi(q_1(\alpha, \dot{q}, t), \dot{q}, t)$$
$$\equiv \chi(\alpha, \dot{q}, t),$$

so that second

$$\beta \equiv \chi(\alpha, \mathring{q}_1(\alpha, \beta, t), t) \equiv \chi(\alpha, d^{-1}(a\dot{q}_1(\alpha, \beta, t) + b), T),$$

from which it follows that

$$\dot{q}_1(\alpha, \beta, T) \equiv d^{-1}(a\dot{q}_1(\alpha, \beta, t) + b). \tag{23}$$

Finally we get for the solution

$$q(\alpha, \beta, T) \equiv q_1(\alpha, \mathring{q}_1(\alpha, \beta, T), T)$$
$$\equiv q_1(\alpha, d^{-1}(a\dot{q}_1(\alpha, \beta, t) + b), T) \text{ [from (23)]}$$
$$\equiv aq_1(\alpha, \dot{q}_1(\alpha, \beta, t), t) + bt + \lambda, \text{ [from (22)]}$$

or

$$q(\alpha, \beta, T) \equiv aq(\alpha, \beta, t) + bt + \lambda. \tag{24}$$

Now remember that we would like to call a symmetric first integral the constant of the motion implied by the invariance of the equation of motion. Perhaps such a terminology is only justified when this special constant of the motion is unique (apart from functionally dependent solutions). To see if this is the case for some value of a, b, λ, d, and μ, it will be enough to control whether the identity (24) is satisfied or not. However, we prefer to maintain our terminology in all circumstances and consider the cases where (24) holds as degenerate ones. An example of such a degenerate case is easily found when f is τ periodic in t and all solutions of (10) are periodic with the same period.

Returning to the main problem, we do not think that anything can be said about it in all its generality. Therefore, we will separately investigate the ten basic transformations contained in (14). These basic transformations are obtained when the influence of only one parameter is taken into account, once as a free parameter and once as a fixed one. So we distinguish as follows:

(i) time translation:

$$a = d = 1,\ b = \lambda = 0,\ \mu \text{ free;} \tag{25}$$

(ii) time-scale transformation:

$$a = 1,\ b = \lambda = \mu = 0,\ d \text{ free;} \tag{26}$$

(iii) coordinate translation:

$$a = d = 1,\ b = \mu = 0,\ \lambda \text{ free;} \tag{27}$$

(iv) coordinate-scale transformation:

$$d = 1,\ b = \lambda = \mu = 0,\ a \text{ free;} \tag{28}$$

(v) Galilean transformation (or "velocity-translation"):

$$a = d = 1,\ \lambda = \mu = 0,\ b \text{ free;} \tag{29}$$

(vi) time periodicity:

$$a = d = 1,\ b = \lambda = 0,\ \mu = \tau \text{ arbitrary fixed;} \tag{30}$$

(vii) fixed time-scale transformation:

$$a = 1,\ b = \lambda = \mu = 0,\ d \text{ arbitrary fixed;} \tag{31}$$

(viii) coordinate periodicity:

$$a = d = 1,\ b = \mu = 0,\ \lambda = P \text{ arbitrary fixed;} \tag{32}$$

(ix) fixed coordinate-scale transformation:

$$d = 1,\ b = \lambda = \mu = 0,\ a \text{ arbitrary fixed;} \tag{33}$$

(x) "velocity periodicity":

$$a = d = 1,\ \lambda = \mu = 0,\ b = P \text{ arbitrary fixed.} \tag{34}$$

In studying these cases in more detail in the next section, we will conclude for the existence of a symmetric L or ϕ whenever we are reduced to a partial differential equation for which well-known existence theorems can be proved under fairly general regularity conditions.

3. EXAMINATION OF THE BASIC TRANSFORMATIONS

(i) Time translation

Suppose that f satisfies (15) which with (25) simply means that f is independent of t. We look for a solution of (11) and (16), satisfying the invariance properties (7) and (17) which again reduce to the requirement of time-independency. As a result we only have to find in this case a function $L(q, \dot{q})$ and a function $\phi(q, \dot{q})$ satisfying, respectively, the equations

$$\frac{\partial^2 L}{\partial q \partial \dot{q}}\dot{q} - \frac{\partial L}{\partial q} = f(q, \dot{q})\frac{\partial^2 L}{\partial \dot{q}^2}, \tag{35}$$

$$\frac{\partial \phi}{\partial q}\mathring{q} - f(q, \dot{q})\frac{\partial \phi}{\partial \mathring{q}} = 0. \tag{36}$$

Such solutions are assured by classical existence theorems. The symmetric solution for ϕ is here unique. This can be seen directly from (36) or from (24), taking in this case the form

$$q(\alpha, \beta, t + \mu) \equiv q(\alpha, \beta, t).$$

In order to violate the unicity of the symmetric ϕ, this identity should have to hold for all μ, which is not possible (isolated constant solutions of (10) are excluded since α and β are arbitrary free constants).

Incidentally, the existence of a time-independent Lagrangian in this case was already mentioned by Havas.[7]

(ii) Time-scale transformation

From (15) and (26) we can easily deduce that in this case f is of the form (see also Denman[2])

$$f(q, \dot{q}, t) = (1/t^2) f_1(t\mathring{q}, q), \tag{37}$$

while the desired solutions for L and ϕ must have the form

$$L(q, \dot{q}, t) = L_1(t\dot{q}, q), \tag{38}$$

$$\phi(q, \dot{q}, t) = \phi_1(t\mathring{q}, q). \tag{39}$$

Putting $u = t\dot{q}$ and taking into account (37), (38), and (39), Eqs. (11) and (16) can be written as partial differential equations in the two independent variables q and u:

$$u\frac{\partial^2 L_1}{\partial u \partial q}+u\frac{\partial^2 L_1}{\partial u^2}-\frac{\partial L}{\partial q}=f_1(u,q)\frac{\partial^2 L_1}{\partial u^2}, \tag{40}$$

$$u\left(\frac{\partial \phi_1}{\partial q}+\frac{\partial \phi_1}{\partial u}\right)-f_1(u,q)\frac{\partial \phi_1}{\partial u}=0. \tag{41}$$

So again solutions for L and ϕ of the desired form can be found. The identity (24) here becomes

$$q(\alpha,\beta,dt)\equiv q(\alpha,\beta,t) \text{ for all } d,$$

which cannot hold so that the implied constant of the motion is unique.

(iii) Coordinate translation

Now f is independent of q and we look for solutions for L and ϕ of the same form. They can be found as solutions of the reduced equations

$$\frac{\partial^2 L}{\partial \dot{q}\partial t}=f(\dot{q},t)\frac{\partial^2 L}{\partial \dot{q}^2}, \tag{42}$$

and

$$-f(\dot{q},t)\frac{\partial \phi}{\partial \dot{q}}+\frac{\partial \phi}{\partial t}=0. \tag{43}$$

The identity (24), here of the form

$$q(\alpha,\beta,t)\equiv q(\alpha,\beta,t)+\lambda,$$

cannot be satisfied for all λ so that the implied constant of the motion again is unique.

(iv) Coordinate-scale transformation

According to (15) and (28) f is now a homogeneous function of the first degree in q and $\dot{q}$,

$$f(aq,a\dot{q},t)\equiv af(q,\dot{q},t), \quad \text{for all } a, \tag{44}$$

while L and ϕ are required to be functions of z and t, where (see also Denman[2])

$$z=q^{-1}\dot{q}. \tag{45}$$

Such solutions again can be found since with (44) and (45), Eqs. (11) and (16) can be written as

$$-z^2\frac{\partial^2 L}{\partial z^2}+\frac{\partial^2 L}{\partial z\partial t}=f(1,z,t)\frac{\partial^2 L}{\partial z^2}, \tag{46}$$

$$-[z^2+f(1,z,t)]\frac{\partial \phi}{\partial z}+\frac{\partial \phi}{\partial t}=0. \tag{47}$$

The solution for ϕ is unique because the identity

$$q(\alpha,\beta,t)\equiv aq(\alpha,\beta,t)$$

is not true for all a.

(v) Galilean transformation

In this case f is given to be of the form (see also Denman[2])

$$f(q,\dot{q},t)=F(w,t), \tag{48}$$

with

$$w=t\dot{q}-q. \tag{49}$$

A Langrangian L and a constant of the motion $\phi=c$ are called symmetric here if they have the same functional form as f. Starting from such a structure, Eqs. (11) and (16) become

$$2\frac{\partial L}{\partial w}=F(w,t)\cdot t^2\frac{\partial^2 L}{\partial w^2}, \tag{50}$$

$$-F(w,t)\cdot t\frac{\partial \phi}{\partial w}+\frac{\partial \phi}{\partial t}=0, \tag{51}$$

so that symmetric solutions for L and ϕ here too exist and the one for ϕ is unique since the identity

$$q(\alpha,\beta,t)\equiv q(\alpha,\beta,t)+bt$$

cannot hold for all b.

(vi) Time periodicity

Given a function which is τ periodic in t, we were not able to prove the existence of a periodic L and ϕ satisfying (11) and (16). However, we believe that they exist in fairly general circumstances. More particularly, concerning the first integral ϕ, many examples can be found such as damped linear oscillators with a periodic forcing term, for which the equation of motion has no periodic solutions and still a periodic first integral can be constructed. The only paper known to us in which some attention is paid to the problem of finding periodic solutions of linear first order partial differential equations with periodic coefficients is due to Levi-Civita.[10]

(vii) Fixed time-scale transformation

Suppose that f has the invariance property

$$f(q,\dot{q},t)\equiv d^2 f(q,\dot{q}d^{-1},dt) \tag{52}$$

for some arbitrary but fixed d. According to (7), (17), and (31) we look for an L and ϕ with the property

$$\mathcal{J}(q,\dot{q},t)\equiv \mathcal{J}(q,d^{-1}\dot{q},dt). \tag{53}$$

Assume now that there exists a positive integer n such that

$$d^n=1 \tag{54}$$

and put for a given function $F(q,\dot{q},t)$

$$F_k(q,\dot{q},t)\equiv F(q,d^{-k}\dot{q},d^k t). \tag{55}$$

Then if L and ϕ are arbitrary solutions of their determining equations

$$\mathcal{L}=\sum_{k=0}^{n-1}L_k(q,\dot{q},t) \text{ and } \Phi=\sum_{k=0}^{n-1}\phi_k(q,\dot{q},t) \tag{56}$$

produce the desired symmetric solutions.

Indeed, from the general properties of the preceding section, L_k and ϕ_k are known to be solutions too of (11) and (16), so the same is true for $\mathcal{L}$ and Φ by the linearity of the equations. Furthermore, $\mathcal{L}$ and Φ trivially satisfy a relation of the form (53) in view of (54).

Here a degeneration can occur (in the sense of nonunique symmetric constant of the motion) if all solutions of (10) satisfy the identity

$$q(\alpha,\beta,dt)\equiv q(\alpha,\beta,t).$$

We were not able to prove the desired results in the case where we cannot appeal to a finite group structure as implied by (54). Remark, however, that the case of time inversion, treated by Denman[2] is a special case of (54) corresponding to $n=2$.

(viii) Coordinate periodicity

Analogous considerations can be made as in the case of time periodicity.

(ix) Fixed coordinate-scale transformation:

See point (vii).

(x) "Velocity periodicity"

See points (vi) and (viii).

4. EXAMPLE

In the introduction we have treated the example of a linearly damped harmonic oscillator from the point of view of time-translation invariance. The symmetric (here time-independent) solutions for L and ϕ were given by (3) and (6).

Now Eq. (1) also satisfies the requirement (44) for coordinate-scale invariance. Solutions for L and ϕ which are invariant for coordinate-scale transformations will have to be obtained from the respective equations

$$-z^2\frac{\partial^2 L}{\partial z^2}+\frac{\partial^2 L}{\partial z\,\partial t}=(2\gamma z+\omega_0^2)\frac{\partial^2 L}{\partial z^2} \tag{57}$$

$$-(z^2+2\gamma z+\omega_0^2)\frac{\partial \phi}{\partial z}+\frac{\partial \phi}{\partial t}=0, \tag{58}$$

where z is given by (45).

Remark that in order to get a regular (i.e., a nondegenerate) Lagrangian from (57), it certainly must depend on t. This illustrates that in our approach different invariances may not be combined; it will in general not be possible to find Lagrangians (and constants of the motion) corresponding to more than one symmetry.

From classical elementary methods we can obtain the following particular solutions of (57) and (58):

$$L=\frac{z+\gamma}{\omega}\left[\arctan\left(\frac{z+\gamma}{\omega}\right)+\omega t\right]-\tfrac{1}{2}\ln\left[1+\left(\frac{z+\gamma}{\omega}\right)^2\right], \tag{59}$$

$$\phi=\arctan\left(\frac{z+\gamma}{\omega}\right)+\omega t, \tag{60}$$

with ω defined by (4). Remark that (60) is identical to Denman's implied constant of the motion for this case.[3] Let us finally consider the special case $\omega_0=0$. Regarding time-translation invariance and coordinate-scale invariance we then can calculate the symmetric L and ϕ by replacing ω by $i\gamma$ in the general expression for $\omega_0\neq 0$. But the equation

$$\ddot{q}+2\gamma\dot{q}=0 \tag{61}$$

now has a supplementary symmetry, namely coordinate-translation invariance. The determining equations for the corresponding symmetric L and ϕ are given by (42) and (43). They give rise to the following particular solutions,

$$L=\tfrac{1}{2}e^{2\gamma t}\dot{q}^2, \tag{62}$$

$$\phi=e^{2\gamma t}\dot{q}. \tag{63}$$

Remark that (62) corresponds to the Lagrangian (2) for $\omega_0=0$. So we could say that it is merely by a lucky coincidence that the symmetric Lagrangian implied by the coordinate-translation invariance of Eq. (61) produces a handy Lagrangian for the extended case $\omega_0\neq 0$.

5. CONCLUSION

Perhaps the most systematic approach to the problem of symmetry and constants of the motion is contained in Noether's theorem and its inverse. There the existence of a constant of the motion is linked to an invariance of the Lagrangian for some one-parameter family of transformations. The transformations (14) are not necessarily of the one-parameter type. For some of these transformations Denman has reduced the existence of a first integral to an invariance of the equation of motion not connected with a Lagrangian governing this motion. By some elementary considerations we have tried to re-establish the role of the Lagrangian and to launch a new idea concerning the implied constant of the motion, which is required to inherit the symmetry property. There seems to be a notable distinction between the cases where we could prove the existence of such a symmetrical first integral and the other ones. Indeed, the favorable transformations were those generating a finite group or a one-parameter continuous group, while the unfavorable ones were generating a discrete infinite-dimensional group. As a final remark we can stipulate that an eventual generalization to more degrees of freedom is not obvious for the simple reason that even the existence of a Lagrangian for a general system of second-order differential equations (the so-called inverse problem in Newtonian mechanics) is not such an easy matter. Remark that we want the system to be directly derivable from a variational principle (eventually after multiplication with integrating factors, see, e.g., Ref. 7). Havas[11] proved that there always exists a Lagrangian for the equivalent first-order system corresponding to any second-order system, but this is not the type of Lagrangian we are interested in.

Recently Santilli[12,13] has been working on two extensive monographs on the inverse problem in Newtonian mechanics. In his second volume[13] Santilli also deals with the general problem of invariances and conservation laws, more or less in the same spirit as in this paper, but for cases in which Noether's theorem applies.

[1] H.H. Denman, J. Math. Phys. **6**, 1611—1616 (1965).
[2] H.H. Denman, J. Math. Phys. **7**, 1910—1915 (1966).
[3] H.H. Denman, Am. J. Phys. **36**, 516—519 (1968).
[4] H.H. Denman and L.H. Buch, J. Math. Phys. **14**, 326—329 (1973).
[5] D.G. Currie and E.J. Saletan, J. Math. Phys. **7**, 967—974 (1966).
[6] H. Bateman, Phys. Rev. **38**, 815 (1931).
[7] P. Havas, Nuovo Cimento (Suppl.) **5**, 363—388 (1957).

[8] This result can be generalized in the following way. Consider instead of (5) a general transformation of both the independent and the dependent variable:

$$\begin{matrix} Q = Q(q,t) \\ T = T(q,t) \end{matrix} \quad \text{with } J = \left| \frac{\partial(Q,T)}{\partial(q,t)} \right| \neq 0. \tag{5'}$$

Suppose again that $L(q,\dot{q},t) \equiv L(Q,Q',T)$. The condition [replacing (12)] which must be satisfied by f in order that the equation of motion (10) would be invariant for the transformation (5') is more involved here. But it can be shown that the invariance of L implies the invariance of (10) if and only if $\partial T/\partial q = \partial^2 T/\partial t^2 = 0$. So T must be of the form admitted in (14), but no restriction must be imposed on the function $Q(q,t)$.

[9]G. H. Katzin and J. Levine, J. Math. Phys. **15**, 1460—1470 (1974).
[10]T. Levi-Civita, C. R. Acad. Sci. Paris **128**, 978—981 (1899).
[11]P. Havas, Acta Phys. Austr. **38**, 145—167 (1973).
[12]R. M. Santilli, *The Inverse Problem in Newtonian Mechanics* [MIT-CTP publication No. 606 (1977); Springer-Verlag, Heidelberg, to be published].
[13]R. M. Santilli, *Generalizations of the Inverse Problem in Newtonian Mechanics* [MIT-CTP publication No. 607 (1977); Springer-Verlag, Heidelberg, to be published].

On the normalization of one-dimensional lattice statistics

R. B. McQuistan

Department of Physics and Laboratory for Surface Studies, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201
(Received 22 November 1977)

It is shown that the normalization of the statistics associated with the distribution of λ-bell particles [λ = 1 for simple particles; λ = 2 for dumbbell particles, etc.] on a one-dimensional lattice space, is given by a simple, two-term recursion relation. This normalization, for a lattice consisting of N sites, is shown to be $(1/x_+)^N$, where x_+ is the real positive root of the equation $x^\lambda + x - 1 = 0$. A general recursion expression is also developed that describes the higher order moments of the statistics.

I. NORMALIZATION

It has been shown[1] that $A_\lambda(q,N)$, the occupational degeneracy for the distribution of q indistinguishable λ-bell particles $[\lambda = 1, 2, 3, \cdots]$ on a one-dimensional lattice space of N equivalent sites, is given by

$$A_\lambda(q,N) = \binom{N - q(\lambda-1)}{q}. \tag{1}$$

Here λ is the number of contiguous sites occupied by each λ-bell particle.

Equation (1) results from the following considerations: There are q indistinguishable λ-bell particles and $N - q\lambda$ indistinguishable vacant lattice sites, or a total of $N - q(\lambda - 1)$ objects. These objects can be arranged in the number of ways described by the binomial coefficient in Eq. (1).

One of the purpose of the present paper is to show that the normalization, ${}_\lambda\Delta_N$, of $A_\lambda(q,N)$ satisfies a simple, two-term recursion relation.

A recursion for $A_\lambda(q,N)$ may be written

$$\binom{N-q(\lambda-1)}{q} = \binom{N-1-q(\lambda-1)}{q} + \binom{N-1-q(\lambda-1)}{q-1}. \tag{2}$$

Thus, the normalization of $A_\lambda(q,N)$ becomes

$$\begin{aligned} {}_\lambda\Delta_N &= \sum_{q=0}^{[N/\lambda]} \binom{N-q(\lambda-1)}{q} \\ &= \sum_{q=0}^{[(N-1)/\lambda]} \binom{N-1-q(\lambda-1)}{q} \\ &\quad + \sum_{q=0}^{[(N-\lambda)/\lambda]} \binom{N-\lambda-q(\lambda-1)}{q} \\ &= {}_\lambda\Delta_{N-1} + {}_\lambda\Delta_{N-\lambda}, \end{aligned} \tag{3}$$

where $[N/\lambda]$ is the largest integer contained in N/λ. In the recursion relation Eq. (3), we take

$$N = \lambda,\ \lambda+1,\ \lambda+2, \cdots, \tag{4}$$

with the initial conditions

$${}_\lambda\Delta_0 = {}_\lambda\Delta_1 = {}_\lambda\Delta_2 = \cdots {}_\lambda\Delta_{\lambda-1} = 1. \tag{5}$$

The initial conditions described in Eq. (5) represents the fact that no λ-bell particles can be placed on a space for which $N \leq \lambda - 1$.

To determine ${}_\lambda\Delta_N$ we multiply it by x^N and sum over $N = \lambda$ to ∞. This yields

$$\sum_{N=\lambda}^{\infty} {}_\lambda\Delta_N x^N = \sum_{N=\lambda}^{\infty} {}_\lambda\Delta_{N-1} x^N + \sum_{N=\lambda}^{\infty} {}_\lambda\Delta_{N-\lambda} x^N. \tag{6}$$

This yields

$$G_\lambda(x) = xG_\lambda(x) + x^\lambda G_\lambda(x) + 1 \tag{7}$$

or

$$G_\lambda(x) = (1 - x - x^\lambda)^{-1}, \tag{8}$$

where

$$G_\lambda(x) \equiv \sum_{N=0}^{\infty} {}_\lambda\Delta_N x^N. \tag{9}$$

If the roots of Eq. (8), ${}_\lambda S_j^{-1} (j = 1, 2, 3, \ldots, \lambda)$, are all unequal, then

$$G_\lambda(x) = \sum_{j=0}^{\lambda} \frac{k_j}{1 - {}_\lambda S_j x^j} \sum_{N=0}^{\infty} \sum_{j=0}^{\lambda} k_j {}_\lambda S_j^N x^N, \tag{10}$$

i.e.,

$${}_\lambda\Delta_N = \sum_{j=0}^{\lambda} k_j {}_\lambda S_j^N.$$

We note that when λ is odd, Eq. (8) has one positive real root, $(\lambda - 1)/2$ pairs of complex roots; when λ is even, there are two real roots, one positive and one negative, and $(\lambda - 2)/2$ pairs of complex roots. When there are two real roots, the absolute value of the positive root is always less than the absolute value of negative root. Thus

$$\lim_{N\to\infty} {}_\lambda\Delta_N = k_0 {}_\lambda S_0^N, \tag{11}$$

where ${}_\lambda S_0$ is the reciprocal of the positive root of Eq. (8) and where

$$k_0 = \lim_{S_0 x \to 1} (1 - {}_\lambda S_0 x) G_\lambda(x). \tag{12}$$

Thus, for $\lambda = 1$, ${}_1\Delta_N = 2^N$ and for $\lambda = 2$, ${}_2\Delta_N = \mu_g^N$, where μ_g is the golden proportion, $\frac{1}{2}[1 + \sqrt{5}]$, etc.

It is interesting to note that ${}_\lambda N_N(x)$ satisfies the following recursion relation:

$${}_\lambda\Delta_{M+N} = {}_\lambda\Delta_M {}_\lambda\Delta_N + \sum_{j=0}^{\lambda-2} {}_\lambda\Delta_{M-\lambda+1+j} \cdot {}_\lambda\Delta_{N-1-j} \tag{13}$$

 0022-2488/78/1905-1055$1.00

which may be obtained by iterating Eq. (3) with the initial conditions expressed in Eq. (5).

II. HIGHER MOMENTS

We define the mth moment of these statistics to be

$$ {}_\lambda\mu_N(m) \equiv \frac{1}{{}_\lambda\Delta_N} \sum_{q=0}^{[N/\lambda]} q^m A_\lambda[q, N]. \tag{14} $$

Then from Eq. (1) we write

$$ \begin{aligned} {}_\lambda\mu_N(m) &= \frac{1}{{}_\lambda\Delta_N} \sum_{q=0}^{[N/\lambda]} q^m A_\lambda[q, N-1] + A_\lambda[q-1, N-1] \\ &= \frac{1}{{}_\lambda\Delta_N} {}_\lambda\mu_{N-1}(m)\, {}_\lambda\Delta_{N-1} \\ &\quad + \frac{1}{{}_\lambda\Delta_N} \sum_{q=0}^{[N/\lambda]} q^m A_\lambda[q-1, N-1] \\ &= \frac{{}_\lambda\Delta_{N-1}}{{}_\lambda\Delta_N} {}_\lambda\mu_{N-1}(m) + \frac{1}{{}_\lambda\Delta_N} \\ &\quad \times \sum_{q=0}^{[N/\lambda]} [1 + (q-1)]^m A_\lambda[q-1, N-1] \end{aligned} $$

$$ {}_\lambda\mu_N(m) = \frac{{}_\lambda\Delta_{N-1}}{{}_\lambda\Delta_N} {}_\lambda\mu_{N-1}(m) + \sum_{j=0}^{m} \binom{m}{j} {}_\lambda\mu_{N-1}(j). $$

CONCLUSION

Starting with an expression for the occupation degeneracy for λ-bell particles distributed on a one-dimensional lattice, we have shown that the normalization satisfies an algebraic equation of order λ. We have also developed a relation which yields the higher moments of these statistics.

[1]R. B. McQuistan, Nuovo Cimento 58, 86—92 (1968).

On Lie algebraic properties of the step operators acting on P or confluent P functions

Masao Mori[a)]

School of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455
(Received 3 September 1976; revised manuscript received 20 July 1977)

P functions and confluent P (CP) functions are classified into two and five groups respectively according to the types of the step operators (SO's) intrinsic to the respective classes of functions. The correspondence between the types of the SO's and the realizations of the Lie algebras $\mathcal{G}(a,b)$ and $\mathcal{T}_6$ is established as follows. The modified SO's acting on P functions (SOP's) belong to either of the type A and E realizations of $\mathcal{G}(1,0)$ and $\mathcal{T}_6$ respectively. The modified SOC's, namely the SO's acting on CP functions, belong to one of the type B, C', C'', F, D' realizations of $\mathcal{G}(1,0)$, $\mathcal{G}(0,1)$, $\mathcal{G}(0,0)$, or $\mathcal{T}_6$.

1. INTRODUCTION

As is well known, the second order differential equations of mathematical physics are classified into three groups by analyticity.[1,2] (1) The first consists of Fuchsian differential equations of second order with three singular points. The general solution of this group is expressed in terms of Riemann's P function. (2) The second consists of the equations closely related to the confluent form of the preceding type. The general solution of this group is expressed in terms of Hukuhara's CP function.[3] (3) The third consists of the equations with higher singularity such as Mathieu's equation and Lamê's equation.

The factorization method for these equations originates from Schrödinger.[4-6] It yields the SO's which shift up or down the parameters appearing in a solution by some integers, and therefore leads to recurrence relations. For the equations and their solutions grouped into (1) and (2), Inui established a unified theory of factorizations and recurrence formulas.[7,8] This theory is characterized by simplicity which results from utilization of Hukuhara's differentiation formulas. Making use of miscellaneous transformations of dependent and independent variables about individual equations, Infeld and Hull classified the types of factorizations into six.[9] As to the solutions of the third group, however, there are some disputes about the existence of recurrence relations, which have not been settled yet.[10,11]

From the viewpoint of Lie algebra, Weisner took the initiative of discussion about some of these equations,[12-14] and Miller set up the comprehensive Lie theory of special functions in which the SO's are expressed in terms of realizations of the Lie algebras $\mathcal{G}(a, b)$ or $\mathcal{T}_6$,[15,16] where the algebraic background of the classification of Infeld and Hull is ascertained. Now a compreshensive survey of the Lie theory of special functions from the standpoint of symmetry and the R-separability of the complex wave equation $(\partial_{00} - \partial_{11} - \partial_{22} - \partial_{33})\Psi = 0$ is carried out by Kalnins and Miller.[17-21]

In the present work, we aim to fill the gap existing between Inui's and Miller's unified theories of special functions. Our results make clear the connection between the function of differentiation formulas for P or CP functions, namely the behavior of the hypergeometric parameters (HGP's) when the SO's are at work, and the realization of the Lie algebras $\mathcal{G}(a, b)$ or $\mathcal{T}_6$.

We begin with preliminary definitions in Sec. 2. In Secs. 3 and 4, the classification of P and CP functions and SO's are stated. A modification of Inui's theory of SO's is shown in Sec. 5. Concrete correspondence between the SO's and the realizations are given in Sec. 6.

2. DEFINITIONS

We denote by $\mathcal{P}$ the set of Riemann's P functions, that is,

$$\mathcal{P} = \mathrm{P}\begin{bmatrix} 0 & 1 & \infty \\ \alpha_1 & \beta_1 & \lambda_1 z \\ \alpha_2 & \beta_2 & \lambda_2 \end{bmatrix}, \quad \alpha_i, \beta_i, \lambda_i \in \mathbf{C}, \ i = 1, 2, \tag{2.1}$$

and by $\mathrm{C}\mathcal{P}$ the set of Hukuhara's CP functions[3,7,8,22]

$$\mathrm{C}\mathcal{P} = \mathrm{P}\begin{bmatrix} \infty & 0 & \\ \rho_1 & \kappa_1 & \sigma_1 \ z \\ \rho_2 & \kappa_2 & \sigma_2 \end{bmatrix},$$

$$\rho_i, \kappa_i, \sigma_i \in \mathbf{C}, \ i = 1, 2, \ \rho_1 \neq \rho_2. \tag{2.2}$$

The parameters in the definitions (2.1) and (2.2) are restricted to satisfy Fuchs' relation

$$\alpha_1 + \alpha_2 + \beta_1 + \beta_2 + \lambda_1 + \lambda_2 = 1 \tag{2.3}$$

and the Fuchs–Hukuhara relation

$$\kappa_1 + \kappa_2 + \sigma_1 + \sigma_2 = 1, \tag{2.4}$$

respectively.

Replacement of the independent variable by one of the expressions

$$z, \ 1 - z, \ \frac{1}{z}, \ \frac{1}{1-z}, \ \frac{1}{z-1}, \ \frac{z-1}{z}, \tag{2.5}$$

defines a one-to-one mapping from $\mathcal{P}$ onto itself.[2,3] Let us call this monodromy transformation the M transformation. Replacing the parameters in the definitions (2.1) or (2.2) by their linear combinations made so as not to violate the conditions (2.3) or (2.4), we obtain transformations on $\mathcal{P}$ or $\mathrm{C}\mathcal{P}$. We shall call this replacement of parameters the π transformation.

Furthermore, if we multiply an element of $\mathcal{P}$ by the

[a)] On leave of absence from Chuo University, Tokyo, Japan.

 0022-2488/78/1905-1057$1.00

factor $\tau(z)=z^a(z-1)^b$, or an element of $C\mathcal{P}$ by $\tau(z)=\exp(az)z^b$, where a and b are complex constants, Fuchs' or the Fuchs–Hukuhara relation will remain unchanged. For we have the following formulas[3,7,8,22]:

$$z^a(z-1)^b \mathrm{P}\begin{bmatrix} 0 & 1 & \infty & \\ \alpha_1 & \beta_1 & \lambda_1 & z \\ \alpha_2 & \beta_2 & \lambda_2 & \end{bmatrix} = \mathrm{P}\begin{bmatrix} 0 & 1 & \infty & \\ \alpha_1+a & \beta_1+b & \lambda_1-a-b & z \\ \alpha_2+a & \beta_2+b & \lambda_2-a-b & \end{bmatrix}, \tag{2.6}$$

$$\exp(az)z^b \mathrm{P}\begin{bmatrix} \overbrace{\quad}^{\infty} & & 0 & \\ \rho_1 & \kappa_1 & \sigma_1 & z \\ \rho_2 & \kappa_2 & \sigma_2 & \end{bmatrix} = \mathrm{P}\begin{bmatrix} \overbrace{\quad}^{\infty} & & 0 & \\ \rho_1+a & \kappa_1-b & \sigma_1+b & z \\ \rho_2+a & \kappa_2-b & \sigma_2+b & \end{bmatrix}. \tag{2.7}$$

Let us call this multiplication by $\tau(z)$ the τ transformation of the element, which is apparently a bijection.

3. THE DIFFERENTIATION FORMULAS FOR P AND CP FUNCTIONS

In view of the relations (2.6) and (2.7), it is known that a τ transformation gives rise to change of two of the parameters appearing in the operand; therefore, it can be a kind of SO. However, other parameters can change only through differentiation of the function.[3] In the present work, we consider the SO's in the form of differential operator; hence our attention will be restricted to the latter parameters. Thus there arises a need to consider the derivatives of the elements of $\mathcal{P}$ and $C\mathcal{P}$. Unfortunately, it is known that the differentiation could cause violation of the relations (2.3) and (2.4).[3] So we should define some proper subsets within which the first derivatives of the elements remain.

Let $\mathfrak{P}$ and $C\mathfrak{P}$ be the sets

$$\mathfrak{P}=\{f(z)\,|\,Df(z)\in\mathcal{P}\}, \tag{3.1}$$

$$C\mathfrak{P}=\{f(z)\,|\,Df(z)\in C\mathcal{P}\}, \tag{3.2}$$

where $D=d/dz$.

Corresponding to the location of zeros in the parameters of $\mathfrak{P}$ or $C\mathfrak{P}$, they separate into several subsets.[3,23,24] When we take the invariance of $\mathcal{P}$ under M transformations into account, $\mathfrak{P}$ is decomposed into two disjoint subsets $\mathfrak{P}_1$ and $\mathfrak{P}_2$ as follows[23]:

$$\mathfrak{P}=\mathfrak{P}_1\cup\mathfrak{P}_2, \tag{3.3}$$

where

$$\mathfrak{P}_1=\left\{\mathrm{P}\begin{bmatrix} 0 & 1 & \infty & \\ 0 & 0 & \lambda_1 & z \\ \alpha & \beta & \lambda_2 & \end{bmatrix}\,\middle|\, \lambda_1\lambda_2\neq 0,\ \alpha+\beta+\lambda_1+\lambda_2=1\right\}, \tag{3.4}$$

$$\mathfrak{P}_2=\left\{\mathrm{P}\begin{bmatrix} 0 & 1 & \infty & \\ 0 & \beta_1 & \lambda_1 & z \\ \alpha & \beta_2 & \lambda_2 & \end{bmatrix}\,\middle|\, \beta_1\beta_2=\lambda_1\lambda_2\neq 0,\ \alpha+\beta_1+\beta_2+\lambda_1+\lambda_2=1\right\}. \tag{3.5}$$

The same consideration about $C\mathfrak{P}$ leads us to the relation[3,24]

$$C\mathfrak{P}=C\mathfrak{P}_1\cup C\mathfrak{P}_2\cup C\mathfrak{P}^{(3)}, \tag{3.6}$$

where the subset with superscript will suffer a further decomposition as is shown later. The components are distinguished by the location of zeros, namely,

$$C\mathfrak{P}_1=\left\{\mathrm{P}\begin{bmatrix} \overbrace{\quad}^{\infty} & & 0 & \\ 0 & 0 & \sigma_1 & z \\ \rho & \kappa & \sigma_2 & \end{bmatrix}\,\middle|\, \sigma_1\sigma_2\neq 0,\ \sigma_1+\sigma_2+\kappa=1,\ \rho\neq 0\right\}, \tag{3.7}$$

$$C\mathfrak{P}_2=\left\{\mathrm{P}\begin{bmatrix} \overbrace{\quad}^{\infty} & & 0 & \\ 0 & \kappa_1 & 0 & z \\ \rho & \kappa_2 & \sigma & \end{bmatrix}\,\middle|\, \kappa_1\kappa_2\neq 0,\ \kappa_1+\kappa_2+\sigma=1,\ \rho\neq 0\right\}, \tag{3.8}$$

$$C\mathfrak{P}^{(3)}=\left\{\mathrm{P}\begin{bmatrix} \overbrace{\quad}^{\infty} & & 0 & \\ \rho_1 & \kappa_1 & 0 & z \\ \rho_2 & \kappa_2 & \sigma & \end{bmatrix}\,\middle|\, \rho_1\kappa_2=-\kappa_1\rho_2\neq 0,\ \kappa_1+\kappa_2+\sigma=1\right\}. \tag{3.9}$$

The differentiation formulas which describe the behavior of the parameters under differentiation are given as follows[3]:

$$D\mathrm{P}\begin{bmatrix} 0 & 1 & \infty & \\ 0 & 0 & \lambda_1 & z \\ \alpha & \beta & \lambda_2 & \end{bmatrix} = \mathrm{P}\begin{bmatrix} 0 & 1 & \infty & \\ 0 & 0 & \lambda_1+1 & z \\ \alpha-1 & \beta-1 & \lambda_2+1 & \end{bmatrix}, \tag{3.10}$$

$$D\mathrm{P}\begin{bmatrix} 0 & 1 & \infty & \\ 0 & \beta_1 & \lambda_1 & z \\ \alpha & \beta_2 & \lambda_2 & \end{bmatrix} = \mathrm{P}\begin{bmatrix} 0 & 1 & \infty & \\ 1 & \beta_1-1 & \lambda_1+1 & z \\ \alpha-1 & \beta_2-1 & \lambda_2+1 & \end{bmatrix} = z\mathrm{P}\begin{bmatrix} 0 & 1 & \infty & \\ 0 & \beta_1-1 & \lambda_1+2 & z \\ \alpha-2 & \beta_2-1 & \lambda_2+2 & \end{bmatrix}. \tag{3.11}$$

$$D\mathrm{P}\begin{bmatrix} \overbrace{\quad}^{\infty} & & 0 & \\ 0 & 0 & \sigma_1 & z \\ \rho & \kappa & \sigma_2 & \end{bmatrix} = \mathrm{P}\begin{bmatrix} \overbrace{\quad}^{\infty} & & 0 & \\ 0 & 2 & \sigma_1-1 & z \\ \rho & \kappa & \sigma_2-1 & \end{bmatrix} = z^{-2}\mathrm{P}\begin{bmatrix} \overbrace{\quad}^{\infty} & & 0 & \\ 0 & 0 & \sigma_1+1 & z \\ \rho & \kappa-2 & \sigma_2+1 & \end{bmatrix}, \tag{3.12}$$

$$D\mathrm{P}\begin{bmatrix} \overbrace{\quad}^{\infty} & & 0 & \\ 0 & \kappa_1 & 0 & z \\ \rho & \kappa_2 & \sigma & \end{bmatrix} = \mathrm{P}\begin{bmatrix} \overbrace{\quad}^{\infty} & & 0 & \\ 0 & \kappa_1+1 & 0 & z \\ \rho & \kappa_2 & \sigma-1 & \end{bmatrix}, \tag{3.13}$$

$$
D\mathrm{P}\begin{bmatrix} \overbrace{\infty} & 0 & \\ \rho_1 & \kappa_1 & 0 & z \\ \rho_2 & \kappa_2 & \sigma & \end{bmatrix}
$$

$$
= \mathrm{P}\begin{bmatrix} \overbrace{\infty} & 0 & \\ \rho_1 & \kappa_1 & 1 & z \\ \rho_2 & \kappa_2 & \sigma-1 & \end{bmatrix}
$$

$$
= z\mathrm{P}\begin{bmatrix} \overbrace{\infty} & 0 & \\ \rho_1 & \kappa_1+1 & 0 & z \\ \rho_2 & \kappa_2+1 & \sigma-2 & \end{bmatrix}. \tag{3.14}
$$

Let us consider the conditions for keeping the first derivatives within $\mathfrak{P}$ or $C\mathfrak{P}$ respectively. The differentiation formula (3.10) tells us that if the product $(\lambda_1+1)(\lambda_2+1)$ does not vanish, the right-hand side remains within the set $\mathfrak{P}$ in particular $\mathfrak{P}_1$. Hence D is in general an invariant operator in $\mathfrak{P}_1$. Successive application of D to an element of $\mathfrak{P}_1$ will construct a ladder of P functions indexed by parameters of integral difference. This ladder could consist of finite or infinite steps according to the relation between the parameters λ_1, λ_2 and the integer Z. For example, the ladder takes the form "bounded above" when one of λ_1 and λ_2 is a nonnegative integer.[15]

In the set $\mathfrak{P}_2$, the operator $z^{-1}D$ is invariant when the parameters satisfy the relations

$$
\beta_1\beta_2 = \lambda_1\lambda_2, \tag{3.15}
$$

$$
(\beta_1-1)(\beta_2-1) = (\lambda_1+2)(\lambda_2+2); \tag{3.16}
$$

hence

$$
\beta_1+\beta_2+2(\lambda_1+\lambda_2) = -3. \tag{3.17}
$$

In this case, $z^{-1}D$ is applicable again to admit the relation

$$
z^{-1}D\mathrm{P}\begin{bmatrix} 0 & 1 & \infty & \\ 0 & \beta_1-1 & \lambda_1+2 & z \\ \alpha-2 & \beta_2-1 & \lambda_2+2 & \end{bmatrix}
$$

$$
= \mathrm{P}\begin{bmatrix} 0 & 1 & \infty & \\ 0 & \beta_1-2 & \lambda_1+4 & z \\ \alpha-4 & \beta_2-2 & \lambda_2+4 & \end{bmatrix}. \tag{3.18}
$$

Even if the condition (3.17) is fulfilled, we are prohibited from constructing the third step, since the equation

$$
(\beta_1-2)(\beta_2-2) = (\lambda_1+4)(\lambda_2+4) \tag{3.19}
$$

conflicts with the condition (3.17). Consequently, the ladder has at most three steps in this case. For the operator $z^{-1}(z-1)D$ instead of $z^{-1}D$, the situation does not change. However, Lie algebraic significance does not lie in the short ladder constructed in $\mathfrak{P}_2$ but in the ladder constructed in $\mathfrak{P}_1$ as follows. Expressing an element of $\mathfrak{P}_2$ as a τ-transformation of an element of $\mathfrak{P}_1$, for example,

$$
\mathrm{P}\begin{bmatrix} 0 & 1 & \infty & \\ 0 & \beta_1 & \lambda_1-\beta_1 & z \\ \alpha & \beta+\beta_1 & \lambda_2-\beta_1 & \end{bmatrix}
$$

$$
= (z-1)^{\beta_1}\mathrm{P}\begin{bmatrix} 0 & 1 & \infty & \\ 0 & 0 & \lambda_1 & z \\ \alpha & \beta & \lambda_2 & \end{bmatrix},
$$

where $\beta_1 = \lambda_1\lambda_2/(\beta+\lambda_1+\lambda_2)$, we can apply the differentiation formula (3.11) to obtain

$$
D(z-1)^{\beta_1}\mathrm{P}\begin{bmatrix} 0 & 1 & \infty & \\ 0 & 0 & \lambda_1 & z \\ \alpha & \beta & \lambda_2 & \end{bmatrix}
$$

$$
= z(z-1)^{\beta_1-1}\mathrm{P}\begin{bmatrix} 0 & 1 & \infty & \\ 0 & 0 & \lambda_1+1 & z \\ \alpha-2 & \beta & \lambda_2+1 & \end{bmatrix}.
$$

Hence we have SO $(z-1)^{-\beta_1}z^{-1}(z-1)D(z-1)^{\beta_1}$ acting on the element of $\mathfrak{P}_1$. The ladder constructed by this routine differs from the ladder constructed by the formula (3.10). Precise relations between the shape of this and the following ladders of CP functions as well, and the irreducible representations of Lie algebras are shown in Ref. 16, Chap. 3–6.

Similar consideration about the sets $C\mathfrak{P}_1$ and $C\mathfrak{P}_2$ give the ladders constructed by the operators z^2D or D.[25] Investigation of $C\mathfrak{P}^{(3)}$, however, leads to the decomposition

$$
C\mathfrak{P}^{(3)} = C\mathfrak{P}_3 \cup C\mathfrak{P}_4,
$$

$$
C\mathfrak{P}_3 = \left\{ \mathrm{P}\begin{bmatrix} \overbrace{\infty} & 0 & \\ -\rho & \kappa & 0 & z \\ \rho & \kappa & \sigma & \end{bmatrix} \middle| \; 2\kappa+\sigma = 1, \; \rho\kappa \neq 0 \right\}
$$

$$
C\mathfrak{P}_4 = \left\{ \mathrm{P}\begin{bmatrix} \overbrace{\infty} & 0 & \\ \rho_1 & \kappa_1 & 0 & z \\ \rho_2 & \kappa_2 & \sigma & \end{bmatrix} \middle| \; \kappa_1+\kappa_2+\sigma = 1, \; \rho_1\kappa_2 = -\rho_2\kappa_1 \neq 0 \right\}.
$$

These subsets are distinguished according as they admit long ladders or not. In fact, we can see that the set $C\mathfrak{P}_3$ admits the invariant operator $z^{-1}D$, while for $C\mathfrak{P}_4$, the situation resembles the case $\mathfrak{P}_2$.[25]

Let Λ be an SO or SOC acting on a function belonging to $\mathfrak{P}$ or $C\mathfrak{P}$; then those acting on a function transformed by τ will be given in the form

$$
\tau\Lambda\tau^{-1}. \tag{3.20}
$$

We call the operator in this form the τ transformation of Λ. It will be easily seen that a τ transformation induces a Lie algebra isomorphism if Λ's are the elements of some Lie algebra.

4. REPRESENTATIVES OF SO'S AND LADDERS

According to Inui,[7,8] all SOP's can be derived from those for the general solution

$$
\mathrm{f}(\alpha,\beta;\gamma;z) = \mathrm{P}\begin{bmatrix} 0 & 1 & \infty & \\ 0 & 0 & \alpha & z \\ 1-\gamma & \gamma-\alpha-\beta & \beta & \end{bmatrix} \tag{4.1}
$$

of the HGE of Gauss

$$
z(1-z)\frac{d^2\mathrm{f}}{dz^2} + \{\gamma-(\alpha+\beta+1)z\}\frac{d\mathrm{f}}{dz} - \alpha\beta\mathrm{f} = 0. \tag{4.2}
$$

And all SOC's can be derived from those for the general solution

$$
\mathrm{f}(\alpha;\gamma;z) = \mathrm{P}\begin{bmatrix} \overbrace{\infty} & 0 & \\ 0 & \alpha & 0 & z \\ 1 & \gamma-\alpha & 1-\gamma & \end{bmatrix} \tag{4.3}
$$

of Kummer's confluent HGE

$$z\frac{d^2f}{dz^2}+(\gamma-z)\frac{df}{dz}-\alpha f=0. \tag{4.4}$$

Taking M, τ, and π transformations into account, all SOP's are equivalent to either of the following operators[24]:

SOP's of the first kind:

$$\begin{aligned}&\overset{(\alpha)}{T}=z^{-\alpha}zDz^{\alpha},\\ &\underset{(\alpha)}{T}=z^{-(\gamma-\alpha)}(z-1)^{-(\alpha+\beta-\gamma)}z(z-1)Dz^{\gamma-\alpha}(z-1)^{\alpha+\beta-\gamma},\end{aligned} \tag{4.5}$$

SOP's of the second kind:

$$\begin{aligned}&\overset{(\alpha,\beta,2\gamma)}{T}=(z-1)^{-\alpha\beta/\gamma}z^{-1}(z-1)D(z-1)^{\alpha\beta/\gamma},\\ &\underset{(\alpha,\beta,2\gamma)}{T}=\tau(z)z(z-1)D\tau(z)^{-1}\end{aligned} \tag{4.6}$$

where $\tau(z)=z^{1-\gamma}(z-1)^{-(\alpha-\gamma+1)(\beta-\gamma+1)/(2-\gamma)}$.[7] A couple of other SOP's of the second kind

$$\begin{aligned}&\overset{(\beta)}{T}=\tau(z,\alpha,\beta)z(z-1)D\tau(z,\alpha,\beta)^{-1}\\ &\overset{(\alpha)}{T}=\tau(z,\beta,\alpha)z(z-1)D\tau(z,\beta,\alpha)^{-1}\end{aligned} \tag{4.6'}$$

where $\tau(z,\alpha,\beta)=z^{-\beta(\gamma-\alpha)/(\beta-\alpha+1)}(z-1)^{\beta(\gamma-\beta-1)/(\beta-\alpha+1)}$, will be used in Sec. 6. These operate on the associated Legendre function, and are equivalent to (4.6).

Let us define the ladders

$$L_1=\mathrm{P}\begin{bmatrix}0&1&\infty\\0&0&\alpha_m&z\\1-\gamma&\gamma-\alpha_m-\beta&\beta\end{bmatrix},\quad \alpha_m=\alpha_0+n, \tag{4.7}$$

$$L_2=\mathrm{P}\begin{bmatrix}0&1&\infty\\0&0&\alpha_m&z\\1-\gamma_m&\gamma_m-\alpha_m-\beta_m&\beta_m\end{bmatrix},$$

$$\alpha_m=\alpha_0+n,\quad \beta_m=\beta_0+n,\quad \gamma_m=\gamma_0+2n, \tag{4.8}$$

where n is an integer and the real parts of α_0, β_0, γ_0 are nonnegative constants smaller than one. These are generated from the P function (4.1) by the SOP's (4.5) and (4.6) respectively.

Taking τ and π transformations into account, all SOC's are equivalent to one of the following SO's [25]:

SOC's of the first kind:

$$\begin{aligned}&\overset{(\alpha)}{T}=z^{-\alpha}zDz^{\alpha},\\ &\underset{(\alpha)}{T}=z^{\alpha-\gamma}e^{z}zDe^{-z}z^{\gamma-\alpha}.\end{aligned} \tag{4.9}$$

SOC's of the second kind:

$$\begin{aligned}&\overset{(\gamma)}{T}=e^{z}De^{-z},\\ &\underset{(\gamma)}{T}=z^{1-\gamma}zDz^{\gamma-1}.\end{aligned} \tag{4.10}$$

SOC'S of the third kind:

$$\begin{aligned}&\overset{(\alpha)}{T}=e^{z/2}z^{-1}De^{-z/2},\\ &\underset{(\alpha)}{T}=e^{z/2}z^{1-2\alpha}zDe^{-z/2}z^{2\alpha-1}.\end{aligned} \tag{4.11}$$

SOC's of the fourth kind:

$$\begin{aligned}&\overset{(\alpha,2\gamma)}{T}=e^{\alpha z/\gamma}z^{-1}De^{-\alpha z/\gamma},\\ &\underset{(\alpha,2\gamma)}{T}=\tau(z)zD\tau(z)^{-1},\end{aligned} \tag{4.12}$$

where $\tau(z)=z^{1-\gamma}\exp[(\gamma+1-\gamma)z/(2-\gamma)]$.

In addition to the SOC's above, other SO's which change α by $\frac{1}{2}$ exist.[3,25] From Eq. (4.4) and the definition (4.9), the relation

$$\overset{(\alpha)}{T}=zD^2+\gamma D=z^{1-\gamma}D(z^{\gamma}D) \tag{4.4'}$$

holds for $f(\alpha;\gamma;z)$. According to the differentiation formula (3.13), we have

$$\begin{aligned}(z^{\gamma}D)f(\alpha;\gamma;z)&=z^{\gamma}D\mathrm{P}\begin{bmatrix}\overbrace{\infty}&0\\0&\alpha&0&z\\1&\gamma-\alpha&1-\gamma\end{bmatrix}\\&=\mathrm{P}\begin{bmatrix}\overbrace{\infty}&0\\0&\alpha-\gamma+1&0&z\\1&-\alpha&\gamma\end{bmatrix}.\end{aligned}$$

If we consider the special case $\gamma=\frac{1}{2}$, then the relations $\overset{(\alpha)}{T}=(\sqrt{z}D)^2$ and

$$\begin{aligned}&\sqrt{z}D\mathrm{P}\begin{bmatrix}\overbrace{\infty}&0\\0&\alpha&0&z\\1&\frac{1}{2}-\alpha&\frac{1}{2}\end{bmatrix}\\&=\mathrm{P}\begin{bmatrix}\overbrace{\infty}&0\\0&\alpha+\frac{1}{2}&0&z\\1&-\alpha&\frac{1}{2}\end{bmatrix}\end{aligned}$$

follow. This means that $\sqrt{z}\,Df(\alpha;\frac{1}{2};z)=f(\alpha+\frac{1}{2};\frac{1}{2};z)$, that is, $(d/dx)f(\alpha;\frac{1}{2};x^2)=f(\alpha+\frac{1}{2};\frac{1}{2};x^2)$, where we have put $z=x^2$. Hence we have the SOC's of the fifth kind

$$\begin{aligned}&\overset{(\alpha/2)}{T}=\sqrt{z}D,\\ &\underset{(\alpha/2)}{T}=e^{z}\sqrt{z}\,De^{-z}.\end{aligned} \tag{4.13}$$

These operators come from the accidental relation (4.4') between Eq. (4.4) and the SOC (4.9).

Let us define ladders of CP functions as follows:

$$CL_1=\mathrm{P}\begin{bmatrix}\overbrace{\infty}&0\\0&\alpha_m&0&z\\1&\gamma-\alpha_m&1-\gamma\end{bmatrix},\quad \alpha_m=\alpha_0+n, \tag{4.14}$$

$$CL_2=\mathrm{P}\begin{bmatrix}\overbrace{\infty}&0\\0&\alpha&0&z\\1&\gamma_m-\alpha&1-\gamma_m\end{bmatrix},\quad \gamma_m=\gamma_0+n, \tag{4.15}$$

$$CL_3=\mathrm{P}\begin{bmatrix}\overbrace{\infty}&0\\0&\alpha_m&0&z\\1&\alpha_m&1-2\alpha_m\end{bmatrix},\quad \alpha_m=\alpha_0+n, \tag{4.16}$$

$$CL_4=\mathrm{P}\begin{bmatrix}\overbrace{\infty}&0\\0&\alpha_m&0&z\\1&\gamma_{m'}-\alpha_m&1-\gamma_{m'}\end{bmatrix},$$

$$\alpha_m=\alpha_0+n,\quad \gamma_{m'}=\gamma_0+2n, \tag{4.17}$$

$$CL_5=\mathrm{P}\begin{bmatrix}\overbrace{\infty\quad\alpha'_m}^{} & 0 & z\\ 0 & 0 & \\ 1 & -\alpha'_m+\frac{1}{2} & \frac{1}{2}\end{bmatrix},\quad \alpha'_m=\alpha_0+\tfrac{1}{2}n, \tag{4.18}$$

where n is an integer, and the real parts of α_0 and γ_0 are nonnegative constants smaller than one. These ladders are generated from the CP function (4.3) by the SOC's (4.9)—(4.13) respectively.

5. A MODIFICATION OF INUI'S THEORY

We express by $\mathfrak{f}(m,z)$ one of the ladders L_1, L_2, CL_1, CL_2, CL_3, CL_4, or CL_5, where m denotes the index of the step. Let $m'(m)$ be a π transformation of m. Then any ladder $p_{m'(m)}(z)$ of P or CP functions whose steps are indexed by $m'(m)$ is expressed as a result of a τ and/or π transformation of $\mathfrak{f}(m,z)$[24,25]:

$$p_{m'(m)}(z)=\tau_m(z)\,\mathfrak{f}(m,z). \tag{5.1}$$

Let us define the following particular product,

$$f_{m'}(z,t)=p_{m'}(z)t^{m'}, \tag{5.2}$$

and call this a P_t or CP_t function according as p is the element of $\mathfrak{P}$ or $\mathrm{C}\mathfrak{P}$ respectively, where t is the newly introduced independent variable.

Let $T(m)$ be the SO which transforms $\mathfrak{f}(m,z)$ to $\mathfrak{f}(m\pm i,z)$ and $S(m')$ be the SO which transforms $f_{m'}(z,t)$ the P_t or CP_t function to $f_{m'\pm j}(z,t)$. Then $S(m')$ will be defined by the relation

$$S(m')=\tau_{m\pm i}(z)S(m)\tau_m(z)^{-1}, \tag{5.3}$$

and will be interpreted by the diagram

$$\begin{array}{ccc}\mathfrak{f}(m,z)t^m & \xrightarrow{\tau_m(z)} & f_{m'}(z,t)\\ \Big\downarrow S(m)=t^{\pm i}T(m) & & \Big\downarrow S(m')\\ \mathfrak{f}(m\pm i,z)t^{m\pm i} & \xrightarrow[\tau_{m\pm i}(z)]{} & f_{m'\pm j}(z,t).\end{array} \tag{5.4}$$

As the occasion arises, replacing t by a product for example, $\zeta=[z(z-1)]^{1/2}t$ [see the relation (6.7)], we can eliminate the parameter dependence of τ, and get

$$S(m')=\tau(\zeta)S(m)\tau(\zeta)^{-1}. \tag{5.5}$$

This and the relation (3.20) mean that the m' SO acting on the P_t or CP_t function is obtained as a τ transformation of $S(m)$.

6. THE SO'S AND THE REALIZATIONS OF LIE ALGEBRAS

A. The SOP's of the first and the second kind and the realization of the Lie algebras $\mathcal{G}(1,0)$ and $\mathcal{T}_6$

For the ladder (4.7), let us make the π transformation

$$\alpha_m=m-l,\quad \beta=-l,\quad \gamma=-2l, \tag{6.1}$$

so that

$$\mathrm{P}\begin{bmatrix}0 & 1 & \infty\\ 0 & 0 & m-l\\ 1+2l & -m & -l\end{bmatrix}\begin{matrix}z\\ \\ \\ \end{matrix}=\mathfrak{f}(m-l,-l;-2l;z). \tag{6.2}$$

If we think of the steps as indexed by m, this is exactly the ladder L_1, but if we think them as indexed by l, (6.2) is seen to be the ladder L_2.

Now the m SO's for the ladder

$$t^m\mathrm{P}\begin{bmatrix}0 & 1 & \infty\\ 0 & 0 & m-l\\ 1+2l & -m & -l\end{bmatrix}\begin{matrix}z\\ \\ \\ \end{matrix} \tag{6.3}$$

of P_t functions will be obtained from (4.5) by the π transfromation (6.1) succeeded by the replacements of D and m by $\partial/\partial z$ and $t\partial/\partial t$ respectively. Thus we have

$$\begin{aligned}&J^+=t\left(z\frac{\partial}{\partial z}+t\frac{\partial}{\partial t}-l\right),\\ &J^-=t^{-1}\left(z(z-1)\frac{\partial}{\partial z}+t\frac{\partial}{\partial t}-lz+l\right),\quad J^3=t\frac{\partial}{\partial t}.\end{aligned} \tag{6.4}$$

This is the type A realization of the nontrivial part $\mathrm{sl}(2,\mathbf{C})$ of $\mathcal{G}(1,0)$.[26]

On the other hand, the l SO's for the ladder (6.3) are obtained from the SO's of the second kind (4.6) through the π transformation (6.1) and the same replacements. However, the most familiar form of the SO's of the second kind are those which shift the parameter l of the associated Legendre function $P_l^m(\cos\theta)$. So it would be better to write them down in the form operating on the general solution

$$\Theta_l^m(\xi)=\mathrm{P}\begin{bmatrix}1 & -1 & \infty\\ -\frac{1}{2}m & -\frac{1}{2}m & -l\\ \frac{1}{2}m & \frac{1}{2}m & l+1\end{bmatrix}\begin{matrix}\xi\\ \\ \\ \end{matrix} \tag{6.5}$$

of Legendre's associated differential equation

$$(1-\xi^2)\frac{d^2\Theta}{d\xi^2}-2\frac{d\Theta}{d\xi}+\left(l(l+1)-\frac{m^2}{1-\xi^2}\right)\Theta=0, \tag{6.6}$$

where l and m attached to Θ are abbreviated.

Consider the ladder of P_t functions

$$\begin{aligned}\Theta_l^m(\xi)t^m&=[z(z-1)]^{m/2}t^m\mathrm{P}\begin{bmatrix}0 & 1 & \infty\\ 0 & 0 & -l+m\\ m & -m & l+m+1\end{bmatrix}\begin{matrix}z=\frac{1}{2}(1-\xi)\\ \\ \\ \end{matrix},\\ &=\mathfrak{f}(-l+m,l+m+1;m+1;z)\xi^m,\end{aligned} \tag{6.7}$$

where $\xi=[z(z-1)]^{1/2}t$. The l SO's of this ladder are seen to be obtainable from $\left[\begin{smallmatrix}\alpha\\ \beta\end{smallmatrix}\right]$ SO's (4.6)'. By the π transformation $\alpha=-l+m$, $\beta=l+m+1$, $\gamma=m+1$, they are expressed in terms of the E operators of Miller as follows:

$$\begin{aligned}\underset{(l)}{S}&=(\xi^2-1)\frac{\partial}{\partial\xi}+(l+1)\xi=P^-J^+ +P^3J^3+(l+1)P^3\\ \underset{(l)}{S}&=-(\xi^2-1)\frac{\partial}{\partial\xi}+l\xi=-P^-J^+-P^3J^3+lP^3,\end{aligned} \tag{6.8}$$

where J^3, $J^\pm$ are the A operators defined by

$$J^3=t\frac{\partial}{\partial t},\quad J^\pm=t^{\pm1}\left\{(z^2-1)^{1/2}\frac{\partial}{\partial z}\pm z(z^2-1)^{-1/2}t\frac{\partial}{\partial t}\right\},^{27} \tag{6.9}$$

acting on the ladder (6.5). The operators

$$P^3=\xi,\quad P^-=t^{-1}(\xi^2-1)^{1/2}, \tag{6.10}$$

and $P^+=t(\xi^2-1)^{1/2}$ constitute the E realization of

$\mathcal{T}_6 = \mathrm{sl}(2,\mathbf{C}) \oplus \mathbf{C}^3$, together with the A operators (6.9).[28]

Thus the SOP's of the second kind are expressed by the E operators of the Lie algebra $\mathcal{T}_6$.

B. The SOC's and the realizations of the Lie algebras $\mathcal{G}(1,0), \mathcal{G}(0,1), \mathcal{G}(0,0)$, and $\mathcal{T}_6$

One can see from the preceding subsection that the main point of discussion lies in the location of the index with respect to which we consider the SO's. The circumstances do not change for the ladders of CP functions as well. So we restrict ourselves to an outline of results.[25]

We enumerate the basis of CP function for each realization[16]:

The basis of the B realization:

$$f^m{}_l(z,t) = t^m \mathrm{P}\begin{bmatrix} \overbrace{\infty}^{} & 0 \\ -\frac{1}{2} \quad -m & -l \quad z \\ \frac{1}{2} \quad m & l+1 \end{bmatrix} = e^{-z/2} z^{-l}\, \mathrm{f}(-l-m;-2l;z) t^m. \tag{6.11}$$

The basis of the C′ realization:

$$f_m(z,t) = t^m \mathrm{P}\begin{bmatrix} \overbrace{\infty}^{} & 0 & \\ -\frac{1}{2} \quad -q & 0 & s=\frac{1}{4}z^2 \\ \frac{1}{2} \quad m+1 & -m+q & \end{bmatrix} = e^{-s/2} s^{-q/2}\, \mathrm{f}(-q;m-q+1;s)\xi^m, \quad \xi = s^{-q/2}t. \tag{6.12}$$

The basis of the D′ realization:

$$f_m(z,t) = e^{-s/2}\, \mathrm{f}(-\tfrac{1}{2}m;\tfrac{1}{2};s=\tfrac{1}{2}z^2) t^m. \tag{6.13}$$

The basis of the C″ realization:

$$f_m(z,t) = \mathrm{f}(\tfrac{1}{2}+m;1+2m;2iz)\xi^m, \quad \xi = 2izt. \tag{6.14}$$

The basis of the F realization:

$$f^m{}_l(z,t) = e^{-z/2} t^m\, \mathrm{f}(l-m;2l;z) z^l, \tag{6.15}$$

which is indexed by l.

By appropriate π and/or τ transformations, these basis functions (6.11)–(6.15) reduce to the extended form of the ladders (4.14), (4.15), (4.18), (4.16), and (4.17) respectively. The m SO's for the ladders CL_1, CL_2, CL_3, and CL_5 are obtainable from the SOC's (4.9), (4.10), (4.11), and (4.13) with the help of appropriate π and/or τ transformations respectively. The l SO's for the ladder (6.15) which associate with SOC's of the fourth kind (4.12) are expressed as

$$\overset{(l)}{S} = P^- J^+ + P^3 J^3 + (l+1)P^3 - [c/(l+1)]J^3 - cI,$$

$$\overset{(l)}{S} = -P^- J^+ - P^3 J^- + lP^3 - (c/l)J^3 + cI,$$

where the operators

$$P^- = -2ct^{-1}z^{-1}, \quad P^3 = 2cz^{-1}$$

and $P^+ = -2ctz^{-1}$ constitute the F realization of the Lie algebra $\mathcal{T}_6$ together with the B realization

$$J^\pm = t^{\pm 1}\left(z\frac{\partial}{\partial z} \pm t\frac{\partial}{\partial t} \mp \frac{1}{2}z\right), \quad J^3 = t\frac{\partial}{\partial t},$$

of subalgebra $\mathrm{sl}(2,\mathbf{C})$ of $\mathcal{T}_6$.[29]

Thus a direct correspondence has established between the types of SO's intrinsic to the respective classes of P or CP functions and the types of the realizations of the Lie algebras $\mathcal{G}(a,b)$ or $\mathcal{T}_6$.

ACKNOWLEDGMENTS

The author is sincerely grateful to Professor Teturo Inui who taught him to turn his attention to the Lie theory of special functions, especially the works of Willard Miller, Jr., and has given him continuous guidance and encouragement. He is very happy to express his sincere gratitude to Professor Willard Miller, Jr. of the University of Minnesota who read the manuscript and gave him kind advice and helpful corrections. He owes his present work entirely to their prominent achievements in the unified theory of special functions.

[1] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Vol. I, Chap. 5.
[2] T. Inui, *Special Functions* (Iwanami, Tokyo, 1962), Chap. 4, in Japanese.
[3] M. Hukuhara, *The Method of Solutions of Ordinary Differential Equations II* (Iwanami, Tokyo, 1941), in Japanese.
[4] E. Schrödinger, "A method of determining quantum mechanical eigenvalues and eigenfunctions," Proc. Roy. Irish Acad. A 46, 9 (1941).
[5] E. Schrödinger, Further studies on solving eigenvalue problems by factorization," Proc. Roy. Irish Acad. A 46, 183 (1941).
[6] E. Schrödinger, "The factorization of the hypergeometric equation," Proc. Roy. Irish Acad. A 46, 53 (1941).
[7] T. Inui, "Unified theory of recurrence formulas I," Progr. Theoret. Phys. (Kyoto) 3, 168 (1948).
[8] T. Inui, "Unified theory of recurrence formulas II," Progr. Theoret. Phys. (Kyoto) 3, 244 (1948).
[9] L. Infeld and T. E. Hull, "The factorization method," Rev. Mod. Phys. 23, 21 (1951).
[10] E. Whittaker, "On the recurrence formulae of the Mathieu functions," J. London Math. Soc. 4, 88 (1929).
[11] I. Marx, "Recurrence relations for prolate spheroidal wave functions," J. Math. & Phys. 32, 269 (1954).
[12] L. Weisner, "Group theoretic origin of certain generating functions," Pacific J. Math. 5, 1033 (1955).
[13] L. Weisner, "Generating functions for Hermite functions," Can. J. Math. 11, 141 (1959).
[14] L. Weisner, "Generating functions for Bessel functions," Can. J. Math. 11, 148 (1959).
[15] W. Miller, Jr., "On Lie algebras and some special functions of mathematical physics," AMS Mem. No. 50 (Am. Math. Soc., Providence, R.I., 1964).
[16] W. Miller, Jr., *Lie Theory and Special Functions* (Academic Press, New York and London, 1968).
[17] E. G. Kalnins and W. Miller, Jr., "Lie theory and the wave equation in space time II. The group SO(4,C)," SIAM J. Math. Anal. (to appear).
[18] E. G. Kalnins and W. Miller, Jr., "Lie theory and the wave equation in space–time. 3. Semisubgroup coordinates," J. Math. Phys. 18, 271 (1977).
[19] E. G. Kalnins and W. Miller, Jr., "Lie theory and the wave equation in space–time. 4. The Klein–Gordon equation and the Poincaré group," J. Math. Phys. 19 (1978).
[20] E. G. Kalnins and W. Miller, Jr., "Lie theory and the wave equation in space–time. 5. R-separable solutions of the wave equation," J. Math. Phys. 19 (1978) (to be published).

[21]E. G. Kalnins and W. Miller, Jr., "Orthogonal separable coordinates for the Laplace—Beltrami operator on the complex four-sphere" (preprint).
[22]M. Iwano, "Riemann's P-function," Lectures given at the University of Minnesota, Spring Quarter 1967 (Lecture notes University of Minnesota).
[23]M. Mori, "Morphology of Step-Operators I," Bull. Facul. Sci. Eng. Chuo Univ. 15, 15 (1972).
[24]M. Mori, "On the algebraic properties of Step-Operators resulting from a classification of P-functions," 90th Anniversary Bull. of Chuo Univ. 19 (1975).
[25]M. Mori, "On the algebraic properties of the Step-Operators resulting from a classification of confluent *P*-functions," Bull. Fac. Sci. Eng. Chuo Univ. 18, 79 (1975).
[26]Reference 16, p. 199.
[27]Reference 16, p. 195.
[28]Reference 16, p. 249.
[29]Reference 16, p. 253.

A Green's function for a cubic lattice[a)]

F. T. Hioe

Institute for Fundamental Studies, Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627
(Received 30 November 1977)

The following Green's function for a cubic lattice is evaluated exactly and expressed in terms of the complete elliptic integrals of the first kind:
$G(E)=(1/\pi^3)\int\int\int_0^\pi dxdydz/[E-a_1\cos x-a_2\cos y-a_3\cos z-a_2\cos y\cos z-a_1\cos z\cos x]$.

1. LATTICE GREEN'S FUNCTIONS FOR CUBIC LATTICES

The Green's functions for the cubic lattices are of interest in many physical problems and have been the subject of extensive studies for many years. The Green's functions for the ordinary three types of cubic lattices, the simple cubic (sc), the body-centered cubic (bcc), and the face-centered cubic (fcc), are defined by

$$G(E)=\frac{1}{\pi^3}\int_0^\pi\int_0^\pi\int_0^\pi \frac{dx\,dy\,dz}{E-\omega(x,y,z)}, \tag{1}$$

where

$$\omega(x,y,z)=\begin{cases}\cos x+\cos y+\cos z & \text{for the sc,} \quad (2)\\ \cos x\cos y\cos z & \text{for the bcc,} \quad (3)\\ \cos x\cos y+\cos y\cos z+\cos z\cos x & \text{for the fcc.} \quad (4)\end{cases}$$

The Green's functions for the above cases have been evaluated exactly and expressed in terms of the complete elliptic integrals of the first kind; the sc case for the first time by Joyce,[1] the bcc case by Maradudin *et al.*,[2] and the fcc case by Iwata.[3] In addition, the Green's functions for the following cases have also been expressed exactly in terms of the complete elliptic integrals of the first kind:

(i) $E=2+\alpha^2$, $\omega(x,y,z)=\cos x+\cos y+\alpha^2\cos z$ by Montroll,[4] (5)

(ii) $\omega(x,y,z)=(2+\alpha^2)^{-1}(\cos x\cos y+\cos y\cos z+\alpha^2\cos z\cos x)$ by Joyce,[5] (6)

and

(iii) $\omega(x,y,z)=\cos x+\cos y+\cos z+\cos x\cos y\cos z+\cos x\cos y+\cos y\cos z+\cos z\cos x$ by Glasser.[6] (7)

In this paper, we shall add to this list the following case:

$$\omega(x,y,z)=\cos x+\cos y+\cos z+\cos y\cos z+\cos z\cos x. \tag{8}$$

In fact we shall obtain an analytic result for the following more general case:

$$\omega(x,y,z)=a_1\cos x+a_2\cos y+a_3\cos z+a_2\cos y\cos z+a_1\cos z\cos x. \tag{9}$$

Consider the Green's function

$$G(E)=\frac{1}{\pi^3}\times\int_0^\pi\int_0^\pi\int_0^\pi \frac{dx\,dy\,dz}{E-\cos x-\cos y-\cos z-\cos y\cos z-\cos z\cos x}, \tag{10}$$

and let us begin by assuming that $E>5$. We have

$$G(E)=\frac{1}{\pi^3}\times\int_0^\pi\int_0^\pi\int_0^\pi \frac{dx\,dy\,dz}{(E-\cos x-\cos y)-(1+\cos x+\cos y)\cos z}$$
$$=\frac{1}{\pi^2}\int_0^\pi\int_0^\pi \frac{dx\,dy}{[(E-\cos x-\cos y)^2-(1+\cos x+\cos y)^2]^{1/2}}, \tag{11}$$

by using

$$\int_0^\pi \frac{dz}{a+b\cos z}=\frac{\pi}{(a^2-b^2)^{1/2}}, \quad a>|b|.$$

$$\therefore G(E)=\frac{1}{\pi^2}\times\int_0^\pi\int \frac{dx\,dy}{[(E^2-1-2E\cos x-2\cos x)-2(E+1)\cos y]^{1/2}}$$
$$=\frac{1}{\pi^2}\int_0^\pi \frac{2}{[(E+1)^2-2(E+1)\cos x]^{1/2}}K\times\left(\frac{2}{(E+1-2\cos x)^{1/2}}\right)dx, \tag{12}$$

by using

$$\int_0^\pi \frac{dy}{(a-b\cos y)^{1/2}}=\frac{2}{(a+b)^{1/2}}K\times\left[\left(\frac{2b}{a+b}\right)^{1/2}\right], \quad a>b,$$

where $K(k)$ is the complete elliptic integral of the first kind. Thus

$$G(E)=\frac{2}{\pi^2(E+1)^{1/2}}\int_0^\pi \frac{1}{[(E+1)-2\cos x]^{1/2}}K\times\left(\frac{2}{(E+1-2\cos x)^{1/2}}\right)dx. \tag{13}$$

[a)]Research partially supported by the Army Research Grant No. DAAG29-77-G-0060.

 0022-2488/78/1905-1064$1.00

Expanding

$$K(k)=\tfrac{1}{2}\pi F(\tfrac{1}{2},\tfrac{1}{2};1;k^2)=\tfrac{1}{2}\pi\sum_{n=0}^{\infty}\frac{(\frac{1}{2})_n(\frac{1}{2})_n}{(1)_n(1)_n}\,k^{2n},$$

where $F(\alpha,\beta;\gamma;z)$ is the Gauss hypergeometric function, we get

$$G(E)=\frac{2}{\pi^2(E+1)^{1/2}}\,\frac{\pi}{2}\sum_{n=0}^{\infty}\frac{(\frac{1}{2})_n(\frac{1}{2})_n}{(1)_n(1)_n}\,4^n \times\int_0^{\pi}\frac{dx}{(E+1-2\cos x)^{n+1/2}}. \tag{14}$$

Now

$$\int_0^{\pi}\frac{dx}{(E+1-2\cos x)^{n+1/2}}=\int_0^{\pi/2}\frac{2d\theta}{(E+3-4\cos^2\theta)^{n+1/2}} =\frac{\pi}{(E+3)^{n+1/2}}F\left(\tfrac{1}{2},n+\tfrac{1}{2};1;\frac{4}{E+3}\right) \tag{15}$$

from the formula[7]

$$\int_0^{\pi/2}\frac{\sin^{\mu}x\cos^{\nu}x}{(a-b\cos^2x)^{\rho}}\,dx =\frac{1}{2a^{\rho}}B\left(\frac{\mu+1}{2},\frac{\nu+1}{2}\right)F\left(\frac{\nu+1}{2},\rho;\frac{\mu+\nu}{2}+1;\frac{b}{a}\right),$$
$$\mathrm{Re}\,\mu>-1,\ \mathrm{Re}\,\nu>-1,\ a>|b|\geq 0.$$

Substituting (15) into (14) and expanding out the Gauss hypergeometric function F, we get

$$\begin{aligned}G(E)&=\frac{1}{[(E+1)(E+3)]^{1/2}}\sum_{n=0}^{\infty}\frac{(\frac{1}{2})_n(\frac{1}{2})_n}{(1)_n(1)_n}\left(\frac{4}{E+3}\right)^n \times\sum_{m=0}^{\infty}\frac{(\frac{1}{2})_m(n+\frac{1}{2})_m}{(1)_m(1)_m}\left(\frac{4}{E+3}\right)^m\\ &=\frac{1}{[(E+1)(E+3)]^{1/2}}\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\frac{(\frac{1}{2})_{m+n}(\frac{1}{2})_m(\frac{1}{2})_n}{(1)_m(1)_n(1)_m(1)_n} \times\left(\frac{4}{E+3}\right)^m\left(\frac{4}{E+3}\right)^n\\ &=\frac{1}{[(E+1)(E+3)]^{1/2}}F_2\left(\tfrac{1}{2};\tfrac{1}{2},\tfrac{1}{2};1,1;\frac{4}{E+3},\frac{4}{E+3}\right),\end{aligned} \tag{16}$$

where

$$F_2(\alpha;\beta,\beta';\gamma,\gamma';x,y)=\sum_m\sum_n\frac{(\alpha)_{m+n}(\beta)_m(\beta')_n}{(\gamma)_m(\gamma')_n m!n!}\,x^my^n,$$

is the second hypergeometric series in two variables of Appell.[8]

It is known that[9]

$$F_2\left(\alpha,\alpha+\tfrac{1}{2}-\beta,\beta,\gamma,2\beta,\frac{x}{(1+\sqrt{y})^2},\frac{4\sqrt{y}}{(1+\sqrt{y})^2}\right) =(1+\sqrt{y})^{2\alpha}F_4(\alpha,\alpha+\tfrac{1}{2}-\beta,\gamma,\beta+\tfrac{1}{2},x,y), \tag{17}$$

where

$$F_4(\alpha,\beta,\gamma,\gamma',x,y)=\sum_m\sum_n\frac{(\alpha)_{m+n}(\beta)_{m+n}}{(\gamma)_m(\gamma')_n m!n!}\,x^my^n,$$

is the fourth Appell series; and according to a theorem of Bailey[10]

$$F_4(\alpha,\gamma+\gamma'-\alpha-1,\gamma,\gamma';z(1-Z),Z(1-z)) =F(\alpha,\gamma+\gamma'-\alpha-1;\gamma;z)F(\alpha,\gamma+\gamma'-\alpha-1;\gamma';Z), \tag{18}$$

i.e., F_4 is the product of two Gauss hypergeometric functions when $\gamma+\gamma'=\alpha+\beta+1$. Thus let us do the following manipulation:

Firstly, from (16), we deduce that since F_2 converges only if $|x|+|y|<1$, we must have $E>5$. Secondly, to transform F_2 to F_4 according (17), we let

$$\frac{4}{E+3}=\frac{x}{(1+\sqrt{y})^2},\quad \frac{4}{E+3}=\frac{4\sqrt{y}}{(1+\sqrt{y})^2}$$

and solve for x and y. We find, remembering that $F_4(\alpha,\beta,\gamma,\gamma',x,y)$ converges only if $x^{1/2}+y^{1/2}<1$, the acceptable solution is

$$x=2\{(E+1)-[(E+1)^2-4]^{1/2}\},\quad y=\tfrac{1}{16}x^2. \tag{19}$$

Thus we find

$$F_2\left(\tfrac{1}{2};\tfrac{1}{2},\tfrac{1}{2};1,1;\frac{4}{E+3},\frac{4}{E+3}\right)=(1+\sqrt{y}) \times F_4(\tfrac{1}{2},\tfrac{1}{2},1,1,x,y), \tag{20}$$

where x and y are given by (19). Thirdly, to transform F_4 into the product of two F's according to (18), we let

$$x=z(1-Z),\quad y=Z(1-z),$$

and solve for z and Z. We find

$$z=\frac{1-x+y-[(1+x-y)^2-4x]^{1/2}}{2},\qquad Z=\frac{1+x-y-[(1+x-y)^2-4x]^{1/2}}{2}. \tag{21}$$

Thus we get

$$F_4(\tfrac{1}{2},\tfrac{1}{2},1,1,x,y)=F(\tfrac{1}{2},\tfrac{1}{2};1;z)F(\tfrac{1}{2},\tfrac{1}{2};1;Z) =(4/\pi^2)K(k_-)K(k_+), \tag{22}$$

where

$$k_{\mp}^2=\tfrac{1}{2}\{1\mp(x-y)-[(1+x-y)^2-4x]^{1/2}\}. \tag{23}$$

Thus finally we get, for $G(E)$ of Eq. (10),

$$G(E)=\frac{4}{\pi^2[(E+1)(E+3)]^{1/2}}(1+\sqrt{y})K(k_-)K(k_+),\quad E>5, \tag{24}$$

where k_-, k_+ and x,y are given by (23) and (19).

Using the property that $K'(\sqrt{2}-1)=\sqrt{2}K(\sqrt{2}-1)$, we find

$$G(5)=\frac{4(\sqrt{2}-1)}{\pi^2\sqrt{3}}K^2(\sqrt{2}-1), \tag{24'}$$

which complements Watson's results[11] for the three cubic lattices with the usual nearest neighbors consideration.

The analytic continuation to the case $E<5$ can be done in a similar way as that for other cases,[12] and it results in $G(E)$ being expressed in terms of the product of two $K(k)$'s with complex moduli. We have not been able to express the real and imaginary parts of $G(E)$ separately in terms of the complete elliptic integrals of real moduli.

It can be shown in the same way that the following Green's function which is a generalization of (10) is given by

$$G(E)=\frac{1}{\pi^3}\int_0^\pi\int_0^\pi\int_0^\pi dx\,dy\,dz\ (E-a_1\cos x-a_2\cos y-a_3\cos z-a_2\cos y\cos z-a_1\cos z\cos x)^{-1} \tag{25}$$

$$=\frac{4(1+\sqrt{y})}{\pi^2[(E+a_3)(E-a_3+2a_1+2a_2)]^{1/2}}\times K(k_-)K(k_+),\quad E>a_3+2a_1+2a_2, \tag{26}$$

where

$$k_\mp^2=\tfrac{1}{2}\{1\mp(x-y)-[(1+x-y)^2-4x]^{1/2}\}, \tag{27}$$

$$x=(4\alpha/\beta^2)[2-\beta-2(1-\beta)^{1/2}], \tag{28}$$

$$y=(1/\beta^2)[2-\beta-2(1-\beta)^{1/2}]^2,$$

$$\alpha=4a_2/(E-a_3+2a_1+2a_2), \tag{29}$$

$$\beta=4a_1/(E-a_3+2a_1+2a_2),$$

where a_1, a_2, a_3 are positive numbers, and where we assumed, without loss of generality, $a_1\neq 0$. It is easy to check that when $a_1=a_2=a_3=1$, we get back the result given by (24), while if $a_1=a_3=1$, $a_2=0$, remembering that

$$(1+k)K(k)=K(2k^{1/2}/(1+k)),$$

we get

$$G(E)=\frac{2}{\pi(E+1)}K\left(\frac{2}{(E+1)^{1/2}}\right),\quad E>3,$$

which can be verified independently by directly evaluating the integral

$$\frac{1}{\pi^2}\int_0^\pi\int\frac{dx\,dy}{E-\cos x-\cos y-\cos x\cos y}.$$

Similarly, it can be verified that when $a_1=1$, $a_2=a_3=0$, we get

$$G(E)=\frac{2}{\pi[E(E+2)]^{1/2}}K\left(\frac{2}{(E+2)^{1/2}}\right),\quad E>2.$$

The particular value of $G(E)$ in (25) for $E=2a_1+2a_2+a_3$ is given by

$$G(2a_1+2a_2+a_3)=\frac{\sqrt{2}\{a_1^2+a_1a_2+2a_2^2-2a_2[a_2(a_1+a_2)]^{1/2}\}}{\pi^2a_1^2[(a_1+a_2)(a_1+a_2+a_3)]^{1/2}}\times K\left(\frac{(a_1+a_2)^{1/2}-a_2^{1/2}}{a_1^{1/2}}\right)\times K'\left(\frac{(a_1+a_2)^{1/2}-a_2^{1/2}}{a_1^{1/2}}\right), \tag{29'}$$

where $K'(k)=K((1-k^2)^{1/2})$. Equation (29′) above is particularly illuminating in exhibiting the way $G(2a_1+2a_2+a_3)$ diverges as $a_2\to 0$, namely as the transition from three dimensions to two dimensions takes place at the "critical" value of $E=2a_1+2a_2+a_3$. This is relevant to understanding the transition which takes place from the two-dimensional spherical model of a ferromagnet[13] which does not exhibit phase transitions to the three-dimensional model which does; and it is also relevant to understanding the transition from a random walker on a two-dimensional lattice who is certain to return to the origin to one on a three-dimensional lattice who has a certain nonzero probability of not returning to the origin.[14]

2. THE GREEN's FUNCTION OF GLASSER IN THE RANGE $-1<E<7$

The following Green's function has been evaluated exactly by Glasser[6]:

$$G(E)=\frac{1}{\pi^3}\int_0^\pi\int_0^\pi\int_0^\pi dx\,dy\,dz\ (E-\cos x-\cos y-\cos z-\cos x\cos y\cos z-\cos x\cos y-\cos y\cos z-\cos z\cos x)^{-1}. \tag{30}$$

Glasser obtained the following result[15] for $E>7$:

$$G(E)=\frac{4}{\pi^2(E+1)}K(k)^2, \tag{31}$$

where

$$k^2=\frac{1}{2}\left[1-\left(\frac{E-7}{E+1}\right)^{1/2}\right]. \tag{32}$$

If we replace E in (30) by $E-i\epsilon$ and take the limit $\epsilon\to 0$, and let

$$\lim_{\epsilon\to 0+}G(E-i\epsilon)=G_R(E)+iG_I(E), \tag{33}$$

the real and imaginary parts G_R and G_I for the case $-1<E<7$ can be expressed, as we show in this section, in terms of the complete elliptic integrals of real moduli. The only other case for the cubic lattices where this can be done was given by Joyce[16] for the body-centered cubic lattice (for $-1<E<1$).

Let us write k^2 in (32) as

$$k^2=\tfrac{1}{2}[1-(1-\xi^{-1})^{1/2}], \tag{34}$$

where

$$\xi=(E+1)/8. \tag{35}$$

We note that $\xi>1$ and $k^2<\frac{1}{2}$. Kummer's relation gives us[17]

$$F(\tfrac{1}{2},\tfrac{1}{2};1;k^2)=F(\tfrac{1}{4},\tfrac{1}{4};1;\xi^{-1}). \tag{36}$$

The following transformation gives us the analytic continuation to the case $\xi<1$:

$$F(\tfrac{1}{4},\tfrac{1}{4};1;\xi^{-1})=(\xi)^{1/4}\times\left(\frac{\Gamma(\frac{1}{4})^2}{2\pi^{3/2}}F(\tfrac{1}{4},\tfrac{1}{4};\tfrac{1}{2};1-\xi)-\frac{2\pi^{1/2}}{\Gamma(\frac{1}{4})^2}(\xi-1)^{1/2}F(\tfrac{3}{4},\tfrac{3}{4};\tfrac{3}{2};1-\xi)\right),\quad |\arg\xi|<\pi, \tag{37}$$

$$=\tfrac{1}{2}\xi^{1/4}\{F(\tfrac{1}{2},\tfrac{1}{2};1;k_+^2)+F(\tfrac{1}{2},\tfrac{1}{2};1;k_-^2)+i[F(\tfrac{1}{2},\tfrac{1}{2};1;k_+^2)-F(\tfrac{1}{2},\tfrac{1}{2};1;k_-^2)]\}, \tag{38}$$

where

$$k_\pm^2 = \tfrac{1}{2} \pm \tfrac{1}{2}(1-\xi)^{1/2} = \tfrac{1}{2}\{1 \pm [(7-E)/8]^{1/2}\}, \quad -1 < E < 7. \tag{39}$$

Equation (38) was obtained from Eq. (37) using relations no. (7) and (9) in Ref. 8, p. 111, and noting the relation

$$(z-1)^{1/2} = -i(1-z)^{1/2} \quad \text{if } z < 1. \tag{40}$$

Substituting (38) into (31) gives us the G_R and G_I in (33) separately in terms of the complete elliptic integrals of the first kind with real moduli:

$$\left.\begin{aligned} G_R(E) &= [\sqrt{2}/\pi^2(E+1)^{1/2}]K(k_+)K(k_-) \\ G_I(E) &= [\sqrt{2}/2\pi^2(E+1)^{1/2}][K(k_+)^2 - K(k_-)^2] \end{aligned}\right\}, \quad -1 < E < 7, \tag{41}$$

where $k_\pm$ are given by (39).

The Green's function $G_I(E)$ is directly related to the density of states in lattice dynamics,[4] in the tight-binding approximation for electrons,[18] and in many other theories of excitations in solids.

ACKNOWLEDGMENT

I am very grateful to Professor Elliott Montroll for stimulating my interest in this research.

Note added in manuscript: Professor Larry Glasser kindly told me that $K(\sqrt{2}-1)$ can be expressed in terms of the gamma functions as

$$K(\sqrt{2}-1) = (\sqrt{2}+1)^{1/2}\,\frac{\Gamma(\frac{1}{8})\Gamma(\frac{3}{8})}{2^{13/4}\pi^{1/2}}.$$

Thus, $G(5)$ of Eq. (24′) can be expressed as

$$G(5) = \frac{\{\Gamma(\frac{1}{8})\Gamma(\frac{3}{8})\}^2}{16\sqrt{6}\pi^3} = 0.262455579.$$

For the question of when $K(k)$ can be expressed in terms of gamma functions, see the interesting paper of Glasser and Zucker, Proc. Natl. Acad. Sci. (USA) **74**, 1800 (1977) {two misprints in this paper: the lower integration limits in Eq. [8(a)] should be $-\pi$, and Eq. [8(b)] should read

$$k^2 = \frac{1}{2}\left[1 - \left(\frac{z-7}{z+1}\right)^{1/2}\right]\Bigg\}.$$

[1]G.S. Joyce, Trans. Roy. Soc. A **273**, 46 (1973).
[2]A.A. Maradudin, E.W. Montroll, G.H. Weiss, R. Herman, and W.H. Miles, Acad. Roy. de Belgique, Bruxelles (1960).
[3]G. Iwata, Natural Science Report (Ochanomizu Univ.) **20**, No. 2, p. 13 (1969).
[4]E.W. Montroll, Proc. 3rd Berkeley Symp. **3**, 209 (1956).
[5]G.S. Joyce, J. Phys. C **4**, L53 (1971).
[6]M.L. Glasser, J. Math. Phys. **13**, 1145 (1972).
[7]I.S. Gradshteyn and I.M. Ryzhik, *Tables of Integrals, Series, and Products* (Academic, New York, 1965), p. 389.
[8]Bateman Manuscript Project, *Higher Transcendental Functions, Vol. I*, edited by A. Erdélyi *et al.* (McGraw-Hill, New York, 1953), p. 224.
[9]P. Appell and J. Kampé de Fériet, *Fonctions hypergéométriques et hypersphériques* (Gauthier-Villars, Paris, 1926), p. 27.
[10]Ref. 8, p. 238.
[11]G.N. Watson, Quart, J. Math. (Oxford) **10**, 266 (1939).
[12]T. Morita and T. Horiguchi, J. Math. Phys. **12**, 986 (1971).
[13]T.H. Berlin and M. Kac, Phys. Rev. **86**, 821 (1952).
[14]C. Domb, Proc. Camb. Phil. Soc. **50**, 589 (1954); E.W. Montroll, Proc. Symp. Appl. Math. Am. Math. Soc. **16**, 193 (1964).
[15]Eq. (6) in Ref. 6 is correct while the right-hand side of Eq. (7) should be divided by 4.
[16]G.S. Joyce, J. Math. Phys. **12**, 1390 (1971).
[17]Ref. 8, p. 111, No. (2).
[18]R.J. Jelitto, J. Phys. Chem. Solids **30**, 609 (1969); E.W. Montroll, J. Math. Phys. **11**, 635 (1970).

Three-dimensional solitons[a)]

Roger G. Newton

Physics Department, Indiana University, Bloomington, Indiana 47401
(Received 29 August 1977)

Using an inverse Radon transform we generate an integro–differential evolution equation in three space dimensions that has soliton solutions which vanish at large distances in all directions. The equation is of second order in time and invariant under translations and rotations. The shapes of the solitons are generally changed by their nonlocal interactions, but their number and velocities are conserved. The method lends itself to other equations.

1. INTRODUCTION

The existence of nonlinear evolution equations with soliton solutions is of great physical interest and has many mathematically fascinating aspects.[1,2] These equations have been more or less confined to one spatial dimension and none but a few examples of two- and three-dimensional equations have been found,[3] whose solitons, moreover, always extend to infinity in some directions. If such equations are to have a significant bearing on the physics of elementary particles it would seem to be important to find evolution equations in three space dimensions which have not merely solitary-wave solutions that vanish at large distances in all directions but produce *solitons* with their remarkable stability properties under collisions.[4] Such equations have not yet been found.

Since there appear to be major difficulties in discovering soliton generating differential equations in three space dimensions, it seems worthwhile to search for more general equations with such solutions. A suitable class may be that of integrodifferential equations, in which one may think of the soliton–soliton interaction as nonlocal. The present paper represents an attempt in that direction. From a known one-dimensional evolution equation we generate a three-dimensional one that is not restricted to a line. The equation is rotationally invariant and the asymptotic directions of motion of the solitons are determined by the initial conditions. Our method is, in principle, applicable to other evolution equations and to any dimension of space, but in this paper we restrict ourselves to one three-dimensional equation.

The tool we will use is the *Radon transform*.[5] Since this integral transform is not widely known among physicists we define it and derive the results we need in Sec. 2. In Sec. 3 we apply its inverse to the iterated Korteweg–de Vries equation. This equation is of second order in the time and does not have the unidirectional character of the KdV equation. Its inverse Radon transform is an integrodifferential equation that is rotationally and translationally invariant. We show that it has N-soliton solutions in which the solitons vanish asymptotically in all spatial direction and move with largely arbitrary (both in magnitude and direction) velocities. Their collisions preserve their number and velocities but, in general, will alter their shapes.

[a)]This material is based upon work supported by the National Science Foundation under Grant No. PHY 75-18942 A01.

Whether the three-dimensional soliton-generating equation (3.6) has any physical application is not known. This question may depend on whether there are physical phenomena that are describable by its solutions, and it remains to be investigated. The method used here can easily be transferred to other equations to generate intergrodifferential equations in higher dimensions with soliton solutions from one-dimensional differential equations that are known to have them. It does not work for the nonlinear Schrödinger equation, because there the one-dimensional solitons are not really solitary waves.[6] While their magnitudes are of the traveling-wave form, their complex phases oscillate. As a result the inverse Radon transform washes them out in the asymptotic region and they disappear.

There is an appendix which contains a direct verification, without use of the inverse-scattering machinery, of the pure N-soliton solution of the KdV equation.

2. THE RADON TRANSFORM

Let us consider two functions, $f(\hat{\mathbf{n}},x)$ and $\hat{f}(\mathbf{z})$, where $x\in\mathbb{R}^1$, $\mathbf{z}\in\mathbb{R}^3$, and $\hat{n}\in\mathbb{R}^3$ with $|\hat{\mathbf{n}}|=1$. We shall not, in this paper, specify exactly in what spaces f and $\hat{f}$ should be. (The reader may consult Ref. 5.) Let $\tilde{f}(\mathbf{k})$ be the three-dimensional Fourier transform of $\hat{f}(\mathbf{z})$:

$$\tilde{f}(\mathbf{k})=(F_3\hat{f})(\mathbf{k})=\int d^3\mathbf{z}\exp(i\mathbf{k}\cdot\mathbf{z})\hat{f}(\mathbf{z}), \tag{2.1}$$

and let $\mathring{f}(\hat{\mathbf{n}},k)$ be the one-dimensional Fourier transform of $f(\hat{\mathbf{n}},x)$ as a function of x for fixed $\hat{\mathbf{n}}$:

$$\mathring{f}(\hat{\mathbf{n}},k)=(F_1 f)(\hat{\mathbf{n}},k)=\int_{-\infty}^{\infty}dx\exp(ikx)f(\hat{\mathbf{n}},x). \tag{2.2}$$

We consider $\hat{f}(\mathbf{z})$ to be the R transform of $f(\hat{\mathbf{n}},x)$,

$$\hat{f}(\mathbf{z})=(Rf)(\mathbf{z}) \tag{2.3}$$

if

$$\tilde{f}(\mathbf{k})=\mathring{f}(\hat{\mathbf{k}},k), \tag{2.4}$$

where $\mathbf{k}=\hat{\mathbf{k}}k$. For consistency it will be necessary to require that $f(\hat{\mathbf{n}},x)$ have the symmetry property

$$f(-\hat{\mathbf{n}},-x)=f(\hat{\mathbf{n}},x) \tag{2.5}$$

so that

$$\mathring{f}(-\hat{\mathbf{n}},-k)=\mathring{f}(\hat{\mathbf{n}},k). \tag{2.6}$$

It should be noted that if we required $\tilde{f}(\mathbf{k})$ to be analytic as a function of each component of $\mathbf{k}$ at $k=0$ then (2.4) would imply a specific $\hat{\mathbf{n}}$-dependence of $\mathring{f}(\hat{\mathbf{n}},0)$

 0022-2488/78/1905-1068$1.00

and of each derivative of $\mathring{f}(\hat{n},k)$ with respect to k at $k=0$.[7] For example it would imply that

$$\mathring{f}(\hat{n},0)=\int_{-\infty}^{\infty}dx\,f(\hat{n},x)$$

be independent of $\hat{n}$. More generally, it would require that

$$\int d\hat{n}\,Y_l^m(\hat{n})\int_{-\infty}^{\infty}dx\,x^p f(\hat{n},x)=0$$

for all $l>p$, if Y_l^m is a spherical harmonic. *We shall however, make no such demands* on $f(\hat{n},x)$ and as a result we end up with more general functions $\tilde{f}(\mathbf{k})$, and hence a larger class of functions $\hat{f}(\mathbf{z})$.

Insertion of (2.1) and use of (2.4) leads to the direct relation between $\hat{f}(\mathbf{z})$ and $f(\hat{n},x)$,

$$\begin{aligned}(Rf)(\mathbf{z})=\hat{f}(z)&=\frac{1}{(2\pi)^3}\int d\hat{\mathbf{k}}\int_0^{\infty}dk\,k^2\exp(-ik\hat{\mathbf{k}}\cdot\mathbf{z})\mathring{f}(\hat{\mathbf{k}},k)\\&=-\Delta\,\frac{1}{16\pi^3}\int d\hat{\mathbf{k}}\int_{-\infty}^{\infty}dk\exp(-ik\hat{\mathbf{k}}\cdot\mathbf{z})\mathring{f}(\hat{\mathbf{k}},k)\\&=-\frac{1}{8\pi^2}\int d\hat{n}\,f''(\hat{n},\hat{n}\cdot\mathbf{z}),\end{aligned}\tag{2.7}$$

because of (2.6) and inversion of (2.2). Here $f''(\hat{n},x)\equiv\partial^2 f(\hat{n},x)/\partial x^2$. The relation R between $f(\hat{n},x)$ and $f(\mathbf{z})$ expressed by (2.7) is the *inverse Radon transform*. We shall refer to it as the R transform.

We can also formally express $f(\hat{n},x)$ in terms of $\hat{f}(\mathbf{z})$ and thereby invert the R transform. By inverting (2.2) and using (2.1) we obtain

$$f(\hat{n},x)=(R^{-1}\hat{f})(\hat{n},x)=\int d^3\mathbf{z}\,f(\mathbf{z})\delta(x-\hat{n}\circ\mathbf{z}).\tag{2.8}$$

This is the Radon transform. [However, it has to be realized that this transform is not uniquely defined,[5] and the result of (2.8) is not necessarily equal to the function $f(\hat{n},x)$ on the right-hand side of (2.7).]

The important properties of R under differentiation immediately follow from (2.4) or (2.7),

$$\nabla\hat{f}(\mathbf{z})=-\frac{1}{8\pi^2}\int d\hat{n}\,f'''(\hat{n},\hat{n}\cdot\mathbf{z})\hat{n},$$

which means

$$R\left(\hat{n}\,\frac{\partial f}{\partial x}\right)=\nabla(Rf),\tag{2.9}$$

and, by repetition,

$$R\left(\frac{\partial^2 f}{\partial x^2}\right)=\Delta(Rf).\tag{2.10}$$

Let us now look at the R transform of a product. Using (2.4) we get

$$R(fg)(\mathbf{z})=\int d^3\mathbf{x}d^3\mathbf{y}\,\Gamma(\mathbf{z},\mathbf{x},\mathbf{y})f(\mathbf{x})g(\mathbf{y}),\tag{2.11}$$

where

$$\begin{aligned}\Gamma(\mathbf{z},\mathbf{x},\mathbf{y})&=(2\pi)^{-4}\int d^3\mathbf{k}\int_{-\infty}^{\infty}d\alpha\exp[i\alpha\hat{\mathbf{k}}\circ(\mathbf{x}-\mathbf{y})\\&\qquad+i\mathbf{k}\circ(\mathbf{y}-\mathbf{z})]\\&=-\Delta_z\gamma(\mathbf{x},\mathbf{y},\mathbf{z}),\end{aligned}\tag{2.12}$$

where Δ_z is the Laplacian with respect to $\mathbf{z}$, and

$$\begin{aligned}\gamma(\mathbf{x},\mathbf{y},\mathbf{z})&=\tfrac{1}{2}(2\pi)^{-4}\int d\hat{\mathbf{k}}\int_{-\infty}^{\infty}d\alpha\int_{-\infty}^{\infty}d\beta\exp[i\alpha\hat{\mathbf{k}}\circ(\mathbf{x}-\mathbf{y})]\\&\qquad\times\exp[(i\beta\hat{\mathbf{k}}\circ(\mathbf{y}-\mathbf{z})]\\&=\frac{1}{8\pi^2}\int d\hat{\mathbf{k}}\,\delta[\hat{\mathbf{k}}\cdot(\mathbf{x}-\mathbf{y})]\delta[\hat{\mathbf{k}}\cdot(\mathbf{y}-\mathbf{z})]\\&=\frac{1}{4\pi^2}\,\frac{1}{|(\mathbf{x}-\mathbf{y})\times(\mathbf{y}-\mathbf{z})|}\\&=\frac{1}{4\pi^2}\,\frac{1}{|(\mathbf{x}\times\mathbf{y}+\mathbf{y}\times\mathbf{z}+\mathbf{z}\times\mathbf{x}|}.\end{aligned}\tag{2.13}$$

Thus γ is symmetric in all three variables,

$$\gamma(\mathbf{x},\mathbf{y},\mathbf{z})=\gamma(\mathbf{x},\mathbf{z},\mathbf{y})=\gamma(\mathbf{y},\mathbf{x},\mathbf{z}).\tag{2.14}$$

Furthermore it is invariant under simultaneous translation, rotation, or reflection of all three variables.

If we define the symbol $\circ$ by

$$\begin{aligned}(f\circ g)(\mathbf{x})&\equiv-\Delta\int d^3\mathbf{y}d^3\mathbf{z}\,f(\mathbf{y})g(\mathbf{z})\gamma(\mathbf{x},\mathbf{y},\mathbf{z})\\&=-\Delta\int d^3\mathbf{y}d^3\mathbf{z}\,\frac{f(\mathbf{y}+\mathbf{x})g(\mathbf{z}+\mathbf{x})}{|\mathbf{y}\times\mathbf{z}|},\end{aligned}\tag{2.15}$$

then we can write our result (2.11) in the simple form

$$R(fg)=(Rf)\circ(Rg).\tag{2.16}$$

Since the left-hand product is commutative and associative, so is the $\circ$-product.

In Fourier-transform language we have

$$F_3(f\circ g)(\mathbf{k})=(2\pi)^{-1}\int_{-\infty}^{\infty}d\alpha\,\tilde{f}(\alpha\mathbf{k})\tilde{g}[(|\mathbf{k}|-\alpha)\hat{\mathbf{k}}]\tag{2.17}$$

if $\tilde{f}=F_3 f$ and $\tilde{g}=F_3 g$.

Consider now a family of functions $f(\hat{n},x,t)$ that form solitary waves[8] traveling along x with velocities c which depend on $\hat{n}$ in such a way that there exists a fixed vector $\mathbf{v}$ so that $c=\hat{n}\circ\mathbf{v}$,

$$f(\hat{n},x,t)=f(\hat{n},x-\hat{n}\circ\mathbf{v}t,0).\tag{2.18}$$

If $f(\hat{n},x,0)$ satisfies the symmetry (2.5), then so does $f(\hat{n},x,t)$. The R transform of the family is then given by

$$\hat{f}(\mathbf{z},t)=-\frac{1}{8\pi^2}\int d\hat{n}\,f''(\hat{n},\hat{n}\circ\mathbf{z}-\hat{n}\circ\mathbf{v}t,0)=\hat{f}(\mathbf{z}-\mathbf{v}t,0)$$

and hence $\hat{f}(\mathbf{z},t)$ forms a solitary wave moving with the velocity $\mathbf{v}$. What is more, if for all $\hat{n}$

$$\lim_{x\to\pm\infty}f''(\hat{n},x,0)=0,\tag{2.19}$$

then it follows that

$$\lim_{|\mathbf{z}|\to\infty}\hat{f}(\mathbf{z},0)=0.\tag{2.20}$$

This is easily proved by writing

$$\begin{aligned}g(\mathbf{z})&=\int d\hat{n}\,f''(\hat{n},\hat{n}\cdot\mathbf{z},0)=2\int_0^{2\pi}d\phi\int_0^1 d\alpha\,f''(\alpha,\phi,|\mathbf{z}|\alpha,0)\\&=\int_0^1 d\alpha\,h(\alpha,|\mathbf{z}|\alpha)=\left(\int_0^a+\int_a^1\right)d\alpha\,h(\alpha,|\mathbf{z}|\alpha)\end{aligned}$$

by (2.5) and setting

$$2\int_0^{2\pi} d\phi f''(\alpha,\phi,|\mathbf{z}|\alpha,0)\equiv h(\alpha,|\mathbf{z}|\alpha).$$

For any given $\epsilon>0$ we choose a so that

$$\left|\int_0^a d\alpha h\right|<\tfrac{1}{2}\epsilon$$

and then, using (2.19), we choose R so that for all $|\mathbf{z}|>R$

$$\left|\int_a^1 d\alpha h\right|<\tfrac{1}{2}\epsilon.$$

Thus (2.20) follows. Therefore, if $f(\hat{\mathbf{n}},x,t)$ is a family of solitary waves for which not only f but also f'' vanishes for large $|x|$ (and each $\hat{\mathbf{n}}$), then $f(\mathbf{z},t)$ is a solitary wave that vanishes for large $|\mathbf{z}|$ in all directions.

3. A NONLINEAR EVOLUTION EQUATION

The KdV equation

$$\frac{\partial u}{\partial t}=6u\frac{\partial u}{\partial x}+\frac{\partial^3 u}{\partial x^3} \tag{3.1}$$

has soliton solutions which move from right to left with velocities c_n that are related to the (discrete) eigenvalues of the associated one-dimensional Schrödinger equation

$$\frac{\partial^2\psi_n}{\partial x^2}+u\psi_n=\lambda_n\psi_n \tag{3.2}$$

by $c_n=4\lambda_n$. The undirectional nature of the motion of the solitons of (3.1) is the result of the odd character of the KdV equation, and the fact that the discrete eigenvalues of (3.2) are all nonnegative. Before we transfer (3.1) to three dimensions it will be useful to remove its unidirectional character by iteration.

Differentiating (3.1) with respect to t and eliminating first t-derivatives by means of (3.1) yields

$$\frac{\partial^2 u}{\partial t^2}=\frac{\partial^2}{\partial x^2}\left(12u^3+9u\frac{\partial^2 u}{\partial x^2}+\frac{3}{2}\frac{\partial^2 u^2}{\partial x^2}+\frac{\partial^4 u}{\partial x^4}\right). \tag{3.3}$$

We may call this the iterated Korteweg–deVries (IKdV) equation. Any solution of (3.1) or of the equation obtained from (3.1) by changing the sign of t, is a solution of (3.3). Thus (3.3) has soliton solutions in which all solitons move to the left, and also soliton solutions in which they all move to the right. Whether it has solutions in which some solitons move to the right and some to the left is unknown. [A pure N-soliton solution of (3.1) in which some but not all of the velocities are reversed does not solve (3.3); see the Appendix.]

We translate the IKdV equation (3.3) into three dimensions by considering $u(x,t)$ as a function of two angle parameters, or a unit vector $\hat{\mathbf{n}}$, is such a way that

$$u(\hat{\mathbf{n}},x,t)=u(-\hat{\mathbf{n}},-x;t), \tag{3.4}$$

and then subjecting it to the R transformation (2.7),

$$\Phi(\mathbf{z},t)=(Ru)(\mathbf{z},t). \tag{3.5}$$

According to (2.16) and (2.9), (3.3) implies that Φ satisfies the integrodifferential equation

$$\frac{\partial^2\Phi}{\partial t^2}=\Delta[12\Phi\circ\Phi\circ\Phi+9\Phi\circ\Delta\Phi+\tfrac{3}{2}\Delta(\Phi\circ\Phi)+\Delta\Delta\Phi], \tag{3.6}$$

where the $\circ$-product is defined by (2.15). Let us now look at the soliton solutions of (3.3).

To start with,

$$u(c,x,t)=\sigma(|c|,x-ct),$$

$$\sigma(|c|,y)=\tfrac{1}{2}|c|\operatorname{sech}^2(\tfrac{1}{2}|c|^{1/2}y), \tag{3.7}$$

is a solitary wave solution of (3.3). The velocity c may be taken positive or negative. We choose an arbitrary vector $\mathbf{v}$ so that $|\mathbf{v}|\geq|c|$ and write $c=\hat{\mathbf{n}}\circ\mathbf{v}$. As we vary $\hat{\mathbf{n}}$, keeping $\mathbf{v}$ fixed, this generates a family of solutions

$$u(\mathbf{v},\hat{\mathbf{n}},x,t)\equiv u(\hat{\mathbf{n}}\circ\mathbf{v},x,t)=\sigma(|\hat{\mathbf{n}}\circ\mathbf{v}|,x-\hat{\mathbf{n}}\circ\mathbf{v}t) \tag{3.8}$$

of (3.3) depending on $\hat{\mathbf{n}}$. Because (3.7) is even in y, it follows that

$$u(\mathbf{v},-\hat{\mathbf{n}},-x,t)=u(\mathbf{v},\hat{\mathbf{n}},x,t) \tag{3.9}$$

and we may take its R transform as in (3.5). According to our discussion at the end of Sec. 2, then Φ is a solitary wave of velocity $\mathbf{v}$,

$$\Phi(\mathbf{v},\mathbf{z},t)=\Phi(\mathbf{v},\mathbf{z}-\mathbf{v}t,0). \tag{3.10}$$

Thus (3.6) has solitary wave solutions whose velocities have arbitrary directions and magnitudes.

In fact, the family of solitary wave solutions of (3.6) is larger than the R transforms of the functions given by (3.8). For $\hat{\mathbf{n}}$ in some set of directions, say $\hat{\mathbf{n}}\in\Omega$, we may choose u to be zero,

$$u(\mathbf{v},\hat{\mathbf{n}},x,t)=\begin{cases}0, & \hat{\mathbf{n}}\in\Omega,\\ \sigma(|\hat{\mathbf{n}}\circ\mathbf{v}|,x-\hat{\mathbf{n}}\circ\mathbf{v}t), & \hat{\mathbf{n}}\notin\Omega.\end{cases} \tag{3.11}$$

If Ω is such that whenever $\hat{\mathbf{n}}\in\Omega$ then $-\hat{\mathbf{n}}\in\Omega$, it follows that u satisfies (3.9) and its R transform is a solitary wave. Of course, the function (3.11) will generally have steplike discontinuities as a function of $\hat{\mathbf{n}}$, but that does not prevent Φ from being continuous and differentiable.

An even larger class of solitrary waves is generated by using the translational invariance of the KdV equation. We may choose, for $\hat{\mathbf{n}}\notin\Omega$,

$$u(\hat{\mathbf{n}}\circ\mathbf{v},x,t)=\sigma(|\hat{\mathbf{n}}\circ\mathbf{v}|,x-\mathbf{n}\circ\mathbf{v}t-\delta) \tag{3.11'}$$

and make δ an arbitrary function of $\hat{\mathbf{n}}$. While this changes only the position of the one-dimensional solitary wave, it generally changes the *shape* of its R transform.

The next step is to consider an N soliton solution of (3.3),

$$u(c_1,\ldots,c_N;x,t)=\sigma_N(c_1,\ldots,c_N;x,t), \tag{3.12}$$

where[9]

$$\sigma_N(c_1,\ldots,c_N;x,t)=2\frac{\partial^2}{\partial x^2}\log\det M. \tag{3.13}$$

The $N\times N$ matrix M is given by

$$M_{ij}=\delta_{ij}+\frac{2\kappa_i}{\kappa_i+\kappa_j}\exp[\kappa_i(c_it-x+x_i)], \tag{3.14}$$

where the x_i are arbitrary constants, and $\kappa_i = |c_i|^{1/2}$. The velocities c_i must all be of the same sign. Suppose that all $c_i > 0$.

The asymptotic form of σ_N is given by

$$\sigma_N \underset{t\to\pm\infty}{\approx} \sum_{i=1}^{N} \sigma(|c_i|, x - c_i t - \delta_i^{\pm}) \tag{3.15}$$

in the sense that for each i

$$\lim_{t\to\pm\infty} \sigma_N(c_1, \ldots, c_N; x + c_i t, t) = \sigma(|c_i|, x - \delta_i^{\pm}), \tag{3.16}$$

where σ is defined by (3.7), and[10]

$$\delta_i^- = x_i,$$

$$\delta_i^+ = x_i + \frac{1}{\kappa_i}\left[\sum_{j<i} \log\left(\frac{\kappa_i - \kappa_j}{\kappa_i + \kappa_j}\right) + \sum_{j>i} \log\left(\frac{\kappa_j + \kappa_i}{\kappa_j - \kappa_i}\right)\right], \tag{3.17}$$

with the understanding that $0 < \kappa_1 < \cdots < \kappa_N$.

We choose an arbitrary set of N vectors $\mathbf{v}_i$, $i = 1, \ldots, N$, and a unit vector $\hat{\mathbf{n}}$ such that

$$c_i = \hat{\mathbf{n}} \cdot \mathbf{v}_i, \quad i = 1, \ldots, N.$$

Now vary $\hat{\mathbf{n}}$, keeping the $\mathbf{v}_i$ fixed. Let Ω be a set of directions $\hat{\mathbf{n}}$ that includes all those for which not all $\hat{\mathbf{n}} \cdot \mathbf{v}_i$, $i = 1, \ldots, N$, have the same sign. (Ω may contain directions for which all $\hat{\mathbf{n}} \cdot \mathbf{v}_i$ have the same sign, but its complement must have positive Lebesgue measure.[11]) Furthermore, let Ω be such that if $\hat{\mathbf{n}} \in \Omega$, then $-\hat{\mathbf{n}} \in \Omega$. Let Ω^+ (Ω^-) be the set of directions $\hat{\mathbf{n}} \notin \Omega$ such that all $\hat{\mathbf{n}} \cdot \mathbf{v}_i > 0$ ($\hat{\mathbf{n}} \cdot \mathbf{v}_i < 0$).

We now define a family of functions

$$u(\hat{\mathbf{n}}, s, t) \equiv \begin{cases} \sigma_N(\hat{\mathbf{n}} \cdot \mathbf{v}_1, \ldots, \hat{\mathbf{n}} \cdot \mathbf{v}_N, x, t), & \text{if } \hat{\mathbf{n}} \in \Omega^+, \\ 0, & \text{if } \hat{\mathbf{n}} \in \Omega, \\ \sigma_N(\hat{\mathbf{n}} \cdot \mathbf{v}_1, \ldots, \hat{\mathbf{n}} \cdot \mathbf{v}_N, -x, -t), & \hat{\mathbf{n}} \in \Omega^-, \end{cases} \tag{3.18}$$

which, for each $\hat{\mathbf{n}}$, satisfy (3.3). It is clear from (3.14) that

$$u(-\hat{\mathbf{n}}, -x, t) = u(\hat{\mathbf{n}}, x, t)$$

for all t, $\hat{\mathbf{n}}$, x. We may therefore take its R transform as in (3.5), and the resulting function $\Phi(\mathbf{z}, t)$ will satisfy (3.6).

We have (indicating x-derivatives by primes)

$$\begin{aligned} \Phi(\mathbf{z}, t) &= -\frac{1}{8\pi^2} \int d\hat{\mathbf{n}}\, u''(\hat{\mathbf{n}}, \hat{\mathbf{n}} \cdot \mathbf{z}, t) \\ &= -\frac{1}{8\pi^2} \int_{\Omega^+} d\hat{\mathbf{n}}\, \sigma''(\hat{\mathbf{n}} \cdot \mathbf{v}_1, \ldots, \hat{\mathbf{n}} \cdot \mathbf{v}_N; \hat{\mathbf{n}} \cdot \mathbf{z}, t) \\ &\quad - \frac{1}{8\pi^2} \int_{\Omega^-} d\hat{\mathbf{n}}\, \sigma_N''(\hat{\mathbf{n}} \cdot \mathbf{v}_1, \ldots, \hat{\mathbf{n}} \cdot \mathbf{v}_N; -\hat{\mathbf{n}} \cdot \mathbf{z}, -t) \\ &= -\frac{1}{8\pi^2} \int_{\Omega^+} d\hat{\mathbf{n}} [\sigma_N''(\hat{\mathbf{n}} \cdot \mathbf{v}_1, \ldots, \hat{\mathbf{n}} \cdot \mathbf{v}_N, \hat{\mathbf{n}} \cdot \mathbf{z}, t) \\ &\quad + \sigma_N''(-\hat{\mathbf{n}} \cdot \mathbf{v}_1, \ldots, -\hat{\mathbf{n}} \cdot \mathbf{v}_N; \hat{\mathbf{n}} \cdot \mathbf{z}, -t)] \\ &= -\frac{1}{4\pi^2} \int_{\Omega^+} d\hat{\mathbf{n}}\, \sigma_N''(\hat{\mathbf{n}} \cdot \mathbf{v}_1, \ldots, \hat{\mathbf{n}} \cdot \mathbf{v}_N; \hat{\mathbf{n}} \cdot \mathbf{z}, t) \end{aligned} \tag{3.19}$$

because (3.14) shows that σ_N is invariant under a simultaneous sign change of t and all c_i. The asymptotic form of Φ is now determined from (3.16),

$$\Phi(\mathbf{z} + \mathbf{v}_i t, t) = -\frac{1}{4\pi^2} \int_{\Omega^+} d\hat{\mathbf{n}}\, \sigma''(\hat{\mathbf{n}} \cdot \mathbf{v}_i, \hat{\mathbf{n}} \cdot \mathbf{z} - \delta_i^{\pm}) + o(1) \tag{3.20}$$

as $t \to \pm\infty$. In this sense, then, we have

$$\Phi(\mathbf{z}, t) \underset{t\to\pm\infty}{\approx} \sum_{i=1}^{n} \eta_i^{\pm}(\mathbf{v}_i, \mathbf{z} - \mathbf{v}_i t), \tag{3.21}$$

where the solitons are given by

$$\eta_i^{\pm}(\mathbf{v}_i, \mathbf{z}) = -\frac{1}{4\pi^2} \int_{\Omega^+} d\hat{\mathbf{n}}\, \sigma''(\hat{\mathbf{n}} \cdot \mathbf{v}_i, \hat{\mathbf{n}} \cdot \mathbf{z} - \delta_i^{\pm}). \tag{3.22}$$

Since (3.17) shows that $\delta_i^+ - \delta_i^-$ depends on the c_j, it depends on $\hat{\mathbf{n}}$. Consequently the shapes of the solitons in the infinite past generally are different from those in the infinite future. In this case the collisions produce not only positional shifts of the solitons, but changes in their shapes as well.

Equation (3.22) can be written in the form

$$\eta_i^{\pm}(\mathbf{v}, z) = Ru(\mathbf{v}_i, \hat{\mathbf{n}}, x - \delta_i^{\pm})$$

in terms of the function u defined in (3.11). But it follows from (2.4) that

$$\int d^3\mathbf{z}\, |Rf|^2 = 2\pi^2 \int d\hat{\mathbf{n}} \int_{-\infty}^{\infty} dx\, |f'|^2, \tag{3.23}$$

and hence

$$\begin{aligned} \int d^3\mathbf{z}\, |\eta_i^{\pm}(\mathbf{v}_i, \mathbf{z})|^2 &= (2\pi)^2 \int_{\hat{\mathbf{n}}\in\Omega^+} d\hat{\mathbf{n}} \int_{-\infty}^{\infty} dx\, |u'(\mathbf{v}_i, \hat{\mathbf{n}}, x - \delta_i^{\pm})|^2 \\ &= (2\pi)^2 \int_{\hat{\mathbf{n}}\in\Omega^+} d\hat{\mathbf{n}} \int_{-\infty}^{\infty} dx\, |u'(\mathbf{v}_i, \hat{\mathbf{n}}, x)|^2. \end{aligned}$$

Since the right-hand side has the same value for + and −, it follows that in spite of its changed shape, the volume of the square of each individual soliton at $t \to -\infty$ equals the volume of the square of the corresponding one at $t \to +\infty$. This guarantees, specifically, that no solitons can disappear and all solitons present at $t \to -\infty$ are again present, with equal "strength" but generally altered shape, at $t \to +\infty$.

We have thus demonstrated that the integrodifferential equation (3.6) has solutions which break up into solitons in the infinite past and future. Moreover, each soliton is confined in the sense that it vanishes at large distances in all directions. The initial number of solitons and their velocities are equal to the final ones, but their shapes are generally changed. These solutions, moreover, are expressible by quadrature, being the R transforms, (3.19), of functions explicitly given by (3.13) and (3.14). Whether (3.6) has other soliton solutions is not known. Since we do not have a general solution of the initial-value problem for the IKdV equation (3.3), except for the special cases in which u satisfies either (3.1) or its time reversed, we do not have a general solution of the initial-value problem for (3.6). Furthermore it is not obvious that the Radon transforms of all solutions of (3.6) must solve (3.3).

The important question of possible physical applications of (3.6) has not yet been investigated.

APPENDIX

We want to present here a simple, direct, algebraic verification of the N-soliton solution of the KdV equation that does not use any of the inverse scattering machinery.

We write the matrix M of (3.14) in matrix form

$$M=1+EL, \tag{A1}$$

where E is the diagonal matrix

$$E=2K\exp(K^3t-Kx+X); \tag{A2}$$

in terms of the matrices X and K,

$$X_{ij}=\delta_{ij}x_i, \quad K_{ij}=\delta_{ij}K_i \tag{A3}$$

and

$$L_{ij}=L_{ji}=(K_i+K_j)^{-1}. \tag{A4}$$

We have

$$KL+LK=Q, \tag{A5}$$

where Q is the matrix $Q_{ij}=1$, and

$$\frac{\partial E}{\partial x}=-KE, \quad \frac{\partial E}{\partial t}=K^3E. \tag{A6}$$

Now, by (A6), and writing $N\equiv M^{-1}$,[12]

$$\psi\equiv 2\frac{\partial}{\partial x}\log\det M=2\mathrm{tr}\left(\frac{\partial M}{\partial x}N\right) \tag{A7}$$

$$=-2\mathrm{tr}(KELN)=-2(ELKN)$$

since EL commutes with N. Also[13]

$$\psi=\mathrm{tr}\left(\frac{\partial\tilde{M}}{\partial x}\tilde{N}\right)=-2\mathrm{tr}(LEK\tilde{N})=-2\mathrm{tr}(EKLN)$$

and hence

$$\psi=\mathrm{tr}QF=\sum_{ij}F_{ij}, \tag{A8}$$

where

$$F=-NE=-(E^{-1}+L)^{-1}. \tag{A9}$$

Now using (A6), we get

$$\frac{\partial N}{\partial x}=NK(1-N), \quad \frac{\partial N}{\partial t}=NK^3(N-1),$$

$$\frac{\partial F}{\partial x}=-NKF, \quad \frac{\partial F}{\partial t}=NK^3F,$$

$$\frac{\partial^2F}{\partial x^2}=NK(2N-1)KF,$$

$$\frac{\partial^3F}{\partial x^3}=NK(3NK+3KN-6NKN-K)KF,$$

and, using (A5),

$$\frac{\partial F}{\partial x}Q\frac{\partial F}{\partial x}=-NK(NK+KN-2NKN)KF.$$

Consequently the matrix F satisfies the equation

$$-\frac{\partial F}{\partial t}=3\frac{\partial F}{\partial x}Q\frac{\partial F}{\partial x}+\frac{\partial^3F}{\partial x^3}. \tag{A10}$$

But since Q is such that for any A and B

$$\mathrm{tr}(QAQB)=\mathrm{tr}(QA)\mathrm{tr}(QB), \tag{A11}$$

multiplication of (A10) by Q, taking traces, and using (A8) gives

$$-\frac{\partial\psi}{\partial t}=3\left(\frac{\partial\psi}{\partial x}\right)^2+\frac{\partial^3\psi}{\partial x^3}. \tag{A12}$$

Differentiation with respect to x and setting $u=\partial\psi/\partial x$ finally shows that u satisfies the KdV equation

$$-\frac{\partial u}{\partial t}=6u\frac{\partial u}{\partial x}+\frac{\partial^3u}{\partial x^3} \tag{A13}$$

which is the time-reversed version of (3.1). Thus u satisfies (3.3). It is of some interest that (A13) is the result of the fact that the symmetric matrix F of (A9) satisfies (A10), or more explicitly

$$-\frac{\partial F_{ij}}{\partial t}=3\sum_{k,l}\frac{\partial F_{ik}}{\partial x}\frac{\partial F_{jl}}{\partial x}+\frac{\partial^3F_{ij}}{\partial x^3}.$$

We may use this technique to check whether the function defined by (3.13) and (3.14), but with the c_i not all of the same sign, satisfies (3.3). For that case, instead of (A2),

$$E=2K\exp(K^3St-Kx+X),$$

where S is a diagonal matrix with some of its diagonal entries -1, and the others, $+1$. Since the x-derivatives do not involve S, it is only necessary to check if S disappears from $\partial^2\psi/\partial t^2$. One readily finds that

$$\frac{\partial^2\psi}{\partial t^2}=-\mathrm{tr}[QNSK^3(2N-1)SK^3NE]$$

in which S disappears only if $S=1$ or $S=-1$. Thus the pure soliton function (3.13), with some solitons coming in from the left and some from the right, does not satisfy the IKdV equation.

[1]C.S. Gardner, J.M. Greene, M.D. Kruskal, and R.M. Miura, Phys. Rev. Lett. **19**, 1095 (1967); P. Lax, Commun. Pure Appl. Math. **21**, 467 (1968).

[2]A.C. Scott, F.Y.F. Chu, and D.W. McLaughlin, Proc. IEEE **61**, 1443 (1973); R.M. Miura, SIAM Rev. **18**, 412 (1976); *Bäcklund Transformations, the Inverse Scattering Method, Solitons, and Their Applications*, edited by R.M. Miura (Springer, New York, 1976); and references in these papers.

[3]See, for example, R. Hirota, J. Phys. Soc. Jpn. **35**, 1566 (1973), pp. 40—68, in *Bäcklund Transformations, the Inverse Scattering Method, Solitons, and Their Applications*, Ref. 1; M.J. Ablowitz and R. Haberman, Phys. Rev. Lett. **35**, 1185 (1975); V.S. Dryuma, Zh. Eksp. Teor. Fiz. Pis'ma Red. **19**, 753 (1974) [JETP Lett. **19**, 387 (1974)]; V.E. Zhakharov and A.B. Shabat, Funkts. Anal. i Ego Prilosh. **8**, 43 (1974) [Func. Anal. Appl. **8**, 226 (1974)].

[4]In this paper we shall confine the name *soliton* to solutions that tend to zero at infinity in all directions of space and that have certain stability properties under collisions.

[5]For a detailed treatment see, for example, I.M. Gel'fand, M.I. Graev, and N.Ya. Vilenkin, *Generalized Functions* (Academic, New York, 1966), Vol. 5.

[6]V. E. Zakharov and A. B. Shabat, Zh. Eksp. Teor. Fiz. **61**, 118 (1971) and **64**, 1627 (1973) [Sov. Phys. JETP **34**, 62 (1972); **37**, 823 (1974)].
[7]These requirements correspond to condition 4 of Ref. 5.
[8]We shall use the name *solitary wave* for any travelling wave that preserves its shape and vanishes at infinity in all directions.
[9]R. Hirota, Phys. Rev. Lett. **27**, 1192 (1971); M. Wadati and M. Toda, J. Phys. Soc. Jpn. **32**, 1403 (1972); C. S. Gardner *et al.*, Commun. Pure Appl. Math. **27**, 97 (1974).
[10]Wadati and Toda, Ref. 6; S. Tanaka, Kyoto Univ. Publ. Res. Inst. Math. Sci. **8**, 419 (1972/73).
[11]This implies a minor restriction on the $\mathbf{v}_i$. For example, two of them must not point in exactly opposite directions.
[12]tr stands for *trace*.
[13]Here, and only here, the tilde indicates *transpose* of a matrix.

On the relaxation to quantum-statistical equilibrium of the Wigner–Weisskopf atom in a one-dimensional radiation field. VIII. Emission in an infinite system in the presence of an extra photon

Russell Davidson

Department of Economics, Queen's University, Kingston, K7L 3N6, Canada

John J. Kozak

Department of Chemistry and Radiation Laboratory, [a] *University of Notre Dame, Notre Dame, Indiana 46556*

(Received 28 November 1977)

In this paper we study the emission of a two-level atom in a radiation field in the case where one mode of the field is assumed to be excited initially, and where the system is assumed to be of infinite extent. (The restriction to a one-dimensional field, which has been made throughout this series, is *not* essential: It is made chiefly for ease of presentation of the mathematical methods.) An *exact* expression is obtained for the probability $\rho(t)$ that the two-level quantum system is in the excited state at time t. This problem, previously unsolved in radiation theory, is tackled by reformulating the expression found in VII [J. Math. Phys. **16**, 1013 (1975)] of this series for the time evolution of $\rho(t)$ in a finite system in the presence of an extra photon, and then constructing the infinite-system limit. A quantitative assessment of the role of the extra photon and of the coupling constant in influencing the dynamics is obtained by studying numerically the expression derived for $\rho(t)$ for a particular choice of initial condition. The study presented here casts light on the problem of time-reversal invariance and clarifies the sense in which exponential decay is universal; in particular, we find that: (1) It is the infinite-system limit which converts the time-reversible solutions of VII into the irreversible solution obtained here, and (2) it is the weak-coupling limit that imposes exponential form on the time dependence of the evolution of the system. The anticipated generalization of our methods to more complicated radiation-matter problems is discussed, and finally, several problems in radiation chemistry and physics, already accessible to exact analysis given the approach introduced here, are cited.

I. INTRODUCTION

The work begun in the previous papers of this series[1–7] (the papers in the series will hereafter be referred to as I–VII, respectively) is continued here. An exact solution was found in VII for a model of the interaction of a two-level atom (or other two-level quantum system) with a one-dimensional radiation field, for that "sector" of the problem where, when the atom is excited, exactly one photon of the radiation field is present. This exact solution was for a system finite in extent, and here the limit of the solution for infinite systems is found. The taking of the limit is not exactly trivial, and involves a reformulation of the result of VII before any progress can be made.

The Hamiltonian of the model is as follows:

$$H=\epsilon_1\alpha\alpha^*+\epsilon_2\alpha^*\alpha+\sum_\lambda \tfrac{1}{2}\hbar\omega_\lambda(a_\lambda^* a_\lambda+1)+\sum_\lambda (h_\lambda^*\alpha^* a_\lambda+h_\lambda\alpha a_\lambda^*),$$

where ϵ_1 and ϵ_2 are the energies of the ground state $|1\rangle$ and excited state $|2\rangle$ of the two-level atom, and where the operators are defined by

$$\alpha=|1\rangle\langle 2|,\quad \alpha^*=|2\rangle\langle 1|$$

$$\langle n_\lambda|a_\lambda|m_\lambda\rangle=[2(n_\lambda+1)]^{1/2}\delta^{\mathrm{Kr}}(m_\lambda-n_\lambda-1)=\langle m_\lambda|a_\lambda^*|n_\lambda\rangle.$$

The state $|n_\lambda\rangle$ has n_λ $(=0,1,2,\cdots)$ photons in the λth mode of the radiation field. The $\delta^{\mathrm{Kr}}(\cdots)$ is the Kronecker delta. Also, $\hbar\omega_\lambda$ is the energy of each photon in the λth mode, and we define $\hbar E=\epsilon_2-\epsilon_1$.

A basis for the Hilbert space of the system is given by the product states

$$|i;\{n_\lambda\}\rangle\equiv|i\rangle\prod_\lambda|n_\lambda\rangle$$

with $i=1,2$ and $n_\lambda=0,1,2,\cdots$. The "sectors" of the Hamiltonian are the eigenspaces of the operator

$$N=\alpha^*\alpha+\tfrac{1}{2}\sum_\lambda a_\lambda^* a_\lambda-1,$$

which commutes with H. The eigenvalue of N associated with the state $|i;\{n_\lambda\}\rangle$ is $(i-1)+\sum_\lambda n_\lambda$, and so N measures the number of photons present when the atom is de-excited $(i=1)$. Papers I–VI dealt exclusively with the sector where $N=1$, while in VII the solution for $N=2$ was obtained for a finite system. We remark in passing that the operator N was defined in VII simply as

$$N=\alpha^*\alpha+\tfrac{1}{2}\sum_\lambda a_\lambda^* a_\lambda.$$

The redefinition here seems more sensible.

[a] The research described herein was supported by the Division of Basic Energy Sciences of the Department of Energy. This is Document No. NDRL-1830 from the Notre Dame Radiation Laboratory.

 0022-2488/78/1905-1074$1.00

If the state of the system at time t is denoted by $|\Psi(t)\rangle$, then by $\varphi_\lambda(t)$ we mean the projection, $\langle 2|\langle 1_\lambda| \times \langle\{0_\mu\}|\Psi(t)\rangle$, on the state where the atom is excited and one photon in mode λ is present. The initial condition for the problem discussed in VII, where $N=2$, is

$$\langle 2|\langle 1_\lambda|\langle\{0_\mu\}|\Psi(0)\rangle = \psi_\lambda.$$

Since we shall assume that at time $t=0$ the atom is in its excited state, normalization imposes the requirement

$$\sum_\lambda |\psi_\lambda|^2 = 1. \tag{1}$$

The solution of this problem is given in Eq. (29) of VII, and it is as follows:

$$\varphi_\lambda(t) = \frac{4h_\lambda^*}{\hbar^2} \sum_\nu \frac{\exp(-i\xi_\nu t)}{H_1'(\xi_\nu)} \times \sum_\mu \frac{1}{(\xi_\mu - \omega_\lambda)H'(\xi_\mu)H(\xi_\nu - \xi_\mu)} \times \sum_\kappa \frac{G(\xi_\kappa)}{H'(\xi_\kappa)H(\xi_\nu - \xi_\kappa)}. \tag{2}$$

The functions appearing in Eq. (2) are defined as follows:

$$H(\xi) \equiv \frac{1}{2\pi i}\left(E - \xi - \sum_\mu \frac{2|h_\mu|^2}{\hbar^2(\omega_\mu - \xi)}\right). \tag{3}$$

(Recall that $\hbar E$ is the energy separating the two levels of the atom; $\hbar\omega_\lambda$ is the energy of a photon in mode λ; h_μ is the coupling parameter in the Hamiltonian.) The function H in Eq. (3) is a meromorphic function of the complex variable ξ, with (simple) poles at $\xi=\omega_\mu$ and interlacing zeros (see IV) at $\xi=\xi_\mu$, say. The function $(H(\xi))^{-1}$ therefore has (simple) poles at $\xi=\xi_\mu$, with residues $(H'(\xi_\mu))^{-1}$ (the prime here denotes differentiation). Next

$$G(\xi) \equiv \frac{1}{(2\pi i)^2} \sum_\lambda \frac{h_\lambda \psi_\lambda}{\omega_\lambda - \xi}. \tag{4}$$

The function G embodies the information about the initial state of the system. Further, we have

$$H_1(\zeta) \equiv \sum_\mu \frac{1}{H'(\xi_\mu)H(\zeta - \xi_\mu)}. \tag{5}$$

This function, too, is meromorphic in the complex variable ζ, with poles at the points $\zeta = \xi_\mu + \xi_\kappa$ for any μ and κ. The poles are interlaced by zeros, at the points $\zeta=\zeta_\lambda$, say. Equation (2) is now completely specified.

In Sec. II, an expression is found for

$$\rho(t) \equiv \sum_\lambda |\varphi_\lambda(t)|^2,$$

the probability at time t that the atom is in its excited state. This expression is then cast in a form from which it is possible to find the limit of $\rho(t)$ as the system size tends to infinity. Then this limit is taken in Sec. III, and an expression suitable for numerical computation is found. Numerical results are presented in Sec. IV along with comparisons of results from the $N=1$ sector. In Sec. V, there is discussion of the work of this paper and of future work to which it should lead.

II. REFORMULATION OF THE FINITE SYSTEM RESULT

It is convenient to introduce at once the dimensionless variables which have been used in all the previous papers of the series. These variables are necessary for the numerical work to follow, and also make limit-taking, whether of infinite size or weak coupling, much easier. It has been argued in VI [Eq. (8)] that a general form for the coupling parameter h_λ is as follows:

$$|h_\lambda|^2 = \frac{\alpha\hbar^2 Ec}{L} f\left(\frac{c|k_\lambda|}{E}\right), \tag{6}$$

where α is a dimensionless coupling constant (a fine-structure constant for one-dimensional systems in fact), c is the speed of light, L is the length of the system, and f is a dimensionless nonnegative real function whose argument involves k_λ, the wavenumber of the λth mode of the field. As usual for radiation,

$$\omega_\lambda = c|k_\lambda| \quad \text{and} \quad k_\lambda = 2\pi n/L$$

for some nonzero integer n. It is always possible to set $f(1)=1$ by a suitable choice of α. We may now introduce the dimensionless variables by making the following definitions:

$$\tau = \alpha E t, \quad \beta_\lambda = \frac{\omega_\lambda}{\alpha E}, \quad \gamma_\mu = \frac{\xi_\mu}{\alpha E}, \quad \delta_\nu = \frac{\zeta_\nu}{\alpha E}. \tag{7}$$

The above are time and frequency variables. The length of the system is given by

$$\sigma^2 = \alpha EL/c,$$

and then, from Eq. (6),

$$|h_\lambda|^2 = \frac{(\alpha E)^2\hbar^2}{\sigma^2} f(\alpha\beta_\lambda).$$

Some dimensionless functions are needed, and so we make the following definitions:

$$H(\alpha E\xi) = \frac{\alpha E}{2\pi i}\hat{H}(\xi), \tag{8a}$$

$$H_1(\alpha E\xi) = \frac{(2\pi i)^2}{\alpha E}\hat{H}_1(\xi), \tag{8b}$$

with derivatives

$$H'(\alpha E\xi) = \frac{1}{2\pi i}\hat{H}'(\xi), \tag{8c}$$

$$H_1'(\alpha E\xi) = \left(\frac{2\pi i}{\alpha E}\right)^2 \hat{H}_1'(\xi). \tag{8d}$$

Corresponding to the function G of Eq. (4) we define

$$\hat{G}(\xi) \equiv \frac{(2\pi i)^2\sqrt{\pi}}{\hbar} G(\alpha E\xi) = \frac{\sqrt{2}}{\sigma} \sum_\lambda \frac{f^{1/2}(\alpha\beta_\lambda)\psi_\lambda}{\beta_\lambda - \xi}. \tag{9}$$

There is no reason here for the function denoted by $f^{1/2}$ to be real, since its relation with f will be taken to be

$$|f^{1/2}(x)|^2 = f(x).$$

With these definitions, then, Eq. (2) assumes the form

$$\varphi_\lambda(t) = \varphi_\lambda\left(\frac{\tau}{\alpha E}\right) \equiv \hat{\varphi}_\lambda(\tau)$$

$$= \frac{4}{\sigma\sqrt{2}} f^{1/2}(\alpha\beta_\lambda) \sum_\nu \frac{\exp(-i\delta_\nu\tau)}{\hat{H}_1'(\delta_\nu)} \times \sum_\mu \frac{1}{(\gamma_\mu - \beta_\lambda)\hat{H}'(\gamma_\mu)\hat{H}(\delta_\nu - \gamma_\mu)} \times \sum_\kappa \frac{\hat{G}(\gamma_\kappa)}{\hat{H}'(\gamma_\kappa)\hat{H}(\delta_\nu - \gamma_\kappa)} . \quad (10)$$

The above expression has several summations over the various discrete spectra of the problem, denoted severally by the β_λ, γ_μ, δ_ν. It is the aim of the remainder of this section to express these summations as contour integrals where the contours are away from the positive real line. By this device, new functions can be defined in terms of which $\hat{\varphi}_\lambda(\tau)$ can be expressed without reference to the discrete spectra. Then taking the infinite-system limit becomes reasonably simple.

First, let us consider the ν summation

$$\sum_\nu \frac{\exp(-i\delta_\nu\tau)}{\hat{H}_1'(\delta_\nu)\hat{H}(\delta_\nu - \gamma_\mu)\hat{H}(\delta_\nu - \gamma_\kappa)} \quad (11)$$

along with the contour integral

$$-\frac{1}{2\pi i}\int_C d\xi \frac{\exp(-i\xi\tau)}{\hat{H}_1(\xi)\hat{H}(\xi - \gamma_\mu)\hat{H}(\xi - \gamma_\kappa)}$$

where C is depicted in Fig. 1. As the radius of the large semicircle tends to infinity, it can be seen by Jordan's lemma[8] that the contribution from the semicircle as opposed to the straight-line part of C vanishes. Consequently, the integral could as well be taken along the contour B shown in Fig. 2, since the integrand has no singularities except on the positive real line. These singularities are at two classes of points: where $\xi = \delta_\nu$, that is, from the zeros of $\hat{H}_1$, and at points like $\xi = \gamma_\lambda + \gamma_\rho$ for some choices of λ and ρ. For $\xi = \delta_\lambda$, the residue is easily seen to be just the summand in the expression (11). Now in Eq. (27) of VII the following expression for $H_1(z)$ is given:

$$H_1(z) = \sum_\mu \sum_\kappa \frac{1}{H'(\xi_\mu)H'(\xi_\kappa)(z - \xi_\mu - \xi_\kappa)}$$

and in dimensionless variables this becomes

$$\hat{H}_1(\xi) = \sum_\mu \sum_\kappa \frac{1}{\hat{H}'(\gamma_\mu)\hat{H}'(\gamma_\kappa)(\xi - \gamma_\mu - \gamma_\kappa)} . \quad (12)$$

Use of this relation shows that, if $\kappa \neq \mu$, the only singularity occurs where $\xi = \gamma_\mu + \gamma_\kappa$, where the residue can be found by a simple calculation to be

$$\tfrac{1}{2}\exp[-i(\gamma_\mu + \gamma_\kappa)\tau].$$

If, on the other hand, $\kappa = \mu$, then there are singularities at each point of the form $\xi = \gamma_\mu + \gamma_\rho$, with residue

$$\tfrac{1}{2}\exp[-i(\gamma_\mu + \gamma_\rho)\tau][\hat{H}'(\gamma_\mu)/\hat{H}'(\gamma_\rho)].$$

At the point $\xi = 2\gamma_\mu$, the residue is

$$\exp(-2i\gamma_\mu\tau).$$

This information leads to the following result:

$$\sum_\lambda \frac{\exp(-i\delta_\lambda\tau)}{\hat{H}_1'(\delta_\lambda)\hat{H}(\delta_\lambda - \gamma_\mu)\hat{H}(\delta_\lambda - \gamma_\kappa)} = -\frac{1}{2\pi i}\int_C d\xi \frac{\exp(-i\xi\tau)}{\hat{H}_1(\xi)\hat{H}(\xi - \gamma_\mu)\hat{H}(\xi - \gamma_\kappa)} - \tfrac{1}{2}\exp[-i(\gamma_\mu + \gamma_\kappa)\tau] + \tfrac{1}{2}\delta_{\kappa\mu}\exp(-i\gamma_\mu\tau) \times \hat{H}'(\gamma_\mu)f_1(\tau), \quad (13)$$

where we have introduced the function

$$f_1(\tau) \equiv -\sum_\kappa \frac{\exp(-i\gamma_\kappa\tau)}{\hat{H}'(\gamma_\kappa)} \quad (14)$$

which can itself be expressed as a contour integral:

$$f_1(\tau) = \frac{1}{2\pi i}\int_C d\xi \exp(-i\xi\tau)\frac{1}{\hat{H}(\xi)} , \quad (15)$$

since the zeros of $\hat{H}$ are at the points γ_κ. Comparison of this expression with Eq. (12) of III shows that $f_1(\tau)$ is nothing but the solution of the spontaneous emission problem in the sector $N = 1$, that is, the quantity

$$\langle 2, \{0_\lambda\}|\Psi(t)\rangle \text{ when } |\Psi(0)\rangle = |2, \{0_\lambda\}\rangle.$$

The infinite-system limit of $f_1(\tau)$ is of course well-defined (see V).

To make use of Eq. (13), let us return to Eq. (10) and compute the probability of the atom's being excited:

$$\rho(\tau) \equiv \sum_\lambda |\hat{\varphi}_\lambda(\tau)|^2.$$

The λ summation in this expression can be performed when one has evaluated the sum:

$$\frac{2}{\sigma^2}\sum_\lambda \frac{f(\alpha\beta_\lambda)}{(\gamma_\mu - \beta_\lambda)(\gamma_\rho - \beta_\lambda)} .$$

If $\mu \neq \rho$, then this sum is equal to the expression

$$\frac{1}{(\gamma_\rho - \gamma_\mu)}\left(\frac{2}{\sigma^2}\sum_\lambda \frac{f(\alpha\beta_\lambda)}{\gamma_\mu - \beta_\lambda} - \frac{2}{\sigma^2}\sum_\lambda \frac{f(\alpha\beta_\lambda)}{\gamma_\rho - \beta_\lambda}\right) .$$

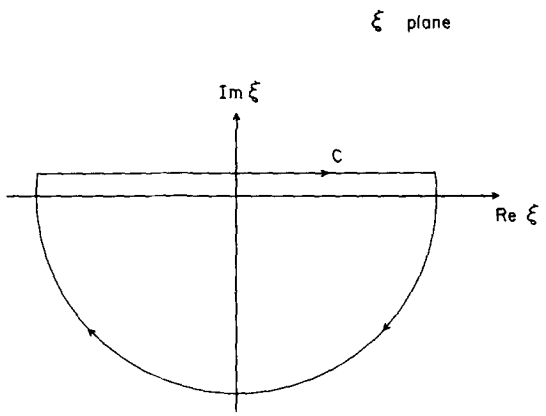

FIG. 1. The contour C chosen in evaluating the integral representation of Eq. (11).

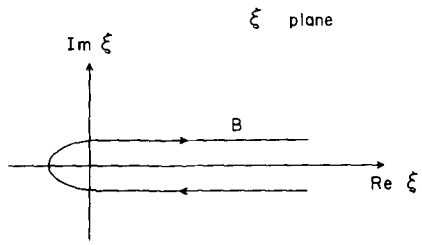

FIG. 2. The contour B chosen in evaluating the expression (27) for $h(\tau)$.

But, from the definitions of Eqs. (8) and (3),

$$\hat{H}(\xi)=\frac{1}{\alpha}-\xi-\frac{2}{\sigma^2}\sum_\lambda\frac{f(\alpha\beta_\lambda)}{\beta_\lambda-\xi} \tag{16}$$

[see Eq. (VII.36)] and (so that we may treat the case $\mu=\rho$)

$$\hat{H}'(\xi)=-1-\frac{2}{\sigma^2}\sum_\lambda\frac{f(\alpha\beta_\lambda)}{(\beta_\lambda-\xi)^2}$$

so that finally

$$\frac{2}{\sigma^2}\sum_\lambda\frac{f(\alpha\beta_\lambda)}{(\gamma_\mu-\beta_\lambda)(\gamma_\rho-\beta_\lambda)}=-1-\delta_{\mu\rho}\hat{H}'(\gamma_\mu)$$

since

$$\hat{H}(\gamma_\mu)=0.$$

It is now a short algebraic manipulation to see that

$$\rho(\tau)=-4\sum_\mu\frac{1}{\hat{H}'(\gamma_\mu)}\left|\sum_\nu\frac{\exp(-i\delta_\nu\tau)}{\hat{H}_1'(\delta_\nu)\hat{H}(\delta_\nu-\gamma_\mu)}\times\sum_\kappa\frac{\hat{G}(\gamma_\kappa)}{\hat{H}'(\gamma_\kappa)\hat{H}(\delta_\nu-\gamma_\kappa)}\right|^2.$$

Use of Eq. (13) provides the following expression:

$$\begin{aligned}\rho(\tau)=-4\sum_\mu\frac{1}{\hat{H}'(\gamma_\mu)}\Bigg|&-\tfrac{1}{2}\exp(-i\gamma_\mu\tau)\\&\times\sum_\kappa\frac{\exp(-i\gamma_\kappa\tau)\hat{G}(\gamma_\kappa)}{\hat{H}'(\gamma_\kappa)}+\tfrac{1}{2}\exp(-i\gamma_\mu\tau)\hat{G}(\gamma_\mu)f_1(\tau)\\&-\frac{1}{2\pi i}\int_B d\xi\,\frac{\exp(-i\xi\tau)}{\hat{H}_1(\xi)\hat{H}(\xi-\gamma_\mu)}\\&\times\sum_\kappa\frac{\hat{G}(\gamma_\kappa)}{\hat{H}'(\gamma_\kappa)\hat{H}(\xi-\gamma_\kappa)}\Bigg|^2.\end{aligned} \tag{17}$$

Now another new function can be defined as follows:

$$f_2(\tau)\equiv\sum_\kappa\frac{\exp(-i\gamma_\kappa\tau)\hat{G}(\gamma_\kappa)}{\hat{H}'(\gamma_\kappa)} \tag{18}$$

with contour integral representation

$$f_2(\tau)=-\frac{1}{2\pi i}\int_C d\xi\,\frac{\exp(-i\xi\tau)\hat{G}(\xi)}{\hat{H}(\xi)} \tag{19}$$

just as for $f_1(\tau)$. The function $f_2(\tau)$ turns out to be the solution of the $N=1$ sector problem with initial condition

$$|\Psi(0)\rangle=\sum_\lambda\psi_\lambda|1;1_\lambda,\{0_\mu\}_{\mu\neq\lambda}\rangle.$$

It is not worth proving this assertion in detail, for it is easily demonstrable by a slight modification of the calculation of Sec. II of III.

The definition given in Eq. (19) permits another useful manipulation. The κ summation in Eq. (17), that is

$$\sum_\kappa\frac{\hat{G}(\gamma_\kappa)}{\hat{H}'(\gamma_\kappa)\hat{H}(\xi-\gamma_\kappa)}$$

can be expressed as a contour integral

$$-\frac{1}{2\pi i}\int_C d\zeta\,\frac{\hat{G}(\zeta)}{\hat{H}(\zeta)\hat{H}(\xi-\zeta)}, \tag{20}$$

provided that ξ lies *above* the contour C. (We have in effect introduced two C-type contours, one below the other.) The expression (20) is a *convolution*, and we can accordingly make use of the following standard facts:

If two functions a, b, defined on the positive real line, have Laplace transforms A, B defined as follows:

$$A(\xi)=i\int_0^\infty d\tau\exp(i\xi\tau)a(\tau) \tag{21a}$$

$$B(\xi)=i\int_0^\infty d\tau\exp(i\xi\tau)b(\tau) \tag{21b}$$

then the Laplace transform of the function ab is the convolution, $A*B$, of A and B, defined by the formula:

$$(A*B)(\xi)\equiv\frac{1}{2\pi i}\int_D d\zeta\,A(\zeta)B(\xi-\zeta). \tag{22}$$

Either A or B, regarded as a function of the complex variable ξ, is defined and holomorphic in the region $\mathrm{Im}\,\xi>k$, for some k. The integrand in Eq. (22) will have singularities ζ with $\mathrm{Im}\,\zeta<k$ associated with the function A and singularities with $\mathrm{Im}\,\zeta>-k$ associated with B. The contour D is defined as one which passes from $-\infty$ to $+\infty$ leaving the singularities associated with A on the left and those associated with B on the right, provided that this is feasible. The inversion formula for Eq. (21) is

$$a(\tau)=\frac{1}{2\pi i}\int_C d\xi\exp(-i\xi\tau)A(\xi). \tag{23}$$

Now, from Eqs. (15) and (19), we may see that $1/\hat{H}(\xi)$ is the Laplace transform of $f_1(\tau)$, so that

$$\frac{1}{\hat{H}(\xi)}=i\int_0^\infty d\tau\exp(i\xi\tau)f_1(\tau), \tag{24}$$

and that $-\hat{G}(\xi)/\hat{H}(\xi)$ is the Laplace transform of $f_2(\tau)$. Since the expression (20) is simply

$$\left(-\frac{\hat{G}}{\hat{H}}*\frac{1}{\hat{H}}\right)(\xi)$$

(C satisfies the requirements on D), it follows that the expression equals

$$i\int_0^\infty d\tau\exp(i\xi\tau)f_1(\tau)f_2(\tau). \tag{25}$$

When the results of Eqs. (18) and (25) are put into Eq. (17), and the whole quantity whose squared modulus is being taken multiplied by $\exp(i\gamma_\mu\tau)$, the result is

$$\begin{aligned}\rho(\tau)=-\sum_\mu\frac{1}{\hat{H}'(\gamma_\mu)}\Bigg|&-f_2(\tau)+\hat{G}(\gamma_\mu)f_1(\tau)-\frac{\exp(i\gamma_\mu\tau)}{\pi}\\&\times\int_C d\xi\,\frac{\exp(-i\xi\tau)}{\hat{H}_1(\xi)\hat{H}(\xi-\gamma_\mu)}\\&\times\int_0^\infty d\tau'\exp(i\xi\tau')f_1(\tau')f_2(\tau')\Bigg|^2.\end{aligned}$$

One last substitution frees this expression from the function $\hat{H}$. By use of Eq. (24), we obtain

$$\begin{aligned}\rho(\tau)=-\sum_\mu\frac{1}{\hat{H}'(\gamma_\mu)}\Bigg|&-f_2(\tau)+\hat{G}(\gamma_\mu)f_1(\tau)+\frac{\exp(i\gamma_\mu\tau)}{\pi i}\\&\times\int_0^\infty d\theta\exp(-i\gamma_\mu\theta)f_1(\theta)\int_0^\infty d\tau'f_1(\tau')f_2(\tau')\\&\times\int_C d\xi\,\frac{\exp[-i\xi(\tau-\tau'-\theta)]}{\hat{H}_1(\xi)}\Bigg|^2.\end{aligned}$$

It is tempting at this point simply to define a new function by the contour integral expression

$$\int_C d\xi \frac{\exp(-i\xi\tau)}{\hat{H}_1(\xi)}.$$

But, since $\hat{H}_1(\xi) \to 0$ as $\xi \to \infty$ [Eq. (12) makes this clear], it is difficult to work with the integral as it stands. However, an integration by parts yields the following result:

$$\int_C d\xi \frac{1}{\hat{H}_1(\xi)} \int_0^\infty d\tau' \exp[-i\xi(\tau-\tau'-\theta)] f_1(\tau') f_2(\tau')$$
$$= i \int_C d\xi \frac{1}{\xi \hat{H}_1(\xi)} \int_0^\infty d\tau' \exp[-i\xi(\tau-\tau'-\theta)] F(\tau'),$$

where

$$F(\tau') = \frac{d}{d\tau'}(f_1(\tau') f_2(\tau')). \tag{26}$$

The ξ that has appeared in the denominator is useful, as will be seen. First, it follows from Eq. (12) and the relation

$$\sum_\kappa \frac{1}{\hat{H}'(\gamma_\kappa)} = -1$$

[set $\tau = 0$ in Eq. (14)] that

$$\lim_{\xi \to \infty} \frac{1}{\xi \hat{H}_1(\xi)} = 1.$$

Thus

$$1/\xi \hat{H}_1(\xi) - 1$$

tends to zero at infinity. This means that, for $\tau > 0$,

$$\int_C d\xi \exp(-i\xi\tau)\left(\frac{1}{\xi \hat{H}_1(\xi)} - 1\right)$$
$$= \int_B d\xi \exp(-i\xi\tau)\left(\frac{1}{\xi \hat{H}_1(\xi)} - 1\right)$$

by Jordan's lemma[8] (with B as in Fig. 2). But since B is a closed contour,

$$\int_B d\xi \exp(-i\xi\tau) = 0.$$

We shall then make the definition:

$$h(\tau) \equiv \frac{1}{2\pi i} \int_B d\xi \frac{\exp(-i\xi\tau)}{\xi \hat{H}_1(\xi)} \quad (\tau \geq 0) \tag{27}$$

and

$$h(\tau) = 0 \quad \text{for } \tau < 0.$$

It follows that

$$h(\tau) = -\sum_\nu \frac{\exp(-i\delta_\nu\tau)}{\delta_\nu \hat{H}_1'(\delta_\nu)} - \frac{1}{\hat{H}_1(0)} \quad (\tau \geq 0)$$

which is well defined. On the other hand,

$$\frac{1}{2\pi i} \int_0^\infty d\tau' F(\tau') \int_C d\xi \exp[-i\xi(\tau-\tau')]$$
$$= \begin{cases} -iF(\tau) & (\tau \geq 0), \\ 0 & (\tau < 0), \end{cases}$$

a result obtained by making use of Eqs. (21) and (23). Therefore, we obtain

$$\int_C d\xi \frac{1}{\hat{H}_1(\xi)} \int_0^\infty d\tau' \exp[-i\xi(\tau-\tau'-\theta)] f_1(\tau') f_2(\tau')$$
$$= -2\pi \int_0^{\tau-\theta} d\tau' F(\tau') h(\tau-\tau'-\theta) + 2\pi i F(\tau-\theta)$$

if $\tau - \theta \geq 0$, and 0 otherwise. It follows that

$$[\exp(i\gamma_\mu\tau)/\pi i] \int_0^\infty d\theta \exp(-i\gamma_\mu\theta) f_1(\theta) \int_0^\infty d\tau' f_1(\tau') f_2(\tau')$$
$$\times \int_C d\xi \exp[-i\xi(\tau-\tau'-\theta)]/\hat{H}_1(\xi)$$
$$= \int_0^\tau d\theta \exp[i\gamma_\mu(\tau-\theta)] f_1(\theta)$$
$$\times [2F(\tau-\theta) + 2i \int_0^{\tau-\theta} d\tau' F(\tau') h(\tau-\tau'-\theta)]$$
$$= \int_0^\tau d\theta \exp(i\gamma_\mu\theta) f_1(\tau-\theta)$$
$$\times [2F(\theta) + 2i \int_0^\theta d\tau' F(\tau') h(\theta-\tau')].$$

The last equality is the result of changing the integration variable θ to $\tau - \theta$. Finally, then

$$\rho(\tau) = -\sum_\mu \frac{1}{\hat{H}'(\gamma_\mu)} \Bigg| -f_2(\tau) + \hat{G}(\gamma_\mu) f_1(\tau)$$
$$+ 2 \int_0^\tau d\theta \exp(i\gamma_\mu\theta) f_1(\tau-\theta)$$
$$\times \left[F(\theta) + i \int_0^\theta d\tau' F(\tau') h(\theta-\tau')\right] \Bigg|^2. \tag{28}$$

This concludes our reformulation of $\rho(\tau)$. Only one discrete spectrum, namely that of the γ_μ, is still involved, and it will not be hard to deal with that in taking the infinite-system limit.

III. THE INFINITE SYSTEM LIMIT

There are a few steps left in the elimination of the γ_μ from Eq. (28). First of all, we may note that (bar denotes complex conjugate)

$$-\sum_\mu \frac{1}{\hat{H}'(\gamma_\mu)} \exp(i\gamma_\mu\theta) \overline{\exp(i\gamma_\mu\theta')}$$
$$= -\sum_\mu \frac{\exp[-i\gamma_\mu(\theta'-\theta)]}{\hat{H}'(\gamma_\mu)}$$
$$= \frac{1}{2\pi i} \int_B d\xi \frac{\exp[-i\xi(\theta'-\theta)]}{\hat{H}(\xi)}. \tag{29}$$

Now from Eq. (16) and the definition of β_λ, an explicit expression for $\hat{H}(\xi)$ can be obtained (see also Sec. V of VII):

$$\hat{H}(\xi) = \frac{1}{\alpha} - \xi + \frac{4}{\sigma^2} \sum_{n=1}^\infty \frac{f(2\pi n\alpha/\sigma^2)}{\xi - 2\pi n/\sigma^2}.$$

As $\sigma^2 \to \infty$, the system size tends to infinity, and one obtains in the limit

$$\hat{H}(\xi) = \frac{1}{\alpha} - \xi + \frac{2}{\pi} \int_0^\infty d\lambda \frac{f(\alpha\lambda)}{\xi - \lambda}$$

for ξ not a positive real number. This limiting expression is no longer a meromorphic function of ξ, but rather it has a branch point at $\xi = 0$ and an associated discontinuity along the positive real axis of ξ. There exist limiting values of $\hat{H}$ as ξ tends to a point on this axis:

$$\hat{H}^\pm(\xi) = \frac{1}{\alpha} - \xi + \frac{2}{\pi} P \int_0^\infty d\lambda \frac{f(\alpha\lambda)}{\xi - \lambda} \mp 2if(\alpha\xi), \tag{30}$$

where the two limiting values correspond to an approach to ξ from above the axis and from below. The symbol P indicates that the following integral is a Cauchy principal part. In the $\sigma^2 \to \infty$ limit, then, the expression in Eq. (29) becomes:

$$\frac{1}{2\pi i}\int_0^\infty d\zeta \exp[-i\zeta(\theta-\theta')]\left(\frac{1}{\hat{H}^+(\zeta)}-\frac{1}{\hat{H}^-(\zeta)}\right). \tag{31}$$

The validity of this last step is not guaranteed unless our system is "ergodic" in the sense discussed in Sec. IV of V. This means that the coupling function f must satisfy the condition $f(x)\to 0$ as $x\to 0+$ [see Eq. (9) of VI] and that the coupling constant α must be sufficiently small, in fact, that

$$\alpha < \left(\frac{2}{\pi}\int_0^\infty \frac{f(\lambda)}{\lambda}\,d\lambda\right)^{-1}$$

(see V). These conditions ensure that $\hat{H}$ has no zeros on the negative real axis. Further, f must be such that, for $\zeta \geq 0$,

$$|\hat{H}^+(\zeta)| = |\hat{H}^-(\zeta)| > \epsilon \tag{32}$$

for some positive ϵ independent of α. Equation (31) is then justified. The independence of ϵ and α, as we shall see shortly, guarantees the existence of a weak-coupling limit. From now on, ergodicity, as defined here, will be assumed. It is convenient to make the definition

$$\begin{aligned}\mathcal{H}(\zeta) &\equiv \frac{1}{2\pi i}\left(\frac{1}{\hat{H}^+(\zeta)}-\frac{1}{\hat{H}^-(\zeta)}\right)\\ &= \frac{2}{\pi}f(\alpha\zeta)\left[\left(\frac{1}{\alpha}-\zeta+\frac{2}{\pi}P\int_0^\infty \frac{d\lambda f(\alpha\lambda)}{\zeta-\lambda}\right)^2+4f^2(\alpha\zeta)\right]^{-1}.\end{aligned} \tag{33}$$

The infinite-system limit of $f_1(\tau)$ is readily expressed in terms of $\mathcal{H}$: from Eq. (15),

$$\begin{aligned}f_1(\tau) &= \frac{1}{2\pi i}\int_C d\xi \exp(-i\xi\tau)\frac{1}{\hat{H}(\xi)}\\ &= \int_0^\infty d\zeta \exp(-i\zeta\tau)\mathcal{H}(\zeta).\end{aligned} \tag{34}$$

Similarly, Eq. (29) becomes

$$\begin{aligned}&-\sum_\mu \frac{1}{\hat{H}'(\gamma_\mu)}\exp(i\gamma_\mu\theta)\exp(\overline{i\gamma_\mu\theta'})\\ &= \int_0^\infty d\zeta \exp(i\zeta\theta)\exp(\overline{i\zeta\theta'})\mathcal{H}(\zeta).\end{aligned} \tag{35}$$

Our next step is to consider

$$\begin{aligned}&-\sum_\mu \frac{1}{\hat{H}'(\gamma_\mu)}\hat{G}(\gamma_\mu)\exp(\overline{i\gamma_\mu\theta})\\ &= -\sum_\mu \exp(-i\gamma_\mu\theta)\frac{\hat{G}(\gamma_\mu)}{\hat{H}'(\gamma_\mu)}\\ &= \frac{1}{2\pi i}\int_B d\xi \exp(-i\xi\theta)\frac{\hat{G}(\xi)}{\hat{H}(\xi)}.\end{aligned} \tag{36}$$

From Eq. (9), one can obtain the infinite-system limit of $\hat{G}(\xi)$:

$$\hat{G}(\xi) = \frac{2\sqrt{2}}{\sigma}\sum_{n=1}^\infty \frac{f^{1/2}(2\pi n\alpha/\sigma^2)\psi_{2\pi n/\sigma^2}}{2\pi n/\sigma^2-\xi}.$$

Normalization [Eq. (1)] imposes the condition

$$2\sum_{n=1}^\infty |\psi_{2\pi n/\sigma^2}|^2 = 1,$$

and so it is sensible to make the definition

$$\psi(\lambda) = \frac{\sigma}{\sqrt{\pi}}\psi_{2\pi n/\sigma^2}, \tag{37}$$

where, as $\sigma^2\to\infty$, $2\pi n/\sigma^2\to\lambda$. The normalization condition is then just

$$\int_0^\infty |\psi(\lambda)|^2 d\lambda = 1.$$

In the limit

$$\hat{G}(\xi) = \left(\frac{2}{\pi}\right)^{1/2}\int_0^\infty \frac{f^{1/2}(\alpha\lambda)\psi(\lambda)\,d\lambda}{\lambda-\xi}. \tag{38}$$

This is another function with a branch point at $\xi=0$ and a cut along the positive real axis. The limiting values are

$$\hat{G}^\pm(\zeta) = \left(\frac{2}{\pi}\right)^{1/2}P\int_0^\infty \frac{f^{1/2}(\alpha\lambda)\psi(\lambda)}{\lambda-\zeta} \pm i\sqrt{2\pi}f^{1/2}(\alpha\zeta)\psi(\zeta). \tag{39}$$

The expression in Eq. (36) can be written as

$$\frac{1}{2\pi i}\int_0^\infty d\zeta \exp(-i\zeta\theta)\left(\frac{\hat{G}^+(\zeta)}{\hat{H}^+(\zeta)}-\frac{\hat{G}^-(\zeta)}{\hat{H}^-(\zeta)}\right).$$

Now in view of the definition of $\mathcal{H}$, Eq. (33), it is useful to define

$$g(\zeta) \equiv \frac{\hat{G}^-(\zeta)\hat{H}^+(\zeta)-\hat{G}^+(\zeta)\hat{H}^-(\zeta)}{\hat{H}^+(\zeta)-\hat{H}^-(\zeta)} \tag{40}$$

for then we have

$$\frac{1}{2\pi i}\left(\frac{\hat{G}^+(\zeta)}{\hat{H}^+(\zeta)}-\frac{\hat{G}^-(\zeta)}{\hat{H}^-(\zeta)}\right) = g(\zeta)\mathcal{H}(\zeta).$$

Equation (36) is then

$$\begin{aligned}&\lim_{\sigma^2\to\infty}\left\{-\sum_\mu \frac{1}{\hat{H}'(\gamma_\mu)}\hat{G}(\gamma_\mu)\exp(\overline{i\gamma_\mu\theta})\right\}\\ &= \int_0^\infty d\zeta \mathcal{H}(\zeta)g(\zeta)\exp(\overline{i\zeta\theta}).\end{aligned} \tag{41}$$

The last step needed in our treatment of Eq. (28) is to verify that

$$\begin{aligned}&\lim_{\sigma^2\to\infty}\left\{-\sum_\mu \frac{1}{\hat{H}'(\gamma_\mu)}\hat{G}(\gamma_\mu)\overline{\hat{G}(\gamma_\mu)}\right\}\\ &= \int_0^\infty d\zeta \mathcal{H}(\zeta)|g(\zeta)|^2.\end{aligned} \tag{42}$$

(The proof is lengthy, and is relegated to Appendix A.) For then, by use of Eqs. (42), (35), and (41), we obtain the infinite-system limit of Eq. (28), as follows:

$$\begin{aligned}\rho(\tau) = \int_0^\infty d\zeta \mathcal{H}(\zeta)\big|&-f_2(\tau)+g(\zeta)f_1(\tau)\\ &+2\int_0^\tau d\theta \exp(i\zeta\theta)f_1(\tau-\theta)\\ &\times[F(\theta)+i\int_0^\theta d\tau' F(\tau')h(\theta-\tau')]\big|^2,\end{aligned} \tag{43}$$

where $\mathcal{H}(\zeta)$ and $g(\zeta)$ are given by Eqs. (33) and (40), and the functions f_1, f_2, F, and h are all infinite-system limits. Equation (40) can be made explicit

$$\begin{aligned}g(\zeta) = \frac{1}{(2\pi)^{1/2}}\Bigg[&\frac{\pi\psi(\zeta)}{f^{1/2}(\alpha\zeta)}\left(\frac{1}{\alpha}-\zeta-\frac{2}{\pi}P\int_0^\infty \frac{f(\alpha\lambda)\,d\lambda}{\lambda-\zeta}\right)\\ &+2P\int_0^\infty \frac{f^{1/2}(\alpha\lambda)\psi(\lambda)\,d\lambda}{\lambda-\zeta}\Bigg].\end{aligned}$$

The limit of f_1 has already been given in Eq. (34). Next

$$f_2(\tau) = -\int_0^\infty d\zeta \exp(-i\zeta\tau)g(\zeta)\mathcal{H}(\zeta). \tag{44}$$

As before

$$F(\theta) = \frac{d}{d\theta}\Big(f_1(\theta)f_2(\theta)\Big).$$

For $h(\tau)$, we obtain

$$h(\tau)=\frac{1}{2\pi i}\int_B d\xi\frac{\exp(-i\xi\tau)}{\xi\hat{H}_1(\xi)} \quad \text{[Eq. (27)]}$$
$$=-\frac{1}{\hat{H}_1(0)}+\frac{1}{2\pi i}\int_0^\infty d\zeta\exp(-i\zeta\tau)\frac{1}{\zeta}\left(\frac{1}{\hat{H}_1^+(\zeta)}-\frac{1}{\hat{H}_1^-(\zeta)}\right). \tag{45}$$

Now the definition of $\hat{H}_1$ [Eqs. (8) and (9)] is

$$\hat{H}_1(\zeta)=\sum_\mu\frac{1}{\hat{H}'(\gamma_\mu)\hat{H}(\zeta-\gamma_\mu)}.$$

Then it follows that

$$\hat{H}_1(\zeta)=-\frac{1}{2\pi i}\int_B\frac{d\xi}{\hat{H}(\xi)\hat{H}(\zeta-\xi)}$$
$$=-\int_0^\infty\frac{H(\xi)\,d\xi}{\hat{H}(\zeta-\xi)}$$

in the limit, and so

$$\hat{H}_1^\pm(\zeta)=-\int_0^\infty\frac{H(\xi)\,d\xi}{\hat{H}^\pm(\zeta-\xi)}.$$

Finally, then, after using these results in Eq. (45), we find

$$h(\tau)=\frac{1}{\int_0^\infty d\xi H(\xi)/\hat{H}(-\xi)}$$
$$+\int_0^\infty d\xi\frac{\exp(-i\zeta\tau)\int_0^\xi d\xi\,H(\xi)H(\xi-\xi)}{\zeta\left|\int_0^\infty d\xi\,H(\xi)/\hat{H}^+(\xi-\xi)\right|^2}. \tag{46}$$

The integral which forms the second term of this expression is of course proper at $\zeta=0$.

We have now given an expression for each function appearing in Eq. (43), and have therefore fully specified the infinite-system limit of our problem.

It was seen in III that in the $N=1$ sector the function $\rho(\tau)$ for an infinite system took on the simple form

$$\rho(\tau)=\exp(-4\tau)$$

when the coupling constant α became zero (the weak-coupling limit). Although ρ is not analytic in α at $\alpha=0$ for any nonzero τ, the limit is well-defined. We may now show that this result persists in the $N=2$ sector. First we make a definition that introduces a new representation of the initial condition of our problem:

$$\varphi(\lambda)=(1/\sqrt{\alpha})\psi(\lambda/\alpha).$$

It can be seen at once from Eqs. (7) and (37) that the initial condition is unchanged as α varies so long as $\varphi(\lambda)$ does not depend on α. (The photon wavefunction depends on energy alone.) Further, normalization is preserved

$$\int_0^\infty|\varphi(\lambda)|^2\,d\lambda=\int_0^\infty|\psi(\lambda)|^2\,d\lambda=1.$$

Now let us consider

$$\int_0^\infty d\xi\exp(-i\xi\tau)H(\xi).$$

By Eq. (34), this is $f_1(\tau)$. It can be written

$$\exp(-i\tau/\alpha)\int_{-1/\alpha}^\infty d\lambda\exp(-i\lambda\tau)H(\lambda+1/\alpha). \tag{47}$$

The expression $H(\lambda+1/\alpha)$ has a simple limit when $\alpha\to0$:

$$H\left(\lambda+\frac{1}{\alpha}\right)=\frac{2}{\pi}f(1+\alpha\lambda)$$
$$\times\left[\left(-\lambda-\frac{2}{\pi}P\int_0^\infty\frac{f(\mu)\,d\mu}{\mu-1-\alpha\lambda}\right)^2+4f^2(1+\alpha\lambda)\right]^{-1}$$
$$-\frac{2}{\pi}\left[\left(\lambda+\frac{2}{\pi}P\int_0^\infty\frac{f(\mu)\,d\mu}{\mu-1}\right)^2+4\right]^{-1} \tag{48}$$

as $\alpha\to0$ since $f(1)=1$. We now prove that

$$\lim_{\alpha\to0}\int_{-1/\alpha}^\infty d\lambda\exp(-i\lambda\tau)H(\lambda+1/\alpha)$$
$$=\int_{-\infty}^\infty d\lambda\exp(-i\lambda\tau)\lim_{\alpha\to0}H(\lambda+1/\alpha). \tag{49}$$

If we put $f(x)=0$ for $x<0$, then $H(\lambda+1/\alpha)=0$ for $\lambda<-1/\alpha$, and so the range of integration in Eq. (47) can be extended down to $-\infty$. Then the result (49) follows by the Lebesgue convergence theorem[9] if there is a function $M(\lambda)$ such that

$$|\exp(-i\lambda\tau)H(\lambda+1/\alpha)|=H(\lambda+1/\alpha)\le M(\lambda) \tag{50}$$

and

$$\int_{-\infty}^\infty M(\lambda)\,d\lambda<\infty.$$

A suitable M can be defined, in view of Eq. (32), as as follows:

$$M(\lambda)=\begin{cases}q/(\lambda+p)^2 & \text{for } \lambda<-2p,\\ q/\epsilon^2 & \text{for } -2p\le\lambda\le2p,\\ q/(\lambda-p)^2 & \text{for } \lambda>2p\end{cases} \tag{51}$$

where

$$q=\frac{2}{\pi}\sup_x f(x)$$

and

$$p=\sup_{x\ge0}\left|\frac{2}{\pi}P\int_0^\infty\frac{f(\mu)\,d\mu}{\mu-x}\right|.$$

It is easy to see that the conditions of Eq. (50) are satisfied.

From Eqs. (47), (48), and (49), then, we obtain

$$\lim_{\alpha\to0}\exp(i\tau/\alpha)f_1(\tau)=\exp(iP\tau)\exp(-2\tau), \tag{52}$$

where

$$P=\frac{2}{\pi}P\int_0^\infty\frac{f(\mu)\,d\mu}{\mu-1}.$$

We have verified that, for the $N=1$ sector,

$$\rho(\tau)=|f_1(\tau)|^2=\exp(-4\tau)\quad\text{as }\alpha\to0.$$

By an exactly similar calculation, one can see that

$$\lim_{\alpha\to0}\frac{\exp(i\tau/\alpha)}{\alpha^{1/2}}f_2(\tau)=\lim_{\alpha\to0}-\frac{\exp(i\tau/\alpha)}{\alpha^{1/2}}$$
$$\times\int_0^\infty d\xi\ \exp(-i\xi\tau)g(\xi)H(\xi)$$
$$=\left(\frac{1}{2\pi}\right)^{1/2}\exp(iP\tau)\exp(-2\tau)Q, \tag{53}$$

where

$$-Q=2\pi i\varphi(1)+2P\int_0^\infty\frac{f^{1/2}(\lambda)\varphi(\lambda)\,d\lambda}{\lambda-1}.$$

In fact, because τ appears in the defining integrals for f_1 and f_2 only through the factor $\exp(-i\lambda\tau)$, the convergence in Eqs. (52) and (53) is uniform with respect to τ. It is also readily seen that away from $\tau=0$, the derivatives with respect to τ of

$$\exp(i\tau/\alpha)f_1(\tau) \quad \text{and} \quad (1/\sqrt{\alpha})\exp(i\tau/\alpha)f_2(\tau)$$

converge uniformly, and therefore converge to the derivatives of the expressions on the right-hand sides of Eqs. (52) and (53). From the definition of the function $F(\tau)$, Eq. (26), it follows that

$$\lim_{\alpha\to 0}\sqrt{\alpha}\,\exp(2i\tau/\alpha)F(\tau)=i\sqrt{2/\pi}\,Q\exp(2iP\tau)\exp(-4\tau). \tag{54}$$

It is shown in Appendix B that

$$\int_0^\tau d\theta\exp(i\zeta\theta)f_1(\tau-\theta)[F(\theta)+i\int_0^\theta d\tau' F(\tau')h(\theta-\tau')] \leq N\sqrt{\alpha} \tag{55}$$

for all $\zeta\geq 0$, with N independent of both ζ and α. Thus, finally, since we know that, for any α,

$$\int_0^\infty H(\zeta)\,d\zeta=f_1(0)=1;$$

$$\int_0^\infty H(\zeta)|g(\zeta)|^2\,d\zeta=1, \quad \text{(Appendix A)}$$

and since $|f_1|$ and $|f_2|$ are bounded by 0 and 1, we obtain, with a further appeal to the Lebesgue convergence theorem, that

$$\lim_{\alpha\to 0}\rho(\tau)=\lim_{\alpha\to 0}|f_1(\tau)|^2=\exp(-4\tau).$$

IV. NUMERICAL RESULTS

In the previous section an exact expression was derived [viz., Eq. (43)] describing the emission of an excited two-level quantum system in a (one-dimensional) field of electromagnetic radiation in the limit where the system size becomes of infinite extent and the mode spectrum becomes continuous. In the formulation of the problem (Sec. I), it was assumed that one quantum of the field was excited initially, in contrast to the situation described in an earlier paper of the series (V), where the infinite-system limit was constructed for the problem of an excited two-level atom in a field of radiation assumed to be *de-excited* initially. Having thus at our disposal the *exact* solution for $\rho(\tau)$ for both choices of initial condition, it is of great interest to study the effect of an extra photon on the time evolution of an excited two-level system, and for this reason the results of a few representative calculations will now be reported.

In the work presented in papers V and VI of this series, calculations were carried out for several different choices of the coupling function $f(x)$, viz., $f(x)=x^{-1/2}$, $f(x)=x^{-1/4}$ and

$$f(x)=4x/(1+x)^2. \tag{56}$$

It was found in V that the first two choices of coupling function led to "ghost states," and nonergodic behavior in the time evolution of $\rho(\tau)$ for sufficiently large values of the coupling parameter α. (We shall return to this point in the following section.) Here, however, we shall focus on the results generated using Eq. (43) with the "ergodic" coupling function (56). Given the choice (56) three different values of the coupling parameter α were selected for study, viz., $\alpha=0.05$, 0.10, and 0.20; these three values of α are sufficiently separated to allow changes in the time evolution of $\rho(\tau)$ with increasing coupling strength to be displayed conveniently. To complete the specification for the numerical study, we recall that the initial photon state $\psi(\lambda)$ must be chosen such that the normalization

$$\int_0^\infty|\psi(\lambda)|^2\,d\lambda=1$$

is preserved, and here we take

$$\psi(\lambda)=\begin{cases}\sqrt{\alpha/2}, & 0\leq\lambda\leq 2/\alpha\\ 0, & \lambda>2/\alpha.\end{cases} \tag{57}$$

[Since in the limit of an infinite system, the modes $2\pi n/\sigma^2\to\lambda$, the choice $\lambda=1/\alpha$ in the coupling function $f(\alpha\lambda)$ corresponds to resonance.] Our choice of $\psi(\lambda)$ has the further property that the state depends on photon energy alone, and not on the coupling constant α.

So as not to interrupt the continuity of the subsequent discussion, the results of several numerical checks will now be reported. First, the reformulation of the finite-system result (carried out in Sec. II), resulted in an expression for $\rho(\tau)$ [Eq. (28)] which could be computed numerically and checked against the results of corresponding calculations reported in VII. In particular, numerical calculations were performed on Eq. (28) for the same values $\alpha=0.1$ and $\sigma^2=1.0$, as were used in the calculations of VII, based on Eq. (10). The agreement between computed values of $\rho(\tau)$ is always within 1%. Secondly, in the reformulation of the problem, in the infinite-system limit it was noted (Sec. II) that $f_1(\tau)$ is the solution of the spontaneous emission problem in the $N=1$ sector. Accordingly, the function $f_1(\tau)$, evaluated for the particular choice (56) of coupling function, can be compared directly with a previously derived result, Eq. (VI-19); a check of this point for each choice of α considered here revealed that the two expressions were in numerical correspondence to within 1% for all τ. Finally, the contribution $f_2(\tau)$ is the lower-sector solution for an initial condition of the atom de-excited and a photon represented by ψ_ρ present, and hence should be strictly zero at $\tau=0$. Numerical calculation of $f_2(\tau)$ at $\tau=0$ showed that the magnitude of $f_2(0)$ was <0.01; concomitantly, calculation of the full expression (43) at $\tau=0$ gave an initial value >0.99 [the integral contribution to the expression (43) vanishes identically at $\tau=0$], and agreement with the formally exact result, $\rho(0)=1$, was judged satisfactory.

Turning now to a discussion of the results displayed in Figs. 3, 4, and 5, we have plotted the time evolution of $\rho(\tau)$ assuming the presence initially of an extra photon in the field [calculated via Eq. (43)], the evolution of $\rho(\tau)$ assuming the field to be de-excited initially [calculated via Eq. (VI-19)], and the weak-coupling solution to the dynamical problem for three representative values of the coupling constant, respectively, $\alpha=0.05$, 0.10, and 0.20. From and analysis of these data, the following conclusions can be drawn. First, for a given choice of α, de-excitation of the two-level quantum system is facilitated when there exists an extra photon initially in the radiation field. Secondly, for each choice of α, the time evolution of the "extra

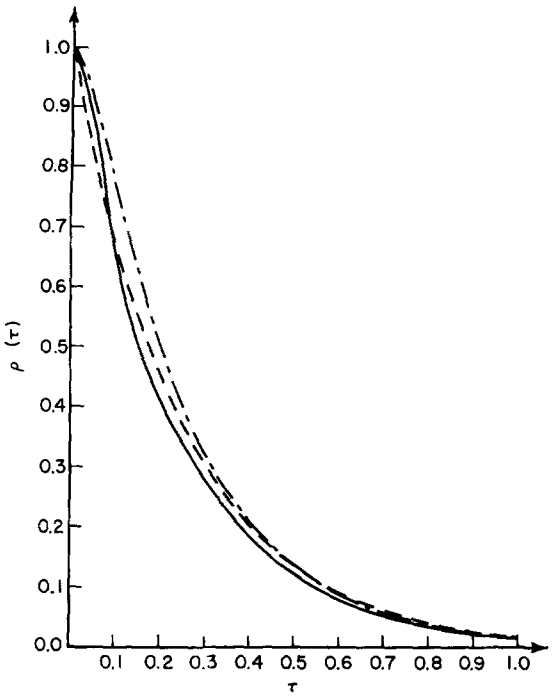

FIG. 3. The time evolution of $\rho(\tau)$ for the choice of coupling constant $\alpha = 0.05$. The solid line (—) displays the evolution assuming the presence initially of an extra photon in the field [calculated via Eq. (43)], the hyphenated line (— — — —) displays the evolution assuming the field to be de-excited initially (calculated via Eq. (VI-19), and the dashed line (- - - - - - - -) displays the weak-coupling (strictly exponential) solution to the dynamical problem.

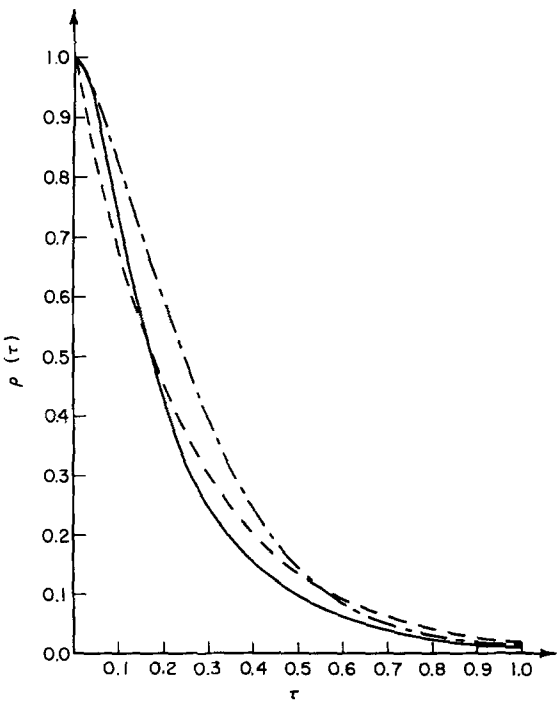

FIG. 4. The time evolution of $\rho(\tau)$ for the choice of coupling constant $\alpha = 0.10$. The conventions in this figure are the same as in Fig. 3.

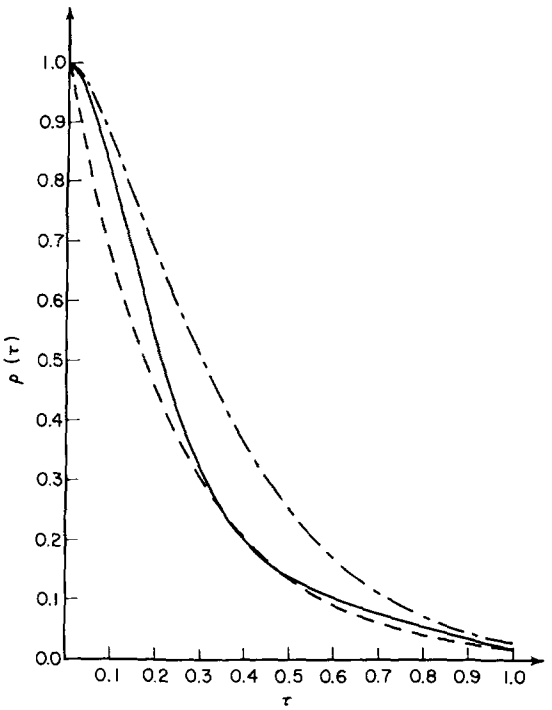

FIG. 5. The time evolution of $\rho(\tau)$ for the choice of coupling constant $\alpha = 0.20$. The conventions in this figure are the same as in Fig. 3.

photon" solution is closer to the weak-coupling (strictly exponential) solution than is the solution generated assuming the field to be de-excited initially. In this regard, it should be noted that *both* exact solutions tend to converge to the weak-coupling solution as the coupling constant α becomes smaller. [The reader should be cautioned that the apparent increase in half-life of the decay as the coupling constant becomes larger is illusory—remember that τ is a dimensionless variable (viz., $\tau = \alpha E l$) scaled by the coupling constant α.] Lastly, we note that there is a qualitative difference between the behavior of the weak-coupling solution and either of the two exact expressions for $\rho(\tau)$ [viz., Eq. (VI-19) and Eq. (43)] in the neighborhood of $\tau = 0$; the latter time-evolution profiles exhibit a "shoulder" which becomes more pronounced with increasing coupling constant. This effect had been noted in earlier studies on a different class of models. (See Ref. 10 and especially the article by Zwanzig.)

V. CONCLUSIONS

Since the work presented in the series of papers, Refs. 1—7, has spanned a number of different aspects of the radiation-matter problem, it may be of use in this concluding section to place the present study within the context of the overall program of research being pursued by the authors. As was noted in VII, methods for obtaining the exact solution for models involving *only* boson operators (with the coupling linear in these operators) have been available since the introduction of the coherent-state representation for the radiation field by Glauber[11] in 1963. When these models are generalized by including fermion operators in the Hamiltonian,

as in the various models of laser theory, the formal methods developed previously had permitted the consideration of only a *finite* number (usually one) of modes of an interacting boson field, if an exact solution was to be achieved. It is in this sense that the work presented in Refs. (1—7) goes beyond previous work in radiation theory, since *both* fermion and boson operators are involved in the Hamiltonian (the α and a_λ in the expression for H), while at the same time an *infinite* number of modes of the boson field is considered. In particular, the methods developed by the authors allow the exact solution to a Hamiltonian model for the spontaneous emission of an excited two-level quantum system in a field of electromagnetic radiation (with the field de-excited initially), whether the system size is assumed to be of finite (IV) or of infinite (V) extent. Furthermore, our recent studies have shown that exact solutions can be obtained for the same model, even in the presence initially of an extra photon in the radiation field, whether the spectrum of modes of the field is assumed discrete (VII) or continuous (the present study). At this point it should be stressed that although the radiation-matter problems considered in this series were analyzed assuming the field of radiation was one dimensional, this is *not* an essential restriction in our approach. The sole purpose for confining the radiation field to one dimension is to allow the underlying mathematical structure of the problem to be more clearly exposed, and to facilitate the numerical computation of the final expressions obtained. Put differently, the crucial mathematical issues which must be faced if one hopes to obtain the exact solution to the general class of models cited above are already present in the one-dimensional problem; once these issues are resolved, provided one orders the modes λ of the field properly, there is no *a priori* theoretical reason which stands in the way of applying the techniques developed to problems for which the dimensionality of the field is greater than one, and in fact we are now in the process of doing so.

The main task in this paper has been to show that the results presented in VII for the dynamical evolution of our system in the $N=2$ sector can be extended to systems infinite in extent. It is not very pleasing that a procedure so simple in principle as the taking of a limit as one parameter tends to infinity should be so difficult in practice. The presence of so much nonanalyticity in the expressions involved is however one of the chief characteristics of the solutions found in this series of papers, and it is, of course, this nonanalyticity which causes the difficulties. The final answer, Eq. (43), is not of an excessively complicated form, and it may prove possible to find a way of deriving it less tortuous than that used here.

The numerical computations demonstrate clearly that the solution, Eq. (43), has all the properties with which physical intuition would wish to endow it. Since this is so, perhaps it is to the point to say here why we have thought it worth-while to work it out in such detail. Equation (43) is the *exact* solution of the model we have been considering in the appropriate sector. As such, it has enabled us to give, in Sec. III, a quite precise definition of what we mean by ergodicity in problems of this kind, and to point out how pathologies may arise which would remove the "nice" properties of the solution. It has, also, shown just what is involved in the seemingly straightforward process of taking the weak-coupling limit, and has allowed us to see what kinds of convergence, uniform, L^1, and so on, are involved in letting α tend to zero. It is plain, for example, that *it is the infinite-system limit which converts the time-reversible solution of VII into the irreversible solution in Eq.* (43). On the other hand, *it is the weak-coupling limit that imposes exponential form on the time-dependence* of the evolution of the system.

It is not unreasonably our hope that the methods of this paper and of VII will be of use in providing exact solutions to more complicated matter-radiation models. We may draw attention to our paper on the exact dynamics of three-level systems[12] and the projected applications of this work to radiationless transitions in aromatic molecules, processes with phonons in solids, as well as the phenomena of phosphorescence and fluorescence. Here we may claim that since the methods introduced here lead to exact expressions for the time evolution of $\rho(\tau)$ (for different initial conditions), the results generated may be of use in evaluating approximate theories advanced in recent years to describe the above-mentioned phenomena. We have also noted that the nonergodicity uncovered in V in the time evolution of $\rho(\tau)$ for sufficiently large values of the coupling constant α is intimately linked to the onset of super-radiant emission, and we conjecture that this nonergodicity may also be related to the onset of enhanced conductivity in one-dimensional (i.e., polymeric) systems.

Over and above the practical applications noted above, we believe that the results presented in this paper bring us a step closer to realizing the overall goal which motivated the work documented in Refs. 1—7: to cast some light on the nature of irreversibility by working out the *exact* dynamics for Hamiltonian models of excited atomic and molecular systems. In this regard, we may point out that sectors of the two-level problem with N greater than 3 can be treated by the methods introduced here; there is every reason to hope that a genuine thermodynamic system can be discussed in which the system size becomes infinite along with the number N in such a way that a finite density of radiation is maintained over all space. Thus, in conclusion, we should like to claim that, although our mathematical methods are still complicated and delicate, they are at least novel, they permit exact solution of interesting and previously unsolved models, and seem capable of much further extension.

APPENDIX A

It will be sufficient for the proof of Eq. (42) if it can be proved that

$$-\sum_\mu \frac{1}{\hat{H}'(\gamma_\mu)}\hat{G}(\gamma_\mu)\overline{\hat{G}(\gamma_\mu)} = \int_0^\infty d\xi\, H(\xi)\,|g(\xi)|^2 = 1. \tag{A1}$$

The discrete sum is the easier calculation and will be

tackled first. Let us define, analogously to Eq. (9), the function:

$$\tilde{G}(\xi)=\frac{\sqrt{2}}{\sigma}\sum_{\lambda}\frac{\overline{f^{1/2}(\alpha\beta_\lambda)\psi_\lambda}}{\beta_\lambda-\xi},$$

of which the limit as $\sigma^2\to\infty$ is readily seen to be [compare Eq. (38)]

$$\left(\frac{2}{\pi}\right)^{1/2}\int_0^\infty\frac{\overline{f^{1/2}(\alpha\lambda)\psi(\lambda)}\,d\lambda}{\lambda-\xi}.$$

Since γ_μ is real, it follows, that

$$\overline{\hat{G}(\gamma_\mu)}=\tilde{G}(\gamma_\mu).$$

For the $\sigma^2\to\infty$ limit, the corresponding result is that

$$\overline{\tilde{G}^{\pm}(\xi)}=\hat{G}^{\mp}(\xi) \tag{A2}$$

(for real ξ).

Let us now consider the contour integral

$$\frac{1}{2\pi i}\int_S\frac{\hat{G}(\xi)\tilde{G}(\xi)\,d\xi}{\hat{H}(\xi)},$$

where S is a large circle $|\xi|=R$, in the limit $R\to\infty$. Because

$$\lim_{\xi\to\infty}\frac{\xi\hat{G}(\xi)\tilde{G}(\xi)}{\hat{H}(\xi)}=0,$$

the integral vanishes. The sum of the residues at the poles of the integrand in the finite part of the ξ plane is therefore zero. At the zeros γ_μ of $\hat{H}(\xi)$, the residues are the quantities

$$\frac{\hat{G}(\gamma_\mu)\tilde{G}(\gamma_\mu)}{\hat{H}'(\gamma_\mu)}=\frac{\hat{G}(\gamma_\mu)\overline{\hat{G}(\gamma_\mu)}}{\hat{H}'(\gamma_\mu)}.$$

There are other poles at the points $\xi=\beta_\nu$, at which the residues can easily be computed from Eqs. (16) and (9). The result is just $|\psi_\nu|^2$. We obtain then that

$$-\sum_\mu\frac{\hat{G}(\gamma_\mu)\overline{\hat{G}(\gamma_\mu)}}{\hat{H}'(\gamma_\mu)}=\sum_\nu|\psi_\nu|^2=1,$$

by the normalization of the initial state. This is the desired result.

For the integral, we may begin by noting that, from the definition of $g(\zeta)$, Eq. (40) and from Eq. (A2), we have

$$2\pi i\mathcal{H}(\xi)|g(\xi)|^2=\left(\frac{\hat{G}^+}{\hat{H}^+}-\frac{\hat{G}^-}{\hat{H}^-}\right)\left(\frac{\tilde{G}^+\hat{H}^--\tilde{G}^-\hat{H}^+}{\hat{H}^--\hat{H}^+}\right).$$

The argument, ζ, of all the functions has been omitted for ease of notation, and use has been made of the fact that

$$\overline{\hat{H}^{\pm}}=\hat{H}^{\mp},$$

which follows directly from Eq. (30). We have, then, that

$$2\pi i\mathcal{H}(\xi)|g(\xi)|^2=\frac{\{\hat{G}^+\tilde{G}^+(\hat{H}^-)^2+\hat{G}^-\tilde{G}^-(\hat{H}^+)^2-\hat{G}^+\tilde{G}^-\hat{H}^+\hat{H}^--\hat{G}^-\tilde{G}^+\hat{H}^+\hat{H}^-\}}{\hat{H}^+\hat{H}^-(\hat{H}^--\hat{H}^+)}. \tag{A3}$$

The result that

$$\int_S\frac{\hat{G}(\xi)\tilde{G}(\xi)\,d\xi}{\hat{H}(\xi)}=0$$

holds whether or not σ^2 becomes infinite, and may be written, in the limit, as

$$\int_0^\infty\left(\frac{\hat{G}^+\tilde{G}^+}{\hat{H}^+}-\frac{\hat{G}^-\tilde{G}^-}{\hat{H}^-}\right)d\zeta=0.$$

For the purposes of evaluating

$$\int_0^\infty\mathcal{H}(\zeta)|g(\zeta)|^2\,d\zeta,$$

then, the quantity

$$-\frac{1}{2\pi i}\left(\frac{\hat{G}^+\tilde{G}^+}{\hat{H}^+}-\frac{\hat{G}^-\tilde{G}^-}{\hat{H}^-}\right)$$

may be added to the integrand without changing the integral. Then, from Eq. (A3), we have

$$\int_0^\infty\mathcal{H}(\zeta)|g(\zeta)|^2\,d\zeta=\frac{1}{2\pi i}\int_0^\infty\left\{\frac{\hat{G}^+\tilde{G}^+(\hat{H}^-)^2+\hat{G}^-\tilde{G}^-(\hat{H}^+)^2-\hat{G}^+\tilde{G}^-\hat{H}^+\hat{H}^--\hat{G}^-\tilde{G}^+\hat{H}^+\hat{H}^-}{\hat{H}^+\hat{H}^-(\hat{H}^--\hat{H}^+)}-\frac{\hat{G}^+\tilde{G}^+}{\hat{H}^+}+\frac{\hat{G}^-\tilde{G}^-}{\hat{H}^-}\right\}d\zeta.$$

The integrand can be simplified, and may be written as

$$\frac{(\hat{G}^+-\hat{G}^-)(\tilde{G}^+-\tilde{G}^-)}{(\hat{H}^--\hat{H}^+)}.$$

In view of Eqs. (39), (30), and (A2), we have

$$\hat{G}^+-\hat{G}^-=2i\sqrt{2\pi}\,f^{1/2}(\alpha\zeta)\psi(\zeta),$$

$$\tilde{G}^+-\tilde{G}^-=2i\sqrt{2\pi}\,\overline{f^{1/2}(\alpha\zeta)\psi(\zeta)},$$

$$\hat{H}^--\hat{H}^+=4if(\alpha\zeta).$$

Therefore,

$$\int_0^\infty\mathcal{H}(\zeta)|g(\zeta)|^2\,d\zeta=\frac{1}{2\pi i}\int_0^\infty\frac{-8\pi f(\alpha\zeta)|\psi(\zeta)|^2}{4if(\alpha\zeta)}\,d\zeta=\int_0^\infty|\psi(\zeta)|^2\,d\zeta=1$$

by normalization. This completes the proof of Eq. (A1).

APPENDIX B

The aim of this Appendix is to prove the result (55):

$$\int_0^\tau d\theta\exp(i\zeta\theta)f_1(\tau-\theta)\left[F(\theta)+i\int_0^\theta d\tau'\,F(\tau')h(\theta-\tau')\right]\le N\sqrt{\alpha}$$

for all $\zeta\ge0$ and with N independent of both ζ and α. The coupling constant α may be restricted to small positive values.

The first step is to show that for all τ, $\alpha h(\tau)\to2$ as $\alpha\to0$, uniformly in τ. The definition of $h(\tau)$ is given in Eq. (46). Let us begin by considering the integral

$$\int_0^\infty d\xi\,\frac{\mathcal{H}(\xi)}{\alpha\hat{H}^+(\zeta-\xi)}.$$

It can be written as

$$\int_{-\infty}^\infty d\lambda\,\frac{\mathcal{H}(\lambda+1/\alpha)}{\alpha\hat{H}^+(-\lambda-1/\alpha+\zeta)}.$$

Since $\mathcal{H}(\lambda+1/\alpha)$ is dominated by the function $M(\lambda)$ defined in Eq. (51), it is permissible, if we wish to apply

the Lebesgue convergence theorem, to consider the integral

$$\int_{-\infty}^{\infty} d\lambda \frac{M(\lambda)}{\alpha \hat{H}^+(-\lambda - 1/\alpha + \zeta)} .$$

First, let

$$\int_{-\infty}^{\infty} d\lambda M(\lambda) = A.$$

Then let us define

$$\tilde{M}(\lambda) = \begin{cases} M(\lambda), & \lambda < -\frac{1}{2},\ \lambda > \frac{1}{2} \\ M(\lambda) - A, & -\frac{1}{2} \leq \lambda \leq \frac{1}{2} . \end{cases}$$

Clearly

$$\int_{-\infty}^{+\infty} d\lambda \tilde{M}(\lambda) = 0. \tag{B1}$$

Consider next the sectionally holomorphic function (see Muskhelishvili[13])

$$\mathcal{H}(\zeta) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{d\lambda \tilde{M}(\lambda)}{\lambda - \zeta} .$$

The function $\mathcal{H}$ is holomorphic in both the upper and lower half-planes away from the real axis of ζ. Its limiting values are (from above)

$$\mathcal{H}^+(\lambda) = \tfrac{1}{2} \tilde{M}(\lambda) + \frac{1}{2\pi i} P \int_{-\infty}^{\infty} \frac{d\lambda' \tilde{M}(\lambda')}{\lambda' - \lambda}$$

(from below)

$$\mathcal{H}^-(\lambda) = -\tfrac{1}{2} \tilde{M}(\lambda) + \frac{1}{2\pi i} P \int_{-\infty}^{\infty} \frac{d\lambda' \tilde{M}(\lambda')}{\lambda' - \lambda}$$

so that

$$\tilde{M}(\lambda) = \mathcal{H}^+(\lambda) - \mathcal{H}^-(\lambda).$$

Because of Eq. (B1), $\mathcal{H}$ tends to zero faster than $1/\zeta$ as $\zeta \to \infty$. By use of the definition of $\hat{H}^+$, Eq. (30), we obtain

$$\int_{-\infty}^{\infty} d\lambda \frac{\tilde{M}(\lambda)}{\alpha \hat{H}^+(-\lambda - 1/\alpha + \zeta)} = \int_{-\infty}^{\infty} d\lambda [\mathcal{H}^+(\lambda) - \mathcal{H}^-(\lambda)] \left(2 + \alpha\lambda - \alpha\zeta + \frac{2\alpha}{\pi} \times P \int_0^{\infty} \frac{d\mu f(\alpha\mu)}{\lambda + 1/\alpha - \zeta - \mu} - 2i\alpha f(-1 - \alpha\lambda + \alpha\zeta) \right)^{-1}. \tag{B2}$$

The denominator vanishes at a point λ_0, say, in the upper half-plane of λ, this point approaching $\lambda = -2/\alpha$ as α grows smaller. It follows then that the contribution to the integral coming from $\mathcal{H}^-(\lambda)$ vanishes. (Close the contour by a large semicircle in the lower half-plane: the integrand is holomorphic inside.) For the other contribution, we make use of a contour closed in the upper half-plane. There is a pole at $\lambda = \lambda_0$, with residue

$$\frac{\mathcal{H}^+(\lambda_0)}{\alpha + O(\alpha^2)} ,$$

and so the integral (B2) equals $2\pi i$ times this. Now

$$\lim_{\alpha \to 0} \frac{\mathcal{H}^+(\lambda_0)}{\alpha + O(\alpha^2)} = \lim_{\alpha \to 0} \frac{\mathcal{H}^+(-2/\alpha)}{\alpha} = 0,$$

since $\mathcal{H}(\zeta) \to 0$ as $\zeta \to \infty$ faster than $1/\zeta$. But since

$$\lim_{\alpha \to 0} \alpha \hat{H}^+(-\lambda - 1/\alpha + \zeta) = 2,$$

it is immediate that

$$\lim_{\alpha \to 0} \int_{-\infty}^{\infty} \frac{d\lambda \tilde{M}(\lambda)}{\alpha \hat{H}^+(-\lambda - 1/\alpha + \zeta)} = 0 = \int_{-\infty}^{\infty} d\lambda \tilde{M}(\lambda) \lim_{\alpha \to 0} \frac{1}{\alpha \hat{H}^+(-\lambda - 1/\alpha + \zeta)} .$$

One may now replace $\tilde{M}$ by M in the above equation. since

$$\frac{1}{\alpha \hat{H}^+(-\lambda - 1/\alpha + \zeta)}$$

tends *uniformly* for fixed ζ to $\frac{1}{2}$ for $-\frac{1}{2} \leq \lambda \leq \frac{1}{2}$. The conditions for the Lebesgue theorem (see again Ref. 9) are now satisfied, so that

$$\lim_{\alpha \to 0} \int_{-\infty}^{\infty} \frac{d\lambda H(\lambda + 1/\alpha)}{\alpha \hat{H}(-\lambda - 1/\alpha + \zeta)} = \int_{-\infty}^{+\infty} d\lambda \lim_{\alpha \to 0} \frac{H(\lambda + 1/\alpha)}{\alpha \hat{H}^+(-\lambda - 1/\alpha + \zeta)} = 2 \int_{-\infty}^{+\infty} d\lambda \lim_{\alpha \to 0} H(\lambda + 1/\alpha) = 2. \tag{B3}$$

We must now look at the integral

$$\int_0^{\zeta} d\xi H(\xi) H(\zeta - \xi).$$

This may be expressed as

$$\frac{1}{\alpha} \int_0^{\alpha\zeta} d\lambda \mathbf{H}(\lambda) \mathbf{H}(\alpha\zeta - \lambda)$$

where

$$\mathbf{H}(\lambda) \equiv H\left(\frac{\lambda}{\alpha}\right) = \alpha^2 \frac{2}{\pi} f(\lambda) \left[\left(1 - \lambda - \frac{2\alpha}{\pi} \int_0^{\infty} \frac{f(\mu)\, d\mu}{\mu - \lambda} \right)^2 + 4\alpha^2 f^2(\lambda) \right]^{-1}.$$

The mean value theorem gives an estimate of the integral for small α; it is

$$\zeta \frac{4\alpha^4}{\pi^2} f^2(\alpha\zeta\epsilon) \quad \text{for } 0 \leq \epsilon \leq 1.$$

It is now clear that the second term on the right-hand side of Eq. (46) vanishes uniformly as $\alpha \to 0$, and the first term, when multiplied by α, tends to $\frac{1}{2}$, by Eq. (B3). Thus indeed

$$\alpha h(\tau) \to 2 \text{ uniformly in } \tau \text{ as } \alpha \to 0. \tag{B4}$$

For the next stage of the proof, we require a refinement of Eqs. (52) and (54). It has already been stated that the convergence in these results is uniform with respect to τ. We now claim that convergence holds also in the L^1 norm. This means, for Eq. (52), that

$$\lim_{\alpha \to 0} \int_0^{\infty} d\tau \left| \exp(i\tau/\alpha) f_1(\tau) - \exp(iP\tau) \exp(-2\tau) \right| = 0 \tag{B5}$$

and there is a similar result for Eq. (54). Equation (B5) holds because everything is a Fourier transform of something else, as follows:

$$\lim_{\alpha \to 0} \int_0^{\infty} d\tau \left| \exp(i\tau/\alpha) f_1(\tau) - \exp(iP\tau) \exp(-2\tau) \right| = \lim_{\alpha \to 0} \int_0^{\infty} d\tau \left| \int_{-\infty}^{\infty} d\lambda \exp(-i\lambda\tau) \times [H(\lambda + 1/\alpha) - \lim_{\alpha \to 0} H(\lambda + 1/\alpha)] \right|$$

$$\leq \lim_{\alpha\to 0} 2\int_0^\infty d\tau \int_{-\infty}^{\infty} d\lambda \cos\lambda\tau \times \left| \mathcal{H}(\lambda + 1/\alpha) - \lim_{\alpha\to 0} \mathcal{H}(\lambda + 1/\alpha) \right| = \lim_{\alpha\to 0} 2\pi \left| \mathcal{H}(1/\alpha) - \lim_{\alpha\to 0} \mathcal{H}(1/\alpha) \right| = 0.$$

The usual Fourier inversion formula has been used and is valid because

$$\mathcal{H}(\lambda + 1/\alpha) = \left| \mathcal{H}(\lambda + 1/\alpha) \right|$$

is, for each value of α, an L^1 function, i.e.,

$$\int_{-\infty}^{\infty} \left| \mathcal{H}(\lambda + 1/\alpha) \right| d\lambda < \infty.$$

The argument above is clearly valid also for Eq. (54).

Next, let us examine the quantity

$$\sqrt{\alpha}\,[F(\theta) + i\int_0^\theta d\tau'\, F(\tau')h(\theta - \tau')]. \tag{B6}$$

This equals

$$\exp(-2i\theta/\alpha)[\sqrt{\alpha}\exp(2i\theta/\alpha)F(\theta)] + i\int_0^\theta d\tau' \exp(-2i\tau'/\alpha)(1/\alpha)[\sqrt{\alpha}\exp(2i\tau'/\alpha)F(\tau') \times \alpha h(\theta - \tau')]. \tag{B7}$$

Let the expression in square brackets be denoted by $R(\tau', \theta, \alpha)$. Then the integral term in the expression (B7) can be written, after a change of variable $\tau' = \alpha\sigma$, as

$$i\int_0^{\theta/\alpha} d\sigma \exp(-2i\sigma)R(\alpha\sigma, \theta, \alpha).$$

Now, because $\alpha h(\tau)$ converges uniformly in τ to 2 as $\alpha \to 0$, it follows from Eq. (54) that, for any θ, $R(\alpha\sigma, \theta, \alpha)$ can be made, for small enough α, as close in the L^1 norm as we please to $R(\alpha\sigma, \theta, 0)$, which is

$$2i\sqrt{2/\pi}\,Q\exp(2iP\alpha\sigma)\exp(-4\alpha\sigma)$$

for any value of θ. Therefore the integral under consideration can be made as close as we please, uniformly in θ, to

$$i\int_0^{\theta/\alpha} d\sigma \exp(-2i\sigma)2i\left(\frac{2}{\pi}\right)^{1/2} Q\exp(2iP\alpha\sigma)\exp(-4\alpha\sigma) = i\left(\frac{2}{\pi}\right)^{1/2} Q\,\frac{1 - \exp(-2i\theta/\alpha)\exp(2iP\theta)\exp(-4\theta)}{1 - \alpha P - 2i\alpha}.$$

Similarly

$$\exp(-2i\theta/\alpha)(\sqrt{\alpha}\exp(2i\theta/\alpha)F(\theta))$$

is as close as we please, uniformly in θ, to

$$i\sqrt{2/\pi}\,Q\exp(-2i\theta/\alpha)\exp(2iP\theta)\exp(-4\theta),$$

so that the expression (B6) is close, in the same sense, to

$$i\left(\frac{2}{\pi}\right)^{1/2} Q \times \left[\frac{1}{1 - \alpha(P + 2i)} - \frac{\alpha(P + 2i)\exp(-2i\theta/\alpha)\exp(2iP\theta)\exp(-4\theta)}{1 - \alpha(P + 2i)}\right]. \tag{B8}$$

Finally, we may consider

$$\frac{1}{\alpha^{1/2}}\int_0^\tau d\theta \exp(i\zeta\theta)f_1(\tau - \theta)[F(\theta) + i\int_0^\theta d\tau'\, F(\tau')h(\theta - \tau')].$$

This may be written as follows:

$$\frac{\exp(-i\tau/\alpha)}{\alpha}\int_0^\tau d\theta \exp(i\zeta\theta)\exp(i\theta/\alpha)\{\exp[i(\tau - \theta)/\alpha]f_1(\tau - \theta)\} \times \sqrt{\alpha}\left[F(\theta) + i\int_0^\theta d\tau'\, F(\tau')h(\theta - \tau')\right].$$

Here

$$\exp[i(\tau - \theta)]f_1(\tau - \theta) \xrightarrow{L_1} \exp[iP(\tau - \theta)]\exp[-2(\tau - \theta)]$$

and this result, coupled with the uniformly approximating expression (B8), means that everything except the preliminary exponential factors in the integrand is close in L^1 norm to

$$i\left(\frac{2}{\pi}\right)^{1/2} Q\exp[iP(\tau - \theta)]\exp[-2(\tau - \theta)] \times \left[\frac{1 - \alpha(P + 2i)\exp(-2i\theta/\alpha)\exp(2iP\theta)\exp(-4\theta)}{1 - \alpha(P + 2i)}\right].$$

For fixed τ, then, the integral is close, in the ordinary sense, to

$$\frac{\exp(-i\tau/\alpha)}{\alpha}\int_0^\tau d\theta \exp(i\zeta\theta)\exp(i\theta/\alpha) \times i\left(\frac{2}{\pi}\right)^{1/2} Q\exp[iP(\tau - \theta)]\exp[-2(\tau - \theta)] \times \left[\frac{1 - \alpha(P + 2i)\exp(-2i\theta/\alpha)\exp(2iP\theta)\exp(-4\theta)}{1 - \alpha(P + 2i)}\right].$$

This expression may be evaluated, and it equals

$$\frac{i\exp(-i\tau/\alpha)\exp(iP\tau)\exp(-2\tau)\sqrt{2/\pi}\,Q}{1 - \alpha(P + 2i)} \times \left[\frac{\exp[i(\zeta - P)\tau]\exp(2\tau)\exp(i\tau/\alpha) - 1}{2\alpha + i(1 + \alpha(\zeta - P))} + \frac{(P + 2i)\{\exp[i(\zeta + P)\tau]\exp(-2\tau)\exp(-i\tau/\alpha) - 1\}}{2 + i(1/\alpha - \zeta - P)}\right].$$

Provided, then, that $\zeta \geq 0$, and $\alpha < 1/2P$, it is clear that this expression is bounded uniformly in ζ and α for fixed τ. If this bound is called N, then the inequality (55) is proved.

[1] R. Davidson and J.J. Kozak, J. Math. Phys. **11**, 189 (1970).
[2] R. Davidson and J.J. Kozak, J. Math. Phys. **11**, 1420 (1970).
[3] R. Davidson and J.J. Kozak, J. Math. Phys. **12**, 903 (1971).
[4] R. Davidson and J.J. Kozak, J. Math. Phys. **14**, 414 (1973).
[5] R. Davidson and J.J. Kozak, J. Math. Phys. **14**, 423 (1973).
[6] J.J.-W. Yang, R. Davidson, and J.J. Kozak, J. Math. Phys. **15**, 491 (1974).
[7] R. Davidson and J.J. Kozak, J. Math. Phys. **16**, 1013 (1975).
[8] For Jordan's lemma: P. Dennery and A. Krzywicki, *Mathematics for Physicists* (Harper and Row, New York, 1967), pp. 56—57.
[9] H.L. Royden, *Real Analysis* (Macmillan, New York, 1968), pp. 88—89.
[10] *Lectures in Theoretical Physics* (Boulder, 1960), edited by W.E. Brittin, B.W. Downs and J. Downs (Interscience, New York, 1961), Vol. III.
[11] R.J. Glauber, Phys. Rev. **131**, 2766 (1963).
[12] R. Davidson and J.J. Kozak, J. Math. Phys. **17**, 1692 (1976).
[13] N.I. Muskhelishvili, *Singular Integral Equations* (Noordhoff, Groningen, 1953).

On time evolution for large quantum systems in superspace

Nobuhiko Mishima

Department of Physics, Tokyo Gakugei University, Koganei-shi, Tokyo, Japan

Tomio Yamakoshi Petrosky

Department of Physics, Science University of Tokyo, Kagurazaka Shinjuku-ku, Tokyo, Japan
(Received 3 October 1977)

A reformulation in a superspace is given for Van Hove and Janner's perturbation theory based on the two-resolvent method for nonequilibrium quantum statistical mechanics. This is attained by introducing an ordered superoperator in the superspace. As an application of our formalism, the derivation of an asymptotic generalized master equation in a Markovian form is given. Further, it is investigated how a quantum subdynamics theory similar to the Brussels school's one can be formally constructed. Our subdynamics theory is formulated on every eigenspace of a superoperator (the ss-Hamiltonian) of which the eigenvalue is related to the intrinsic parameter E of Van Hove and Janner's generalized master equation.

1. INTRODUCTION

Through the formulation of generalized master equations, the study of irreversible processes in nonequilibrium quantum statistical mechanics has recently made remarkable progress. In this direction there are two different approaches.

One of them, developed by Van Hove[1] and by Janner,[2] is based on the "two-resolvent method," where the von Neumann equation is solved by using the two resolvent operators of the total Hamiltonian in the ordinary Hilbert space and by perturbation theory. They have derived a generalized master equation which gives the microcanonical distribution function as the asymptotic solution in time.

Another approach developed by the Brussels school led by Prigogine[3-8] is based on the "one-resolvent method," where the von Neumann equation is solved by using the single resolvent superoperator of the so-called "super-Liouvillian" in the superspace by perturbation theory. They have derived another generalized master equation which gives the canonical distribution function as the asymptotic solution. One of the interesting results of this method is the discovery of the "subdynamics" theory. They show that the exact time evolution of a large system governed by the von Neumann equation can be decomposed into two independent evolutions occurring in the complementary subspaces; one of these evolutions is exactly governed by the asumptotic master equation and another by the complementary master equation describing the short time behavior of the system. Recently, as an extension of the work of Pytte,[9] Grecos *et al.*[10] has shown that without the perturbation schemes this subdynamics theory can be reformulated on the basis of the eigenvalue problem of the collision kernel of the generalized master equation under an assumption of the analytic properties of the resolvent.

In this paper we give a reformulation of the two-resolvent method given by Van Hove and by Janner into the superspace[11-15] and show how the superspace provides a suitable mathematical background for the two-resolvent method. And, as an application of this reformulation, we show that it is possible to construct formally a quantum subdynamics theory similar to Brussels school under a similar argument to that of Pytte and Grecos *et al.* in the one-resolvent method. Namely, we set up the assumption that in the two-resolvent method the singularities of the product of the two analytically continued resolvents near the real axis are isolated simple poles arising from the eigenvalues of the collision kernel of the generalized master equation.

Our reformulation in the superspace is achieved in a very natural way by introducing an "ordered superoperator" which is a linear operator mapping the superspace on itself. As a result, we can introduce the two basic ordered superoperators, the "symmetrized super-Hamiltonian" and the "antisymmetrized super-Hamiltonian" (the so-called super-Liouvillian), which are related to the two intrinsic parameters of the two-resolvent method, respectively. We derive a master equation for the asymptotic evolution superoperator in a Markovian form under the above-mentioned assumption and show that one of the parameters, E, in this equation is an eigenvalue of the symmetrized super-Hamiltonian. Then, from this asymptotic evolution superoperator, we construct two fundamental projection superoperators of the subdynamics theory. This fact shows that the subspaces projected by these projection superoperators lie completely in an eigenspace of the symmetrized super-Hamiltonian. We further show that one of these subspaces contains the microcanonical stationary solution of the von Neumann equation characterized by an eigenvalue of the total Hamiltonian of the system in the ordinary Hilbert space.

In the next section we give a brief explanation of the superspace. In Sec. 3 the perturbation theory given by Van Hove and by Janner is reformulated in terms of the superspace. In Sec. 4 we give the master equation for the asymptotic evolution superoperator. Section 5 is devoted to deriving the exact stationary solution of this master equation. Then we show that this solution is just the projection superoperator which projects the microcanonical subspace characterized by an eigenvalue of the ordinary total Hamiltonian. Finally, we given the quantum subdynamics theory in Sec. 6.

2. FORMALISM

We start with a brief explanation of the superspace[11-15] which is the mathematical foundation of our theory.

 0022-2488/78/1905-1087$1.00

As in the ordinary formulation of quantum mechanics the following is assumed: The quantum Hilbert space $\mathfrak{H}$ is spanned by the basic set $\{|\alpha\rangle, |\beta\rangle, \cdots\}$, which are eigenstates of an observable with discrete eigenvalues (for simplicity, our discussions are restricted to the case of discrete spectra). This basis forms a complete orthonormal set as usual, that is,

$$\langle \alpha | \beta \rangle = \delta_{\alpha,\beta}, \quad \sum_{\alpha} |\alpha\rangle \langle \alpha| = \mathbf{1}. \tag{2.1}$$

As is well known, linear operators in $\mathfrak{H}$ form a vector space under their addition and multiplication by numbers. Hence, a new vector space, consisting of the operators $\{X, Y, \cdots\}$ in $\mathfrak{H}$, is introduced and these elements are represented by curly ket-vectors $\{|X), |Y), \cdots\}$ as in Dirac's notation. Then, in this vector space,

$$|\lambda X + \mu Y) = \lambda |X) + \mu |Y), \tag{2.2}$$

where λ and μ are numbers. Adjoints of the curly ket-vectors are represented by bra-vectors. $\{(X|, (Y|, \cdots\}$. Further, the inner product between these vectors is defined by

$$(X|Y) = \mathrm{tr}[X^\dagger Y] = (Y^\dagger | X^\dagger), \tag{2.3}$$

where tr means the trace in $\mathfrak{H}$. This has the properties

$$(X|Y)^* = (Y|X), \tag{2.4}$$

where the asterisk means complex conjugate. Thus, we have a unitary space which is called the superspace, $\mathfrak{S}$ (or the operator space), and its element supervector (s-vector) or superstate (s-state).[16] The dyad operator, $|\alpha\rangle\langle\beta|$, constructed from basic set in $\mathfrak{H}$ is particularly denoted by $|\alpha, \beta)$ in $\mathfrak{S}$. It is characterized by (2.1) as orthonormal,

$$(\alpha, \beta | \alpha', \beta') = \delta_{\alpha,\alpha'} \delta_{\beta,\beta'}, \tag{2.5}$$

and for any s-vector, $|X)$, the following expansion holds,

$$|X) = \sum_{\alpha,\beta} |\alpha, \beta)(\alpha, \beta|X). \tag{2.6}$$

Therefore, the set of $|\alpha, \beta)$ forms a complete orthonormal basic set in $\mathfrak{S}$. The trace of the operator X in $\mathfrak{H}$ is expressed in $\mathfrak{S}$ by

$$\mathrm{tr}[X] = \sum_{\alpha} (\alpha, \alpha|X) = (1|X), \tag{2.7}$$

where 1 is the unit operator.

Superoperators (s-operators) $\mathcal{A}, \mathcal{B}, \mathcal{C}, \cdots$, which map the space $\mathfrak{S}$ onto itself can be introduced. The adjoint $\mathfrak{S}$ s-operator $\mathcal{A}^\dagger$ of $\mathcal{A}$ is such that

$$(X|\mathcal{A}^\dagger|Y) = (Y|\mathcal{A}|X)^*, \tag{2.8}$$

for every s-vectors $|X)$ and $|Y)$, and Hermitian and a unitary s-operator are defined by the s-operators satisfying $\mathcal{A}^\dagger = \mathcal{A}$ and $\mathcal{A}^\dagger\mathcal{A} = \mathcal{A}\mathcal{A}^\dagger = I$, respectively, where I is the unit s operator. The eigenvalue problems in $\mathfrak{S}$ are just the same as in $\mathfrak{H}$:

$$\mathcal{A}|A) = a|A), \tag{2.9}$$

where a is a number, then $|A)$ is called a supereigenvector of the s operator $\mathcal{A}$, and a is its eigenvalue. The well-known properties of the eigenvalues and the s-eigenvectors for a Hermitian and a unitary s-operator also hold; that is, the orthogonality of the s-eigenvectors and the spectral resolution of these s-operators, etc. From (2.6), we see that dyads of the orthonormal basic set in $\mathfrak{S}$ form the idemfactor,

$$\sum_{\alpha,\alpha'} |\alpha, \alpha')(\alpha, \alpha'| = I. \tag{2.10}$$

Then, we can express a s-operator $\mathcal{A}$ with its tetradic elements $(\alpha, \alpha'|\mathcal{A}|\beta, \beta')$ by

$$\mathcal{A} = \sum_{\alpha,\alpha'} \sum_{\beta,\beta'} |\alpha, \alpha')(\beta, \beta'|(\alpha, \alpha'|\mathcal{A}|\beta, \beta'). \tag{2.11}$$

We now introduce an ordered s-operator, which plays an essential role throughout this work, as follows: Corresponding to two operators A and B in $\mathfrak{H}$, the ordered s-operator $(A \wedge B)$ is defined such that

$$(A \wedge B)|X) = |AXB), \tag{2.12}$$

for any s-vector $|X)$. From this definition, we can easily obtain the formulas as, for the addition,

$$\begin{aligned} &(A \wedge C) + (B \wedge C) = [(A + B) \wedge C], \\ &(A \wedge B) + (A \wedge C) = [A \wedge (B + C)], \end{aligned} \tag{2.13}$$

and for the product,

$$(A \wedge B)(C \wedge D) = (AC \wedge DB). \tag{2.14}$$

From (2.8), the adjoint of the ordered s-operator has the property that

$$(A \wedge B)^\dagger = (A^\dagger \wedge B^\dagger). \tag{2.15}$$

Further, it is particularly convenient for brevity of our formalism to introduce the one side s-operators, which are defined such that either A or B in an ordered s-operator $(A \wedge B)$ is the unit operator in $\mathfrak{H}$. They are symbolized by the notations as

$$A^> \equiv (A \wedge 1), \quad B^< \equiv (1 \wedge B), \tag{2.16}$$

where $A^>$ is called a left-hand side s-operator (l.s-operator) and $B^<$ a right-hand side s-operator (r.s-operator), respectively. As easily seen, these one side s-operators have the following properties,

$$A^> B^< = B^< A^> = (A \wedge B), \tag{2.17}$$

and these permit us to express any ordered s-operator by the sums and the products of the one side s-operators. Here we summarize the useful formulas satisfied by the one side s-operators without proof [hereafter, the upper symbols > (the lower symbols <) always have to be taken together]:

$$(A + B)^{\gtrless} = A^{\gtrless} + B^{\gtrless}, \tag{2.18}$$

$$A^> B^> = (AB)^>, \quad A^< B^< = (BA)^<, \tag{2.19}$$

$$(A^n)^{\gtrless} = (A^{\gtrless})^n, \tag{2.20}$$

$$(e^A)^{\gtrless} = e^{A^{\gtrless}}, \quad e^{A^{\gtrless} + B^{\gtrless}} = e^{A^{\gtrless}} e^{B^{\gtrless}}, \tag{2.21}$$

$$\frac{1}{A^{\gtrless} - z} = \left(\frac{1}{A - z}\right)^{\gtrless}. \tag{2.22}$$

The commutator $[A, B]_\pm = AB \pm BA$ in $\mathfrak{H}$ and its n-fold commutator can also be expressed by the one side s-operators in the abbreviated form as

$$|[A, B]_\pm) = (A^> \pm A^<)|B) = (B^< \pm B^>)|A), \tag{2.23}$$

$$|[A, [A, \cdots [A, B]_\pm \cdots]_\pm]_\pm) = (A^> \pm A^<)^n |B). \tag{2.24}$$

Now we will give a general formalism necessary to study the irreversibility of many-body systems in the superspace. We consider the spatially homogeneous

system for which the Hamiltonian consists of a sum of an unperturbed Hamiltonian H_0 and a perturbation V, i.e.,

$$H = H_0 + V, \tag{2.25}$$

and assume that the whole set of eigenstates $|\alpha>$ of H_0, i.e.,

$$H_0|\alpha> = \epsilon_\alpha|\alpha>, \tag{2.26}$$

forms a complete orthonormal basic set as (2.1) and the Hilbert space $\mathfrak{H}$ describing the system can be spanned by them. Here, we denote by α a collection of the quantum variables characterizing an eigenstate of H_0, and ϵ_α is the unperturbed energy of the system. To simplify the mathematical arguments, we treat the system to be enclosed in a box as usual and α to be discrete. However, since we are considering a physically very large system having many degrees of freedom, the state density is practically so dense that we may replace the summation over the state by its integration if necessary.

In quantum statistical mechanics, the physical states of the system are described by the density operator which is expressed by a s-state as $|\rho(t))$ (density s-state) in $\mathfrak{S}$ and governed by the von Neumann equation,

$$i\partial_t|\rho(t)) = \tilde{H}|\rho(t)), \tag{2.27}$$

with the definition,

$$\tilde{H} \equiv H^> - H^<. \tag{2.28}$$

(We use the unit $\hbar = 1$). We call the s-operator $\tilde{H}$ antisymmetrized super-Hamiltonian (a.s-Hamiltonian) or super-Liouvillian, and $H^>$ and $H^<$ are called one-side s-Hamiltonians. Since $H^>$ and $H^<$ are Hermitians, $\tilde{H}$ is also Hermitian.

For the unperturbed Hamiltonian H_0, we can also define one-side s-Hamiltonians $H_0^>$, $H_0^<$ and the unperturbed a.s-Hamiltonian $\tilde{H}_0 \equiv H_0^> - H_0^<$. They satisfy the following eigenvalue equations,

$$\begin{aligned} &H_0^>|\alpha,\alpha') = \epsilon_\alpha|\alpha,\alpha'), \quad H_0^<|\alpha,\alpha') = \epsilon_{\alpha'}|\alpha,\alpha'), \\ &\tilde{H}_0|\alpha,\alpha') = (\epsilon_\alpha - \epsilon_{\alpha'})|\alpha,\alpha'), \end{aligned} \tag{2.29}$$

and s-eigenstates $|\alpha,\alpha')$ form a complete basic set in $\mathfrak{S}$.

The density s-state has the following properties:

$$\begin{aligned} &(\alpha,\alpha'|\rho(t)) = (\rho(t)|\alpha,\alpha'), \\ &(\alpha,\alpha|\rho(t)) \geq 0, \quad (1|\rho(t)) = 1, \end{aligned} \tag{2.30}$$

which correspond to hermiticity, positive definite, and normalization, respectively. The expectation value of an observable A is given by

$$\langle A\rangle_t = (\rho(t)|A). \tag{2.31}$$

In order to separate the density s-state into two parts, the diagonal part and the nondiagonal part, it is convenient for us to introduce projection s-operators (projectors) defined by

$$\mathcal{P} \equiv \sum_\alpha p_\alpha^> p_\alpha^<, \quad \mathcal{Q} \equiv I - \mathcal{P}, \tag{2.32}$$

with $p_\alpha \equiv |\alpha><\alpha|$, from which it follows that

$$\mathcal{P}^2 = \mathcal{P}, \quad \mathcal{P}\mathcal{Q} = \mathcal{Q}\mathcal{P}, \quad \mathcal{Q}^2 = \mathcal{Q}, \tag{2.33}$$

$$H_0^\gtrless \mathcal{P} = \mathcal{P} H_0^\gtrless, \quad H_0^\gtrless \mathcal{Q} = \mathcal{Q} H_0^\gtrless. \tag{2.34}$$

By using these projectors, the density s-state is decomposed into two parts:

$$\begin{aligned} &|\rho) = |\rho_d) + |\rho_{nd}), \\ &|\rho_d) \equiv \mathcal{P}|\rho), \quad |\rho_{nd}) \equiv \mathcal{Q}|\rho), \end{aligned} \tag{2.35}$$

where the diagonal part $|\rho_d)$ is called the vacuum component and the nondiagonal part $|\rho_{nd})$ the correlation component.[3]

On the basis of the one-side resolvent s-operators, i.e.,

$$R^\gtrless(z) = \frac{1}{\tilde{H}^\gtrless - z} = \left(\frac{1}{H - z}\right)^\gtrless, \tag{2.36}$$

the formal solution of the von Neumann Eq. (2.27) can be obtained by the two-resolvent method of Van Hove and Janner such as

$$|\rho(t)) = \mathcal{U}(t)|\rho(0)), \tag{2.37}$$

with the time evolution unitary s-operator,

$$\begin{aligned} \mathcal{U}(t) &= e^{-i\tilde{H}t} \\ &= \left(\frac{-1}{2\pi i}\right)^2 \int_\Gamma dz \int_{\Gamma'} dz'\, e^{-i(z-z')t} R^>(z) R^<(z'). \end{aligned} \tag{2.38}$$

Here, the paths of integrations Γ and Γ' indicated in Fig. 1 are any positive contours completely enclosing the real axes in complex z and z' planes, respectively. For $t \neq 0$, the paths of z and z' either on upper or on lower half planes shown in Fig. 1 do not contribute to (2.38). Then, by interchanging the integral variables, it can be written by

$$\mathcal{U}(t) = \int_{-\infty}^{+\infty} dE\, \mathcal{U}_E(t), \tag{2.39}$$

with

$$\mathcal{U}_E(t) = \frac{s(t)}{(2\pi i)^2} \int_\gamma dl\, e^{-ilt} R^>(E - \tfrac{1}{2}l) R^<(E - \tfrac{1}{2}l), \tag{2.40}$$

where $s(t) \equiv t/|t|$ and the path of integration γ is any positive contour completely enclosing the real axis in the complex l plane.

On the other hand, for $t = 0$, it is noticeable that all contours shown in Fig. 1 contribute to (2.38) and hence $\mathcal{U}_E(0)$ can be written with (2.39) by

$$\begin{aligned} &\mathcal{U}_E(0) \\ &= \lim_{\eta \to 0} \frac{1}{2\pi i}\left(\frac{1}{\tilde{H} - E - i\eta} - \frac{1}{\tilde{H} - E + i\eta}\right) \equiv \hat{P}_E, \end{aligned} \tag{2.41}$$

where η is a small positive quantity. Here, a symmetrized super-Hamiltonian (s.s-Hamiltonian) $\tilde{H}$ is defined

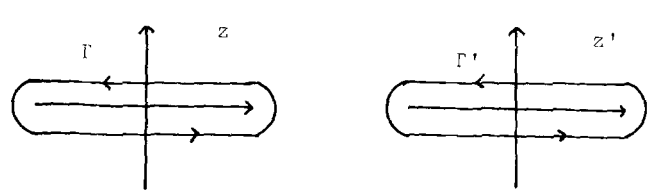

FIG. 1. Contours Γ and Γ'.

by

$$\hat{H} = (H^> + H^<)/2, \tag{2.42}$$

and $\hat{P}_E$ is the projector which belongs to an eigenvalue E of $\hat{H}$. It can satisfy the relations of the spectral resolution,

$$\hat{H} = \int_{-\infty}^{+\infty} dE\, \hat{P}_E E, \tag{2.43}$$

$$\hat{P}_E \hat{P}_{E'} = \delta(E - E')\hat{P}_E, \quad \int_{-\infty}^{+\infty} dE\, \hat{P}_E = I. \tag{2.44}$$

From the fact that $\hat{H}$ is commutable with $\tilde{H}$, $\hat{P}_E$ is also commutable with $\tilde{H}$ and hence we get

$$U(t)\hat{P}_E = \hat{P}_E U(t). \tag{2.45}$$

3. PERTURBATION THEORY

Here, we will discuss a reformulation of the perturbation theory given by Van Hove and Janner[1,2,17] in terms of the superspace.

The one-side resolvents $R^>(z)$ and $R^<(z)$ satisfy the integral equations,

$$R^{\gtrless}(z) = R_0^{\gtrless}(z) - R_0^{\gtrless}(z)V^{\gtrless}R^{\gtrless}(z), \tag{3.1}$$

with the unperturbed one-side resolvents of $H^>$ and $H^<$,

$$R_0^{\gtrless}(z) = \frac{1}{H_0^{\gtrless} - z} = \left(\frac{1}{H_0 - z}\right)^{\gtrless}. \tag{3.2}$$

(We note here that the relations which are given by dropping the symbols > and < from the above and forthcoming relations for one-side s-operators hold among the corresponding operators in $\mathfrak{H}$.) Their iterated solutions are

$$R^{\gtrless}(z) = R_0^{\gtrless}(z)\sum_{k=0}^{\infty}\{-V^{\gtrless}R_0^{\gtrless}(z)\}^k. \tag{3.3}$$

Using the projectors p_α in (2.32) for any one-side s-operators $A^{\gtrless}$, we can define their diagonal parts, $A_d^{\gtrless} \equiv \sum_\alpha p_\alpha^{\gtrless} A^{\gtrless} p_\alpha^{\gtrless}$ and their nondiagonal parts $A_{nd}^{\gtrless} \equiv \sum_\alpha (I - p_\alpha)A\, p_\alpha$. Then, the one-side resolvents are decomposed into the two parts as

$$R^{\gtrless}(z) = D^{\gtrless}(z) + R_{nd}^{\gtrless}(z), \quad D^{\gtrless}(z) \equiv R_d^{\gtrless}(z). \tag{3.4}$$

From (3.1), we have

$$D^{\gtrless} = R_0^{\gtrless} - R_0^{\gtrless}(V^{\gtrless}R^{\gtrless})_d = R_0^{\gtrless} - R_0^{\gtrless}V_d^{\gtrless}D^{\gtrless} - R_0^{\gtrless}(V_{nd}^{\gtrless}R_{nd}^{\gtrless})_d, \tag{3.5}$$

$$R_{nd}^{\gtrless} = -R_0^{\gtrless}(V^{\gtrless}R^{\gtrless})_{nd} = -R_0^{\gtrless}(V^{\gtrless}D^{\gtrless})_{nd} - R_0^{\gtrless}(V^{\gtrless}R_{nd}^{\gtrless})_{nd}. \tag{3.6}$$

(For simplicity, the complex variable z is omitted except when needed for clarity.) Iterating (3.6), we can get

$$R_{nd}^{\gtrless} = \{\sum_{k=1}^{\infty}(-R_0^{\gtrless}V^{\gtrless})^k\}_{tnd}D^{\gtrless}, \tag{3.7}$$

with the definition

$$\{ABC\cdots\}_{tnd} = (A(B(C\cdots)_{nd})_{nd})_{nd}, \tag{3.8}$$

which is called the topological irreducible nondiagonal part.[18] Substituting (3.3) into (3.7) with (3.4), we derive

$$R_{nd}^{\gtrless} = \{-(D^{\gtrless} + R_{nd}^{\gtrless})V^{\gtrless}\}_{tnd}D^{\gtrless}. \tag{3.9}$$

Let us define a s-operator $N^{\gtrless}(z)$ by the integral equation

$$N^{\gtrless}(z) = -V_{nd}^{\gtrless} - \{N^{\gtrless}(z)D^{\gtrless}(z)V^{\gtrless}\}_{tnd}, \tag{3.10}$$

and multiply both sides by $D^{\gtrless}$. Then, if we ignore the contribution from the terms of the occasional equality occurred between the intermediate states (these terms are called interlocking terms by Van Hove and he demonstrated that their contribution is proportional to one over the volume and thus vanishes in the limit of a large system[19]), we can find that $D^{\gtrless}N^{\gtrless}D^{\gtrless}$ satisfies the same Eq. (3.9) as $R_{nd}^{\gtrless}$. Thus, putting $R_{nd}^{\gtrless} = D^{\gtrless}N^{\gtrless}D^{\gtrless}$ in (3.4), we obtain

$$R^{\gtrless} = D^{\gtrless} + D^{\gtrless}N^{\gtrless}D^{\gtrless}. \tag{3.11}$$

By iterating (3.10), a solution for $N^{\gtrless}$ is derived as

$$N^{\gtrless} = \{-\sum_{k=0}^{\infty}V^{\gtrless}(-D^{\gtrless}V^{\gtrless})^k\}_{tnd}, \tag{3.12}$$

where $\{\ \}_{tnd}$ is defined as

$$\{\cdots BCD\}_{tnd} = (((\cdots B)_{tnd}C)_{tnd}D)_{tnd}. \tag{3.13}$$

It stands for the irreducible nondiagonal part of Van Hove and means a nondiagonal part which does not contain any equal state to another among the intermediate states. Further, with defining the irreducible diagonal part $\{\ \}_{id}$ by

$$\{\cdots BCD\}_{id} = (((\cdots B)_{tnd}C)_{tnd}D)_d, \tag{3.14}$$

a s-operator $G^{\gtrless}(z)$ is introduced by

$$G^{\gtrless}(z) = \{-\sum_{k=0}^{\infty}V^{\gtrless}(-D^{\gtrless}(z)V^{\gtrless})^k\}_{id}. \tag{3.15}$$

Then, using (3.12) and (3.15), we can rewrite (3.5) as

$$D^{\gtrless} = R_0^{\gtrless} + R_0^{\gtrless}G^{\gtrless}D^{\gtrless}. \tag{3.16}$$

This has a solution for D,

$$D^{\gtrless}(z) = \frac{1}{H_0^{\gtrless} - G^{\gtrless}(z) - z}, \tag{3.17}$$

which implies that

$$\begin{aligned} D^>(z)\,|\,\alpha,\beta) &= D_\alpha(z)\,|\,\alpha,\beta), \\ D^<(z)\,|\,\alpha,\beta) &= D_\beta(z)\,|\,\alpha,\beta), \end{aligned} \tag{3.18}$$

with the definitions

$$D_\alpha(z) \equiv \langle\alpha\,|\,D(z)\,|\,\alpha\rangle = \frac{1}{\epsilon_\alpha - D_\alpha(z) - z}, \tag{3.19}$$

$$\begin{aligned} G_\alpha(z) \equiv \langle\alpha\,|\{&-V + VD(z)V \\ &- VD(z)VD(z)V + \cdots\}_{id}\,|\,\alpha\rangle. \end{aligned} \tag{3.20}$$

In this perturbation theory, the analytic properties of the function $D_\alpha(z)$, $G_\alpha(z)$, and $\langle\alpha\,|N(z)\,|\beta\rangle$ $[\equiv N_{\alpha\beta}(z)]$ of (3.12) have been investigated in detail by Van Hove[1,19] and by Hugenholtz[20] in the limit of a large system. Their results are the following: Since the Hamiltonian is Hermitian, its resolvent $R(z)$ is holomorphic in z out of the real axis and $R^\dagger(z) = R(z^*)$. Then, the above functions have the same analytic property and hold the reality conditions

$$D_\alpha{}^*(z) = D_\alpha(z^*), \quad G_\alpha^*(z) = G_\alpha(z^*), \quad N_{\alpha\beta}{}^*(z) = N_{\beta\alpha}(z^*). \tag{3.21}$$

$G_\alpha(z)$ and $N_{\alpha\beta}(z)$ are bounded on the real axis and have finite discontinuities across the real axis in a certain interval. As can be seen from (3.19), $D_\alpha(z)$ also has finite discontinuities of the same kind, and in addition it can have poles. That is, for z approaching a point E on the real axis with $\pm i\eta$ where $\eta > 0$, $G_\alpha(z)$ becomes

as

$$\lim_{\eta\to 0} G_\alpha(E\pm i\eta)=K_\alpha(E)\pm iJ_\alpha(E), \tag{3.22}$$

with $K_\alpha(E)$ real and $J_\alpha(E)$ real nonnegative. Then, we get

$$D_\alpha(E\pm i\eta)=[\epsilon_\alpha - E - K_\alpha(E)\mp iJ_\alpha(E)]^{-1}, \tag{3.23}$$

and see that $D_\alpha(E)$ has poles only when the equation, $\epsilon_\alpha - E - K_\alpha(E)=0$, has roots in certain regions of E where $J_\alpha(E)=0$ is satisfied.

In this case where $D_\alpha(E)$ has poles, the state α is called nondissipative and otherwise dissipative. Our purpose is to study irreversible processes of large quantum systems, therefore we restrict ourselves to treating dissipative systems throughout this work.

Now, on the basis of (3.11) we can formulate the perturbation expansion for the product $R^>R^<$ decomposing into four parts, i.e., a diagonal part, a creation part, a destruction part, and a propagation of correlations part, in a similar way to the Brussels school.[3,5] By using projectors $\mathcal{P}$ and $\mathcal{Q}$ in (2.32), $R^>R^<$ can be written as a sum of four terms as

$$R^>R^< = \mathcal{P}R^>R^<\mathcal{P} + \mathcal{Q}R^>R^<\mathcal{P} + \mathcal{P}R^>R^<\mathcal{Q} + \mathcal{Q}R^>R^<\mathcal{Q}, \tag{3.24}$$

and using (3.11), we obtain the following relations:

$$\mathcal{P}R^>R^<\mathcal{P} = \mathcal{P}D^>D^<\mathcal{P} + \mathcal{P}D^>D^<N^>N^<D^>D^<\mathcal{P}, \tag{3.25}$$

$$\mathcal{Q}R^>R^<\mathcal{P} = \mathcal{Q}D^>N^>D^>D^<\mathcal{P} + \mathcal{Q}D^>D^<N^<D^<\mathcal{P} + \mathcal{Q}D^>D^<N^>N^<D^>D^<\mathcal{P}, \tag{3.26}$$

$$\mathcal{P}R^>R^<\mathcal{Q} = \mathcal{P}D^>N^>D^>D^<\mathcal{Q} + \mathcal{P}D^>D^<N^<D^<\mathcal{Q} + \mathcal{P}D^>D^<N^>N^<D^>D^<\mathcal{Q}, \tag{3.27}$$

$$\mathcal{Q}R^>R^<\mathcal{Q} = \mathcal{Q}D^>D^<\mathcal{Q} + \mathcal{Q}D^>N^>D^>D^<\mathcal{Q} + \mathcal{Q}D^>D^<N^<D^<\mathcal{Q} + \mathcal{Q}D^>D^<N^>N^<D^>D^<\mathcal{Q}. \tag{3.28}$$

Further, putting the iterated expansion (3.12) into the above relations and rearranging them, we can derive an integral equation for $\mathcal{P}R^>R^<\mathcal{P}$ from (3.25) and we can represent the other terms by its functionals. The details of the derivation will be explained in Appendix A by the means of a diagramatic method. Here, we will describe only the results

$$\begin{aligned}\mathcal{P}R^>(z)R^<(z')\mathcal{P} &= \mathcal{P}D^>(z)D^<(z')\mathcal{P} + \mathcal{P}D^>(z)D^<(z')\\ &\quad\times\mathcal{P}W(z,z')\mathcal{P}R^>(z)R^<(z')\mathcal{P}\\ &= \mathcal{P}D^>(z)D^<(z')\mathcal{P} + \mathcal{P}R^>(z)R^<(z')\mathcal{P}W(z,z')\mathcal{P}\\ &\quad\times D^>(z)D^<(z')\mathcal{P},\end{aligned} \tag{3.29}$$

$$\begin{aligned}R^>(z)R^<(z') &= [\mathcal{P}+\mathcal{Q}C(z,z')\mathcal{P}]R^>(z)R^<(z')\\ &\quad\times[\mathcal{P}+\mathcal{P}D(z,z')\mathcal{Q}] + \mathcal{Q}\mathcal{G}(z,z')\mathcal{Q},\end{aligned} \tag{3.30}$$

where irreducible s-operators are introduced by

$$W(z,z')=\mathcal{P}\{N^>(z)N^<(z')\}_{irr}\mathcal{P}, \tag{3.31}$$

the creation s-operator by

$$\mathcal{Q}C(z,z')\mathcal{P} = \mathcal{Q}[D^>(z)N^>(z) + D^<(z')N^<(z') + D^>(z)D^<(z')\{N^>(z)N^<(z')\}_{irr}]\mathcal{P}, \tag{3.32}$$

the destruction s-operator by

$$\mathcal{P}D(z,z')\mathcal{Q} = \mathcal{P}[N^>(z)D^>(z) + N^<(z')D^<(z') + \{N^>(z)N^<(z')\}_{irr}D^>(z)D^<(z')]\mathcal{Q}, \tag{3.33}$$

and the propagation of correlations s-operator by

$$\begin{aligned}\mathcal{Q}\mathcal{G}(z,z')\mathcal{Q} &= \mathcal{Q}D^>(z)D^<(z')[I + N^>(z)D^>(z)\\ &\quad + N^<(z')D^<(z') + \{N^>(z)N^<(z')\}_{irr}\\ &\quad\times D^>(z)D^<(z')]\mathcal{Q},\end{aligned} \tag{3.34}$$

with $\{N^>N^<\}_{irr}$ of which the tetradic element is given by

$$\begin{aligned}&(\alpha,\beta\,|\{N^>(z)N^<(z')\}_{irr}|\gamma,\delta)\\ &\quad=\langle\alpha|\{\sum_{k=0}^{\infty}[-V(-D(z)V)^k]_{nd}|\gamma><\delta|\\ &\quad\times\sum_{j=0}^{\infty}[-V(-D(z')V)^j]_{nd}\}_{irr}|\beta\rangle.\end{aligned} \tag{3.35}$$

The subscript "irr" means that all intermediate states are unequal to each other and to the initial and final states. Here, we remark that in our dissipative systems the irreducible s-operators W, $\mathcal{Q}C\mathcal{P}$, $\mathcal{P}D\mathcal{Q}$, and $\mathcal{Q}\mathcal{G}\mathcal{Q}$ have the analytic properties being finite and discontinuous across the real axis in z and z'.

A formal solution of (3.29) can be derived as follows. From (3.17) we have

$$\begin{aligned}D^>(z)-D^<(z') &= [z-z'-H_0^>+H_0^<+G^>(z)\\ &\quad - G^<(z')]D^>(z)D^<(z'),\end{aligned} \tag{3.36}$$

and taking the $\mathcal{P}$–$\mathcal{P}$ component of it, we get

$$\begin{aligned}\Delta(z,z') &= [z-z'+\mathcal{G}(z,z')]\\ &\quad\times\mathcal{P}D^>(z)D^<(z')\mathcal{P},\end{aligned} \tag{3.37}$$

where the s-operators, which are commutable with each other, are defined by

$$\Delta(z,z')\equiv\mathcal{P}[D^>(z)-D^<(z')]\mathcal{P}, \tag{3.38}$$

$$\mathcal{G}(z,z')\equiv\mathcal{P}[G^>(z)-G^<(z')]\mathcal{P}. \tag{3.39}$$

Then, by the aid of (3.29) and (3.37), we obtain the formal solution

$$\begin{aligned}\mathcal{P}R^>(z)R^<(z')\mathcal{P} &= [z-z'-\chi(z,z')]^{-1}\Delta(z,z')\\ &= \Delta(z,z')[z-z'-\tilde{\chi}(z,z')]^{-1},\end{aligned} \tag{3.40}$$

with the definitions of

$$\chi(z,z')\equiv[\Delta(z,z')W(z,z')-\mathcal{G}(z,z')], \tag{3.41}$$

$$\tilde{\chi}(z,z')\equiv[W(z,z')\Delta(z,z')-\mathcal{G}(z,z')], \tag{3.42}$$

which are called collision s-operators.

Thus, on the basis of the expression of (3.30) with (3.40), we can arrive at our basic representation for the evolution s-operator as

$$\begin{aligned}U_E(t) &= \frac{s(t)}{(2\pi i)^2}\int_\gamma dl\, e^{-ilt}\{(\mathcal{P}+\mathcal{Q}C_E(l)\mathcal{P})\\ &\quad\times[(l-\chi_E(l))^{-1}\Delta_E(l)\\ &\quad\text{or}\\ &\quad\Delta_E(l)(l-\tilde{\chi}_E(l))^{-1}](\mathcal{P}+\mathcal{P}D_E(l)\mathcal{Q})\\ &\quad+\mathcal{Q}\mathcal{G}_E(l)\mathcal{Q}\}\quad(\text{for } t\neq 0),\end{aligned} \tag{3.43}$$

where the abbreviated notation $\mathcal{J}_E(l)\equiv\mathcal{J}(E+l/2,\,E-l/2)$ is used.

In concluding this section, we will list some relations satisfied by our s-operators:

$$\Delta(z,z')=-\Delta(z',z),\quad \mathcal{G}(z,z')=-\mathcal{G}(z'z), \tag{3.44}$$

$$(\alpha,\alpha|W(z,z')|\beta,\beta)=(\beta,\beta|W(z'z)|\alpha,\alpha), \tag{3.45}$$

$$(\alpha,\alpha|\rho R^{>}(z)R^{<}(z')\rho|\beta,\beta)=(\beta,\beta|\rho R^{>}(z')R^{<}(z)\rho|\alpha,\alpha), \tag{3.46}$$

$$(\alpha,\alpha|\Delta(z,z')|\alpha,\alpha)=(z-z')\sum_{\beta}(\alpha,\alpha|\rho R^{>}(z)R^{<}(z')\rho|\beta,\beta), \tag{3.47}$$

$$(\alpha,\alpha|\mathcal{G}(z,z')|\alpha,\alpha)=\sum_{\beta}(\alpha,\alpha|W(z,z')\Delta(z,z')|\beta,\beta), \tag{3.48}$$

$$\chi(z,z')=\rho\tilde{H}Q\mathcal{C}(z,z')\rho,\quad \tilde{\chi}(z,z')=\rho\mathcal{D}(z,z')Q\tilde{H}\rho. \tag{3.49}$$

These relations will be needed in the following sections and the proofs for some of them will be given in Appendix B.

4. ASYMPTOTIC EVOLUTION S-OPERATORS

We will now discuss the asymptotic time evolution s-operator for $U_E(t)$ under the assumption of the analytic properties of the functions introduced in Sec. 3 and derive the master equations.

In our dissipative system, the integrand in (2.40) or in (3.43) has a finite discontinuity along the real axis of l in the limit of a large system and can be analytically continued into the upper or lower half-plane corresponding to the time being positive or negative, respectively. Then, the time evolution of $U_E(t)$ may be determined by the structure of its analytically continued s-operator.

Our assumption is that "the analytically continued s-operators for $D^{>}(E+\frac{1}{2}l)$, $N^{>}(E+\frac{1}{2}l)$, and $\{N^{>}(E+\frac{1}{2}l)N^{<}(E-\frac{1}{2}l)\}_{irr}$ are regular near the real axis, and the singularities of the product, $R^{(\pm)>}(E+\frac{1}{2}l)R^{(\mp)<}(E-\frac{1}{2}l)$, near the real axis are isolated simple poles arising $[l-\chi_E^{(\pm)}(l)]^{-1}$ or $[l-\tilde{\chi}_E^{(\pm)}(l)]^{-1}$ as the eigenvalues l_j $(j=0,1,2,\cdots)$, of the collision s-operators $\chi_E^{(\pm)}(l)$ or $\tilde{\chi}_E^{(\pm)}(l)$." [Hereinafter, the symbols (+) and (−) stand for the analytically continued s-operator from the upper and lower half-planes across the cut along the real axis, respectively, and upper (lower) symbols must be taken together.] According to (3.40), the sets of the poles of $[l-\chi_E^{(\pm)}(l)]^{-1}$ and $[l-\tilde{\chi}_E^{(\pm)}(l)]^{-1}$ must be coincident and to (3.47), they include the point $l=0$.

Now we define the asymptotic evolution s-operators $\Sigma_E^{(\pm)}(t)$ and $\Sigma_E^{(-)}(t)$ for $t>0$ and $t<0$, respectively, as

$$\Sigma_E^{(\pm)}(t)=\frac{s(t)}{(2\pi i)^2}\int_{\gamma_0}dl\,e^{-ilt}R^{(\pm)>}(E+\tfrac{1}{2}l)R^{(\mp)<}(E-\tfrac{1}{2}l), \tag{4.1}$$

where the contour γ_0 encircles the poles at $l=l_j$ counter clockwise, by excluding all other singularities in the continued planes.

To determine the supereigenstates of the collision s-operators and to evaluate $\rho\,\Sigma_E^{(\pm)}(t)\rho$, we first rewrite $\rho R^{>}R^{<}\rho$ in the neighborhood of each pole $l=l_j$ as

$$\rho R^{(\pm)>}(E+\tfrac{1}{2}l)R^{(\mp)<}(E-\tfrac{1}{2}l)\rho=[l-\chi_E^{(\pm)}(l_j)]^{-1}\mathcal{A}_{Ej}^{(\pm)}(l)=\mathcal{A}_{Ej}^{(\pm)}(l)[l-\tilde{\chi}_E^{(\pm)}(l_j)]^{-1}. \tag{4.2}$$

With the relation $\chi\Delta=\Delta\tilde{\chi}$ and (3.40), $\mathcal{A}_{Ej}^{(\pm)}(l)$ satisfies

$$\mathcal{A}_{Ej}^{(\pm)}(l)=\frac{l-\chi_E^{(\pm)}(l_j)}{l-\chi_E^{(\pm)}(l)}\Delta_E^{(\pm)}(l)=\Delta_E^{(\pm)}(l)\,\frac{l-\tilde{\chi}_E^{(\pm)}(l_j)}{l-\tilde{\chi}_E^{(\pm)}(l)}, \tag{4.3}$$

and

$$\mathcal{A}_{Ej}^{(\pm)}(l_j)=[I-\partial\chi_E^{(\pm)}(l)]^{-1}\Delta_E^{(\pm)}(l)\big|_{l=l_j}=\Delta_E^{(\pm)}(l)[I-\partial\tilde{\chi}_E^{(\pm)}(l)]^{-1}\big|_{l=l_j}, \tag{4.4}$$

where $\partial\equiv\partial/\partial l$. We may now state our assumption, which leads to l_j being the eigenvalue of the collision s-operators, that

$$\chi_E^{(\pm)}(l_j)\mathcal{A}_{Ej}^{(\pm)}(l_j)=l_j\mathcal{A}_{Ej}^{(\pm)}(l_j)=\mathcal{A}_E^{(\pm)}(l_j)\tilde{\chi}_E^{(\pm)}(l_j), \tag{4.5}$$

i.e., $\mathcal{A}_{Ej}^{(\pm)}(l_j)|X)$ and $(Y|\mathcal{A}_{Ej}^{(\pm)}(l_j)$ are a right and a left supereigenstate with the eigenvalue l_j, respectively, where $|X)$ and $(Y|$ are arbitrary s-states.

To evaluate the integral for $\rho\,\Sigma_E^{(\pm)}(t)\rho$ in (4.1), we expand $\mathcal{A}_{Ej}^{(\pm)}(l_j)$ in (4.2) about the points l_j,

$$\mathcal{A}_{Ej}^{(\pm)}(l)=\mathcal{A}_{Ej}^{(\pm)}(l_j)+\partial\mathcal{A}_{Ej}^{(\pm)}(l)\big|_{l=l_j}(l-l_j)+\cdots, \tag{4.6}$$

and replace $\chi_E^{(\pm)}(l_j)$ or $\tilde{\chi}_E^{(\pm)}(l_j)$ by l_j in the term having the factor $\mathcal{A}_{Ej}^{(\pm)}(l_j)$ in this expansion of (4.2). Then taking the residues of (4.2) at $l=l_j$, we obtain

$$\begin{aligned}\rho\,\Sigma_E^{(\pm)}(t)\rho&=\frac{s(t)}{2\pi i}\sum_j e^{-il_jt}\mathcal{A}_{Ej}^{(\pm)}(l_j)\\&=\frac{s(t)}{2\pi i}\sum_j e^{-i\chi_E^{(\pm)}(l_j)t}\mathcal{A}_{Ej}^{(\pm)}(l_j)\\&=\frac{s(t)}{2\pi i}\sum_j\mathcal{A}_{Ej}^{(\pm)}(l_j)e^{-i\tilde{\chi}_E^{(\pm)}(l_j)t}.\end{aligned} \tag{4.7}$$

Since the poles are in the lower (upper) half-plane for $t>0$ $(t<0)$, the terms in the right-hand side of (4.7) tend to zero except the one belonging to the zero eigenvalue of the collision s-operators $\chi_E^{(\pm)}(0)$ and $\tilde{\chi}_E^{(\pm)}(0)$, when the time increases (decreases).

We will now derive a time evolution equation of $\rho\,\Sigma_E^{(\pm)}(t)\rho$ in a Markovian form. For this purpose we expand the collision s-operators about their regular point $l=0$, that is

$$\chi_E^{(\pm)}(l_j)\mathcal{A}_{Ej}^{(\pm)}(l_j)=\sum_{n=0}^{\infty}\frac{1}{n!}\partial^n\chi_E^{(\pm)}(l)\big|_{l=0}l_j^n\mathcal{A}_{Ej}^{(\pm)}(l_j)=\sum_{n=0}^{\infty}\frac{1}{n!}\partial^n\chi_E^{(\pm)}(l)\big|_{l=0}[\chi_E^{(\pm)}(l_j)]^n\mathcal{A}_{Ej}^{(\pm)}(l_j), \tag{4.8}$$

$$\mathcal{A}_{Ej}^{(\pm)}(l_j)\tilde{\chi}_E^{(\pm)}(l_j)=\sum_{n=0}^{\infty}\frac{1}{n!}\mathcal{A}_{Ej}^{(\pm)}(l_j)[\tilde{\chi}_E^{(\pm)}(l_j)]^n\partial^n\tilde{\chi}_E^{(\pm)}(l)\big|_{l=0}. \tag{4.9}$$

These equations and (4.5) show that $\chi_E^{(\pm)}(l_j)$ and $\tilde{\chi}_E^{(\pm)}(l_j)$ are the solutions of the following equations, respectively,

$$\rho\Gamma_E^{(\pm)}\rho=\sum_{n=0}^{\infty}\frac{1}{n!}\partial^n\chi_E^{(\pm)}(l)\big|_{l=0}(\rho\Gamma_E^{(\pm)}\rho)^n, \tag{4.10}$$

$$\rho\tilde{\Gamma}_E^{(\pm)}\rho=\sum_{n=0}\frac{1}{n!}(\rho\tilde{\Gamma}_E^{(\pm)}\rho)^n\partial^n\tilde{\chi}_E^{(\pm)}(l)\big|_{l=0}, \tag{4.11}$$

with

$$\rho\Gamma_E^{(\pm)}\rho\mathcal{A}_{Ej}^{(\pm)}(l_j)=l_j\mathcal{A}_{Ej}^{(\pm)}(l_j)$$
$$=\mathcal{A}_{Ej}^{(\pm)}(l_j)\rho\tilde{\Gamma}_E^{(\pm)}\rho. \tag{4.12}$$

Using these s-operators, we can rewrite (4.7) as

$$\rho\Sigma_E^{(\pm)}(t)\rho=e^{-i\rho\Gamma_E^{(\pm)}\rho t}\rho\Sigma_E^{(\pm)}(0)\rho$$
$$=\rho\Sigma_E^{(\pm)}(0)\rho e^{-i\rho\tilde{\Gamma}_E^{(\pm)}\rho t} \tag{4.13}$$

Then, taking the time derivative of (4.13), we get the Markovian master equation

$$i\partial_t\rho\Sigma_E^{(\pm)}(t)\rho=\rho\Gamma_E^{(\pm)}\rho\Sigma_E^{(\pm)}(t)\rho=\rho\Sigma_E^{(\pm)}(t)\rho\tilde{\Gamma}_E^{(\pm)}\rho, \tag{4.14}$$

with the initial condition

$$\rho\Sigma_E^{(\pm)}(0)\rho=\frac{\pm1}{2\pi i}\sum_j\mathcal{A}_{Ej}^{(\pm)}(l_j). \tag{4.15}$$

The remaining components of $\Sigma_E^{(\pm)}(t)$ can be evaluated in similar ways. Since $\mathcal{QC}_E^{(\pm)}(l)\rho$, $\rho\mathcal{D}_E^{(\pm)}(l)\mathcal{Q}$, and $\mathcal{QQ}_E^{(\pm)}(l)\mathcal{Q}$ have no singularities at l_j, (4.1) together with (3.30) yields

$$\Sigma_E^{(\pm)}(t)=\frac{s(t)}{2\pi i}\sum_j[\rho+\mathcal{QC}_E^{(\pm)}(l_j)\rho]e^{-il_jt}\mathcal{A}_{Ej}^{(\pm)}(l_j)$$
$$\times[\rho+\rho\mathcal{D}_E^{(\pm)}(l_j)\mathcal{Q}]. \tag{4.16}$$

Expanding $\mathcal{QC}_E^{(\pm)}(l_j)\rho$ and $\rho\mathcal{D}_E^{(\pm)}(l_j)\mathcal{Q}$ about the origin, we get from (4.16)

$$\Sigma_E^{(\pm)}(t)=[\rho+\mathcal{Q}\mathbb{C}_E^{(\pm)}\rho]\Sigma_E^{(\pm)}(t)[\rho+\rho\mathbb{D}_E^{(\pm)}\mathcal{Q}], \tag{4.17}$$

where the time-independent s-operators $\mathcal{Q}\mathbb{C}_E^{(\pm)}\rho$ and $\rho\mathbb{D}_E^{(\pm)}\mathcal{Q}$ are defined by

$$\mathcal{Q}\mathbb{C}_E^{(\pm)}\rho=\sum_{n=0}^{\infty}\frac{1}{n!}\partial^n\mathcal{QC}_E^{(\pm)}(l)\rho\,|_{l=0}(\rho\Gamma_E^{(\pm)}\rho)^n, \tag{4.18}$$

$$\rho\mathbb{D}_E^{(\pm)}\mathcal{Q}=\sum_{n=0}^{\infty}\frac{1}{n!}(\rho\tilde{\Gamma}_E^{(\pm)}\rho)^n\partial^n\rho\mathcal{D}_E^{(\pm)}(l)\mathcal{Q}\,|_{l=0}. \tag{4.19}$$

Equation (4.17) shows that the correlation components of $\Sigma_E^{(\pm)}(t)$ are given by synchronous functionals of the vacuum–vacuum component through the action of $\mathcal{Q}\mathbb{C}_E^{(\pm)}\rho$ and $\rho\mathbb{D}_E^{(\pm)}\mathcal{Q}$.

In order to understand the role of our evolution s-operator, $\Sigma_E^{(\pm)}(t)$, we now rewrite (4.14) in a form with the time-dependent collision s-operator $\chi_E^{(\pm)'}(t)$ defined by the inverse Laplace transformation of $\chi_E^{(\pm)}(l)$ as

$$\chi_E^{(\pm)}(l)=2\pi\int_0^{\pm\infty}dt\,e^{+ilt}\chi_E^{(\pm)'}(t). \tag{4.20}$$

Substituting (4.20) into (4.10), we have

$$\rho\Gamma_E^{(\pm)}\rho=2\pi\int_0^{\pm\infty}\chi_E^{(\pm)'}(\tau)e^{+i\rho\Gamma_E^{(\pm)}\rho\tau}. \tag{4.21}$$

Then, acting this s-operator on the left of (4.13), we obtain

$$i\partial_t\rho\Sigma_E^{(\pm)}(t)\rho=2\pi\int_0^{\pm\infty}d\tau\,\chi_E^{(\pm)'}(\tau)\,\rho\Sigma_E^{(\pm)}(t-\tau)\rho. \tag{4.22}$$

The master equation (4.22) must be compared with Van Hove's generalized master equation

$$i\partial_t\rho U_E(t)\rho$$
$$=f_E^{(\pm)}(t)+2\pi\int_0^t d\tau\,\chi_E^{(\pm)'}(\tau)\rho U_E(t-\tau)\rho, \tag{4.23}$$

where

$$f_E^{(\pm)}(t)=\frac{s(t)}{(2\pi i)^2}\int_\gamma dl\,e^{-ilt}\Delta_E^{(\pm)}(l), \tag{4.24}$$

and this can be derived from taking the time derivative of the ρ–ρ component of (3.43) and using the convolution theorem of the Laplace transformation. The inhomogeneous term, (4.24), vanishes when $|t|\to\infty$ for dissipative systems. Hence, we may say that (4.22) and therefore (4.14) is the asymptotic master equation which describes the long time behavior of $U_E(t)$.

With similar calculations we can easily obtain the other equations symmetrical to (4.22) and (4.23) as

$$i\partial_t\rho\Sigma_E^{(\pm)}(t)\rho=2\pi\int_0^{\pm\infty}d\tau\,\rho\Sigma_E^{(\pm)}(t-\tau)\rho\tilde{\chi}_E^{(\pm)'}(\tau), \tag{4.25}$$

$$i\partial_t\rho U_E(t)\rho=f_E^{(\pm)}(t)$$
$$+2\pi\int_0^t d\tau\,\rho U_E(t-\tau)\rho\tilde{\chi}_E^{(\pm)'}(\tau), \tag{4.26}$$

where $\tilde{\chi}_E^{(\pm)'}(t)$ is the inverse Laplace transformation of $\tilde{\chi}_E^{(\pm)}(l)$.

5. STATIONARY SOLUTION OF $\Sigma_E^{(\pm)}(t)$

In this section, we will derive an exact stationary solution $\Sigma_E^{e(\pm)}$ of $\Sigma_E^{(\pm)}(t)$.

For the stationary solution, $\Sigma_E^{e(\pm)}$, we have

$$\partial_t\rho\Sigma_E^{e(\pm)}\rho=0, \tag{5.1}$$

and then combining (4.15), we get

$$\rho\Gamma_E^{(\pm)}\rho\Sigma_E^{e(\pm)}\rho=\rho\Sigma_E^{e(\pm)}\rho\tilde{\Gamma}_E^{(\pm)}\rho=0. \tag{5.2}$$

By substituting (4.10) and (4.11) into (5.2), it follows that

$$\chi_E^{(\pm)}(0)\rho\Sigma_E^{e(\pm)}\rho=\rho\Sigma_E^{e(\pm)}\rho\tilde{\chi}_E^{(\pm)}(0)=0. \tag{5.3}$$

Equations (5.2) and (5.3) show that $\rho\Sigma_E^{e(\pm)}\rho$ is a solution belonging to the zero eigenvalue of the eigenvalue equations (4.5) or (4.12). Thus, we find that the stationary solution is equivalent to the solution $\Sigma_E^{(\pm)}(t)$ in the limit for $|t|\to\infty$.

Further, by using (4.18) and (4.19) for (4.17) and also (5.2), each components of $\Sigma_E^{e(\pm)}$ is expressed by a functional of $\rho\Sigma_E^{e(\pm)}\rho$ as follows,

$$\Sigma_E^{e(\pm)}=[\rho+\mathcal{QC}_E^{(\pm)}(0)\rho]\Sigma_E^{e(\pm)}[\rho+\rho\mathcal{D}_E^{(\pm)}(0)\mathcal{Q}]. \tag{5.4}$$

In order to find the solution of $\rho\Sigma_E^{e(\pm)}\rho$, we rewrite (5.3) by the definitions of χ and $\tilde{\chi}$ as

$$\mathcal{G}_E^{(\pm)}(0)\rho\Sigma_E^{e(\pm)}\rho=\Delta_E^{(\pm)}(0)W_E^{(\pm)}(0)\rho\Sigma_E^{e(\pm)}\rho, \tag{5.5}$$

$$\rho\Sigma_E^{e(\pm)}\rho\mathcal{G}_E^{(\pm)}(0)=\rho\Sigma_E^{e(\pm)}\rho W_E^{(\pm)}(0)\Delta_E^{(\pm)}(0). \tag{5.6}$$

Then, if we put

$$(\alpha,\alpha|\rho\Sigma_E^{e(\pm)}\rho|\beta,\beta)=(\alpha,\alpha|\Delta_E^{(\pm)}(0)|\alpha,\alpha)F_E^{(\pm)}(\beta), \tag{5.7}$$

the tetradic elements, $(\alpha,\alpha|$ and $|\beta,\beta)$, of the right-hand side of (5.5) becomes

$$(\alpha,\alpha|\Delta_E^{(\pm)}(0)|\alpha,\alpha)\sum_\gamma(\alpha,\alpha|W_E^{(\pm)}(0)\Delta_E^{(\pm)}(0)|\gamma,\gamma)F_E^{(\pm)}(\beta), \tag{5.8}$$

and using (3.48), we can see that (5.8) is just equal to the tetradic element of the left-hand side of (5.5).

Therefore, (5.7) is the solution of (5.5) with unknown function $F_E^{(\pm)}(\beta)$.

To determine $F_E^{(\pm)}(\beta)$, we put both sides of (4.1) between $(\alpha,\alpha|$ and $|\beta,\beta)$ and take the summation over α. Then, if we use the relations (3.46), (3.47), and (3.44), and perform the contour integral around the pole at $l=0$, we can obtain

$$\sum_\alpha (\alpha,\alpha|P\Sigma_E^{e(\pm)}P|\beta,\beta) = \frac{\pm 1}{2\pi i}(\beta,\beta|\Delta_E^{(\pm)}(0)|\beta,\beta). \tag{5.9}$$

Further, by putting (5.7) into the left-hand side of (5.9), the function $F_E^{(\pm)}(\beta)$ is determined as

$$F_E^{(\pm)}(\beta) = \frac{\pm 1}{2\pi i}\frac{(\beta,\beta|\Delta_E^{(\pm)}(0)|\beta,\beta)}{\sum_\gamma(\gamma,\gamma|\Delta_E^{(\pm)}(0)|\gamma,\gamma)}. \tag{5.10}$$

Finally, we get the solution for (5.5) from (5.7),

$$P\Sigma_E^{e(\pm)}P = \frac{P|Q_E)(Q_E|P}{(1|Q_E)}, \tag{5.11}$$

based on (B3) in Appendix B, where Q_E is defined by

$$Q_E = \frac{1}{2\pi i}\lim_{\eta\to 0}[R(E+i\eta)-R(E-i\eta)], \tag{5.12}$$

with $\eta>0$. Q_E is the projection operator in $\mathfrak{H}$ belonging to an eigenvalue E for the total Hamiltonian H. In a similar way we can see that (5.11) is also the solution of (5.6).

In order to get the complete expression for $\sum_E^{e(\pm)}$, we substitute (5.11) into (5.4) and use the relations

$$|R(z)-R(z')) = [P+QC(z,z')P]\times|R(z)-R(z')), \tag{5.13}$$

$$(R(z^*)-R(z'^*)| = (R(z^*)-R(z'^*)|\times[P+PD(z,z')Q], \tag{5.14}$$

which is derived from (B4) in Appendix B, and then we arrive at the same solution for $t>0$ and $t<0$ as

$$\Sigma_E^{e(\pm)} = \lim_{t\to\pm\infty}\Sigma_E^{(\pm)}(t) = \frac{|Q_E)(Q_E|}{(1|Q_E)} \equiv \hat{P}_E^e. \tag{5.15}$$

This solution is equivalent to the one derived by Van Hove[1] and by Janner[2] for $U_E(\pm\infty)$. Further, we note here that this is the projector belonging to an eigenvalue E for the s-Hamiltonian, $\tilde{H}$ in $\mathfrak{S}$ and the value of E coincides with the eigenvalue for the total Hamiltonian H in $\mathfrak{H}$. This is also the projector belonging to the zero eigenvalue for the a.s-Hamiltonian, $\tilde{H}$, and then

$$\tilde{H}\hat{P}_E^e = E\hat{P}_E^e, \quad \tilde{H}\hat{P}_E^e = 0, \tag{5.16}$$

$$\hat{P}_E^e\hat{P}_{E'}^e = \delta(E-E')\hat{P}_E^e, \tag{5.17}$$

Therefore, by the use of this projector we can project the stationary solution of the von Neumann equation, and the expectation value of any observable A in stationary systems can be written as

$$\langle A\rangle^e = \lim_{t\to\pm\infty}\langle A\rangle_t = \int_{-\infty}^{+\infty} dE\,\langle A\rangle_E p_E, \tag{5.18}$$

where

$$\langle A\rangle_E = \frac{(A|Q_E)}{(1|Q_E)}, \tag{5.19}$$

and

$$p_E = (Q_E|\rho(0)). \tag{5.20}$$

Here, $\langle A\rangle_E$ is the microcanonical average of the observable A on an energy shell, and p_E is the probability distribution of the initial state on this energy shell.

6. QUANTUM SUBDYNAMICS

We now introduce a s-operator defined by

$$\Pi_E^{(\pm)} \equiv \Sigma_E^{(\pm)}(0), \tag{6.1}$$

which will play the essential role as a projector in our quantum subdynamics theory.

We first show that this s-operator satisfies one of the basic relations,

$$\Pi_E^{(\pm)}U(t) = \Sigma_E^{(\pm)}(t) = U(t)\Pi_E^{(\pm)}, \tag{6.2}$$

where $U(t)$ is the complete evolution s-operator defined in (2.38). These relations show that $\Pi_E^{(\pm)}$ projects the asymptotic evolution s-operator from the complete evolution s-operator. It can be easily seen from (6.2) that $\Pi_E^{(\pm)}$ is commutable with the a.s-Hamiltonian, $\tilde{H}$, and thus $\Sigma_E^{(\pm)}(t)$ is a solution of the von Neumann equation.

We will here prove the equality in the left-hand side of (6.2). For this purpose it is sufficient, because of (4.17), to prove

$$P\Pi_E^{(\pm)}U(t) = P\Sigma_E^{(\pm)}(t). \tag{6.3}$$

By putting $U(t)=\exp[-i\tilde{H}t]$, (4.16) and (6.1) into (6.3), and then by comparing the terms proportional to the same order on t in both sides of (6.3), we see that if the relation

$$\sum_j A_{Ej}^{(\pm)}(l_j)[P+PD_E^{(\pm)}(l_j)Q]\tilde{H}^n = \sum_j[\chi_E^{(\pm)}(l_j)]^n A_{Ej}^{(\pm)}(l_j)[P+PD_E^{(\pm)}(l_j)Q], \tag{6.4}$$

is satisfied, (6.3) also holds, where we have invoked (4.5). We prove (6.4) by induction: For $n=0$, this is trivially true. Assume that (6.4) is true for given n and consider

$$\sum_j A_{Ej}^{(\pm)}(l_j)[P+PD_E^{(\pm)}(l_j)Q]\tilde{H}^{n+1} = \sum_j[\chi_E^{(\pm)}(l_j)]^n A_{Ej}^{(\pm)}(l_j)[P+PD_E^{(\pm)}(l_j)Q]\tilde{H}. \tag{6.5}$$

If the relation

$$R^{(\pm)>}(E+\tfrac{1}{2}l)R^{(\mp)<}(E-\tfrac{1}{2}l)\tilde{H} = \tilde{H}R^{(\pm)>}(E+\tfrac{1}{2}l)R^{(\mp)<}(E-\tfrac{1}{2}l), \tag{6.6}$$

which leads together with (4.1) to

$$P\Sigma_E^{(\pm)}(t)\tilde{H} = P\tilde{H}\Sigma_E^{(\pm)}(t) \tag{6.7}$$

holds, the right-hand side of (6.5) can be rewritten in the following form,

$$\sum_j P\tilde{H}[P+QC_E^{(\pm)}(l_j)P][\chi_E^{(\pm)}(l_j)]^n \times A_{Ej}^{(\pm)}(l_j)[P+PD_E^{(\pm)}(l_j)Q] = \sum_j[\chi_E^{(\pm)}(l_j)]^{n+1}A_{Ej}^{(\pm)}(l_j)[P+PD_E^{(\pm)}(l_j)Q], \tag{6.8}$$

where we have used the relation obtained by comparing both sides of (6.7) based on (4.16), and used the relations (3.49) and $P\tilde{H}P=0$. The induction hypothesis (6.4) is thus extended to the value $n+1$. In view of

(2.36), (6.6) obviously holds for l in the upper (lower) half-plane, when $t>0$ ($t<0$). The analytically continued s-operators are no longer in the form (2.36), but it can be shown that (6.6) holds as before for all regular points in the continued planes (see Appendix C). Hence, the proof of (6.3) is fininished. The proof of the relation $\Sigma_E^{(\pm)}(t)=U(t)\Pi_E^{(\pm)}$ is quite analogous to the present one.

Our next task is to prove the other important relations in the subdynamics theory, i.e.,

$$\hat{H}\Pi_E^{(\pm)}=E\Pi_E^{(\pm)}=\Pi_E^{(\pm)}\hat{H}, \tag{6.9}$$

$$\Pi_E^{(\pm)}\Pi_{E'}^{(\pm)}=\delta(E-E')\Pi_E^{(\pm)}. \tag{6.10}$$

These relations show that $\Pi_E^{(\pm)}$ is the projector belonging to an eigenvalue E for the s.s- Hamiltonian, $\hat{H}$. This is the very point at which our quantum subdynamics theory is different from that of the Brussels school.[6-8] That is, our projector $\Pi_E^{(\pm)}$ lies completely in the eigenspace $\hat{P}_E$, and hence with (6.2), the asymptotic time evolution described by $\Sigma_E^{(\pm)}(t)$ is entirely contained in a subspace $\Pi_E^{(\pm)}$ of the eigenspace $\hat{P}_E$.

To prove (6.9) and (6.10), we use the relation

$$\begin{aligned}\hat{H}R^{(\pm)>}(E+\tfrac{1}{2}l)R^{(\mp)<}(E-\tfrac{1}{2}l)&=ER^{(\pm)>}(E+\tfrac{1}{2}l)R^{(\mp)<}(E-\tfrac{1}{2}l)\\&\quad+\tfrac{1}{2}[R^{(\pm)>}(E+\tfrac{1}{2}l)+R^{(\mp)<}(E-\tfrac{1}{2}l)],\end{aligned} \tag{6.11}$$

which can be obtained from (2.36) by analytically continuing to the lower or upper half-planes in l (See Appendix C). From (6.11) and (4.1) (for $t=0$), we get

$$\hat{H}\Pi_E^{(\pm)}=E\Pi_E^{(\pm)}+\frac{\pm 1}{(2\pi i)^2}\int_{\gamma_0}dl\,\tfrac{1}{2}[R^{(\pm)>}(E+\tfrac{1}{2}l)+R^{(\mp)<}(E-\tfrac{1}{2}l)], \tag{6.12}$$

but the second term in (6.12) vanishes because of the basic assumption, which leads to the regularity of $R^{(\pm)>}$ and $R^{(\mp)<}$ in the neighborhood of the points $l=l_j$. Hence, the equality in the left-hand side of (6.9) is proved. The proof of the relation $E\Pi_E^{(\pm)}=\Pi_E^{(\pm)}\hat{H}$ is quite analogous to the present one.

Because of the hermiticity of $\hat{H}$, it is evident from (6.9) that the projectors $\Pi_E^{(\pm)}$ belonging to different eigenvalues are orthogonal with each other and hence $\Pi_E^{(\pm)}\Pi_{E'}^{(\pm)}$ is proportional to $\delta(E-E')\Pi_E^{(\pm)}$. Its proportional coefficient is determined as unity under the relations

$$\begin{aligned}\delta(E-E')\hat{P}_E^e&=\lim_{\substack{t\to\pm\infty\\t'\to\pm\infty}}\Pi_E^{(\pm)}U(t)\Pi_{E'}^{(\pm)}U(t')\\&=\lim_{\substack{t\to\pm\infty\\t'\to\pm\infty}}\Pi_E^{(\pm)}\Pi_{E'}^{(\pm)}U(t+t'),\end{aligned} \tag{6.13}$$

which are given by (5.15), (5.17), and (6.2). Hence, (6.10) is proved.

As can be seen from the definition (6.1), $\Pi_E^{(\pm)}$ is constructed from the contribution of a part of the contour integral over l for $\hat{P}_E$. Therefore, we can introduce another important projector $\overline{\Pi}_E^{(\pm)}$ as the remainder after the subtraction of $\Pi_E^{(\pm)}$ from the complete projector $\hat{P}_E$, i.e.,

$$\overline{\Pi}_E^{(\pm)}=\hat{P}_E-\Pi_E^{(\pm)}, \tag{6.14}$$

which projects out the contribution describing the short time evolution from $U(t)$ in the eigenspace $\hat{P}_E$.

Combining (6.14) with (2.43)–(2.45), (6.2), (6.9), and (6.10), we can easily get the following relations:

$$\hat{H}\overline{\Pi}_E^{(\pm)}=E\overline{\Pi}_E^{(\pm)}=\overline{\Pi}_E^{(\pm)}\hat{H}, \tag{6.15}$$

$$\overline{\Pi}_E^{(\pm)}\overline{\Pi}_{E'}^{(\pm)}=\delta(E-E')\overline{\Pi}_E^{(\pm)}, \tag{6.16}$$

$$\overline{\Pi}_E^{(\pm)}U(t)=U(t)\overline{\Pi}_E^{(\pm)}\equiv\overline{\Sigma}_E^{(\pm)}(t), \tag{6.17}$$

$$\Pi_E^{(\pm)}\overline{\Pi}_{E'}^{(\pm)}=\overline{\Pi}_{E'}^{(\pm)}\Pi_E^{(\pm)}=0. \tag{6.18}$$

These relations show that the time evolutions of $\Sigma_E^{(\pm)}(t)$ and $\overline{\Sigma}_E^{(\pm)}(t)$ in the subspace $\Pi_E^{(\pm)}$ and in its complementary subspace $\overline{\Pi}_E^{(\pm)}$ of an eigenspace $\hat{P}_E$ are governed independently with each other by the von Neumann equation, and are entirely contained in each subspace, respectively. From these facts, we may conclude that our quantum subdynamics theory has been constructed on every eigenspace $\hat{P}_E$ belonging to the s.s-Hamiltonian, $\hat{H}$, which is quite different from the one of the Brussels school.

Let us now concentrate our attention on the subspace $\Pi_E^{(\pm)}$. As $\Pi_E^{(\pm)}U(t)$ satisfies the von Neumann equation, we get by the use of (4.17) (for $t=0$)

$$\begin{aligned}i\partial_t\rho\Pi^{(\pm)}U(t)&=\rho\hat{H}\Pi_E^{(\pm)}U(t)\\&=\rho\tilde{H}Q\mathbb{C}_E^{(\pm)}\rho\Pi_E^{(\pm)}U(t).\end{aligned} \tag{6.19}$$

On the other hand, by using (4.10), (4.18), and (3.48), it follows for the kinetic s-operator $\rho\Gamma_E^{(\pm)}\rho$ that

$$\rho\Gamma_E^{(\pm)}\rho=\rho\tilde{H}Q\mathbb{C}_E^{(\pm)}\rho, \tag{6.20}$$

and hence (6.19) together with (6.20) is consistent with (4.14). Similarly, we can easily get the relations symmetrical to (6.19) and (6.20) as

$$i\partial_t U(t)\Pi_E^{(\pm)}\rho=U(t)\Pi_E^{(\pm)}\rho\mathbb{D}_E^{(\pm)}Q\tilde{H}\rho, \tag{6.21}$$

with

$$\rho\tilde{\Gamma}_E^{(\pm)}\rho=\rho\mathbb{D}_E^{(\pm)}Q\tilde{H}\rho, \tag{6.22}$$

which is also consistent with (4.14). In this way, we again arrive at the same equations which have been derived in Sec. 4.

We now turn our attention to the complementary subspace $\overline{\Pi}_E^{(\pm)}$ and show that it has the properties symmetrical to the subspace $\Pi_E^{(\pm)}$. Using (6.18) and (4.17) (for $t=0$), we obtain

$$\begin{aligned}0=\rho\Pi_E^{(\pm)}\overline{\Pi}_E^{(\pm)}&=(\rho\Pi^{(\pm)}\rho+\rho\Pi_E^{(\pm)}Q)\overline{\Pi}_E^{(\pm)}\\&=\rho\Pi_E^{(\pm)}\rho(\rho+\rho\mathbb{D}_E^{(\pm)}Q)\overline{\Pi}_E^{(\pm)},\end{aligned} \tag{6.23}$$

and hence (as $\rho\overline{\Pi}_E^{(\pm)}\rho\neq 0$)

$$\rho\overline{\Pi}_E^{(\pm)}=-\rho\mathbb{D}_E^{(\pm)}Q\overline{\Pi}_E^{(\pm)}; \tag{6.24}$$

similarly

$$\overline{\Pi}_E^{(\pm)}\rho=-\overline{\Pi}_E^{(\pm)}Q\mathbb{C}_E^{(\pm)}\rho. \tag{6.25}$$

Using them together with (6.17), we get the relation analogous to (4.17) as

$$\overline{\Sigma}_E^{(\pm)}(t)=[Q-\rho\mathbb{D}_E^{(\pm)}Q]\overline{\Sigma}_E^{(\pm)}(t)[Q-Q\mathbb{C}_E^{(\pm)}\rho]. \tag{6.26}$$

Hence, we see that the independent component is now the correlation–correlation part, $Q\Sigma_E^{(\pm)}(t)Q$, and other components are its functionals in contrast to

the subspace $\Pi_E^{(\pm)}$.

In order to get the kinetic equation for $Q\overline{\Sigma}_E^{(\pm)}(t)Q$, we put

$$Q\overline{\Sigma}_E^{(\pm)}(t)Q = e^{-i\mathcal{Q}\tilde{\Theta}_E^{(\pm)}\mathcal{Q}t}Q\overline{\Pi}_E^{(\pm)}Q = Q\overline{\Pi}_E^{(\pm)}Qe^{-i\mathcal{Q}\Theta_E^{(\pm)}\mathcal{Q}t}, \quad (6.27)$$

and hence

$$i\partial_t Q\overline{\Sigma}_E^{(\pm)}(t)Q = Q\tilde{\Theta}_E^{(\pm)}Q\overline{\Sigma}_E^{(\pm)}(t)Q = Q\overline{\Sigma}_E^{(\pm)}(t)Q\Theta_E^{(\pm)}Q. \quad (6.28)$$

To derive an expression for $Q\tilde{\Theta}_E^{(\pm)}Q$, we start from

$$\begin{aligned} i\partial_t Q\overline{\Sigma}_E^{(\pm)}(t)Q &= i\partial_t Q\mathcal{U}(t)\overline{\Pi}_E^{(\pm)}Q = Q\tilde{H}\overline{\Sigma}_E^{(\pm)}(t)Q \\ &= Q\tilde{H}Q\overline{\Sigma}_E^{(\pm)}(t)Q + Q\tilde{H}P\overline{\Sigma}_E^{(\pm)}(t)Q \\ &= (Q\tilde{H}Q - Q\tilde{H}P\mathbb{D}_E^{(\pm)}Q)Q\overline{\Sigma}_E^{(\pm)}(t)Q. \end{aligned} \quad (6.29)$$

By comparing this equation with (6.28), we obtain

$$Q\tilde{\Theta}_E^{(\pm)}Q = Q\tilde{H}Q - Q\tilde{H}P\mathbb{D}_E^{(\pm)}Q; \quad (6.30)$$

similarly

$$Q\Theta_E^{(\pm)}Q = Q\tilde{H}Q - Q\mathbb{C}_E^{(\pm)}P\tilde{H}Q. \quad (6.31)$$

The relation (6.30) and (6.31) are comparable with (6.20) and (6.22), by noting $P\tilde{H}P = 0$.

Before closing this section, let us see what the projectors $\hat{P}_E$, $\Pi_E^{(\pm)}$, and $\overline{\Pi}_E^{(\pm)}$ are, when the strength of the interaction is zero. By noting $\mathcal{C}_E^{(\pm)}(l) = \mathcal{D}_E^{(\pm)}(l) = \chi_E^{(\pm)}(l) = \tilde{\chi}_E^{(\pm)}(l) = 0$ when $V = 0$, it is easily seen from (2.41) and (4.16) that

$$\hat{P}_{E0} = \sum_{\alpha,\beta} |\alpha,\beta)\delta\left(E - \frac{\epsilon_\alpha + \epsilon_\beta}{2}\right)(\alpha,\beta|, \quad (6.32)$$

$$\begin{aligned} \Pi_{E0}^{(\pm)} &= \lim_{\eta \to 0} \frac{\pm 1}{2\pi i} P[R_0^>(E \pm i\eta) - R_0^<(E \mp i\eta)]P \\ &= \sum_\alpha |\alpha,\alpha)\delta(E - \epsilon_\alpha)(\alpha,\alpha|, \end{aligned} \quad (6.33)$$

and hence

$$\overline{\Pi}_{E0}^{(\pm)} = \sum_{\substack{\alpha,\beta \\ \alpha \neq \beta}} |\alpha,\beta)\delta\left(E - \frac{\epsilon_\alpha + \epsilon_\beta}{2}\right)(\alpha,\beta|, \quad (6.34)$$

where the suffix "0" means that the interaction is zero. Therefore, we obtain

$$\int_{-\infty}^{+\infty} dE\, \Pi_{E0}^{(\pm)} = P, \quad \int_{-\infty}^{+\infty} dE\, \overline{\Pi}_{E0}^{(\pm)} = Q, \quad (6.35)$$

and conclude that $\Pi_{E0}^{(\pm)}$ and $\overline{\Pi}_{E0}^{(\pm)}$ are the $\hat{P}_{E0}$ components of P and Q belonging to the unperturbed s.s-Hamiltonian, $\hat{H}_0$, respectively.

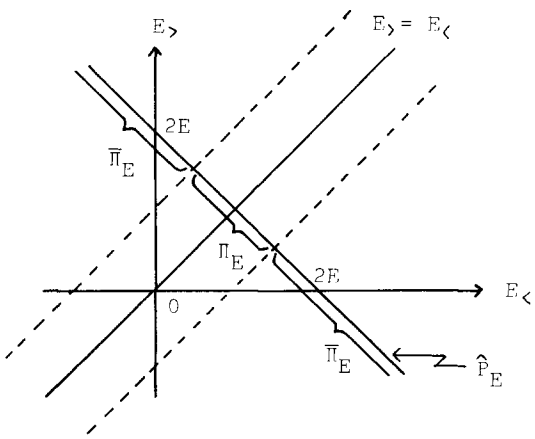

FIG. 2. Geometrical representation of the superspace: abscissa: the eigenvalue of $H^<$; ordinate: the eigenvalue of $H^>$.

7. DISCUSSION

In this section we discuss our results which are obtained in the previous sections. The perturbation theory of Van Hove and Janner has been reformulated into the superspace by introducing the ordered s-operator. Then, under the assumption for the analytic properties of the product of two resolvents, the master equation for the asymptotic evolution s-operator has been derived in a Markovian form on the basis of the eigenvalue problem of the collision s-operator of the Van Hove and Janner generalized master equation.

As we have shown, our reformulation in the superspace has saved the very complicated calculation of the two-resolvent method in the ordinary space, and our ordered s-operator makes it possible to decompose each term in the perturbation theory into a product form in very natural ways. This favorable feature will allow for practical application of the two-resolvent method to physical problems. When the system in which the quantum statistical effects cannot be neglected is treated, the contraction must be considered to deal with the vacuum-to-vacuum transitions due to the quantum statistical effect in the perturbation theory,[4,21] and it can be seen that on the basis of our formalism the calculation can be easily performed in the two-resolvent method. We will discuss this problem elsewhere.

We have further constructed the quantum subdynamics theory on every eigenspace of the s.s- Hamiltonian, $\hat{H}$. Our results may be illustrated briefly by means of two-dimensional Cartesian space as shown in Fig. 2, where the abscissa is the eigenvalue $E_<$ of the r.s-Hamiltonian $H^<$ and the ordinate is the eigenvalue $E_<$ of the l.s-Hamiltonian $H^>$, respectively. There, the eigenspace $\hat{P}_E$ corresponds to the line $E_> + E_< = 2E$. This line is decomposed into the two complementary parts by the projectors $\Pi_E^{(\pm)}$ and $\overline{\Pi}_E^{(\pm)}$. The line $E_> = E_<$ represents the eigenspace belonging to the zero eigenvalue of the as Hamiltonian, $\tilde{H}$. The intersecting point of these lines represents the microcanonical subspace characterized by the eigenvalue E of the ordinary total Hamiltonian, and it is projected out by the projector $\hat{P}_E^e$. This point is contained in the part corresponding to $\Pi_E^{(\pm)}$. After the atomic time scale, the contributions from the points on $\overline{\Pi}_E^{(\pm)}$ immediately vanish. The contributions from the points on $\Pi_E^{(\pm)}$ are exponentially dumped, except the contribution from the intersecting point on $\hat{P}_E^e$, being governed by the asymptotic master equation, (4.14).

Finally, we would like to remark upon the advantage of the two-resolvent method: That is, it makes it pos-

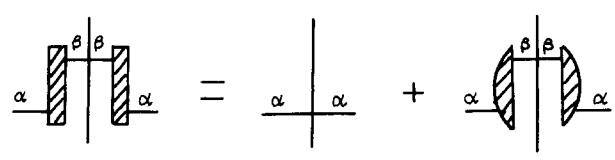

FIG. 3. Diagrams for $P(R \wedge R)P$ corresponding to (A1).

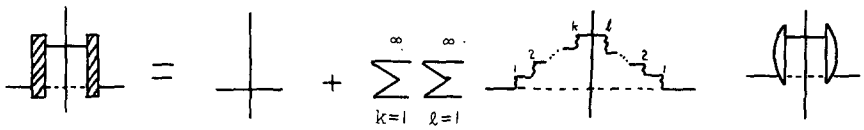

FIG. 4. Diagrams for $P(R\wedge R)P$: expanded form.

sible to discuss various descriptions for quantum systems, like the kinetic description for irreversible processes, the S-matrix formalism for scattering processes and the ground states or metastable states of the systems, as mutually complementary aspects of large quantum systems under a unified criterion of the "dissipativeness" of systems.[1,2,17–20,22] We believe that our formalism offers a suitable background to discuss these problems.

ACKNOWLEDGMENTS

The authors would like to express their thanks to Professor S. Goto for his continuous encouragement and to Dr. M. Nakamura for his valuable discussions. One of the authors (T.Y.P.) also wishes to express his gratitude to Professor R. Suzuki for his hospitality.

APPENDIX A

We will derive (3.29) and (3.30) by means of a diagramatic method. We represent the ordered s-operator $(A\wedge B)$ as the following figure,

$$A\,|\,B,$$

where a vertical solid line is drawn to represent the symbol "$\wedge$" at the corresponding location, and the same letters A and B are used to represent the corresponding diagrams to the operators A and B for the moment. In our diagram, the product is defined in accordance with (2.14) as

$$A\,|\,B*C\,|\,D=AC\,|\,DB.$$

In other words, when we decompose a diagram into a product, we put the innermost fragment of the diagram on the right-hand side in the product. Now, we represent (3.25), i.e.,

$$P(R\wedge R)P=P(D\wedge D)P+P(DND\wedge DND)P, \qquad \text{(A1)}$$

as shown in Fig. 3. That is, a resolvent R is represented by a rectangle with horizontal solid lines. The horizontal solid line shows D. The operator N is represented by a semicircle drawn with oblique lines. In our diagrams the difference of the states is expressed by the difference in height of the horizontal lines. Thus, the projectors P at the left-hand side and the right-hand side in each term of (A1) are expressed by the fact that the outside and inside horizontal lines in the corresponding diagrams are equal in height among themselves, respectively. (We also label the horizontal lines with the same Greek letter to express the same states, if necessary.) Further, a potential, $-V$, is represented by a vertical wavy line. Then, the series expansion of $P(R\wedge R)P$, obtained by substituting (3.12) into (A1), is represented as shown in Fig. 4, where the horizontal broken line is drawn to emphasize that the outside horizontal lines in each diagram are equal in height among themselves. The suffix "*ind*" in N is expressed in this figure by the fact that all horizontal lines belonging to N are different in height from each other.

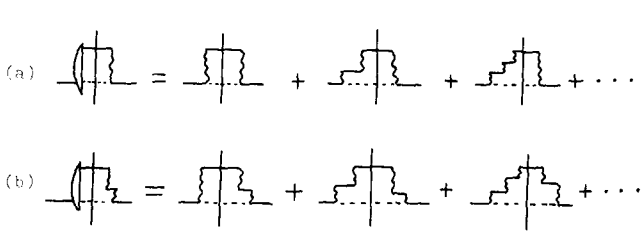

FIG. 5. Diagrams having no common intermediate states in both sides.

FIG. 6. Diagram for $P\{DND\wedge DND\}_{\text{irr}}P$.

Let us now derive (3.29) from (2.25). From the right-hand side of Fig. 4, we first lump together the diagrams having a single potential to the right of the vertical solid line [see Fig. 5(a)], having two potentials to the right of the vertical solid line [see Fig. 5(b)], and so on, but having no intermediate states which are equal to the intermediate states on the opposite side of the vertical solid line. Here, the semicircle expresses that there are no intermediate states equal to the ones on the opposite side of the vertical solid line. Then, summing up all diagrams obtained by the above procedure, we get the diagram shown in Fig. 6.

From the remaining diagrams in Fig. 4, we next lump together, as shown in Fig. 7, the diagrams in which the rightmost intermediate state is equal to a state on the opposite side, the second, the third, and so on. Substituting Figs. 6 and 7 into Fig. 4, we get the relation shown in Fig. 8 which is just the diagrammatic representation of (3.29).

A similar procedure is valid to derive (3.30) for (3.26)–(3.28). Here we only give the outline of its derivation. Equation (3.26) is represented by the diagrams as shown in Fig. 9, where the projector Q is expressed by the fact that the outside horizontal lines in each of the diagrams are different in height. Analogously, the third term in Fig. 9 is rearranged into the diagrams shown in Fig. 10. From Figs. 9, 10, and 8, we get the relation shown in Fig. 11. Similarly, we get the relations (a) and (b) in Fig. 12 from (3.27) and (3.28), respectively. Combining Figs. 8, 11, and 12, we get the diagrammatic representation of (3.30).

APPENDIX B

Proof of (3.45): From (3.25) and (3.29), we get an integral equation for W,

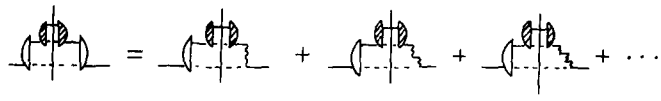

FIG. 7. Diagrams having common intermediate states in both sides.

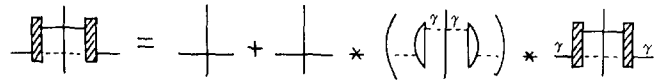

FIG. 8. Diagrams corresponding to (3.29).

$$W(z,z') = \rho N^>(z)N^<(z')\rho - W(z,z')\rho D^>(z)D^<(z')\rho N^>(z)N^<(z')\rho . \tag{B1}$$

Solving this equation by iteration we get

$$W(z,z') = \rho N^>(z)N^<(z')\rho - \rho N^>(z)N^<(z')\rho \times D^>(z)D^<(z')\rho N^>(z)N^<(z')\rho + \cdots . \tag{B2}$$

By this equation, the symmetric property (3.45) can be easily proved.

(3.46) can be proved analogously by using the iterated solution of (3.29) and the symmetric property (3.45).

Proof of (3.47): From the definition of Δ, (3.38), we get

$$\begin{aligned}(\alpha,\alpha|\Delta(z,z')|\alpha,\alpha) &= (\alpha,\alpha|R(z)-R(z')) \\ &= (z-z')(\alpha,\alpha|R(z)R(z')) \\ &= (z-z')(\alpha,\alpha|\rho R^>(z)R^<(z')\rho|1),\end{aligned} \tag{B3}$$

where we have used the following relation,

$$R(z)-R(z') = (z-z')R(z)R(z'). \tag{B4}$$

By (2.7), the right-hand side of (B3) is just equal to the right-hand side of (3.47).

Proof of (3.48): From (3.40) we get

$$\begin{aligned}&\mathcal{G}(z,z')\rho R^>(z)R^<(z')\rho \\ &\quad = \Delta(z,z')W(z,z')\rho R^>(z)R^<(z')\rho \\ &\qquad + \Delta(z,z') - (z-z')\rho R^>(z)R^<(z')\rho .\end{aligned} \tag{B5}$$

Putting this equation between $(\alpha,\alpha|$ and $|1)$, we get

$$\begin{aligned}&(\alpha,\alpha|\mathcal{G}(z,z')|\alpha,\alpha)(\alpha,\alpha|R(z)R(z')) \\ &\quad = \sum_\beta (\alpha,\alpha|\Delta(z,z')|\alpha,\alpha)(\alpha,\alpha|W(z,z')|\beta,\beta) \\ &\qquad \times(\beta,\beta|R(z)R(z')) \\ &\quad = \sum_\beta (\alpha,\alpha|R(z)R(z'))(\alpha,\alpha|W(z,z')|\beta,\beta) \\ &\qquad \times(\beta,\beta|\Delta(z,z')|\beta,\beta),\end{aligned} \tag{B6}$$

where we have used the relation (B3) and (B4). Dividing both sides by $(\alpha,\alpha|RR)$, we just get (3.48).

Proof of (3.49): We prove here the left-hand equality in (3.49). The proof of the right is quite analogous to the present one. From (3.32) and the relation $\rho\tilde{H}_0 Q = 0$, we have

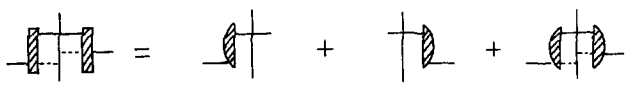

FIG. 9. Diagrams for $Q(R\wedge R)\rho$ corresponding to (3.26).

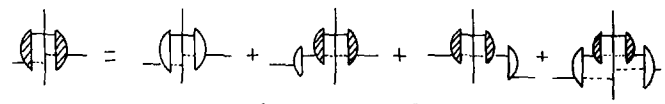

FIG. 10. Diagrams for $Q(DND\wedge DND)\rho$.

$$\begin{aligned}\rho\tilde{H}QC\rho &= \rho(V^>QD^>N^> - V^<QD^<N^<)\rho \\ &\quad + \rho[V^>QD^<N^< + (V^>QD^>N^>D^<N^<)_{irr}]\rho \\ &\quad - \rho[V^<QD^>N^> + (V^<QD^<N^<D^>N^>)_{irr}]\rho .\end{aligned} \tag{B7}$$

Putting the expansion for $N^\gtrless$, (3.12), adding $\rho(V^>-V^<)\rho = 0$ to the first term of (B7), and using the definition of $G^\gtrless$, (3.15), we can reformulate (B7) to

$$\rho\tilde{H}QC\rho = -\rho(G^>-G^<)\rho + \rho(D^>-D^<)\rho(N^>N^<)_{irr}\rho . \tag{B8}$$

The right-hand side of this relation is just equal to the definition of χ, (3.41).

APPENDIX C

Proof of (6.6) and (6.11): Our proof is only for the analytically continued functions from the upper half-plane. The proofs for the ones from the lower half-plane are quite analogous to the present ones. We change the variable l to $2l$ and drop the notations (+) and (−) for simplicity. Let l be a regular point in the lower half-plane and expand $R^>R^<$ about a point l_0 in the upper half-plane,

$$\begin{aligned}&R^>(E+l)R^<(E-l) \\ &\quad = \sum_{n=0}^\infty \frac{1}{n!}\partial_l^n [R^>(E+l)R^<(E-l)]|_{l=l_0}(l-l_0)^n \\ &\quad = \sum_{n=0}^\infty \sum_{m=0}^n \frac{1}{(n-m)!\,m!}[\partial_l^{n-m}R^>(E+l)] \\ &\qquad \times[\partial_l^m R^<(E-l)]|_{l=l_0}(l-l_0)^n \\ &\quad = \sum_{n=0}^\infty \sum_{m=0}^\infty R^>(E+l_0)R^<(E-l_0)[R^>(E+l_0)^{n-m} \\ &\qquad \times[-R^<(E-l_0)]^m(l-l_0)^n .\end{aligned} \tag{C1}$$

Each term in the right-hand side of (C1) is commutable with $\tilde{H}$, and hence (6.6) is proved.

Let us also expand $R^> + R^<$ about the same point,

$$\begin{aligned}&R^>(E+l)+R^<(E-l) \\ &\quad = \sum_{n=0}^\infty \frac{1}{n!}\partial_l^n [R^>(E+l)+R^<(E-l)]|_{l=l_0}(l-l_0)^n \\ &\quad = \sum_{n=0}^\infty \{[R^>(E+l_0)]^{n+1} - [-R^<(E-l_0)]^{n+1}\}(l-l_0)^n \\ &\quad = \sum_{n=0}^\infty \sum_{m=0}^n [R^>(E+l_0)+R^<(E-l_0)][R^>(E+l_0)]^{n-m} \\ &\qquad \times[-R^<(E-l_0)]^n(l-l_0)^n ,\end{aligned} \tag{C2}$$

where we have used the identity

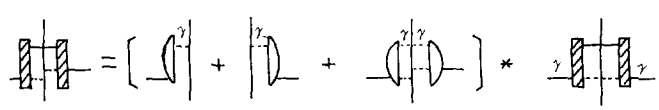

FIG. 11. Diagrams for $Q(R\wedge R)\rho$ with $QC\rho$.

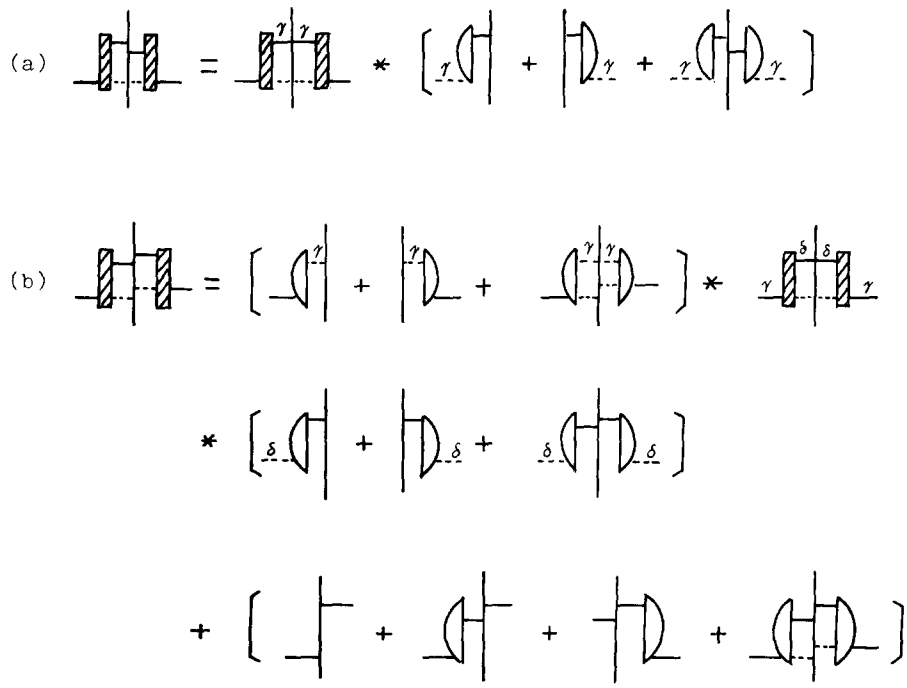

FIG. 12. Diagrams for (a) $P(R\wedge R)Q$, (b) $Q(R\wedge R)Q$.

$$a^{n+1} - b^{n+1} = \sum_{m=0}^{n} (a-b)a^{n-m}b^{m}. \tag{C3}$$

Putting the relations

$$R^{>}(E+l_0) + R^{<}(E-l_0) = 2(\hat{H} - E)R^{>}(E+l_0)R^{<}(E-l_0), \tag{C4}$$

which can be easily obtained from (2.36), and then (C1) into (C2), we get

$$R^{>}(E+l) + R^{<}(E-l) = 2(\hat{H} - E)R^{>}(E+l)R^{<}(E-l), \tag{C5}$$

which is just identical with (6.11).

[1]L. Van Hove, Physica 23, 441 (1957).
[2]A. Janner, Helv. Phys. Acta 35, 47 (1962).
[3]I. Prigogine, *Non-Equilibrium Statistical Mechanics* (Interscience, New York, 1962).
[4]R. Balescu, *Statistical Mechanics of Charged Particles* (Interscience, New York, 1963).
[5]I. Prigogine and P. Resibois, Physica 27, 629 (1961).
[6]I. Prigogine, C. George, and F. Henin, Physica 45, 418 (1969).
[7]R. Balescu and J. Wallenborm, Physica 54, 447 (1971).
[8]C. George, I. Prigogine, and L. Rosenfeld, Mat. Fys. Medd. Dan. Vid. Selsk. 38 (1972).
[9]A. Pytte, Phys. Fluids 11, 522 (1968).
[10]A. Grecos, T. Guo, and W. Guo, Physica A 80, 421 (1975).
[11]U. Fano, Rev. Mod. Phys. 29, 74 (1957).
[12]J. Schwinger, Proc. Nat. Acad. Sci. USA 46, 261 (1960).
[13]H. Primas, Helv. Phys. Acta 34, 331 (1961).
[14]G. Emch, Helv. Phys. Acta 37, 532 (1964).
[15]W. Schieve, *Lecture in Statistical Physics*, edited by W. Schieve and J. Turner (Springer, New York, 1974).
[16]In general, (2.3) does not exist for the case $\mathfrak{H}$ being infinite dimensional. However, we assume that (2.3) exists, since it does not cause any trouble in this work.
[17]R. Swenson, J. Math. Phys. 4, 544 (1963).
[18]P. Resibois, Physica 29, 721 (1963).
[19]L. Van Hove, Physica 21, 901 (1955); 22, 343 (1956).
[20]N. Hugenholtz, Physica 23, 481 (1957).
[21]N. Mishima, T. Petrosky, and M. Yamazaki, J. Stat. Phys. 14, 358 (1976).
[22]M. Prosperi, J. Math. Phys. 3, 329 (1962).

Natural connections on Stiefel bundles are sourceless gauge fields

Jerzy Nowakowski

Center for Theoretical Studies, University of Miami, Coral Gables, Florida 33124

Andrzej Trautman[a)]

Institute for Theoretical Physics, State University of New York, Stony Brook, New York 11794
(Received 12 July 1977)

It is shown that the natural connections defined on real and complex Stiefel bundles over Grassmannian manifolds are sourceless gauge fields corresponding to the gauge groups $G = SO(k)$, $k = 2,3,\cdots$ and $U(k)$, $k = 1,2,\cdots$, respectively. Stiefel bundles and their connections are important in view of their universality: Any gauge field, with group G, defined on a compact manifold may be obtained by embedding the manifold in a Grassmannian of sufficiently high dimension.

INTRODUCTION

Electromagnetism is a gauge theory[1]; gauge fields are believed to play a role in the description of strong[2] and weak[3] interactions; in a somewhat different sense, gravitation also corresponds to a gauge field.[4,5] Until recently, most of the solutions of the gauge field equations considered by physicists were topologically trivial: They could be continuously deformed into the zero field. A notable exception was the electromagnetic field of a magnetic pole.[6] Solutions of the sourceless Yang–Mills equations which are not homotopic to zero have been found by Belavin, Polyakov, Schwartz, and Tyupkin.[7] They have been given a physical interpretation,[8,9] generalized[10,11] and shown to be associated with a problem in algebraic geometry.[12] Another method for obtaining topologically nontrivial solutions of the Maxwell and Yang–Mills equations has been described in Ref. 13. The method is based on the observation that the magnetic pole of the lowest strength and the pseudoparticle solution of the Yang–Mills equations correspond to the natural connections defined on the Hopf bundles $S_3 \to S_2$ and $S_7 \to S_4$, respectively. In the present paper, we generalize the results on the gauge fields associated with Hopf fibrations to Stiefel manifolds considered as principal bundles over Grassmannians. The structure (gauge) group of these bundles is one of the groups $G = SO(n)$, $n = 2,3,\ldots$, $U(n)$ or $Sp(n)$, $n = 1,2,\ldots$. Each of these bundles has a natural connection[14] and its gauge field (curvature) is sourceless. We give the details of the proof only in the real and complex cases, leaving the quaternionic bundles for further study.

GAUGE FIELDS AND CONNECTIONS

To fix the terminology, let us recall[15,16] that a (smooth) principal bundle consists of:

(i) a (smooth) fiber bundle π: $P \to M$;

(ii) a Lie group G which acts on P smoothly and freely to the right: There is a differentiable map δ: $P \times G \to P$ such that $\delta_a \circ \delta_b = \delta_{ba}$ and $\delta_a = id_P \longleftrightarrow a = e$, where $\delta_a(p) = \delta(p,a)$, $p \in P$, $a,b \in G$ and e is the unit element of G; moreover,

(iii) the action of G is compatible with the fiber bundle structure: $\pi \circ \delta_a = \pi$ and any point in M has a neighborhood U such that $\pi^{-1}(U)$ is isomorphic to $U \times G$.

If $U \subset M$ is open, the s: $U \to P$ is called a (local) section if $\pi \circ s = id_U$; if $U = M$, then such an s is called a global section. A principal bundle P is trivial, i.e., isomorphic to the product $M \times G$, if and only if it admits a global section.

Any homomorphism of Lie groups h: $G \to H$ induces a (derived) homomorphism of Lie algebras, h': $G' \to H'$. For example, if $G = H$ and $\mathrm{ad}_a(b) = aba^{-1}$, then $\mathrm{Ad}_a = ad'_a$: $G' \to G'$ defines the adjoint representation of G in its Lie algebra. The action of G on P induces a homomorphism of Lie algebras δ': $G' \to$ Lie algebra of vertical vector fields on P, given by $\delta'(A) =$ vector field tangent to the curve $t \to \delta_{\exp tA}$, $A \in G'$. Let T denote the tangent functor. A connection on P may be defined by a connection form ω, i.e., by a map ω: $TP \to G$ which is linear on the fibers of $TP \to P$, equivariant under the action of G,

$$\omega \circ T\delta_a = \mathrm{Ad}_{a^{-1}} \circ \omega, \quad a \in G, \tag{1}$$

and such that

$$\omega(\delta'(A)) = A, \quad A \in G'. \tag{2}$$

The vector space $\mathrm{hor}_p P = \{u \in T_p P \mid \omega(u) = 0\}$ is called the horizontal space at $p \in P$; $\mathrm{ver}_p P = \{u \in T_p P \mid T\pi(u) = 0\}$ is the vertical space at p, and there is a unique decomposition $u = \mathrm{hor}\, u + \mathrm{ver}\, u$ corresponding to $T_p P = \mathrm{hor}_p P + \mathrm{ver}_p P$. If ρ: $G \to GL(V)$ is a representation of G in the vector space V, then the V-valued k-form ϕ on P is said to be of type ρ if $\delta_a^* \phi = \rho_{a^{-1}} \circ \phi$, where $\delta_a^* \phi$ is the pullback of ϕ by δ_a. The k-form hor ϕ defined by $\mathrm{hor}\,\phi(u_1, u_2, \ldots, u_k) = \phi(\mathrm{hor}\, u_1, \mathrm{hor}\, u_2, \ldots, \mathrm{hor}\, u_k)$ is then also of type ρ and $D\phi = \mathrm{hor}\, d\phi$ is a $(k+1)$-form of type ρ, called the covariant exterior derivative of ϕ. For example, the connection 1-form ω is G' valued, of type Ad, and $\Omega = D\omega = d\omega + \frac{1}{2}[\omega, \omega]$ is the curvature form. The k-form ϕ is called horizontal relative to π if $\phi(u_1, u_2, \ldots, u_k) = 0$ for any $u_2, \ldots, u_k \in T_p P$ and $u_1 \in \mathrm{ver}_p P$.

The covariant exterior derivative of a horizontal V-

[a)]On leave from the Institute of Theoretical Physics, Warsaw University, Hoza 69, Warsaw, Poland.

 0022-2488/78/1905-1100$1.00

valued form of type ρ is given by the formula

$$D\phi = d\phi + \rho'(\omega)\Lambda\phi,$$

where $\rho'(\omega)$: $TP \to \mathcal{L}(V)$ is obtained by composition of ω with the derived map ρ': $G' \to \mathcal{L}(V)$ and the wedge sign implies both the exterior multiplication of forms on P and the evaluation map $\mathcal{L}(V) \times V \to V$. For example, Ω is a G'-valued horizontal 2-form of type Ad, and since $\mathrm{Ad}_A(B) = [A, B]$ $(A, B \in G')$ we have $D\Omega = d\Omega + [\omega, \Omega] = 0$ (the last equality is the "Bianchi identity").

In a gauge theory, the total space P of the bundle is interpreted as the space of phase factors,[17] M is the space–time manifold, G is the gauge group, ω is the gauge potential, and Ω is the gauge field.

In theoretical physics, one usually works with local sections of the bundle and the corresponding pullbacks; $s: U \to P$ is then called a local gauge, $\Gamma = s^*\omega$ is the potential and $s^*\Omega$ is the field strength in gauge s. The map $\delta_a: P \to P$ is interpreted as a gauge transformation of the first kind. If $S: U \to G$, then $s': U \to P$ defined by $s'(x) = \delta(s(x), S(x))$ is another local section. We say that the local gauge s' is obtained from the local gauge s by the gauge transformation of the second kind S. Assuming $G \subset GL(V)$, we may write the relation between Γ and $\Gamma' = s'^*\omega$ as

$$\Gamma' = S^{-1}\Gamma S + S^{-1}dS.$$

To construct a gauge theory, one should

(i) choose a gauge group G and consider principal G-bundles endowed with connections,

(ii) specify the type of particle(s) coupled to the gauge field: this is done by picking out the representation(s) $\rho: G \to GL(V)$; wavefunctions of particles of type ρ are then given by 0-forms of type ρ; $\phi \circ s$ is the wavefunction in gauge s;

(iii) make an assumption about the field equations satisfied by the gauge field.

For example, in electromagnetism the gauge group is U(1), all its irreducible representations are of the form $\rho_n: \mathrm{U}(1) \to \mathrm{U}(1)$, where $\rho_n(u) = u^n$, $n \in Z$, $u \in U(1)$. A particle of type ρ_n is simply a particle of electric charge n and the covariant exterior derivative of its wavefunction is

$$D\phi = d\phi + n\omega\phi.$$

Since $U(1) \ni u = \exp iA$, where $A \in R$, the Lie algebra of U(1) is iR and the form ω is purely imaginary. The Maxwell equations on P are $d\Omega = 0$ and $d^*\Omega = 4\pi^* j$, where j is the 1-form of electric current and the star appearing on the left of a horizontal form denotes its dual with respect to the metric lifted from the base space. In an SU(n) theory, one often takes $\rho = \mathrm{Ad}$; ϕ is then called a Higgs field. In this case, Ω is assumed to fulfill the field equation

$$D^*\Omega = 4\pi^* j.$$

In the sourceless case, which is the only one considered here,

$$D^*\Omega = 0. \tag{3}$$

STIEFEL BUNDLES AND GRASSMANN MANIFOLDS

Stiefel bundles over Grassmann manifolds[18,19] generalize the Hopf fibrations

$$S_{2n-1} \to CP_{n-1}, \tag{4}$$

$$S_{4n-1} \to HP_{n-1}, \tag{5}$$

which are known to admit sourceless, topologically nontrivial gauge fields.[13]

Let F be one of the following: the field R of the reals, the field C of complex numbers, or the division algebra H of quaternions. If $z \in F$, then $\bar{z} = z$ for $F = R$ and $\bar{z}$ is the conjugate of z for $z \in C$ or H. Consider the right vector space F^n and the scalar product defined by

$$(\mathbf{u}|\mathbf{v}) = \bar{u}_\alpha v_\alpha, \quad \mathbf{u} = (u_\alpha), \quad \mathbf{v} = (v_\alpha), \quad \mathbf{v} = (v_\alpha) \in F^n. \tag{6}$$

The unit coordinate vectors in F^n are ϵ_α, $\alpha = 1, \ldots, n$ and thus $\mathbf{u} = \sum u_\alpha \epsilon_\alpha$. Here and in sequel, summation is understood over the range indicated by a pair of repeated indices.

Let $\mathrm{U}_n(F)$ be the connected component containing the unit of the (Lie) group of linear transformations $a: F^n \to F^n$, $a\mathbf{u} = (\sum_{\beta=1}^n a_{\alpha\beta}u_\beta)$, preserving the form (6). If $a = (a_{\alpha\beta})$ and $a^+ = (\bar{a}_{\beta\alpha})$ is a transpose conjugate matrix, then (6) is preserved, $(a\mathbf{u}|a\mathbf{v}) = (\mathbf{u}|\mathbf{v})$ provided that $a^+a = I$. Depending on the field, the group $\mathrm{U}_n(F)$ is the group of rotations, the unitary group, or the symplectic group

$$\mathrm{U}_n(F) = \begin{cases} \mathrm{SO}(n) & \text{for } F = R, \\ \mathrm{U}(n) & \text{for } F = C, \\ \mathrm{Sp}(n) & \text{for } F = H, \end{cases}$$

and its real dimension is $\frac{1}{2}n(n+1)\dim_R F - n$.

For $k = 1, 2, \ldots, n$ one defines a k-frame u in F^n as the ordered set $(\mathbf{u}_1, \ldots, \mathbf{u}_k) = (\mathbf{u}_i) = u$ of k orthonormal vectors,

$$(\mathbf{u}_i|\mathbf{u}_j) = \delta_{ij}, \quad \mathbf{u}_i \in F^n, \quad i, j = 1, \ldots, k.$$

Each vector $\mathbf{u}_i$ is the n-tuple of elements of F, $\mathbf{u}_i = (u_{\alpha i})$, $\alpha = 1, \ldots, n$. We put $l = n - k$ and make the set of all k-frames in F^n into a manifold. The connected component of $(\epsilon_1, \ldots, \epsilon_k)$ in this manifold is called the Stiefel space $\mathrm{V}_{n,k}(F)$. The group $\mathrm{U}_n(F)$ acts on $\mathrm{V}_{n,k}(F)$ to the left by $(a, u) \to au = u'$, $u'_{\alpha i} = \sum_\beta a_{\alpha\beta}u_{\beta i}$. This action is transitive and the stability group at $(\epsilon_1, \ldots, \epsilon_k)$ is the subgroup $\mathrm{U}_l(F)$ of $\mathrm{U}_n(F)$, consisting of all matrices of the form

$$\begin{pmatrix} \delta_{ij} & 0 \\ 0 & a_{AB} \end{pmatrix},$$

where $(a_{AB}) \in \mathbf{U}_l(F)$, $A, B = k+1, \ldots, n$. Therefore, $\mathrm{V}_{n,k}(F)$ may be identified with the left coset space $\mathrm{U}_n(F)/\mathrm{U}_l(F)$ and, under this identification, the canonical projection

$$\pi_0: U_n(F) \to \mathrm{U}_n(F)/\mathrm{U}_l(F) = \mathrm{V}_{n,k}(F) \tag{7}$$

maps

$$a = \begin{pmatrix} a_{ij} & a_{iB} \\ a_{Aj} & a_{AB} \end{pmatrix} \in \mathrm{U}_n(F)$$

into the k-frame $u=(\mathbf{u}_i)$, where $\mathbf{u}_i=\sum_\alpha \epsilon_\alpha a_{\alpha i}$.

The group $U_k(F)$, which may also be considered as a subgroup of $U_n(F)$, acts on $V_{n,k}(F)$ to the right: $(u,a)\to \delta_a(u)=ua=u'$, $u'_{\alpha i}=\sum_j u_{\alpha j}a_{ji}$, $a=(a_{ij})\in U_k(F)$. The quotient of $V_{n,k}(F)$ by this action is a set $G_{n,k}(F)$ which can be given the structure of a differentiable manifold of real dimension equal to $k\cdot l\dim_R F$. This Grassmann manifold may also be described as the set of all (oriented for $F=R$) k-planes through the origin in F^n: If u', $u\in V_{n,k}(F)$, then u and u' span the same (oriented) k-plane if $u'=ua$ for some $a\in U_k(F)$.

The Stiefel bundle

$$\pi: V_{n,k}(F)\to G_{n,k}(F)$$

with group $U_k(F)$ reduces for $k=1$ to (4) or (5), depending on whether $F=C$ or H; moreover, $V_{n,1}(R)=G_{n,1}(R)=S_{n-1}$.

The canonical 1-form on $U_n(F)$

$$\omega=a^+da=(\omega_{\alpha\beta})$$

has values in the Lie algebra of $U_n(F)$, i.e., each of the component forms $\omega_{\alpha\beta}$ $(\alpha,\beta=1,\ldots,n)$ is F-valued and $\bar\omega_{\alpha\beta}+\omega_{\alpha\beta}=0$. The form ω is left invariant, right equivariant,

$$\delta_a^*\omega=\mathrm{Ad}_{a^{-1}}\omega, \text{ for any } a\in U_n(F). \tag{8}$$

and satisfies the Maurer–Cartan equations

$$d\omega_{\alpha\beta}+\omega_{\alpha\gamma}\wedge\omega_{\gamma\beta}=0. \tag{9}$$

The forms $\omega_{\alpha j}$ $(\alpha=1,\ldots,n;\ j=1,\ldots,k)$ are horizontal relative to $\pi_\circ$. It follows from (8) that the forms ω_{ij} are invariant under the action of $U_l(F)$ on $U_n(F)$ to the right; therefore, they project to forms on $V_{n,k}(F)$, which will be denoted by the same symbols. Moreover, the forms (ω_{ij}) satisfy on $V_{n,k}(F)$ conditions (1) and (2): They define the natural connection on the Stiefel bundle.[14]

Similarly, the quadratic form $\sum_{\alpha,j}\bar\omega_{\alpha j}\omega_{\alpha j}$ is invariant under $U_l(F)$ and defines a Riemannian metric dl^2 on $V_{n,k}(F)$,

$$\sum_{\alpha,j}\bar\omega_{\alpha j}\omega_{\alpha j}=\pi_\circ^*\,dl^2.$$

The quadratic form $\sum_{A,j}\bar\omega_{Aj}\omega_{Aj}$ on $U_n(F)$ is invariant under the action of both $U_l(F)$ and $U_k(F)$. Therefore, it is the pullback of, or projects to, a Riemannian metric ds^2 on $G_{n,k}(F)$

$$\sum_{A,j}\bar\omega_{Aj}\omega_{Aj}=(\pi\circ\pi_\circ)^*\,ds^2 \tag{10}$$

and

$$dl^2=\pi^*(ds)^2+\sum_{i,j}\bar\omega_{ij}\omega_{ij} \quad \text{on } V_{n,k}(F).$$

The metric on $V_{n,k}(F)$ is of the type considered in generalized Kaluza–Klein theories.[16] For $k=1$, dl^2 reduces to the natural Riemannian metric of a sphere whereas ds^2 is the Fubini–Study metric of a projective space ($F=C$ or H)[13] or the natural metric of a sphere ($F=R$). The metrics dl^2 and ds^2 are positive definite and invariant under the action of $U_n(F)$ on $V_{n,k}(F)$ and $G_{n,k}(F)$, respectively.

It is clear from (10) that $U_n(F)$ considered as a bundle over $G_{n,k}(F)$ may be viewed as a restriction of the bundle of orthonormal frames of $G_{n,k}(F)$, corresponding to the canonical injection $U_k(F)\times U_l(F)\to U_{k+l}(F)$. The form $\delta_{ij}\omega_{AB}+\delta_{AB}\omega_{ij}$ on $U_n(F)$ defines a Levi-Civita connection for $G_{n,k}(F)$. Its curvature form is $\delta_{ij}\Omega_{AB}+\delta_{AB}\Omega_{ij}$, where, by virtue of (9),

$$\Omega_{AB}=d\omega_{AB}+\omega_{AB}+\omega_{AC}\wedge\omega_{CB}=\bar\omega_{iA}\wedge\omega_{iB}$$

and

$$\Omega_{ij}=d\omega_{ij}+\omega_{ik}\wedge\omega_{kj}=\bar\omega_{Ai}\wedge\omega_{Aj}. \tag{11}$$

The form Ω_{ij} (or Ω_{AB}) is the pullback to $U_n(F)$ of the curvature form of the natural connection on $V_{n,k}(F)$ (or $V_{n,l}(F)$).

PROOF OF $D*\Omega=0$

Consider the following $l=(n-k)$-forms on $U_n(F)$:

$$\psi_{i_1i_2\cdots i_l}=\frac{1}{l!}\epsilon_{A_1\cdots A_l}\omega_{A_1i_1}\wedge\omega_{A_2i_2}\wedge\cdots\wedge\omega_{A_li_l},$$

where $\epsilon_{A_1\cdots A_l}$ is the Levi-Civita symbol. The ψ's are symmetric in the indices $i_1,\ldots,i_l$. These forms are horizontal relative to $\pi_\circ$. For $F=R$ they are also invariant under the action of $SO(l)$ on $SO(n)$. Therefore, they are pullbacks by π of forms on $V_{n,k}(R)$ which will be denoted by the same symbols. For $F=C$, the $2l$-forms

$$\psi_{i_1\cdots i_l}\wedge\psi_{j_1\cdots j_l} \tag{12}$$

are invariant under the action of $U(l)$ on $U(n)$. By using

$$\epsilon_{A_1A_2\cdots A_l}\epsilon^{B_1B_2\cdots B_l}=\delta^{B_1}_{[A_1}\delta^{B_2}_{A_2}\cdots\delta^{B_l}_{A_l]}$$

and Eq. (11), the projections of (12) onto $V_{n,k}(C)$ may be represented as an equivariant homogeneous polynomial of degree l in Ω_{ij}.

Evaluating the covariant exterior derivative of ψ on $V_{n,k}(R)$, using the Maurer-Cartan equation (9) and $\omega_{AB}+\omega_{BA}=0$, one obtains

$$D\psi_{i_1\cdots i_l}=0 \quad \text{on } V_{n,k}(R). \tag{13}$$

It follows directly from the Bianchi identity that covariant exterior derivative of (12) is zero.

The $k\cdot l$ form on $SO(n)$ given by

$$\eta\propto\epsilon_{i_1^1\cdots i_1^k}\cdots\epsilon_{i_l^1\cdots i_l^k}\psi_{i_1^1}\wedge\cdots\wedge\psi_{i_1^k\cdots i_l^k} \tag{14}$$

is invariant under both $SO(k)$ and $SO(l)$; therefore, it projects to a nonzero form of maximal degree on $G_{n,k}(R)$ which may be identified with the volume kl-form. The form (14) considered on $U(n)$ changes by a phase factor under the action of $U(k)$ and $U(l)$. Therefore, the form $\bar\eta\wedge\eta$ projects to a volume element on $G_{n,k}(C)$. In this case, the volume element is an invariant polynomial of degree $k\cdot l$ in Ω.

The volume elements define orientation on the Grassmannian, which, together with the Riemannian metric ds^2, enables us to define and compute the dual of any form on $G_{n,k}(F)$ or of any horizontal form on $V_{n,k}(F)$. In particular, for $F=R$ we obtain

$$*\Omega_{ij}\propto\epsilon_{iji_1^3\cdots i_1^k}\epsilon_{i_2^1i_2^2\cdots i_2^k}\cdots\epsilon_{i_l^1i_l^2\cdots i_l^k}\cdot\Omega_{i_1^1i_1^2}\wedge\Omega_{i_3^1i_3^2}\wedge\cdots\wedge\Omega_{i_l^1i_l^2}\wedge\Psi_{i_1^3\cdots i_l^3}\wedge\cdots\wedge\Psi_{i_1^k\cdots i_l^k}.$$

The Bianchi identities, together with (13), imply that the curvature form Ω_{ij} is sourceless,

$$D{}^*\Omega_{ij}=0. \tag{15}$$

In the complex case, the form ${}^*\Omega_{ij}$ is of even degree and may be represented as an (equivariant) polynomial of degree $k\cdot l-1$ in Ω. Therefore, Eq. (15) holds as a consequence of the Bianchi identities.

EXAMPLES

The simplest examples correspond to $F=R$ and $l=1$. In this case we have the fibration

$$\mathrm{SO}(n)=\mathrm{V}_{n,n-1}(R)\to \mathrm{G}_{n,n-1}(R)=S_{n-1}$$

and the corresponding $\mathrm{SO}(n-1)$-gauge field is simply the Levi-Civita connection of the sphere S_{n-1}. When standard parametrization is used, the metric of the sphere is $ds^2=\omega_{n1}^2+\omega_{n2}^2+\cdots+\omega_{n\,n-1}^2$, $\omega_{n1}=d\theta_1$, $\omega_{n2}=\sin\theta_1 d\theta_2,\ldots,\omega_{n\,n-1}=\sin\theta_1\sin\theta_2\cdots\sin\theta_{n-2}d\theta_{n-1}$ and $\Omega_{ij}=\omega_{ni}\Lambda\omega_{nj}$.

In the complex case the fibration corresponding to $l=1$ is

$$\mathrm{SU}(n)=\mathrm{V}_{n,n-1}(C)\to \mathrm{G}_{n,n-1}(C)=CP_{n-1}$$

with the gauge $\mathrm{U}(n-1)$. For $n=2$ this reduces to the Hopf fibration $S_3\to S_2$. The gauge group is U(1) and the connection corresponds to the magnetic pole of lowest order. For $n=3$ the fibration is $\mathrm{SU}(3)\to CP_2$, the gauge group is $\mathrm{U}(2)=\mathrm{U}(1)\times\mathrm{SU}(2)$ and the connection decomposes into the "electromagnetic instanton" and the solution recently found by C.N. Yang.[20] For instance, with Fubini–Study metric on CP_2

$$ds^2=\bar\omega_{31}\omega_{31}+\bar\omega_{32}\omega_{32},$$

$$\omega_{31}=\sigma e^{-i\mu}\{i\sin\theta\cos\phi[(1/\sigma^2)\,d\mu-\sin^2\theta\sin^2\phi\,d\nu]-\cos\theta\cos\phi\,d\theta+\sin\theta\sin\phi\,d\phi\},$$

$$\omega_{32}=\sigma e^{-i\nu}\{i\cos\theta\sin\theta\sin\phi\,d\nu-\sin\phi\,d\theta-\sin\theta\cos\theta\cos\phi\,d\phi\},$$

$$\sigma=(\sin^2\theta\sin^2\phi+\cos^2\theta)^{-1/2}\quad\text{and}\quad\Omega_{ij}=\bar\omega_{3i}\Lambda\omega_{3j}.$$

The case $k=1$ is trivial when $F=R$ and reduces to Hopf fibration (4) for $F=C$. For low dimensions the isomorphism between different groups or Lie algebras leads to a correspondence between different Grassmannians and the solutions of gauge equations.

CONCLUDING REMARKS

The importance of the Stiefel bundles and of the corresponding natural connections lies in their universality: If G is an orthogonal, unitary, or symplectic group, then any principal G-bundle P with connection ω can be obtained by embedding its base space M in a Grassmann manifold W of sufficiently high dimension.[14] Clearly, the connection on P, induced by the embedding $k: M\to W$, in general will not correspond to a sourceless gauge field. An interesting problem is to characterize the embeddings k which lead to sourceless gauge fields.

A simple example of such an embedding is known in the case of magnetic poles: For any integer m, the embedding.

$$k_m: S_2\to CP_m$$

is given in terms of homogenous coordinates by

$$k_m(z_1,z_2)=(z_1^m,\ (m_1)^{1/2}z_1^{m-1}z_2,\ \ldots,\ (m_p)^{1/2}z_1^{m-p}z_2^p,\ \ldots,\ z_2^m)$$

pulls back the natural connection on the U(1)-bundle $S_{2m+1}\to CP_m$ to $L(m,1)\to S_2$, where $L(m,1)$ is the lens space.

The method described here is easily generalized to noncompact groups. By replacing the positive definite form (6) with

$$\bar u_1 v_1+\cdots+\bar u_p v_p-\bar u_{p+1}v_{p+1}-\cdots-\bar u_{p+q}v_{p+q}$$

one is led to the groups

$$\mathrm{U}_{p,q}(F)=\begin{cases}\mathrm{SO}(p,q) & \text{for } F=R\\ \mathrm{U}(p,q) & \text{for } F=C\\ \mathrm{Sp}(p,q) & \text{for } F=H\end{cases}$$

and fiberings

$$\mathrm{U}_{p,q}(F)\to \mathrm{V}_{p,p;m,n}(F)=\mathrm{U}_{p,q}(F)/\mathrm{U}_{p-m,q-n}(F)$$
$$\to \mathrm{G}_{p,q;m,n}(F)=\mathrm{U}_{p,q}(F)/\mathrm{U}_{p-m,q-n}(F)\times\mathrm{U}_{m,n}(F).$$

The corresponding Stiefel bundles, with connections which are also sourceless, are defined over an indefinite Grassmann manifold in which metric (10) is replaced by metric with signature

$$[\dim_R F(m(p-m)+n(q-n)),\ \dim_R F(n(p-m)+m(q-n))].$$

[1] H. Weyl, Z. Physik **56**, 330 (1929).
[2] C. N. Yang and R. L. Mills, Phys. Rev. **96**, 191 (1954).
[3] S. Weinberg, Phys. Rev. Lett. **19**, 1264 (1967); A. Salam, in *Elementary Particle Theory*, edited by N. Svartholm, (Almqvist and Wiksell, Stockholm, 1968).
[4] T. W. B. Kibble, J. Math. Phys. **2**, 212 (1961).
[5] C. N. Yang, Phys. Rev. Lett. **33**, 445 (1974).
[6] P. A. M. Dirac, Proc. Roy. Soc. A **133**, 60 (1931).
[7] A. A. Belavin, A. M. Polyakov, A. S. Schwartz, and Yu. S. Tyupkin, Phys. Lett. B **59**, 85 (1975).
[8] R. Jackiw and C. Rebbi, Phys. Rev. Lett. **37**, 172 (1976).
[9] G. t'Hooft, Phys. Rev. D **14**, 3432 (1976).
[10] E. Witten, Phys. Rev. Lett. **38**, 121 (1977).
[11] R. Jackiw, C. Nohl, and C. Rebbi, Phys. Rev. D **15**, 1642 (1977).
[12] M. F. Atiyah and R. S. Ward, "Instantons and Algebraic Geometry" (Oxford University preprint, 1977).
[13] A. Trautman, Intern. J. Theor. Phys. (to be published).
[14] M. S. Narasimhan and S. Ramanan, Amer. J. Math. **83**, 563 (1961); **85**, 223 (1963).
[15] S. Sternberg, *Lectures on Differential Geometry* (Prentice-Hall, Englewood Cliffs, N. J., 1965).
[16] A. Trautman, Rep. Math. Phys. (Toruń) **1**, 29 (1970).
[17] T. T. Wu and C. N. Yang, Phys. Rev. D **14**, 437 (1976).
[18] D. Husemoller, *Fibre Bundles* (McGraw-Hill, New York, 1966).
[19] S. S. Chern, *Complex Manifolds without Potential Theory* (Van Nostrand, New York, 1967).
[20] C. N. Yang (private communication).

Radiation (damping) in a universe with topologically closed space sections[a)]

R. Beig

Institut für Theoretische Physik der Universität Wien, Boltzmanngasse 5, 1090 Wien, Austria
(Received 28 March 1977)

A soluble model for an interacting oscillator-field system discussed previously is reconsidered with Minkowski space replaced by a Lorentzian static manifold with topologically closed space sections. It is shown that the general solution to the dynamical equations is—in the technical sense—almost periodic in time. In the case of the Einstein universe a more detailed discussion is presented, containing especially a study of the "thermodynamic limit" in which the radius of the universe tends to infinity.

I. INTRODUCTION

This work concerns itself with the dynamics of a system in which a harmonic one-dimensional oscillator interacts linearly with a scalar massless field in $(3+1)$-dimensional spacetime. It grew out of a study of an analogous sytem in Minkowski space.[1] There it was shown that, provided the Cauchy data have finite energy, the system, though reversible, has a certain "dissipative" property. The energy of the oscillator vanishes in the limit as time $|t| \to \infty$. The finite-energy solutions may hence be described by saving that they correspond to an incident wave which excites an initially (with $t \to -\infty$) quiescent oscillator which then reradiates the energy thus acquired to infinity.[2]

If in this set up one replaces Minkowski space by a spacetime with topologically closed (in particular compact) space slices, a new situation arises. First, the total energy is now always finite. Second, related to the first, energy cannot "escape to infinity." It is hence to be expected, that the dissipative behavior in the sense mentioned above will strictly be absent here.

Preliminary attempts to integrate the equations in special cases by a rather "brute force" method, as outlined in the Appendix, suggested (but failed to prove) that the system might behave quite analogously to a quantum system in a finite volume, the wavefunction of which, due to the discreteness of the energy spectrum, is an almost periodic function of time. (This is, of course, the quantum analog of Poincaré recurrence; see, e.g., Ref. 3.)

Since that is a typical Hilbert space result, it seemed natural to tackle the present problem in a Hilbert space formulation which is introduced in Sec. II. Section III contains the proof of almost periodicity. In Sec. IV the solution to the Cauchy problem is written down more explicitly. To obtain more concrete information, we specialize the underlying manifold to be the Einstein universe in Sec. V. We show that, letting the radius R of the universe tend to infinity, one recovers the time evolution in Minkowski space (provided an obvious identification between the fields in the two models is made). Considering the asymptotic behavior of the oscillator in the two cases, this especially implies non-interchangeability of the limits $t \to \infty$, $R \to \infty$, a fact familiar from nonequilibrium statistical mechanics. From local causality arguments it is clear that the closed topology, though decisive for the asymptotic behavior, cannot have any impact on the oscillator for times $t < R\pi$. It is shown that its motion is (again with suitable modifications) in fact, equal to the one in flat space. This is not unexpected on grounds of the local conformal flatness of the Einstein universe, from which the absence of tails in the radiation follow. In the appendix a heuristic argument is given which should serve to show that the almost periodic motion of the oscillator is turned into damping if the universe contains matter which absorbs radiation. In view of the results in Ref. 4 it seems likely that incorporation of the cosmic expansion in the model has a similar effect.

II. THE MODEL

We consider as spacetime M a smooth manifold of the form $M = R^1 \times N$, where N is a closed, orientable 3-manifold. M is endowed with a line element g of Lorentzian signature and a global timelike Killing vector field $\partial/\partial t$. In terms of the set of equations to be described shortly one can without loss of generality assume g to be conformally rescaled such that

$$ds^2 = g_{\mu\nu}\,dx^\mu\,dx^\nu = dt^2 - \gamma_{ij}(\mathbf{x})\,dx^i\,dx^j \qquad (2.1)$$
$$(\mu,\nu = 0,\ldots,3;\ i,j = 1,\ldots,3),$$

where $\gamma_{ij}(\mathbf{x})$ is a Riemannian metric on N. The model consists of a one-dimensional harmonic oscillator $Q(t)$ coupled to a scalar massless field $\Phi(\mathbf{x}, t)$. The equations are $(c = m = 1)$:

$$\frac{d^2}{dt^2}Q(t) + \omega_0^2 Q(t) = \ddot{Q}(t) + \omega_0^2 Q(t) = \lambda\int dV\,\rho(\mathbf{x})\Phi(\mathbf{x}, t), \qquad (2.2a)$$

$$\left(\Box + \frac{R}{6}\right)\Phi(\mathbf{x}, t) = \left(\frac{\partial^2}{\partial t^2} - \Delta + \frac{{}^3R}{6}\right)\Phi(\mathbf{x}, t) = \lambda\rho(\mathbf{x})Q(t) \qquad \left(dV\rho = 1\right). \qquad (2.2b)$$

Here $dV = (\det\gamma_{ij})^{1/2}\,d^3x$ is the invariant volume element on N. $\Box = g^{\mu\nu}\nabla_\mu\nabla_\nu$, $\Delta = \gamma^{ij}\nabla_i\nabla_j$, R is the scalar curvature on M, ω_0^2 is the spring constant, λ is a coupling

[a)]Work supported by a stipend of the Bundesministerium für Wissenschaft und Forschung.

 0022-2488/78/1905-1104$1.00

constant, and $\rho(\mathbf{x})$ is a prescribed scalar $\in C^\infty(N)$. We further assume the three-scalar curvature ${}^3R$ on N to be positive, $R = {}^3R > 0$.[5] The equations are invariant under conformal transformations which do not involve time.

By Cauchy (initial) data to the system (2.2) we mean a quadruple $|f\rangle = |Q, \dot{Q} = P, \Phi(\mathbf{x}), \dot{\Phi}(\mathbf{x}) = \Pi(\mathbf{x})\rangle$, where $\Phi, \Pi \in C^\infty(N)$ and which, for later convenience, is allowed to attain complex values.

Use will be made of the following (Sobolev) spaces: $H^0(N) = L^2(N)$, consisting of complex-valued square integrable functions $u(\mathbf{x})$ on N: $(u|u) = \int dV |u(\mathbf{x})|^2 < \infty$, $H^i(N)$ $(i = 1, 2)$ of functions whose weak derivatives up to the ith order are in $L^2(N)$. The following known properties of the positive operator $L = -\Delta + R/6$ will be used in a crucial way (see, e.g., Ref. 6):

(A) L with the domain $D(L) = H^2(N)$ is self-adjoint in $H^0(N)$;

(B) L^{-1} is a compact operator which implies that the spectrum of L consists of isolated eigenvalues ν_i with $0 < \nu_1 < \nu_2 < \cdots$ of finite multiplicity.

To introduce a norm into the space H of Cauchy data $|f\rangle = |Q, P, \Phi, \Pi\rangle$, the total conserved energy seems to be a good candidate,

$$E(f) = \tfrac{1}{2}[\omega_0^2 |Q|^2 + |P|^2] + \tfrac{1}{2}[(\Phi|L\Phi) + (\Pi|\Pi)] - \frac{\lambda}{2}[Q(\rho|\Phi) + \bar{Q}(\Phi|\rho)]. \tag{2.3}$$

Let $\tilde{\Phi} = \Phi - \lambda Q L^{-1}\rho$. Assume that ω_0^2, λ_1, and $\rho(\mathbf{x})$ are arranged in such a way that

$$\bar{\omega}^2 = \omega_0^2 - \lambda^2(\rho|L^{-2}\rho) > 0. \tag{2.4}$$

[For given λ, $\rho(\mathbf{x})$ this can always be achieved by making the "bare" spring constant ω_0^2 sufficiently large.] E can now be written

$$E(f) = \tfrac{1}{2}[\bar{\omega}^2 |Q|^2 + |P|^2] + \tfrac{1}{2}[(\tilde{\Phi}|L\tilde{\Phi}) + (\Pi|\Pi)] \tag{2.5}$$

and is hence greater then or equal to 0 and equal to 0 if and only if $|f\rangle = 0$. Henceforth we write $E^{1/2}(f) = \|f\|$. $\|\ \|$ constitutes a norm in H which is obviously induced by

$$\langle f_1|f_2\rangle = \tfrac{1}{2}[\omega_0^2 \bar{Q}_1 Q_2 + \bar{P}_1 P_2] + \tfrac{1}{2}[(\Phi_1|L\Phi_2) + (\Pi_1|\Pi_2)] - \frac{\lambda}{2}[\bar{Q}_1(\rho|\Phi_2) + Q_2(\Phi_1|\rho)]. \tag{2.6}$$

The completion of H in the norm $\|\ \|$ is given by $R^2 \oplus H^1(N) \oplus H^0(N)$ which will also be denoted by H.

We now write (2.2) in "first order form"

$$\frac{\partial f}{\partial t} = iAf, \tag{2.7}$$

where (in obvious notation)

$$A = \frac{1}{i}\left(\begin{array}{cc|cc} 0 & 1 & 0 & 0 \\ -\omega_0^2 & 0 & \lambda(\rho| & 0 \\ \hline 0 & 0 & 0 & 1 \\ \lambda|\rho) & 0 & -L & 0 \end{array}\right). \tag{2.8}$$

Energy conservation implies that A is symmetric with respect to $\langle\ \rangle$ in $D(A) = R^2 \oplus H^2(N) \oplus H^1(N)$. [Contrary to quantum mechanics, which (2.7) reminds one of, A has nothing to do with the energy of the system. It would rather be appropriate to call it the "frequency operator."]

III. DYNAMICS

Theorem 1: A is self-adjoint.

Proof: Define for $\mathrm{Im}z \neq 0$ the following quantities:

$$|\mu(z)) = (z^2 - L)^{-1}|\rho), \tag{3.1}$$

$$D(z) = z^2 - \omega_0^2 - \lambda^2(\rho|\mu(z)) = z^2 - \bar{\omega}^2 - \lambda^2 z^2(\mu(0)|\mu(z)), \tag{3.2}$$

$$B(z) = D(z)(z^2 - L)^{-1} + \lambda^2|\mu(z))(\overline{\mu(z)}|. \tag{3.3}$$

A straightforward calculation gives for the resolvent $R(z) = (A - z)^{-1}$ of A,

$$\frac{1}{i}R(z) = (D(z))^{-1}\left(\begin{array}{cc|cc} iz & 1 & -\lambda iz(\overline{\mu(z)}| & -\lambda(\overline{\mu(z)}| \\ -z^2 + D(z) & iz & \lambda z^2(\overline{\mu(z)}| & -\lambda iz(\overline{\mu(z)}| \\ \hline -\lambda iz|\mu(z)) & \lambda|\mu(z)) & izB(z) & B(z) \\ \lambda z^2|\mu(z)) & -\lambda iz|\mu(z)) & -z^2B(z) + D(z) & izB(z) \end{array}\right). \tag{3.4}$$

Using self-adjointness of L with $D(L) = H^2(N)$ in $H^0(N)$, inspection of the structure of $R(z)$ shows that it maps H into $D(A) = R^2 \oplus H^2(N) \oplus H^1(N)$ for $\mathrm{Im}z \neq 0$ which implies[7] that A is self-adjoint. Q.E.D.

Self-adjointness implies that A generates a strongly continuous one-parameter group of unitary transformations $U(t) = \exp(iAt)$.[8] The unique solution (in Hilbert space) to the Cauchy problem is therefore given by

$$|f(t)\rangle = \exp(iAt)|f(0)\rangle. \tag{3.5}$$

This can also be written as an inverse Laplace transform

$$|f(t)\rangle = \lim_{c\to\infty}\frac{1}{2\pi}\int_{\epsilon - ic}^{\epsilon + ic} ds\, \exp(st)R(-is)|f(0)\rangle \qquad (t > 0, \epsilon > 0) \tag{3.6}$$

which will be useful later on.

It is still to be seen that the time evolution maps C^∞ data into themselves. However, a result of that sort is known for the inhomogeneous wave equation with more general source $\rho(\mathbf{x}, t)$.[9] In our case $\rho(\mathbf{x}, t) \in C^\infty(N)$ for all t. Also, as a consequence of the strong continuity of the evolution, $\rho(\mathbf{x}, t) = \rho(\mathbf{x})Q(t)$ is itself continuous in t. Under these conditions the assumptions of Ref. 9 are valid.

Theorem 2: The spectrum of A consists of isolated eigenvalues of finite multiplicity and is symmetric with respect to the origin, $\cdots \omega_{-2} < \omega_{-1} < 0 < \omega_1 < \omega_2 < \cdots$ with $\omega_{-i} = -\omega_i$.

Proof: Consider $R(0) = A^{-1}$. For what follows it is convenient to set $\lambda = 0$ in the inner product (2.6) and observe that this yields an equivalent norm on H (though, of course, A is now no longer symmetric). We now write A^{-1} in the form

$$A^{-1} = \left(\begin{array}{c|c} x & x \\ \hline x & 0 \end{array}\right) + \left(\begin{array}{c|c} 0 & 0 \\ \hline 0 & x \end{array}\right). \tag{3.7}$$

It is easily seen that both terms in (3.7) are bounded

operators in the "$\lambda = 0$ norm" and hence in the energy norm. The first term in (3.5) is of finite rank and therefore compact. The second term reads

$$i(\omega_0^2)^{-1}\left(\begin{array}{c|cc} 0 & 0 & \\ \hline 0 & 0 & -L^{-1} \\ & 1 & 0 \end{array}\right) + \text{finite rank.} \tag{3.8}$$

One has to worry only about the first term in (3.8). It is bounded since

$$\begin{aligned}(L^{-1}\Pi \,|\, LL^{-1}\Pi) + (\Phi \,|\, \Phi) &\leq \|L^{-1}\|_{L^2}(\Pi \,|\, \Pi) + C(\Phi \,|\, L\Phi) \\ &\leq D[(\Phi \,|\, L\Phi) + (\Pi \,|\, \Pi)].\end{aligned} \tag{3.9}$$

It is also self-adjoint. Furthermore its square is

$$(\omega_0^2)^{-2}\left(\begin{array}{c|cc} 0 & 0 & \\ \hline 0 & L^{-2} & 0 \\ & 0 & L^{-1} \end{array}\right)$$

which is compact by statement (B) in Sec. II. By a well-known theorem this implies compactness of this operator itself. Hence A^{-1}, as a sum of compact operators, is compact, and the spectrum of A is purely discrete and of finite multiplicity. The symmetry of the spectrum with respect to the origin is a reflection of Eqs. (2.2) being real and follows from inspection of the resolvent (3.4) or else from the eigenvalue equations to be investigated later. Q.E.D.

From this theorem it follows that the general solution to (2.2a) and (2.2b) is of the form

$$|f(t)\rangle = \sum_{i=-\infty}^{\infty} \exp(i\omega_i t)a_i |f_i\rangle, \quad a_i \in C, \tag{3.10}$$

$|f_i\rangle$ being the complete set of eigenfunctions. The sum (3.10) converges uniformly in t in the norm $\|\ \|$. This implies uniform convergence of the corresponding sum in the expression for $Q(t)$, $\dot{Q}(t)$ contained in (3.10) and, in the C^∞ case, for the ones for $\Phi(\mathbf{x}, t)$, $\dot{\Phi}(\mathbf{x}, t)$ almost everywhere in N. Thus, by a general theorem,[10] $Q(t)$, $\dot{Q}(t)$ and, almost everywhere in N, $\Phi(\mathbf{x}, t)$, $\dot{\Phi}(\mathbf{x}, t)$ are almost periodic functions of time.

This result exhibits the fact that the system—strictly speaking—does not have any tendency to reach a "state of equilibrium" for arbitrary large times where all the energy has been dissipated into the field degrees of freedom. We say "strictly speaking," because it is physically obvious and it will be shown in Sec. V in the case of the Einstein universe that there are means of escaping the implications of that result. Either one performs a limit, where the "radius" R of the universe (defined suitably) tends to infinity, *before* the limit $t \to \infty$. Or, more generally, one proves that, confining the support of Φ, $\dot{\Phi}$ at $t = 0$ to a region of length $l \ll R$ which contains the support of ρ and provided $\lambda^2 \gg R^{-1}$, the system does come to equilibrium in the above sense if only observed on a time scale $\bar{t} < R\pi$.

As preparation we need more information on the spectral properties of A.

IV. EIGENVALUES AND VECTORS OF A

Writing down the eigenvalue equation for A implies two equations which can be immediately arrived at by formally inserting an exponential ansatz into (2.2), $Q(t) = \exp(i\omega t)Q_\omega$, $\Phi(\mathbf{x}, t) = \exp(i\omega t)\Phi_\omega(\mathbf{x})$:

$$(-\omega^2 + \omega_0^2)Q_\omega = \lambda(\rho \,|\, \Phi_\omega), \tag{4.1}$$

$$(-\omega^2 + L)\,|\,\Phi_\omega) = \lambda Q_\omega \,|\, \rho), \tag{4.2}$$

and noting that

$$|f_\omega\rangle = |Q_\omega, i\omega Q_\omega, \Phi_\omega(\mathbf{x}), i\omega\Phi_\omega(\mathbf{x})\rangle. \tag{4.3}$$

Due to compactness of L^{-1} one can apply Fredholm theory to (4.2). Let Λ_i be the subspace of $H^0(N)$ spanned by the eigenvectors $L\,|\,\nu_{i_j}) = \nu_i\,|\,\nu_{i_j})$. Then the following cases (listed in increasing generality) have to be considered:

(A) $|\rho)$ is orthogonal to Λ_i for some i

This implies that (4.2) may be solved for $\omega^2 = \nu_i$ to give

$$|\Phi_{\pm\sqrt{\nu_i}}) = \lambda Q_{\pm\sqrt{\nu_i}} \sum_{\substack{j=1 \\ j\neq i}}^{\infty} \frac{(\nu_j \,|\, \rho)\,|\,\nu_j)}{\nu_j - \nu_i} + \text{const}\,|\,\nu_i). \tag{4.4}$$

[We assume $\{|\,\nu_i)\}$ to be an orthonormal system in $H^0(N)$.] Equation (4.1) now gives

$$\left[-\nu_i + \bar{\omega}^2 - \lambda^2\nu_i \sum_{\substack{j=1 \\ j\neq i}}^{\infty} \frac{|(\nu_j \,|\, \rho)|^2}{\nu_j(\nu_j - \nu_i)}\right] Q_{\pm\sqrt{\nu_i}} = 0. \tag{4.5}$$

(4.5) is clearly solved by $Q_{\pm\sqrt{\nu_i}} = 0$. If, by accident, the term in brackets in (4.5) vanishes for the considered i, then $Q_{\sqrt{\nu_i}}$ may be not equal to 0, and hence the eigenspaces $H_{\pm i} \subset H$ for eigenvalues $\omega = \pm\sqrt{\nu_i}$ have a degeneracy in addition to the one present in Λ_i.

(B) $|\rho)$ is orthogonal to a proper subset of Λ_i

Now $\pm\sqrt{\nu_i}$ appears as an eigenvalue of A as well, but necessarily $Q_{\pm\sqrt{\nu_i}} = P_{\pm\sqrt{\nu_i}} = 0$. This can be described by saying that the modes of the free field do not couple to the oscillator. They will, however, not occur at all in the coupled system in the general case where

(C) $|\rho)$ is not orthogonal to any $|\,\nu_{i_j})$

Now we have to take $\omega^2 \neq \nu_i$. In this case (4.2) has the unique solution

$$|\Phi_\omega) = \lambda Q_\omega (L - \omega^2)^{-1}\,|\,\rho) \tag{4.6}$$

which, inserted into (4.1), gives

$$[-\omega^2 + \bar{\omega}^2 - \lambda^2\omega^2(\rho \,|\, (L - \omega^2)^{-1}\rho)]Q_\omega = -D(\omega)Q_\omega = 0. \tag{4.7}$$

Hence the eigenvalues ω with $\omega^2 \neq \nu_i$ are given by the solutions to the equation

$$D(\omega) = \omega^2 - \bar{\omega}^2 + \lambda^2\omega^2(\rho \,|\, (L - \omega^2)^{-1}\rho) = 0. \tag{4.8}$$

$D(\omega)$ is real analytic in the open intervals $(\pm\sqrt{\nu_i}, \pm\sqrt{\nu_{i+1}})$ and has poles at $\omega^2 = \nu_i$. Furthermore one notes that

$$\frac{d}{d\omega^2} D(\omega) = 1 + \lambda^2(\rho \,|\, (L - \omega^2)^{-2}\rho) > 0. \tag{4.9}$$

Hence simple graphical argument shows that there is exactly one zero $\pm\omega_i$ of $D(\omega)$ in each of the intervals $(-\sqrt{\nu_i}, -\sqrt{\nu_{i-1}})$, respectively $(\sqrt{\nu_{i-1}}, \sqrt{\nu_i})$, with $i = 1, 2, \cdots$ and $\nu_0 = 0$. All of these are simple zeros.

This result is important because it says that pertur-

bations of the given metric in N such as distortion of symmetry or increase of the volume of N which will increase the density of eigenvalues of the operator L, will do the same to the coupled system. These perturbations will thus, roughly speaking, tend to increase the recurrence time of the system. This is expected on intuitive grounds, since the almost periodic behavior of, say, the oscillator is due to the fact that waves, after emission, agitate the oscillator again after having "travelled round the universe." Perturbations will, of course, in general decrease the efficiency of this process of "self-interaction via topology" and make it negligible for the actual universe (if it turned out to be closed). From the point of view of physical effects the situation is less ridiculous in the case of, for example, a radiating body enclosed between ideal reflecting walls, a problem which may easily be modeled by making minor changes in the present work.

Despite the fact that some more progress in the study of our system might perhaps be attainable on grounds of general theorems, we now prefer to make life much easier by specifying the manifold N to be S^3.

V. THE EINSTEIN UNIVERSE

Now the Lorentz metric is given by

$$ds^2 = dt^2 - \mathcal{R}^2[d\alpha^2 + \sin^2\alpha\, d\Omega^2], \tag{5.1}$$

Here $0 \leq \alpha \leq \pi$, $d\Omega^2 = d\vartheta^2 + \sin^2\vartheta\, d\varphi^2$ is the line element on the 2-sphere. There are coordinate singularities at $\alpha, \vartheta = 0, \pi$.

For ρ we take for simplicity a point distribution at $\alpha = 0$,

$$\rho(\mathbf{x}) = \rho(\alpha) = -\frac{1}{2\pi\mathcal{R}^3}\frac{\delta'(\alpha)}{\sin\alpha}. \tag{5.2}$$

According to (2.4), the singular nature of ρ demands a "renormalization" $\omega_0^2 \to \infty$. Therefore, to be completely rigorous, one would have to take an extended $\rho = \rho_\epsilon$ with,

$$\lim_{\epsilon\to 0}\rho_\epsilon = -\frac{1}{2\pi\mathcal{R}^3}\frac{\delta'(\alpha)}{\sin\alpha}$$

and take the limit $\epsilon \to 0$ in the final results. Such a procedure has been described in detail in Ref. 1 for Minkowski space and will be simply omitted here since it does not add insight to the problem.

For S^3 the eigenfunctions of L are the well-known hyperspherical harmonics[11] which we denote by $|nlm)$ with $n = 1, 2, \cdots$; $0 \leq l \leq n-1$; $-l \leq m \leq l$. One has

$$L|nlm) = \frac{n^2}{\mathcal{R}^2}|nlm). \tag{5.3}$$

In $\mathbf{x}$ space we have

$$|n00) = \frac{1}{(2\pi^2\mathcal{R}^3)^{1/2}}\frac{\sin n\alpha}{\sin\alpha} \tag{5.4}$$

which is all of $|nlm)$ which will be needed. ρ may be expanded in the distributional sense as

$$|\rho) = \sum_{n=1}^{\infty}\frac{n}{(2\pi^2\mathcal{R}^3)^{3/2}}|n00), \tag{5.5}$$

so one is in case (B) of Sec. IV. By eigenfunction summation we obtain

$$\Phi_\omega(\alpha) = (L-\omega^2)^{-1}\lambda|\rho)$$
$$= \frac{\lambda}{4\pi\mathcal{R}\sin\alpha}\frac{\sin\omega\mathcal{R}(\pi-\alpha)}{\sin\omega\mathcal{R}\pi} \quad \left(\omega \neq \pm\frac{n}{\mathcal{R}}\right). \tag{5.6}$$

Since attention will be restricted to the motion of the oscillator, we only have to consider the eigenvalues of A which are not equal to $\pm n/\mathcal{R}$. These are the solutions $\omega_i(\omega_{-i})$ to the equation

$$D(\omega) = \omega^2 - \bar{\omega}^2 - 2\Gamma\omega\cot\omega\mathcal{R}\pi + \frac{2\Gamma}{\mathcal{R}\pi} = 0 \quad \left(2\Gamma = \frac{\lambda^2}{4\pi}\right). \tag{5.7}$$

The associated normalized eigenvectors are

$$|f\omega_i\rangle = c_i^{-1}|1, i\omega_i, \Phi_{\omega_i}, i\omega_i\phi_{\omega_i}\rangle, \tag{5.8}$$

where

$$c_i^2 = \left[\frac{\omega}{2}\frac{d}{d\omega}D(\omega)\right]_{\omega=\omega_i}$$
$$= \omega_i^2 - \Gamma\omega_i\cot\omega_i\mathcal{R}\pi + \Gamma\omega_i^2\frac{\mathcal{R}\pi}{\sin^2\omega_i\mathcal{R}\pi}. \tag{5.9}$$

Using (5.7) this gives

$$c_i^2 = \tfrac{1}{2}(\omega_i^2 + \bar{\omega}^2) + \frac{\Gamma}{\mathcal{R}\pi} + \Gamma\mathcal{R}\pi\omega_i^2\left[1 + \left(\frac{\omega_i^2 - \bar{\omega}^2 + 2\Gamma/\mathcal{R}\pi}{2\Gamma\omega_i}\right)^2\right] \tag{5.10}$$

and (5.6) may be written

$$\Phi_{\omega_i}(\alpha) = \frac{\lambda}{4\pi\mathcal{R}\sin\alpha}\left[\cos\omega_i\mathcal{R}\alpha - \frac{\omega_i^2 - \bar{\omega}^2 + 2\Gamma/\mathcal{R}\pi}{2\Gamma\omega_i}\times\sin\omega_i\mathcal{R}\alpha\right], \tag{5.11}$$

Spectrally decomposing $|f(t)\rangle = \exp(iAt)|f(0)\rangle$ yields

$$|f(t)\rangle = \sum_{i=-\infty}^{\infty}\exp(i\omega_i t)\langle f_{\omega_i}|f(0)\rangle|f_{\omega_i}\rangle. \tag{5.12}$$

This means for $Q(t)$,

$$Q(t) = \sum_{i=-\infty}^{\infty}\frac{\exp(i\omega_i t)}{2c_i^2}[\omega_i^2 Q - i\omega_i\mathring{Q} + \omega_i^2(\Phi_{\omega_i}|\Phi) - i\omega_i(\Phi_{\omega_i}|\mathring{\Phi})], \tag{5.13}$$

where Q, $\mathring{Q}$, Φ, and $\mathring{\Phi}$ are the Cauchy data.

Next we ask what happens to our solution in the limit $\mathcal{R} \to \infty$. To clarify what we mean by this limit we note that the manifold $(M_\mathcal{R}, g_\mathcal{R})$ which we take to be the Einstein manifold with the line $\alpha = \pi$ removed can be regarded as limiting to Minkowski spacetime with $\mathcal{R} \to \infty$. To this end apply the coordinate transformation $r = \alpha/\mathcal{R}$ before taking the limit in the metric (5.1). [A more precise formulation of such a limiting process is afforded by viewing the collection $(M_\mathcal{R}, g_\mathcal{R})$ as a five-manifold to which Minkowski spacetime can be attached as a boundary given by $\mathcal{R} = \infty$ (see Ref. 12).] In the application of $\mathcal{R} \to \infty$ to the initial value solution $(Q(t), \Phi(t))$ it is to be understood that $Q(0)$, $\mathring{Q}(0)$ is kept fixed, while $\Phi(0)$, $\mathring{\Phi}(0)$ are supported in $0 \leq \alpha \leq l/\mathcal{R} < \pi$. We take $\Phi(0)$, $\mathring{\Phi}(0)$ to be functions solely of α, for simplicity. Define $\bar{\Phi}(\mathcal{R}\alpha, 0) = \bar{\Phi}(r, 0) = \Phi(\alpha, 0)$ and $\bar{\mathring{\Phi}}$ analogously and regard them as functions on R^+ with support in $0 \leq r \leq l$, fixed during $\mathcal{R} \to \infty$.

Lemma: With the limit understood in this sense, $(Q(t), \Phi(\mathbf{x}, t))$ tends to the Minkowski space solution with $R \to \infty$.

Proof: Consider for the moment only the $Q, \dot{Q}$ terms in Eq. (5.13). From the general discussion of the zeros of $D(\omega)$ and (5.10) one observes that the sums get replaced by integrals. The surviving terms are

$$Q(t) = \dot{G}(t)Q + G(t)\dot{Q}, \tag{5.14}$$

where

$$G(t) = \frac{2\Gamma}{\pi}\int_{-\infty}^{+\infty} d\omega \frac{-i\omega \exp(i\omega t)}{4\Gamma^2\omega^2 + (\omega^2 - \bar{\omega}^2)^2}. \tag{5.15}$$

The integrand in (5.15) has a typical Lorentzian shape which is characteristic for resonance scattering of light in the presence of damping (see Refs. 13 and 14).

Performing the integration by contour closing gives

$$G(t) = \exp(-\Gamma|t|)\frac{\sin(\bar{\omega}^2 - \Gamma^2)^{1/2}t}{(\bar{\omega}^2 - \Gamma^2)^{1/2}}. \tag{5.16}$$

Therefore, if $\Phi(\mathbf{x}, t) = \dot{\Phi}(\mathbf{x}, t) = 0$ at $t = 0$, the oscillator suffers exponential damping in both time directions. Defining, as in Ref. 1,

$$\bar{\psi}(r) = r\int\frac{d\Omega}{4\pi}\bar{\Phi}(\mathbf{x}, 0),$$
$$\bar{\chi}(r) = r\int\frac{d\Omega}{4\pi}\dot{\bar{\Phi}}(\mathbf{x}, 0), \tag{5.17}$$

one gets for the remaining terms

$$\lambda\int_0^t dr[\dot{G}(t-r)\bar{\psi}(r) + G(t-r)\bar{\chi}(r)]. \tag{5.18}$$

The Eqs. (5.14), (5.15), and (5.18) are in full accordance with Ref. 1. An analogous statement may be shown to be true for $\Phi(\mathbf{x}, t)$ in the limit $R \to \infty$, which proves the lemma.

The result just obtained is, however, just a special case of the next one. (We treat them separately, because the methods of proof are different and interesting in themselves.)

For reasons described in the Introduction, there is to be expected a close similarity of the motion of the oscillator for $t < R\pi$ with the one in flat space ($R = \infty$). More specifically, this will—for causality reasons—hold for the $Q, \dot{Q}$ terms in (5.13) for $t < 2R\pi$ and for the $\Phi, \dot{\Phi}$ terms only for $t < R\pi$.

In order to check this, it is convenient to work in the representation (3.6). Using (3.4), (3.6), and (5.7) we get for the $Q, \dot{Q}$ term

$$Q(t) = \frac{1}{2\pi i}\int_{\epsilon-i\infty}^{\epsilon+i\infty} ds \frac{\exp(st)}{s^2 + 2\Gamma s\coth R\pi s + \bar{\omega}^2 - 2\Gamma/R\pi} \times[sQ + \dot{Q}] \quad (t > 0). \tag{5.19}$$

Now the denominator in the integrand in (5.19) may be split up

$$\frac{1}{s^2 + 2\Gamma s\coth R\pi s + \bar{\omega}^2 - 2\Gamma/R\pi} = \frac{1}{s^2 + 2\Gamma s + \bar{\omega}^2 - 2\Gamma/R\pi} + \text{rest}, \tag{5.20}$$

where the rest consists of $\exp(-2R\pi s)$, $\exp(-4R\pi s)$ multiplied by terms which are regular and vanish with $|s| \to \infty$ in the half plane $\mathrm{Re}\, s \geq 0$. Therefore, for $0 < t < 2R\pi$ we are left with

$$Q(t) = \frac{1}{2\pi i}\int_{\epsilon-i\infty}^{\epsilon+i\infty} ds \frac{\exp(st)}{s^2 + 2\Gamma s + \bar{\omega}^2 - 2\Gamma/R\pi}[sQ + \dot{Q}]. \tag{5.21}$$

Here $\mu^2 = \bar{\omega}^2 - 2\Gamma/R\pi$ plays the role of $\bar{\omega}^2$ in the flat case. We assume that $\mu^2 > 0$. Then (5.21) may be integrated to give

$$Q(t) = \dot{H}(t)Q + H(t)\dot{Q}, \tag{5.22}$$

where

$$H(t) = \exp(-\Gamma t)\frac{\sin(\mu^2 - \Gamma^2)^{1/2}t}{(\mu^2 - \Gamma^2)^{1/2}}. \tag{5.23}$$

Defining

$$\psi(\alpha) = R\sin\alpha\int\frac{d\Omega}{4\pi}\Phi(\mathbf{x}),$$
$$\chi(\alpha) = R\sin\alpha\int\frac{d\Omega}{4\pi}\dot{\Phi}(\mathbf{x}) \tag{5.24}$$

the $\Phi, \dot{\Phi}$ terms become for $0 < t < R\pi$

$$\lambda\int_0^{t/R} d\alpha\,[\dot{H}(t - R\alpha)\psi(\alpha) + H(t - R\alpha)\chi(\alpha)]. \tag{5.25}$$

It should be noted that the process just described may be used to calculate the explicit solution in every interval $nR\pi < t < (n+1)R\pi$. Essentially the expression corresponding to $H(t)$ will pick up terms t^k $(k \leq n)$.

From Eqs. (5.22), (5.23), and (5.25) the statement made at the end of Sec. III may easily be made precise in the following lemma.

Lemma: Let $l/R = O(\epsilon)$, $1/\Gamma R = O(\epsilon)$, and $1 - t/R\pi = O(1)$. Then $Q(t) = O(e^{-1/\epsilon})$ with $\epsilon \to 0$.

ACKNOWLEDGMENTS

I am indebted to Dr. D. Dubin for very helpful conversations. I would like to thank Dr. D. W. Sciama for the hospitality at the Department of Astrophysics, University of Oxford, where most of this work was carried out.

APPENDIX

The formalism described in this work to deal with the Cauchy problem for a specific field-particle system relies heavily upon energy conservation and the related self-adjointness of the operator which generates the time evolution. This means, for example, that the presence of absorbing matter is discarded. This will not be very relevant for the asymptotic motion of the oscillator in the case of open space sections, but will be critical in the closed case. To see this we use the following heuristic argument.

Consider, again, the wave equation

$$\left(\Box + \frac{R}{6}\right)\Phi(\mathbf{x}, t) = -\frac{\lambda}{2\pi R^3}\frac{\delta'(\alpha)}{\sin\alpha}Q(t) \tag{A1}$$

in the Einstein universe. A "retarded" solution of (A1) is given by

$$\Phi(\mathbf{x},t)=\frac{\lambda}{4\pi R\sin\alpha}\left\{Q(t-R\alpha)+\sum_{\nu=1}^{\infty}[Q(t-R\alpha-2\nu R\pi)-Q(t+R\alpha-2\nu R\pi)]\right\}. \tag{A2}$$

It is retarded in the sense that $\Phi(\mathbf{x},t)=0$ for $t<R\alpha$ if $Q=0$ for $t<0$. [The sum in (A2) will not in general converge. Hence our manipulations are to be regarded as purely formal and are justified only by the fact that the pairs (Q,Φ) we end up with are actually solutions of (A1) together with (2.2a).] Inserting (A2) into (2.2a) leads, after renormalization, to

$$\ddot{Q}(t)+2\Gamma\dot{Q}(t)+4\Gamma\sum_{\nu=1}^{\infty}\dot{Q}(t-2\nu R\pi)+\mu^2Q(t)=0 \tag{A3}$$

which may be viewed as an analog to the Lorentz–Dirac equation of electrodynamics in the present circumstances. It looks time asymmetric, but it is not. Using (A3) at time $t-4\pi R$ and subtracting from (A3) yields an equation which is manifestly time symmetric with respect to $t-2\pi R$. Hence, for any $Q(t)$, which solves (A3) for all times, $Q(-t)$ is also a solution. For this to be the case it is crucial that the factor in front of the infinite sum in (A3) is twice the coefficient of $\dot{Q}(t)$. Inserting the ansatz $Q(t)=\exp(i\omega t)$ into (A3) and summing up the series formally, one arrives at the "characteristic equation" $D(\omega)=0$, which is familiar from Secs. III and V. Its having only real solutions is, of course, guaranteed in the preceding treatment by self-adjointness of the operator A.

Imagine now an absorbing medium uniformly distributed in the universe. We model the presence of absorption primitively by replacing (A3) by

$$\ddot{Q}(t)+2\Gamma\dot{Q}(t)+4\Gamma\sum_{\nu=1}^{\infty}\alpha^{-\nu}\dot{Q}(t-2\nu R\pi)+\mu^2Q(t)=0\quad(\alpha>1). \tag{A4}$$

It is not difficult to convince oneself that the zeros of the corresponding characteristic equation now move to the upper half of the complex ω plane.

One should, of course, also try to incorporate the expansion of the universe $R=R(t)$ into the model. Then Eq. (A3) (absorption is discarded now) becomes

$$\ddot{Q}(t)+2\Gamma\dot{Q}(t)+4\Gamma\sum_{\nu=1}^{\infty}\dot{Q}\{f^{-1}[f(t)-2\nu\pi]\}+\mu^2Q(t)=0, \tag{A5}$$

where

$$f(t)=\int^t\frac{dt'}{R(t')}\,.$$

In view of the results obtained in Ref. 4 it seems likely that the cosmic expansion $[\ddot{f}(t)<0]$ also turns the periodic solutions of (A3) into damped ones. However, I have been unable to prove this.

[1] P.C. Aichelburg and R. Beig, Ann. Phys. **98**, 264 (1975).
[2] R. Beig, to appear in Proceedings of the 1976 E. Fermi Summer School, Varenna, LXVII Course.
[3] I.C. Percival, J. Math. Phys. **2**, 235 (1961).
[4] P.C. Aichelburg and R. Beig, Phys. Rev. D **15**, 389 (1977).
[5] For a metric which is already of the form (1.1), all that is necessary for the two theorems in Sec. III is that R is a given positive function on N. Also the number of dimensions of N could be made arbitrary.
[6] R.S. Palais, *Seminar on the Atiyah–Singer Index Theorem* (Ann. of Math. Studies, No. 57, Princeton, N.J., 1965).
[7] See, e.g., M. Reed and B. Simon, *Methods of Modern Mathematical Physics I* (Academic, New York, 1972).
[8] E. Hille, *Functional Analysis and Semi-Groups* (Am. Math. Soc., New York, 1948).
[9] L. Maurin, Stud. Math. **16**, 200 (1958).
[10] A.S. Besicovitch, *Almost Periodic Functions* (Dover, New York, 1954).
[11] E.C. Englefield, *Group Theory and the Coulomb Problem* (Wiley-Interscience, New York, 1972).
[12] R. Geroch, Commun. Math. Phys. **13**, 180 (1969).
[13] O. Loudon, *The Quantum Theory of Light* (Clarendon, Oxford, 1973).
[14] W. Thirring, *Cargese Lectures in Physics*, edited by M. Levy (Gordon & Breach, New York, 1968).

Coherent states and stochastic formulation of quantum mechanics

A. Lonke

Department of Physics, Ben-Gurion University of the Negev, P. O. Box 653, Beer-Sheva 84120 Israel
(Received 10 November 1977)

For physical systems with Hamiltonians analytic in the position and momentum operators the diagonal matrix elements of the density operator in the coherent state representation may be considered as a phase-space probability measure in the stochastic formulation of quantum mechanics. The correspondence rule relating quantum-mechanical observables to ordinary functions in phase space is obtained. It is shown that the dynamics of the time-dependent phase-space probability densities is again representable in the form of a stochastic process.

I. INTRODUCTION

The physical meaning of a stochastic formulation of quantum mechanics may be concisely elucidated by the following problem. Given a quantum-mechanical system in the state described by the density operator ρ and given an observable A, find a phase-space probability measure $\rho(q,p)$ and a corresponding function $A(q,p)$ such that[1]

$$\mathrm{Tr}\rho A = \int dq\, dp \rho(q,p) A(q,p). \tag{1.1}$$

From the pure operational point of view the two approaches are equivalent. Both furnish the same one-to-one correspondence between well-defined mathematical objects and the outcomes of physical measurements. Both lead to the same results once this correspondence is established. Nevertheless, the stochastic formulation of quantum mechanics has several attractive features. The first is a conceptual one. It is the fact that the quantum behavior of a system (like uncertainty relations, etc.) can be explained by the probabilistic nature of the process of measurement in the phase space. More prosaic, but not less important, are the practical implications of putting quantum and classical statistical mechanics upon the same formal basis.

A stochastic formulation of quantum mechanics involves, basically, two interrelated problems. These are the determination of the probability measure $\rho(q,p)$ and the establishment of the proper correspondence between quantum-mechanical observables A and ordinary functions in phase-space, $A(q,p)$. The first attempts in this direction ran into the difficulty of dealing with quasidistributions. Complex distributions had been introduced by Dirac[2] and investigated by Prugovečki[3,4] in the framework of complex probability measures. The Wigner distributions[5] which are not positive definite were exploited by Moyal.[6] It was shown later by Cohen[7,8] that these probability measures are not unique and there exists, in principle, a whole set of distributions satisfying the same conditions. Consequently, no unique correspondence rule can be forced.

A positive definite probability measure in phase-space was recently introduced by Prugovečki.[9] Analyzing the process of measurement[10,11] which in reality is not arbitrarily controllable he introduces the concept of a fuzzy phase space. A point in the ordinary phase space, $(q,p) \in \Gamma$, represents the outcome of two simultaneous infinitely sharp measurements of position and momentum. However, every conceivable act of measurement is accompanied by unavoidable imprecisions. Incorporating them as a part of the observed physical reality, Prugovečki attaches to the outcome of each position and momentum measurement, q and p, the corresponding normalized confidence measures ν_q and ν'_p, which describe the accuracy calibration of the measuring instrument. Thus, a point in the fuzzy phase space is defined as the fourtuple $((q,\nu_q),(p,\nu'_p)) \in (\Gamma,\nu)$. The equivalence of nonrelativistic quantum mechanics based upon sharp and fuzzy measurements was proved by Ali and Doebner.[12]

One encounters here a completely new approach which ascribes to every measuring instrument $I(\chi_q,\hat{\chi}_p)$ the confidence measures[9] χ_q and $\hat{\chi}_p$, depending on the accuracy calibration at q and p. It can be shown[13] that the existence of a fuzzy phase-space representation depends on the precision of the measuring instrument. In fact, such representations do not exist unless the standard deviations of χ_q and $\hat{\chi}_p$ (multiplied by $\sqrt{2}$) satisfy the inequality

$$\sigma_k(\chi_q)\sigma_k(\hat{\chi}_p) \geq \hbar \tag{1.2}$$

for all degrees of freedom k.

Thus, in this formalism the uncertainty relations stem from a postulated quality of the measuring process itself and not from the commutation relations of quantum-mechanical operators. There is a limit, therefore, in the accuracy calibrations of physical instruments. In the optimal case when nature is observed by optimally accurate instruments I_s, the probability density $\rho(q,p)$ turns out to be a diagonal matrix element of the density operator ρ in the coherent state representation.[14] These representations are specified by a positive parameter s, referred to as the instrument characteristic.

In the present note we shall follow a somewhat more conservative path. Our intent is to show that exactly the same family of optimal probability measures may be derived without reference to the limitations of the process of measurement. They follow straightforwardly from the properties of the coherent states. Every coherent state representation depends on the physical

 0022-2488/78/1905-1110$1.00

specifications of the harmonic oscillator which generates it. The later can be described by a single positive parameter equivalent to the mentioned instrument characteristic, s. For every s and for every given operator A, we derive a correspondence rule establishing the explicit form of an ordinary function $A(q,p)$ satisfying (1.1). The rest is a matter of interpretation.

The coherent states of a harmonic oscillator have, in modern terms, a considerably long history. First introduced by Schrödinger[15] for the description of nonspreading wavepackets having oscillatory properties, they reappear in the works of Glauber[16] and Sudarshan and Mehta[17,18] in their investigations of coherent light sources. From this time they are known as the "coherent states," though the importance of these representation spaces, as it was quickly recognized, reaches beyond the quantum-mechanical treatment of radiation fields. One finds them being used, for instance, in the studies of phonons in crystals,[19] superconductivity, and superfluidity[20] and in the quantum-mechanical description of phase and angle variables.[21] The concept of coherent states was suggested as a convenient mathematical tool for various statistical mechanical calculations.[22]

Let us briefly review the main properties of the coherent states which will be used in the subsequent discussion. We start with a single-mode harmonic oscillator with the Hamiltonian

$$H=\hbar\omega(a^\dagger a+\tfrac{1}{2}), \tag{1.3}$$

where, in terms of position and momentum operators,

$$\begin{aligned} a&=[i/(2\hbar m\omega)^{1/2}](p-im\omega q), \\ a^\dagger&=[-i/(2\hbar m\omega)^{1/2}](p+im\omega q), \end{aligned} \qquad [a,a^\dagger]=1 \tag{1.4}$$

The stationary eigenstates of the oscillator are the number states

$$|n\rangle=[(a^\dagger)^n/(n!)^{1/2}]|0\rangle, \tag{1.5}$$

corresponding to the energies

$$E_n=\hbar\omega(n+\tfrac{1}{2}). \tag{1.6}$$

The normalized vacuum state is

$$|0\rangle=(m\omega/\pi\hbar)^{1/4}\exp(-m\omega q^2/2\hbar). \tag{1.7}$$

In the physical Hilbert space spanned by the number states the ladder operators (1.4) are given by

$$a^\dagger|n\rangle=(n+1)^{1/2}|n+1\rangle, \quad a|n\rangle=n^{1/2}|n-1\rangle. \tag{1.8}$$

Obviously, $a|0\rangle=0$. The coherent states are defined as the eigenstates of the annihilation operator,

$$a|\alpha\rangle=\alpha|\alpha\rangle \tag{1.9}$$

where α, the eigenvalue of a non-Hermitian operator, is a complex number. These states can be derived using the Weyl operator

$$D(\alpha)=\exp(\alpha a^\dagger-\alpha^* a), \tag{1.10}$$

which has the following properties,

$$D^\dagger(\alpha)=D^{-1}(\alpha) \quad \text{(unitary)}, \tag{1.11a}$$

$$D^\dagger(\alpha)=D(-\alpha), \tag{1.11b}$$

$$[a,D(\alpha)]=\alpha D(\alpha). \tag{1.11c}$$

Applying the last equation to the vacuum state, we derive, for the functions obeying (1.9),

$$|\alpha\rangle=D(\alpha)|0\rangle. \tag{1.12}$$

Since (1.10) can be rewritten as

$$D(\alpha)=\exp(-\tfrac{1}{2}|\alpha|^2)\exp(\alpha a^\dagger)\exp(-\alpha^* a), \tag{1.13}$$

Eq. (1.12) furnishes the expansion of an arbitrary coherent state in terms of the number states,

$$|\alpha\rangle=\exp(-\tfrac{1}{2}|\alpha|^2)\sum_{n=0}^{\infty}\frac{\alpha^n}{(n!)^{1/2}}|n\rangle. \tag{1.14}$$

The coherent states span a Hilbert space of entire analytic functions. The mathematical properties of these spaces had been studied by Bargmann.[23,24] In general, they arise when one is looking for a natural representation space for complex linear transforms of the position and momentum variables preserving the canonical commutation relations.[25,26] The states (1.14) are not orthogonal for $\alpha\neq\beta$. This is the consequence of dealing with an uncountable overcomplete set of functions. The inner product of two coherent states, as it follows from (1.14), is given by

$$\langle\alpha|\beta\rangle=\exp(\alpha^*\beta-\tfrac{1}{2}|\alpha|^2-\tfrac{1}{2}|\beta|). \tag{1.15}$$

The completeness of this set of states is manifested by the relation

$$\pi^{-1}\int d^2\alpha\,|\alpha\rangle\langle\alpha|=\sum_{n=0}^{\infty}|n\rangle\langle n|=1, \tag{1.16}$$

where the integration is carried out over the entire complex plane, $d^2\alpha=d\,\mathrm{Re}\alpha\,d\,\mathrm{Im}\alpha$. This equation is an immediate result following from (1.14) and the formula

$$\int d^2\alpha\exp(-|\alpha|^2)(\alpha^*)^n\alpha^m=\pi n!\,\delta_{m,n}. \tag{1.17}$$

The coherent state representation has an attractive feature pointed out by Carruthers and Nieto.[21] It consists of the possibility of confining the calculation of measurable physical quantities to operations on diagonal matrix elements exclusively. In order to elucidate this important fact we note, first, that for every function having a Taylor expansion equation (1.17) leads to the following identities:

$$\pi^{-1}\int d^2\beta\exp\{\alpha^*\beta-|\beta|^2\}\beta^n f(\beta^*)=(\partial/\partial\alpha^*)^n f(\alpha^*), \qquad n=0,1,2,\ldots, \tag{1.18a}$$

$$\pi^{-1}\int d^2\beta\exp(\alpha\beta^*-|\beta|^2)(\beta^*)^n f(\beta)=(\partial/\partial\alpha)^n f(\alpha). \tag{1.18b}$$

Consider, now, two operators, say $\rho(a^\dagger,a)$ and $A(a^\dagger,a)$, which by assumption are analytic in $a^\dagger$ and a. This means that they are expandable in terms of these operators and can be represented in the form of a convergent ordered double series (finite or infinite),

$$A(a^\dagger,a)=\sum_{m,n}A_{m,n}(a^\dagger)^m a^n. \tag{1.19}$$

Calculating the diagonal matrix element of the product $A\rho$ in the coherent state representation, we derive

$$\begin{aligned}\langle\alpha|A\rho|\alpha\rangle&=\pi^{-1}\int d^2\beta\langle\alpha|A|\beta\rangle\langle\beta|\rho|\alpha\rangle \\ &=\pi^{-1}\int d^2\beta\,|\langle\alpha|\beta\rangle|^2 A(\alpha^*,\beta)\rho(\beta^*,\alpha)\end{aligned}$$

$$= \pi^{-1} \exp(-|\alpha|^2) \sum_{m,n} A_{m,n} (\alpha^*)^m \int d^2\beta$$
$$\times \exp(\alpha^*\beta - |\beta|^2) \beta^n [\exp(\beta^*\alpha) \rho(\beta^*, \alpha)]$$
$$= \exp(-|\alpha|^2) A(\alpha^*, \partial/\partial\alpha^*) \exp(|\alpha|^2) \rho(\alpha^*, \alpha), \quad (1.20)$$

the last equality following from (1.18a). Thus, the mean value of an arbitrary observable A will be given by

$$\mathrm{Tr}\rho A = \pi^{-1} \int d^2\alpha \exp(-|\alpha|^2) A(\alpha^*, \partial/\partial\alpha^*)$$
$$\times \exp(|\alpha|^2) \langle \alpha | \rho | \alpha \rangle, \quad (1.21)$$

provided, of course, that the analyticity assumption (1.19) holds.

This result of Carruthers and Nieto provides, surprisingly enough, a solution of the original problem. One has the positive definite function $\langle \alpha | \rho | \alpha \rangle$ which may be viewed upon as a phase-space probability density.[27] The fact that it is subjected here to certain differential operations prescribed by $A(\alpha^*, \partial/\partial\alpha^*)$ is hardly an obstacle. One can easily reformulate (1.21) integrating by parts and, thus, establish a proper correspondence between the observable A and an ordinary function $A(\alpha^*, \alpha)$ in the stochastic formulation. Finally, one might prefer to complete the task by the unsignificant change of notation: $\mathrm{Re}\alpha \to q$, $\mathrm{Im}\alpha \to p$, $d^2\alpha \to dq\,dp$.

Equation (1.21) implies that the whole information about the quantum-mechanical system carried by the density operator ρ is conveyed by its diagonal matrix elements in the coherent state representation. Therefore, the off-diagonal matrix elements of ρ in this representation have to be related somehow to the diagonal ones. This connection is found in Sec. II, where we show that the former are integral transforms of the later. In Sec. III this relationship is exploited in order to establish the mentioned correspondence rule between observables and ordinary phase-space functions. As expected, we obtain a result which could be deduced straightforwardly from (1.21). However, having such a "diagonal" representation we are in a position to do more than a mere confirmation of known results. In Sec. IV we investigate the dynamics of the phase-space probability densities. We show that the time propagation of these distributions can also be given a stochastic interpretation.

II. INFORMATIONAL COMPLETENESS OF THE DIAGONAL MATRIX ELEMENTS

An arbitrary coherent state, (1.12), depends, in general, on the complex number α and the physical parameters of the oscillator such as mass and frequency. However, the explicit structure of the creation and annihilation operators (1.4), as well as of the vacuum state (1.7), indicates that every value of mass and frequency of the representative oscillator can be renormalized by a proper choice of the length scale. Therefore, the physical specifications of the harmonic oscillator chosen to generate the coherent state representation can be described by a single parameter.

Let us introduce the dimensionless operator,

$$\mathcal{q} = q/\lambda, \quad (2.1)$$

where λ is the unit of the length scale. Respectively,

$$\mathcal{p} = (\lambda/\hbar)p = -i\partial/\partial\mathcal{q} \quad (2.2)$$

defines a dimensionless momentum operator. With these definitions the creation and annihilation operators (1.4) become

$$a^\dagger(s) = (2)^{-1/2}(s^{-1}\mathcal{p} - is\mathcal{q}),$$
$$a(s) = (2)^{-1/2}(s^{-1}\mathcal{p} + is\mathcal{q}), \quad (2.3)$$

where

$$s = (\hbar/m\omega\lambda^2)^{1/2} > 0 \quad (2.4)$$

is a dimensionless parameter. The properly normalized vacuum state can be written, similarly, as

$$|0, s\rangle = (s\sqrt{\pi})^{-1/2} \exp(-\mathcal{q}^2/2s^2). \quad (2.5)$$

Thus, the dimensionless parameter s solely describes the dependence of the whole representation space on the physical properties of the generating oscillator. This dependence will be inherited, certainly, by the matrix elements of an operator in the coherent state representation. It has nothing in common with the dynamical variable under consideration and reflects exclusively the choice of the representation space. In essence, a matrix element depends on s in the same manner as, for instance, the distance traveled by a particle depends on λ in (2.1). This is the dependence of a quantity representing a physical reality on the adopted scale of measure. In the fuzzy phase-space formulation this parameter s is referred to as the instrument characteristic. Here, however, where the measuring process is not apparent in the formalism the adoption of this terminology would be not appropriate. Bearing in mind that we are dealing with the same parameter, we shall refer here to s as the *scaling parameter* of the coherent state representation space.

Our intent is to use this s dependence of the matrix elements. To this end we calculate, first, the s derivative of an arbitrary coherent state. It is continuous in s and from (1.12) it follows that

$$(\partial/\partial s)|\alpha, s\rangle = [(\partial/\partial s)D(\alpha, s)]|0, s\rangle$$
$$+ D(\alpha, s)(\partial/\partial s)|0, s\rangle. \quad (2.6)$$

This procedure is well defined since the variation of s does not affect the domain of definition of the operators. In fact, the operator $a(s')$ can be written as a linear combination of $a(s)$ and $a^\dagger(s)$ taken with coefficients depending on s and s'. The operator $a(s)$, therefore, is differentiable with respect to s. And so is every function of these operators. The index s in the notation of operators and states will be neglected below for economy of script.

The differentiation of the Weyl operator, (1.10), term by term yields

$$\frac{\partial}{\partial s} D(\alpha) = \sum_{n=1}^{\infty} \frac{1}{n!} \sum_{m=1}^{n} (\alpha a^\dagger - \alpha^* a)^{n-m}$$
$$\times \left(\frac{\partial}{\partial s}(\alpha a^\dagger - \alpha^* a)\right) (\alpha a^\dagger - \alpha^* a)^{m-1}. \quad (2.7)$$

Now, from the explicit expressions for the operators $a^\dagger$ and a, (2.3), we have

$$\frac{\partial}{\partial s}a=-\frac{1}{s}a^\dagger,\quad \frac{\partial}{\partial s}a^\dagger=-\frac{1}{s}a, \tag{2.8}$$

whence

$$\frac{\partial}{\partial s}D(\alpha)=-\frac{1}{s}\sum_{n=1}^{\infty}\frac{1}{n!}\sum_{m=1}^{n}(\alpha a^\dagger-\alpha^* a)^{n-m}$$

$$\times(\alpha a-\alpha^* a^\dagger)(\alpha a^\dagger-\alpha^* a)^{m-1}. \tag{2.9}$$

Taking into account the commutation relation

$$[\alpha a-\alpha^* a^\dagger,\ \alpha a^\dagger-\alpha^* a]=\alpha^2-(\alpha^*)^2, \tag{2.10}$$

Eq. (2.9) can be rewritten as

$$(\partial/\partial s)D(\alpha)=-\frac{1}{s}\sum_{n=1}^{\infty}\frac{1}{n!}\sum_{m=1}^{n}\{(n-m)[\alpha^2-(\alpha^*)^2](\alpha a^\dagger-\alpha^* a)^{n-2}$$

$$+(\alpha a^\dagger-\alpha^* a)^{n-1}(\alpha a-\alpha^* a^\dagger)\}, \tag{2.11}$$

which after simple algebra becomes

$$\frac{\partial}{\partial s}D(\alpha)=-\frac{1}{s}D(\alpha)[\alpha a-\alpha^* a^\dagger+\tfrac{1}{2}\alpha^2-\tfrac{1}{2}(\alpha^*)^2]. \tag{2.12}$$

The second term in (2.6) follows straightforwardly,

$$\frac{\partial}{\partial s}|0\rangle=\frac{1}{s}\left[\left(\frac{\mathcal{P}}{s}\right)^2-\frac{1}{2}\right]|0\rangle. \tag{2.13}$$

Thus, substituting

$$\mathcal{P}=(s/\sqrt{2})(a+a^\dagger), \tag{2.14}$$

we have

$$(\partial/\partial s)|0\rangle=(1/2s)(a^\dagger)^2|0\rangle. \tag{2.15}$$

Together with (2.12) and (2.15) Eq. (2.6) yields

$$(\partial/\partial s)|\alpha\rangle=\frac{1}{2s}D(\alpha)[2(\alpha^* a^\dagger-\alpha a)-\alpha^2+(\alpha^*)^2$$

$$+(a^\dagger)^2]|0\rangle. \tag{2.16}$$

Finally, making use of (1.11c) and (1.4), we derive

$$\frac{\partial}{\partial s}|\alpha\rangle=\frac{1}{2s}[(a^\dagger)^2-\alpha^2]|\alpha\rangle$$

$$=\frac{1}{2s}\left[\left(\frac{\partial}{\partial\alpha}+\tfrac{1}{2}\alpha^*\right)^2-\alpha^2\right]|\alpha\rangle, \tag{2.17a}$$

and similarly

$$(\partial/\partial s)\langle\alpha|=\frac{1}{2s}\left[\left(\frac{\partial}{\partial\alpha^*}+\tfrac{1}{2}\alpha\right)^2-(\alpha^*)^2\right]\langle\alpha|. \tag{2.17b}$$

In these calculations α and α^* have to be considered as independent variables.

The coherent states form a complete space for every s. Considering it as a varying parameter, we have for an arbitrary matrix element of an operator ρ

$$2s\frac{\partial}{\partial s}\langle\alpha|\rho|\beta\rangle=\left[\left(\frac{\partial}{\partial\alpha^*}+\tfrac{1}{2}\alpha\right)^2+\left(\frac{\partial}{\partial\beta}+\tfrac{1}{2}\beta^*\right)^2\right.$$

$$\left.-(\alpha^*)^2-\beta^2\right]\langle\alpha|\rho|\beta\rangle. \tag{2.18}$$

In the following we suppose that ρ is a self-adjoint Hilbert–Schmidt operator belonging to the trace class. Every density operator has to satisfy these requirements.

The solution of the partial differential equation (2.18) is

$$\langle\alpha|\rho|\beta\rangle=\exp[-\tfrac{1}{2}(|\alpha|^2+|\beta|^2)]$$

$$\times\sum_{m,n=0}^{\infty}s^{n-m}W_{m,n}\Psi_m(\alpha^*+\beta)\Psi_n[i(\alpha^*-\beta)], \tag{2.19}$$

where $W_{m,n}$ are constant independent of α^* and β as well as of s and

$$\Psi_n(\gamma)=(n!\,2^n\sqrt{2\pi})^{-1/2}\exp(-\tfrac{1}{4}\gamma^2)H_n(\gamma/\sqrt{2}) \tag{2.20}$$

stands for the analytic continuation of the parabolic cylinder functions[28] satisfying the weighted orthogonality condition

$$(2/\pi)^{1/2}\int d^2\gamma\exp[-|\gamma|^2+\tfrac{1}{2}\gamma^2+\tfrac{1}{2}(\gamma^*)^2]\Psi_m(\gamma)\Psi_n(\gamma)$$

$$=\delta_{m,n}. \tag{2.21}$$

These equations can be easily checked using (1.18) and the integral representation for the Hermite polynomials,

$$H_n(\gamma/\sqrt{2})=(2^{n/2}/\pi)\int d^2z(z^*)^n\exp(-|z|^2-\tfrac{1}{2}z^2+\gamma z). \tag{2.22}$$

Taking the particular case $\alpha=\beta=\alpha_R+i\alpha_I$, we derive for the diagonal matrix elements

$$\langle\alpha|\rho|\alpha\rangle=\exp(-|\alpha|^2)\sum_{m,n=0}s^{n-m}W_{m,n}\Psi_m(2\alpha_R)\Psi_n(2\alpha_I). \tag{2.23}$$

It follows, thus, that the constants $W_{m,n}$ are real. In terms of the diagonal matrix elements they are given by

$$W_{m,n}=4s^{m-n}\int d^2\alpha\exp(|\alpha|^2)$$

$$\times\langle\alpha|\rho|\alpha\rangle\Psi_m(2\alpha_R)\Psi_n(2\alpha_I). \tag{2.24}$$

These quantities, in principle, are measurable. A straightforward calculation (see Appendix) shows that

$$W_{m,n}=(2^{n+m-1}m!\,n!\,\pi)^{1/2}\sum_{j=0}^{[m,n]}(-i/2)^j$$

$$\times[j!(m-j)!(n-j)!]^{-1/2}\operatorname{Tr}\{\rho\mathcal{P}^{m-j}\mathcal{P}^{n-j}\}, \tag{2.25}$$

where $[m,n]$ stands for the smaller of the two positive integers m and n.

We encounter here a remarkable property of the coherent state representation. Every off-diagonal matrix element of ρ can be expressed in terms of the diagonal ones. Substituting (2.24) into (2.19), we derive

$$\langle\alpha|\rho|\beta\rangle=\pi^{-1}\int d^2\gamma\langle\gamma|\rho|\gamma\rangle K(\gamma,\gamma^*|\alpha^*,\beta)$$

$$\times\exp[-\tfrac{1}{2}(|\alpha|^2+|\beta|^2)], \tag{2.26}$$

where

$$K(\gamma,\gamma^*|\alpha^*,\beta)=4\pi\exp(|\gamma|^2)\left(\sum_{m=0}^{\infty}\Psi_m(2\gamma_R)\Psi_m(\alpha^*+\beta)\right)$$

$$\times\left(\sum_{n=0}^{\infty}\Psi_n(2\gamma_I)\,\Psi_n[i(\alpha^*-\beta)]\right). \tag{2.27}$$

The sums in this expression can be evaluated. For a real x and complex η we have

$$\sum_{m=0}^{\infty} \Psi_m(x)\,\Psi_m(\eta)$$
$$= (2\pi)^{-1/2} \exp[-\tfrac{1}{4}(x^2+\eta^2)] \sum_{m=0}^{\infty} (2^{-m}/m!)\, H_m(x/\sqrt{2})\, H_m(\eta/\sqrt{2})$$
$$= (2\pi)^{-1/2} \exp[-\tfrac{1}{4}(x^2+\eta^2)]\pi^{-2} \int d^2z_1 d^2z_2 \times \exp(-|z_1|^2 - |z_2|^2 - \tfrac{1}{2}z_1^2 - \tfrac{1}{2}z_2^2 + xz_1 + \eta z_2 + z_1^* z_2^*)$$
$$= (2\pi)^{-1/2} \exp[-\tfrac{1}{4}(x^2+\eta^2)]\pi^{-1} \int d^2z \times \exp(-|z|^2 - \tfrac{1}{2}z^2 - \tfrac{1}{2}z^{*2} + xz + \eta z^*)$$
$$= (2\pi)^{-1} \exp[-(x-\eta)^2/8] \int dy \exp[iy(x-\eta)], \qquad (2.28)$$

after the integration over Rez has been carried out. Making use of this expression, we write, finally,

$$K(\gamma,\gamma^*|\alpha^*,\beta) = \pi^{-1} \exp[\tfrac{1}{2}(|\gamma|^2 + \alpha^*\gamma + \gamma^*\beta - \alpha^*\beta)] \times \int d^2z \exp[(\alpha^* - \gamma^*)z^* - (\beta - \gamma)z]. \qquad (2.29)$$

Thus, it has been shown that every off-diagonal matrix element of the operator ρ in the coherent state representation (2.26) is nothing but an integral transform of the diagonal ones. The kernel of this transformation given by (2.29) is holomorphic with respect to the complex variables α^* and β defining the representative states. The whole information about the quantum-mechanical state of the physical system as well as the chosen representation is conveyed by the real and positive definite diagonal matrix elements, $\langle\gamma|\rho|\gamma\rangle$. Neither the kernel nor the exponential factor in (2.26) depend on the scaling parameter of the representation space, s. They are, in this sense, universal, and the matrix diagonal of the density operator in this representation is informationally complete. These facts prove to be very instrumental in the stochastic formulation of quantum mechanics.

We would like to note at the end that, for $\alpha = \beta$, the kernel, as might have been expected, furnishes an identity transformation:

$$K(\gamma,\gamma^*|\alpha^*,\alpha) = \pi \exp(|\alpha|^2)\delta(\gamma_R - \alpha_R)\delta(\gamma_I - \alpha_I) = \pi \exp(|\alpha|^2)\delta^{(2)}(\gamma - \alpha). \qquad (2.30)$$

III. THE CORRESPONDENCE RULE

Consider a quantum-mechanical observable, $A(\mathcal{f}, \mathcal{p})$, which, by assumption, is an analytic function of the operators $\mathcal{f}$ and $\mathcal{p}$. By this we mean that in terms of the operators a and $a^\dagger$, (2.3), the operator A may be represented as given by (1.19). Our aim is to express the expectation value of this observable in the quantum-mechanical state ρ,

$$\mathrm{Tr}(\rho A) = \pi^{-2} \int d^2\alpha d^2\beta \langle\alpha|\rho|\beta\rangle\langle\beta|A|\alpha\rangle, \qquad (3.1)$$

in the form of a stochastic average, (1.1).

To this end let us substitute into (3.1) the integral transform (2.26) and

$$\langle\beta|A|\alpha\rangle = \langle\beta|\alpha\rangle \sum_{m,n} A_{m,n}(\beta^*)^m \alpha^n, \qquad (3.2)$$

following from (1.19) and (1.9). With these substitutions (3.1) yields

$$\mathrm{Tr}(\rho A) = \pi^{-1} \int d^2\gamma \langle\gamma|\rho|\gamma\rangle A(\gamma^*,\gamma), \qquad (3.3)$$

where

$$A(\gamma^*,\gamma) = \pi^{-2} \int d^2\alpha d^2\beta K(\gamma,\gamma^*|\alpha^*,\beta) \times \exp\{\beta^*\alpha - |\alpha|^2 - |\beta|^2\} \sum_{m,n} A_{m,n}(\beta^*)^m \alpha^n. \qquad (3.4)$$

The integral in (3.3) has acquired already the desired form. In order to complete the task one has to evaluate explicitly the integral (3.4) which does not depend on the particular quantum-mechanical state of the system. Substituting (2.29), we observe that the β integration can be carried out immediately using (1.18a). The result is

$$A(\gamma^*,\gamma) = \exp(\tfrac{1}{2}|\gamma|^2) \sum_{m,n} A_{m,n}\pi^{-2} \int d^2z d^2\alpha (\tfrac{1}{2}\gamma^* - \tfrac{1}{2}\alpha^* - z)^m \alpha^n \times \exp[-3|\alpha|^2/2 + \tfrac{1}{2}(\alpha\gamma^* + \alpha^*\gamma) + (\alpha^* - \gamma^*)z^* - (\alpha - \gamma)z]. \qquad (3.5)$$

Now, introducing the change of variables, $\alpha \to \alpha - 2z^*$, the integration over z becomes feasible. We have

$$A(\gamma^*,\gamma) = \exp(\tfrac{1}{2}|\gamma|^2) \sum_{m,n} A_{m,n} 2^{-m-1} 3^{-n-1} \pi^{-1} \int d^2\alpha (\gamma^* - \alpha^*)^m \alpha^n \times \exp[-(|\alpha|^2 + \alpha\gamma^* - 3\alpha^*\gamma)/6]. \qquad (3.6)$$

In order to calculate the remaining integral,

$$I_{m,n} = \pi^{-1} \int d^2\alpha (\gamma^* - \alpha^*)^m \alpha^n \exp[-(|\alpha|^2 + \alpha\gamma^* - 3\alpha^*\gamma)/6], \qquad (3.7)$$

it is convenient to introduce the generating function,

$$Y_m(\zeta) = \sum_{n=0}^{\infty} I_{m,n}(\zeta^n/n!) = \pi^{-1} \int d^2\alpha (\gamma^* - \alpha^*)^m \times \exp[-(|\alpha|^2 + \alpha\gamma^* - 3\alpha^*\gamma)/6 + \zeta\alpha]$$
$$= 6 \times 2^m \exp(-\tfrac{1}{2}|\gamma|^2 + 3\zeta\gamma)(\gamma^* - 3\zeta)^m, \qquad (3.8)$$

whence

$$I_{m,n} = (\partial/\partial\zeta)^n Y_m(\zeta)|_{\zeta=0} = 2^{m+1} 3^{n+1} \exp(-\tfrac{1}{2}|\gamma|^2) \sum_{k=0}^{[m,n]} (-1)^k \times \binom{m}{k}\binom{n}{k} k!\, (\gamma^*)^{m-k} \gamma^{n-k}. \qquad (3.9)$$

Substituting this expression into (3.6) we derive, finally,

$$A(\gamma^*,\gamma)=\sum_{m,n}A_{m,n}(\gamma^*)^m\gamma^n\sum_{k=0}^{[m,n]}(-1)^k \times\binom{m}{k}\binom{n}{k}k!\,(1/|\gamma|^2)^k. \quad (3.10)$$

This is the "classical" function corresponding in the stochastic formulation to the quantum-mechanical operator,

$$A(a^\dagger,a)=\sum_{m,n}A_{m,n}(a^\dagger)^m a^n. \quad (3.11)$$

It might be convenient at this stage to recover the usual phase-space notations. The complex number γ is the eigenvalue of the annihilation operator a, (2.3). For this reason let us write

$$\gamma=2^{-1/2}(s^{-1}q+isp), \quad (3.12)$$

where $q=s\sqrt{2}\,\mathrm{Re}\gamma$ and $p=s^{-1}\sqrt{2}\,\mathrm{Im}\gamma$ are c-number. In this notation

$$d^2\gamma=d\,\mathrm{Re}\gamma\,d\,\mathrm{Im}\gamma=\tfrac{1}{2}dq\,dp, \quad (3.13)$$

and the integral (3.3) attains the form of (1.1) with

$$\rho(q,p)=(2\pi)^{-1}\langle\gamma|\rho|\gamma\rangle>0, \quad (3.14)$$

and $A(q,p)$ following from (3.10) after the substitution of (3.12).

It can be easily seen that when the quantum-mechanical observable is a function of the operator $s^{-1}\mathcal{q}+is\mathcal{p}$ (or of its Hermitian conjugate, $s^{-1}\mathcal{q}-is\mathcal{p}$) exclusively then the corresponding ordinary function in the stochastic formulation preserves the same functional form. The correspondence rule in this case (and in this case only) turns out to be merely a replacement of the operators by c-numbers,

$$A(s^{-1}\mathcal{q}+is\mathcal{p})\to A(s^{-1}q\pm isp). \quad (3.15)$$

However, an operator which depends on $\mathcal{q}$ and $\mathcal{p}$ in a different manner will alter its functional form as prescribed by (3.10).

Let us rewrite this expression in terms of q and p. Suppose that we are looking for the correspondence rule for an observable

$$A(\mathcal{q},\mathcal{p})=\sum_{m,n}A_{m,n}\mathcal{q}^m\mathcal{p}^n. \quad (3.16)$$

Then, taking into account (2.24) and the relation between $\mathrm{Tr}\{\rho\mathcal{q}^m\mathcal{p}^n\}$ and the coefficients $W_{m,n}$ [see Appendix Eq. (A3)] it follows that

$$A(\mathcal{q},\mathcal{p})\to A^{(s)}(q,p)=\sum_{m,n}A_{m,n}2^{-m-n}s^{m-n} \times\sum_{j=0}^{[m,n]}(2i)^j\binom{m}{j}\binom{n}{j}j!\,H_{m-j}(q/s)\,H_{n-j}(sp), \quad (3.17)$$

where the superscript s indicates that the average has to be taken with the probability distribution $\rho(q,p;s)$. Consequently, functions of the position or momentum operators exclusively ($m=0$ or $n=0$) are subject to the following correspondence rules:

$$A(\mathcal{q})\to A^{(s)}(q)=\sum_n A_{n,0}(s/2)^nH_n(q/s), \quad (3.18)$$

$$A(\mathcal{p})\to A^{(s)}(p)=\sum_n A_{0,n}(2s)^{-n}H_n(sp). \quad (3.19)$$

IV. TIME-DEPENDENT PHASE-SPACE DISTRIBUTIONS

In the preceding sections ρ was considered to be an arbitrary density operator. In general, for a physical system not necessarily in the state of thermal equilibrium it is time dependent. The dynamics of this operator is governed by the well-known von Neumann equation,

$$i\hbar(\partial/\partial t)\rho_t=[H,\rho_t]. \quad (4.1)$$

This equation determines, in principle, the state of the system at an arbitrary time t provided the initial state (say, at $t=0$) ρ_0 is known. Speaking in operational terms, it predicts the outcome of a determinative measurement performed on a quantum-mechanical system after a preparatory measurement[11] had established the state ρ_0.

The same, obviously, applies to the probability densities in the stochastic formulation of quantum mechanics which are nothing but the diagonal elements of this operator in the coherent state representation. Here, however, due to the existence of the integral transform, (2.26), the variation in time of the phase space probability density, $\langle\gamma|\rho_t|\gamma\rangle$, attains the form of a stochastic process again. The time-dependent probability density becomes an average of a certain time- and system-dependent function taken over the whole phase space with the initial distribution $\langle\gamma|\rho_0|\gamma\rangle$.

In fact, the formal solution of von Neumann's equation is

$$\rho_t=U_t\rho_0U_t^\dagger, \quad (4.2)$$

where

$$U_t=\exp(-iHt/\hbar)=U_{-t}^\dagger \quad (4.3)$$

is the unitary time evolution operator. It follows, thus,

$$\begin{aligned}\langle\gamma|\rho_t|\gamma\rangle&=\pi^{-2}\int d^2\alpha d^2\beta\langle\gamma|U_t|\alpha\rangle\langle\alpha|\rho_0|\beta\rangle\langle\beta|U_t^\dagger|\gamma\rangle\\&=\pi^{-1}\int d^2\eta\langle\eta|\rho_0|\eta\rangle\,\pi^{-2}\int d^2\alpha d^2\beta\\&\quad\times\exp(-\tfrac{1}{2}|\alpha|^2-\tfrac{1}{2}|\beta|^2)\\&\quad\times\langle\gamma|U_t|\alpha\rangle K(\eta,\eta^*|\alpha^*,\beta)\langle\beta|U_t^\dagger|\gamma\rangle,\end{aligned} \quad (4.4)$$

where the second equality results from the substitution of (2.26).

In the subsequent we shall confine the discussion to the case of system Hamiltonians, $H(\mathcal{q},\mathcal{p})$, which are analytic in $\mathcal{q}$ and $\mathcal{p}$. The time evolution operators of such systems undergo the expansion (1.19) and

$$\langle\alpha|U_t|\beta\rangle=\langle\alpha|\beta\rangle U_t(\alpha^*,\beta), \quad (4.5)$$

where $U_t(\alpha^*,\beta)$ is a holomorphic function.

Let us investigate in some more detail the integrand of (4.4). To this end we observe, first, that the function K, (2.29), can be written formally as a matrix element of a certain operator in the coherent state representation. It is easy to check that

$$\exp(-\tfrac{1}{2}|\alpha|^2-\tfrac{1}{2}|\beta|^2)K(\eta,\eta^*|\alpha^*,\beta)$$
$$=\exp(\tfrac{1}{2}|\eta|^2)\pi^{-1}\int d^2z\sum_{n=0}^{\infty}[(-\tfrac{3}{2})^n/n!]\exp(\eta z-\eta^* z^*)$$
$$\times\langle\alpha|\exp(\tfrac{1}{2}\eta+z^*)a^\dagger](a^\dagger)^n a^n\exp[(\tfrac{1}{2}\eta^*-z)a]|\beta\rangle. \quad (4.6)$$

This expression can be simplified. We have

$$\langle\alpha|\exp\{(\tfrac{1}{2}\eta+z^*)a^\dagger\}(a^\dagger)^n a^n\exp\{(\tfrac{1}{2}\eta^*-z)a\}|\beta\rangle$$
$$=(\partial/\partial z^*)^n(-\partial/\partial z)^n\langle\alpha|\exp[(\tfrac{1}{2}\eta+z^*)a^\dagger]$$
$$\times\exp[(\tfrac{1}{2}\eta^*-z)a]|\beta\rangle. \quad (4.7)$$

Now, substituting (4.7) into (4.6) and integrating by parts, we derive

$$\exp[-\tfrac{1}{2}|\alpha|^2-\tfrac{1}{2}|\beta|^2]K(\eta,\eta^*|\alpha^*,\beta)$$
$$=\exp(-|\eta|^2)\langle\alpha|\Omega(\eta,\eta^*)|\beta\rangle, \quad (4.8)$$

where the operator

$$\Omega(\eta,\eta^*)=\pi^{-1}\int d^2z\exp[(\tfrac{1}{2}\eta+z^*)a^\dagger-\eta^* z^*]$$
$$\times\exp[(\tfrac{1}{2}\eta^*-z)a+\eta z] \quad (4.9)$$

depends on the complex parameter η and is self-adjoint. This becomes clear after changing the sign of the integration variable, $z\to -z$, in the expression for the Hermitian conjugate,

$$\Omega(\eta,\eta^*)=[\Omega(\eta,\eta^*)]^\dagger. \quad (4.10)$$

Furthermore, from (4.8) and (2.30) it follows that

$$\mathrm{Tr}\Omega(\eta,\eta^*)=1\quad\forall\ \eta. \quad (4.11)$$

Thus, substituting (4.8) into (4.4), we derive

$$\langle\gamma|\rho_t|\gamma\rangle=\pi^{-1}\int d^2\eta\langle\eta|\rho_0|\eta\rangle\langle\gamma|U_t\Omega(\eta,\eta^*)U_t^\dagger|\gamma\rangle. \quad (4.12)$$

In this equation the variation in time of the phase-space distribution function itself may be given a stochastic interpretation. The probability measure is defined by the initial distribution and is independent of time. The sample space consists of the elementary events which are the diagonal matrix elements of the self-adjoint operator $\Omega(\eta,\eta^*)$ brought into the right time by the unitary transformation U_t reflecting the properties of the physical system under consideration. At $t=0$ this function is concentrated infinitely sharp around the point γ. With time, however, it propagates over the whole η-space. The phase-space probability density at γ and $t>0$ becomes a stochastic average of this function taken with the initial probability measure.

APPENDIX: THE DERIVATION OF EQ. (2.25)

Let us start with the calculation of the trace of ρ. From (2.23) and noting that

$$\exp(-|\alpha|^2)=(2\pi)^{1/2}\Psi_0(2\alpha_R)\Psi_0(2\alpha_I) \quad (A1)$$

[see (2.20)], it follows immediately

$$\mathrm{Tr}\rho=\pi^{-1}\int d^2\alpha\langle\alpha|\rho|\alpha\rangle$$
$$=(2/\pi)^{1/2}\sum_{m,n}W_{m,n}s^{n-m}\int d^2\alpha\Psi_0(2\alpha_R)\Psi_0(2\alpha_I)$$
$$\times\Psi_m(2\alpha_R)\Psi_n(2\alpha_I)$$
$$=(8\pi)^{-1/2}W_{0,0}. \quad (A2)$$

Thus, the first element of the double series $W_{m,n}$ represents the trace of ρ. We turn, now, to the proof of a more general statement, namely,

$$\mathrm{Tr}\{\rho\mathscr{g}^M\mathscr{f}^N\}=(2^{M+N}8\pi)^{-1/2}\sum_{j=0}^{[M,N]}i^j M!N!(j!)^{-1}$$
$$\times[(M-j)!(N-j)!]^{-1/2}W_{M-j,N-j}, \quad (A3)$$

where $[M,N]=\min(M,N)$.

In order to prove this relation, consider the matrix element $\langle\alpha|\mathscr{g}^M\mathscr{f}^N\rho|\alpha\rangle$. From (1.14) and (2.3) it follows that

$$\langle\alpha|\mathscr{g}^M\mathscr{f}^N\rho|\alpha\rangle$$
$$=i^N 2^{-(M+N)/2}s^{M-N}\exp(-|\alpha|^2)(\alpha^*+\partial/\partial\alpha^*)^M$$
$$\times(\alpha^*-\partial/\partial\alpha^*)^N\exp(|\alpha|^2)\langle\alpha|\rho|\alpha\rangle, \quad (A4)$$

and, therefore,

$$\mathrm{Tr}\{\rho\mathscr{g}^M\mathscr{f}^N\}$$
$$=i^N 2^{-1(M+N)/2}s^{M-N}\pi^{-1}$$
$$\times\int d^2\alpha\exp(-|\alpha|^2)(\alpha^*+\partial/\partial\alpha^*)^M(\alpha^*-\partial/\partial\alpha^*)^N$$
$$\times\exp(|\alpha|^2)\langle\alpha|\rho|\alpha\rangle. \quad (A5)$$

At this stage it might be convenient to introduce the ladder operators

$$a_k=\alpha_k+\tfrac{1}{2}(\partial/\partial\alpha_k),$$
$$a_k^\dagger=\alpha_k-\tfrac{1}{2}(\partial/\partial\alpha_k), \qquad (k=R,I) \quad (A6)$$

satisfying the usual commutation relations,

$$[a_k,a_l^\dagger]=\delta_{k,l},$$
$$[a_k,a_l]=[a_k^\dagger,a_l^\dagger]=0, \quad (A7)$$

and such that

$$a_k\Psi_m(2\alpha_k)=m^{1/2}\Psi_{m-1}(2\alpha_k), \quad (A8a)$$

$$a_k^\dagger\Psi_m(2\alpha_k)=(m+1)^{1/2}\Psi_{m+1}(2\alpha_k). \quad (A8b)$$

Now, substituting

$$\partial/\partial\alpha^*=\tfrac{1}{2}(\partial/\partial\alpha_R)+\tfrac{1}{2}i(\partial/\partial\alpha_I)$$

and (2.23) into (A5), we derive, using the notation (A6),

$$\mathrm{Tr}(\rho\mathscr{g}^M\mathscr{f}^N)=i^N 2^{-1(M+N)/2}s^{M-N}\sum_{m,n}s^{n-m}W_{m,n}\pi^{-1}\int d^2\alpha$$
$$\times\exp(-|\alpha|^2)(a_R-ia_I)^M(a_R^\dagger-ia_I)^N$$
$$\times\Psi_m(2\alpha_R)\Psi_n(2\alpha_I). \quad (A9)$$

Many terms in this sum vanish. Since the trace of an operator is an invariant, it is representation independent and, thus, the final result has to be independent of the scaling parameter s. It is clear, therefore, that in the double sum of (A9) survive only those terms for which

$$m-n=M-N. \quad (A10)$$

Hence, the operators a_R and $a_I^\dagger$ acting on different subspaces commute. Consequently, one may apply straightforwardly the binomial theorem expanding (A9) into

$$\begin{aligned}\mathrm{Tr}(\rho \mathcal{g}^M \mathcal{f}^N) &= 2^{-(M+N)/2}(2/\pi)^{1/2}\sum_{m,n} W_{m,n}\sum_{k=0}^{M}\sum_{j=0}^{N}(-i)^{M-k-j}\binom{M}{k}\binom{N}{j}\\ &\times\int d\alpha_R \Psi_0(2\alpha_R)a_R^k(a_R^\dagger)^j\Psi_m(2\alpha_R)\\ &\times\int d\alpha_I \Psi_0(2\alpha_I)(a_I^\dagger)^{M-k}a_I^{N-j}\Psi_n(2\alpha_I).\end{aligned} \tag{A11}$$

Consider the last integral. It vanishes in the following cases:

(i) $N-j>n$: Because the successive application of a_I to $\Psi_n(2\alpha_I)$ will by (A8a) inevitably render the factor $a_I\Psi_0(2\alpha_I)=0$.

(ii) $n>N-j-M+k$: Because the joint action of the ladder operators in the integrand transforms $\Psi_n(2\alpha_I)$ into $\Psi_{n-N+j+M-k}(2\alpha_I)$, which is orthogonal to the vacuum state for all n when this inequality is satisfied.

Thus, the contribution of the last integral is zero unless

$$N-j\le n\le(N-j)-(M-k). \tag{A12}$$

Since $N-j$ and $M-k$ are by definition nonnegative integers, (A12) is satisfied if and only if

$$k=M,\quad n=N-j, \tag{A13}$$

and, consequently, the integral is

$$\tfrac{1}{2}\int d(2\alpha_I)\Psi_0(2\alpha_I)a_I^{N-j}\Psi_{N-j}(2\alpha_I)=\tfrac{1}{2}[(N-j)!\,]^{1/2}. \tag{A14}$$

Furthermore, from (A13) and (A10) it follows that

$$m=M-j\ge 0, \tag{A15}$$

whence

$$\begin{aligned}\mathrm{Tr}(\rho \mathcal{g}^M \mathcal{f}^N) &= (2^{M+N}2\pi)^{-1/2}\sum_{j=0}^{[M,N]} i^j W_{M-j,N-j}\binom{N}{j}[(N-j)!\,]^{1/2}\\ &\times\int d\alpha_R \Psi_0(2\alpha_R)a_R^M(a_R^\dagger)^j\Psi_{M-j}(2\alpha_R).\end{aligned} \tag{A16}$$

The summation here is restricted by the smaller of the two positive integers M and N since, by (A15), j cannot exceed M.

In order to complete the proof of (A3) we have to calculate in (A16) the remaining integral. This is simple enough. Using (A8), we have

$$\begin{aligned}&\tfrac{1}{2}\int d(2\alpha_R)\Psi_0(2\alpha_R)a_R^M(a_R^\dagger)^j\Psi_{M-j}(2\alpha_R)\\ &\quad=\tfrac{1}{2}M!/[(M-j)!\,]^{1/2}.\end{aligned} \tag{A17}$$

Substituting this value into (A16) and simplifying the coefficients, one readily obtains (A3).

As a corollary it follows that

$$\begin{aligned}\mathrm{Tr}(\rho \mathcal{g}^M) &= (M!/2^M 8\pi)^{1/2}W_{M,0},\\ \mathrm{Tr}(\rho \mathcal{f}^N) &= (N!/2^N 8\pi)^{1/2}W_{0,N}.\end{aligned} \tag{A18}$$

Now, consider Eq. (A3) with $M=m-k$ and $N=n-k$ for all values of k such that $0\le k\le[m,n]$. This leads to a system of $[m,n]+1$ linear equations in $W_{m-j,n-j}$ with a triangular matrix which can be easily inverted. The solution of this system yields Eq. (2.25).

[1]For brevity we shall consider one-dimensional systems. The generalization to higher dimensions is trivial. The limits of integration are $(-\infty,\infty)$.
[2]P. A. M. Dirac, Rev. Mod. Phys. 17, 195 (1945).
[3]E. Prugovečki, J. Math. Phys. 7, 1054 (1966).
[4]E. Prugovečki, Can. J. Phys. 45, 2173 (1967).
[5]E. Wigner, Phys. Rev. 40, 749 (1932).
[6]J. E. Moyal, Proc. Cambridge Philos. Soc. 45, 99 (1949).
[7]L. Cohen, J. Math. Phys. 7, 781 (1966).
[8]L. Cohen, J. Math. Phys. 17, 1863 (1976).
[9]E. Prugovečki, Found. Phys. 17, 517 (1976).
[10]E. Prugovečki, Found. Phys. 4, 9 (1974).
[11]E. Prugovečki, Found. Phys. 5, 557 (1975).
[12]S. T. Ali and H. D. Doebner, J. Math. Phys. 17, 1105 (1976).
[13]S. T. Ali and E. Prugovečki, J. Math. Phys. 18, 219 (1977).
[14]E. Prugovečki, J. Math. Phys. 17, 1673 (1976).
[15]E. Schrodinger, Z. Physik 14, 664 (1926).
[16]R. J. Glauber, Phys. Rev. 130, 2529 (1963); 131, 2766 (1963).
[17]E. C. G. Sudarshan, Phys. Rev. Lett. 10, 227 (1963).
[18]C. L. Mehta and E. C. G. Sudarshan, Phys. Rev. 138, B274 (1965).
[19]P. Carruthers and M. M. Nieto, Phys. Rev. Lett. 14, 387 (1965).
[20]F. W. Cummings and J. R. Johnson, Phys. Rev. 151, 105 (1966).
[21]P. Carruthers and M. M. Nieto, Rev. Mod. Phys. 40, 411 (1968).
[22]M. Rasetti, Int. J. Theor. Phys. 14, 1 (1975).
[23]V. Bargmann, Comm. Pure Appl. Math. 14, 187 (1961).
[24]V. Bargmann, Comm. Pure Appl. Math. 20, 1 (1967).
[25]K. B. Wolf, J. Math. Phys. 15, 1295 (1974).
[26]K. B. Wolf, "The Heisenberg-Weyl Ring in Quantum Mechanics," in *Group Theory and Its Applications*, edited by E. M. Loebl (Academic, New York), Vol. III.
[27]ρ is a self-adjoint Hilbert—Schmidt operator belonging to the trace class.
[28]See, e.g., M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965).

The group structure of pseudo-Riemannian curvature spaces[a)]

Hans Tilgner

Mathematische Physik, Freie Universität Berlin, 1 Berlin 33, F. R. Germany
(Received 5 August 1977)

The linear conformal group **G** (pseudo-orthogonal automorphisms and dilatations) on a pseudo-orthogonal vector space induces an action in the space of pseudo-orthogonal curvature structures, which leaves Singer and Thorpe's direct decomposition in trivial, non-Einsteinian and Weyl curvature structures invariant. It is shown that the condition of a curvature to be homogeneous, reductive, or symmetric is **G**-invariant. A condition for a non-Weyl curvature to be symmetric is formulated explicitly. Nomizu's injection of the Jordan algebra of Lorentz-self-adjoint matrices is used to describe some **G**-orbits of non-Einsteinian curvatures. The Duffin–Kemmer–Petiau meson triple allows the construction of a cosmological model with trivial curvature.

1. INTRODUCTION

Given a pseudo-Riemannian manifold $(\mathcal{M}, \tau)$, the unique torsion free Levi-Civita connection ∇ induces the pseudo-Riemannian curvature structure R by

$$R(X, Y) = \nabla_X \nabla_Y - \nabla_Y \nabla_X - \nabla_{[X,Y]}$$

for vector fields X, Y on $\mathcal{M}$. R is skew symmetric in X and Y, fulfills the first Bianchi identity, (C.3) below, and defines a section $R(X, Y)$ in the pseudo-orthogonal Lie algebra bundle over $\mathcal{M}$. Singer and Thorpe[1] conversely defined a curvature structure on a pseudo-orthogonal vector space as a (1, 3) tensor with the three C axioms. The linear space spanned by these curvature structures can be taken as a typical fibre of a vector bundle over $\mathcal{M}$ in which the Levi-Civita curvature tensor field is a section. This *curvature bundle* is the associated bundle of the principal fibre bundle of pseudo-orthogonal frames over $\mathcal{M}$ with respect to the action of the pseudo-orthogonal group on the (1, 3) tensors on a pseudo-orthogonal vector space. The special types of pseudo-Riemannian manifolds (of constant curvature, Einsteinian, etc.) usually are defined in terms of this Levi-Civita section. Petrov was the first to give a classification of curvature structures on four-dimensional Einstein spaces.

In Sec. 2 we repeat the general theory of pseudo-orthogonal curvature structures in a formulation which was first given by Singer and Thorpe. Additional results can be found in Refs. 2—5 and in a little different notation in Refs. 6—9. We follow this latter notation and use especially Nomizus' construction of curvature structures by means of elements in the Jordan algebra [with respect to the anticommutator $\{A, B\} = \frac{1}{2}(AB + BA)$] $JA(\mathbb{V}, \langle\,,\rangle)$ of $\langle\,,\rangle$-symmetric endomorphisms on the pseudo-orthogonal vector space $(\mathbb{V}, \langle\,,\rangle)$. In Sec. 3 Nomizus' injection of this Jordan algebra into the curvature space is used to decompose this space with the help of the Weyl and the Einstein projector. The linear conformal group G, i.e., the group of pseudo-orthogonal transformations $O(\mathbb{V}, \langle\,,\rangle) = \{G \in Gl(\mathbb{V}, \mathbb{R}) / \langle Gx, Gy\rangle = \langle x, y\rangle$ for all $x, y \in \mathbb{V}\}$ and dilatations, operates on the curvature space with kernel Z_2 in such a way that this decomposition is invariant. In Sec. 4 curvature structures which are defined by decompositions of Lie algebras and a $\mathbb{G}$-invariant chain of specializations: homogeneous, reductive, Lie algebraic, symmetric, Lie algebra are discussed. It is shown that to be homogeneous, reductive, $\cdots$, is a common property of all structures in a $\mathbb{G}$-orbit. A necessary and sufficient condition for a curvature with vanishing Weyl component to be symmetric is derived. The action of $O(\mathbb{V}, \langle\,,\rangle)$ on the Jordan algebra $JA(\mathbb{V}, \langle\,,\rangle)$ by automorphisms is used to apply Helwig's results on idempotents in real simple Jordan algebras to the Lorentz case and to describe a class of orbits with vanishing Weyl component. The $\mathbb{G}$-orbit classification is quite simple in the one-dimensional space of trivial curvature structures which decomposes into the point zero and two orbits of positive, resp., negative trivial curvature. In Sec. 5 we use the well-known matrix representation of the Duffin—Kemmer—Petiau triple over $(\mathbb{V}, \langle\,,\rangle)$ to construct two examples of pseudo-Riemannian symmetric manifold with canonical curvature structure in these two orbits. It was shown in Ref. 10 that these are covered by suitable hyperboloids, in the Lorentz case twice by the de Sitter resp. anti-de Sitter spaces.

In the following $(\mathbb{V}, \langle\,,\rangle)$ will be a real pseudo-orthogonal vector space of dimension $n > 2$. The Lie algebra of $O(\mathbb{V}, \langle\,,\rangle)$ is $so(\mathbb{V}, \langle\,,\rangle)$. In the Lorentz case we write $O(1,3;\mathbb{R})$ for the Lorentz group and $SO_0(1,3;\mathbb{R})$ for its connectivity component of the identity. The *linear conformal group* $\mathbb{G}$ is the group of linear transformations in $\mathbb{V}$ subject to $\langle Gx, Gy\rangle = \lambda_G^2 \langle x, y\rangle$, $0 \neq \lambda_G \in \mathbb{R}$, its Lie algebra $L\mathbb{G}$ is given by all Q subject to $\langle Qx, y\rangle + \langle x, Qy\rangle = 2\mu_Q \langle x, y\rangle$, $\mu_Q \in \mathbb{R}$, for all x, y in $\mathbb{V}$. A Lie bracket will be denoted by $[\,,\,]$, the commutator in an associative algebra by $[\,,\,]_-$.

The whole discussion of curvature structures is local, i.e., describes tensors on a tangent Minkowski space. Since the linear conformal group $\mathbb{G}$ acts on the space of curvature structures as a transformation group, the globalization is immediate: The principal $\mathbb{G}$-bundle over a pseudo-Riemannian manifold $\mathbb{M}$ allows the construction of the associated curvature bundle with respect to this $\mathbb{G}$ action.

It remains to study the relation of linear conformal transformations on curvature spaces to conformal transformations of the base manifolds.

[a)]Dedicated to Günther Ludwig on the occasion of his 60th birthday.

 0022-2488/78/1905-1118$1.00

2. CURVATURE SPACES

A pseudo-orthogonal curvature structure on the pseudo-orthogonal vector space $(\mathbb{V}, \langle\,,\,\rangle)$ is a bilinear mapping R: $\mathbb{V}\times\mathbb{V} \to \text{end}\,\mathbb{V}$, subject to

(C.1) $R(y,x) = -R(x,y)$ (skew symmetry),

(C.2) $\langle R(x,y)z, w\rangle + \langle z, R(x,y)w\rangle = 0$,

i.e., $R(x,y) \in \text{so}(\mathbb{V}, \langle\,,\,\rangle)$.

(C.3) $R(x,y)z + R(z,x)y + R(y,z)x = 0$

(first Bianchi identity)

for x,y,z,w in $\mathbb{V}$. From these axioms one has

(C.4) $\langle R(x,y)z, w\rangle = \langle R(z,w)x, y\rangle$.

The linear space of these curvature structures will be denoted by $\text{curv}(\mathbb{V}, \langle\,,\,\rangle)$. An example for a curvature structure can be constructed in terms of a semisimple Lie algebra structure $[\,,\,]$ on $\mathbb{V}$: Taking for $\langle\,,\,\rangle$ the (invariant) Killing form, $R(x,y)z = \text{ad}([x,y])z = [[x,y],z]$ defines an element R in $\text{curv}(\mathbb{V}, \langle\,,\,\rangle)$.

The *curvature space* $\text{curv}(\mathbb{V}, \langle\,,\,\rangle)$ always contains the *trivial* curvature structure R_0 given by

$$R_0(x,y)z = \langle y,z\rangle x - \langle x,z\rangle y.$$

More generally, given A, B in $JA(\mathbb{V}, \langle\,,\,\rangle)$,

$$R_0^{A,B}(x,y) = \tfrac{1}{2}\{R_0(Ax, By) + R_0(Bx, Ay)\} = R_0^{B,A}(x,y)$$

defines a new curvature structure $R_0^{A,B}$. The *Ricci form* of $R \in \text{curv}(\mathbb{V}, \langle\,,\,\rangle)$

$$\rho_R(x,y) = \text{tr}(z \mapsto R(z,x)y)$$

is a symmetric bilinear form on $\mathbb{V}$. The *Ricci transformation* L_R of R is defined by

$$\langle L_R x, y\rangle = \rho_R(x,y).$$

It is ρ_R and $\langle\,,\,\rangle$ self-adjoint, i.e., in $JA(\mathbb{V}, \langle\,,\,\rangle)$. The *curvature scalar* $\text{Sc}(R)$ of R is the trace of L_R. Since the trace of $z \mapsto \langle x,z\rangle y$ is $\langle x,y\rangle$ one has

$$L_{R_0^{A,B}} = -\tfrac{1}{2}(AB + BA - A\,\text{tr}B - B\,\text{tr}A),\quad L_{R_0} = (n-1)\,\text{id}_n,$$
$$\text{Sc}(R_0^{A,B}) = \text{tr}AB - \text{tr}A\,\text{tr}B,\quad \text{Sc}(R_0) = n(n-1), \tag{1}$$

and $\rho_{R_0}(x,y) = (n-1)\langle x,y\rangle$. The *Ricci curvature structure* R^{ic} of $R \in \text{curv}(\mathbb{V}, \langle\,,\,\rangle)$ is defined by $R_0^{L_R, \text{id}_n}$, i.e., by

$$R^{ic}(x,y)z = \tfrac{1}{2}\{\rho_R(y,z)x - \rho_R(x,z)y + \langle y,z\rangle L_R x - \langle x,z\rangle L_R y\}. \tag{2}$$

Writing $L_{R_0^{A,B}} = C$, one has

$$(R_0^{A,B})^{ic} = R_0^{C,\text{id}_n},\quad R_0^{ic} = (n-1)R_0, \tag{3}$$

$$\rho_{R^{ic}}(x,y) = \frac{n-2}{2}\rho_R(x,y) + \frac{\text{Sc}(R)}{2}\langle x,y\rangle,$$
$$L_{R^{ic}} = \frac{n-2}{2}L_R + \frac{\text{Sc}(R)}{2}\text{id}_n,\quad \text{Sc}(R^{ic}) = (n-1)\text{Sc}(R). \tag{4}$$

The *sectional curvature* of R is defined by $\langle R(x,y)y, x\rangle / \langle R_0(x,y)y, x\rangle$. It is easy to see that Ricciform and Riccitransformation are linear in their indices, that the *Ricci map* $\text{Ri}: R \mapsto R^{ic}$ is an endomorphism and Sc is a linear form on the curvature space. In the following we use the linear map

$$\Omega: A \to R_0^{A,\text{id}_n},\quad \Omega: JA(\mathbb{V}, \langle\,,\,\rangle) \to \text{curv}(\mathbb{V}, \langle\,,\,\rangle).$$

It is injective and for all $R \in \text{curv}(\mathbb{V}, \langle\,,\,\rangle)$

$$\Omega(L_R) = \text{Ri}(R). \tag{5}$$

The linear conformal group $\mathbb{G}$, resp. its Lie algebra $L\mathbb{G}$, acts on $\text{curv}(\mathbb{V}, \langle\,,\,\rangle)$ by

$$(G\cdot R)(x,y) = GR(G^{-1}x, G^{-1}y)G^{-1},\quad \text{resp.}$$
$$(Q\cdot R)(x,y) = [Q, R(x,y)]_- - R(Qx,y) - R(x,Qy).$$

These actions make the curvature space a $\mathbb{G}$-, resp., a $L\mathbb{G}$-module. The kernel of this *curvature representation* of $\mathbb{G}$ is $\mathbb{Z}_2$, i.e., the group acting on the curvature space actually is $\mathbb{G}/\mathbb{Z}_2$; the kernel of the Lie algebra representation is trivial. It is easy to verify that

$$G\cdot(Q\cdot R) = (G\cdot Q)\cdot(G\cdot R), \tag{6}$$

which shows that if Q annihilates R, then $G\cdot Q$ annihilates $G\cdot R$. Given A in $JA(\mathbb{V}, \langle\,,\,\rangle)$, then $G\cdot A = GAG^{-1}$, resp., $Q\cdot A = [Q,A]_-$, again are in $JA(\mathbb{V}, \langle\,,\,\rangle)$, and

$$G\cdot R_0^{A,B} = \frac{1}{\lambda_G^2}R_0^{G\circ A, G\circ B} \tag{7}$$
$$\text{resp. } Q\cdot R_0^{A,B} = R_0^{Q\circ A,B} + R_0^{A,Q\circ B} - 2\mu_Q R_0^{A,B},$$

$$G\cdot\Omega(A) = \frac{1}{\lambda_G^2}\Omega(G\cdot A), \tag{8}$$
$$\text{resp. } Q\cdot\Omega(A) = \Omega(Q\cdot A - 2\mu_Q A),$$

$$G\cdot R_0 = \frac{1}{\lambda_G^2}R_0, \tag{9}$$
$$\text{resp. } Q\cdot R_0 = -2\mu_Q R_0.$$

From $\rho_{G\cdot R}(x,y) = \rho_R(G^{-1}x, G^{-1}y)$, resp. $\rho_{Q\circ R}(x,y) = -\rho_R(Qx,y) - \rho_R(x,Qy)$, one shows

$$G\cdot L_R = \lambda_G^2 L_{G\cdot R},\quad \text{resp. } Q\cdot L_R = L_{Q\circ R} + 2\mu_Q L_R, \tag{10}$$
$$\text{Sc}(G\cdot L_R) = \lambda_G^{-2}\text{Sc}(R),\quad \text{resp. } \text{Sc}(Q\cdot R) = -2\mu_Q\text{Sc}(R). \tag{11}$$

Hence Sc is a $\mathbb{G}$- resp. $L\mathbb{G}$-equivariant, linear form on the curvature space which is even invariant resp. annihilated, under the pseudo-orthogonal subgroup, resp. subalgebra. Using (5), (8), and (10) one shows that the Ricci map commutes with the multiplication, i.e.,

$$G\cdot\text{Ri}(R) = \text{Ri}(G\cdot R),\quad \text{resp., } Q\cdot\text{Ri}(R) = \text{Ri}(Q\cdot R). \tag{12}$$

3. INVARIANT DECOMPOSITION OF CURVATURE SPACES

Given R in $\text{curv}(\mathbb{V}, \langle\,,\,\rangle)$, its *Weyl curvature structure* is

$$\mathcal{W}R = R - \frac{2}{n-2}R^{ic} + \frac{\text{Sc}(R)}{(n-1)(n-2)}R_0 \in \text{curv}(\mathbb{V}, <\,,\,>)$$

and its *Einstein curvature* structure is

$$\mathcal{E}R = R - \frac{\text{Sc}(R)}{n(n-1)}R_0 \in \text{curv}(\mathbb{V}, \langle\,,\,\rangle).$$

Lemma: (a) $\rho_{\mathcal{W}R} = 0$, $L_{\mathcal{W}R} = 0$, $\text{Sc}(\mathcal{W}R) = 0$,

$$\rho_{\mathcal{E}R}(x,y) = \rho_R(x,y) - \frac{\text{Sc}(R)}{n}\langle x,y\rangle,$$

$$L_{\mathcal{E}R} = L_R - \frac{\mathrm{Sc}(R)}{(n-1)n} L_{R_0}, \quad \mathrm{Sc}(\mathcal{E}R) = 0,$$

(b) $\mathcal{W}R^{A,\mathrm{id}} = 0$, i.e., $\mathcal{W} \circ \Omega = 0$ especially $\mathcal{W}R_0 = 0$,

$\mathcal{E}R^{A,\mathrm{id}} = \mathcal{E}\Omega(A) = \Omega(A - (\mathrm{tr}A/n)\mathrm{id}_n)$ especially $\mathcal{E}R_0 = 0$, (13)

(c) $\mathcal{W}$, $\mathcal{E}$, $\mathcal{E} \circ \mathcal{W}$, and $\mathcal{E} - \mathcal{W}$ are projectors on $\mathrm{curv}(\mathbb{V}, \langle\,,\rangle)$ with

$\mathcal{E} \circ \mathcal{W} = \mathcal{W} = \mathcal{W} \circ \mathcal{E}$,

(d) $\mathcal{W} \circ \mathrm{Ri} = \mathrm{Ri} \circ \mathcal{W} = 0$ and $\mathcal{E} \circ \mathrm{Ri} = \mathrm{Ri} \circ \mathcal{E}$.

The proof of (a), (b), and (c) is technical, for (d) consider the complementary projector $\mathcal{W}^\perp$ of $\mathcal{W}$ on to $\ker\mathcal{W}$, given by

$$\mathcal{W}^\perp R = (\mathrm{id} - \mathcal{W})R = \frac{1}{n-2}\left(2R^{ic} - \frac{\mathrm{Sc}(R)}{n-1} R_0\right).$$

Using (a) we get

$$\mathcal{W}^\perp R = \mathcal{W}^\perp \mathcal{W}^\perp R = \frac{1}{n-2}\left(2[(\mathrm{id} - \mathcal{W})R]^{ic} - \frac{\mathrm{Sc}(R - \mathcal{W}R)}{n-1} R_0\right)$$

$$= \mathcal{W}^\perp - \frac{1}{n-2}(\mathcal{W}R)^{ic},$$

i.e., $(\mathcal{W}R)^{ic} = 0$; the rest is straightforward again.

Lemma: (a) $\mathcal{W}R = 0$ implies

$$\Omega\left(2L_R - \frac{\mathrm{Sc}(R)}{n-1}\mathrm{id}_n\right) = (n-2)R,$$

hence $\ker\mathcal{W} \subset \mathrm{Im}\Omega$,

(b) $\Omega \circ L \circ \mathcal{W}^\perp = \mathrm{Ri}$,

(c) $L \circ \Omega$ is injective, hence bijective, (14)

(d) Ω is bijective onto $\ker\mathcal{W}$ and the restriction of L to $\mathrm{Ker}\mathcal{W}$ is bijective.

The proof is straightforward. Summarizing we have the commutative diagram of short exact sequences[7]

$$\begin{array}{ccccc}
\ker\mathcal{W} & \underset{\mathcal{W}^\perp}{\overset{\longrightarrow}{\longleftarrow}} & \mathrm{curv}(\mathbb{V}, \langle\,,\rangle) & \underset{}{\overset{\mathcal{W}}{\rightleftarrows}} & \mathrm{Im}\mathcal{W} \\
{\scriptstyle L^{-1}}\uparrow\downarrow{\scriptstyle L} & & \downarrow{\scriptstyle \mathrm{Ri}} & & \downarrow{\scriptstyle 0} \\
JA(\mathbb{V}, \langle\,,\rangle) & \underset{\Omega^1}{\overset{\Omega}{\rightleftarrows}} & \mathrm{curv}(\mathbb{V}, \langle\,,\rangle) & \overset{\mathcal{W}}{\longrightarrow} & \mathrm{Im}\mathcal{W}
\end{array} \tag{15}$$

which shows how L and Ri are related to each other via Ω and $\mathcal{W}^\perp$. Obviously

$$L^{-1}: A \mapsto \frac{1}{n-2}(2R_0^{A,\mathrm{id}} - \frac{\mathrm{tr}A}{n-1} R_0),$$

$$L^{-1}: JA(\mathbb{V}, \langle\,,\rangle) \to \ker\mathcal{W},$$

$$\Omega^{-1}: R \mapsto \frac{1}{n-2}\left(2L_R - \frac{\mathrm{Sc}(R)}{n-1}\mathrm{id}_n\right),$$

$$\Omega^{-1}: \ker\mathcal{W} \to JA(\mathbb{V}, \langle\,,\rangle),$$

with $\Omega \circ \Omega^{-1} = \mathcal{W}^\perp$ and $L^{-1} \circ L = \mathcal{W}^\perp$. Ω^{-1} is called the *deviation* map in Ref. 7. $JA(\mathbb{V}, \langle\,,\rangle)$ can be decomposed directly into the space of multiples of id_n and the $\frac{1}{2}n(n+1) - 1$-dimensional space of traceless endomorphisms. Hence $\ker\mathcal{W}$ can be decomposed directly into $\mathbb{R}R_0$ and the Ω-image of these traceless endomorphisms which is exactly $\mathrm{Im}(\mathcal{E} - \mathcal{W})$. From $\mathrm{id} = \mathcal{E}^\perp \oplus \mathcal{E} - \mathcal{W} \oplus \mathcal{W}$ we get

$$\mathrm{curv}(\mathbb{V}, \langle\,,\rangle) = \mathbb{R}R_0 \oplus \mathrm{Im}(\mathcal{E} - \mathcal{W}) \oplus \mathrm{Im}\mathcal{W}. \tag{16}$$

Moreover, like Ri, the projectors $\mathcal{E}$ and $\mathcal{W}$ commute with the action of $\mathbb{G}$, resp. $L\mathbb{G}$, which shows that (15) is a commutative diagram of $\mathbb{G}$- resp. $L\mathbb{G}$-modules and (16) is a $\mathbb{G}$- resp. $L\mathbb{G}$-invariant decomposition. The meaning of decomposition (16) was described in Ref. 1 by the following theorem.

Theorem: (a) R in $\mathbb{R}R_0$ <==> the sectional curvature of R is constant,

(b) R in $\mathrm{Im}\mathcal{W}$ <==> the Ricci form of R vanishes,

(c) R in $\mathbb{R}R_0 \oplus \mathrm{Im}\mathcal{W}$ <==> L_R is a multiple of id_n (17)
(*Einsteinian* case),

(d) R in $\mathrm{Im}(\mathcal{E} - \mathcal{W}) \oplus \mathrm{Im}\mathcal{W}$ <==> the scalar curvature of R vanishes.

Since (16) is $\mathbb{G}$-invariant the classification of $\mathbb{G}$-orbits reduces to that in the three subspaces.

Examples of Weyl curvature structures can be constructed in the four-dimensional Lorentz case as follows: Given α in $\mathbb{R}$, the traceless nilpotent matrices

$$\begin{bmatrix} \alpha & \pm\alpha & 0 & 0 \\ \mp\alpha & -\alpha & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{18}$$

are in $JA(\mathbb{V}, \langle\,,\rangle)$ and $R_0^{A,A} = \mathcal{W}R_0^{A,A} \neq 0$. Since $R_0^{A,B}$ is linear in A and B such a curvature structure can be decomposed according to

$$R_0^{A,B} = \tfrac{1}{2}(R_0^{A+B,A+B} - R_0^{A,A} - R_0^{B,B}),$$

i.e., into "pure" elements $R_0^{C,C}$, which from (7) lie in "pure" $\mathbb{G}$-orbits.

4. $\mathbb{G}$-ORBITS WHICH ARE DEFINED BY LIE ALGEBRAS

Let $\mathbb{L}$ be a real Lie algebra, $\mathbb{L}_0$ a subalgebra in $\mathbb{L}$, and $\mathbb{V}$ a complementary subapace to $\mathbb{L}_0$ in $\mathbb{L}$ with a bilinear form $\langle\,,\rangle$ as above. Writing the typical element of $\mathbb{L}$ as $K \oplus x$ with K in $\mathbb{L}_0$ and x in $\mathbb{V}$, one has $[K \oplus x, L \oplus y] = ([K,L] + [x,y]_0 + [x,L]_0 - [y,K]_0) \oplus ([K,y]_\mathbb{V} - [L,x]_\mathbb{V} + [x,y]_\mathbb{V})$. Introducing bilinear maps

$$C(x,y) = [x,y]_0, \quad C: \mathbb{V} \times \mathbb{V} \to \mathbb{L}_0,$$

$$t(x,y) = [x,y], \quad t: \mathbb{V} \times \mathbb{V} \to \mathbb{V},$$

a linear map Θ: $\mathbb{V} \to \mathrm{end}\,\mathbb{L}_0$ by

$$\Theta_x L = [x,L]_0, \quad \Theta: x \mapsto \Theta_x,$$

and denoting the adjoint action of $\mathbb{L}_0$ on $\mathbb{V}$ by $[L,x]_\mathbb{V} = Lx$, the Lie bracket in $\mathbb{L}$ can be written

$$[K \oplus x, L \oplus y] = ([K,L] + C(x,y) + \Theta_x L - \Theta_y K) \oplus (Ky - Lx + t(x,y)). \tag{19}$$

The Lie algebra axioms in $\mathbb{L}$ then imply

(CT.1) $C(x,y) = -C(y,x)$ and $t(x,y) = -t(y,x)$ (skew symmetry),

(CT.2) $C(x,y)z + C(y,z)x + C(z,x)y + t(t(x,y),z)$
$+ t(t(y,z),x) + t(t(z,x),y)$
(Bianchi—Jacobi identity),

(CT.3) $[K, C(y,z)] = C(Kx,y) + C(x,Ky)$
$+ \Theta_{t(y,z)}K - [\Theta_y, \Theta_z]_- K,$

(CT.4) $Kt(y,z) = t(Kx,y) + t(x,Ky) + (\Theta_y K)z - (\Theta_z K)y,$

(CT.5) $[K,L]z = KLz - LKz,$

(CT.6) $C(t(x,y),z) + C(t(y,z),x) + C(t(z,x),y)$
$= \Theta_x C(y,z) + \Theta_y C(z,x) + \Theta_z C(x,y),$

(CT.7) $\Theta_z[K,L] = [\Theta_z K, L] + [K, \Theta_z L] + \Theta_{Kz}L - \Theta_{Lz}K.$

We call such a decomposition *homogeneous*, and *pseudo-orthogonal homogeneous* if in addition $\langle\ ,\ \rangle$ is $C(x,y)$- and $t(x,\cdot)$-invariant, i.e., if

(POH.1) $\langle C(x,y)z, w\rangle + \langle z, C(x,y)w\rangle = 0,$

(POH.2) $\langle t(x,z), w\rangle + \langle z, t(x,w)\rangle = 0.$

(CT.3) shows that Θ is a morphism of the algebra $(\mathbb{V}, t)$ into the Lie algebra end $\mathbb{L}_0$ if and only if $K \cdot C = 0$; (CT.4) shows that if Θ vanishes, $\mathbb{L}_0$ acts by derivations on the algebra $(\mathbb{V}, t)$. On $\mathbb{V}$ we have

$$[[x,y],z] = [C(t(x,y),z) - \Theta_z C(x,y)] \oplus [C(x,y)z + t(t(x,y),z)]. \tag{20}$$

Introducing a (trilinear) triple structure on $\mathbb{V}$ by

$$[x,y,z] = C(x,y)z + t(t(x,y),z) = \mathrm{ad}(x,y)z \tag{21}$$

from (CT.1) and (CT.2), we get the first two axioms of a Lie triple[11,12]

(LT.1) $[y,x,z] = -[x,y,z],$

(LT.2) $[x,y,z] + [y,z,x] + [z,x,y] = 0.$

If in addition to the CT-axioms the POH-axioms hold, then $\mathrm{ad}(x,y)$ obviously is an element of $\mathrm{curv}(\mathbb{V}, \langle\ ,\ \rangle)$. We call such a curvature structure *homogeneous*.

Conversely, given bilinear mappings $C: \mathbb{V}\times\mathbb{V} \to \mathrm{end}\mathbb{V}$, $t: \mathbb{V}\times\mathbb{V} \to \mathbb{V}$, and linear mappings Θ_x: end $\mathbb{V} \to$ end $\mathbb{V}$ such that the seven CT-axioms hold for the commutator $[\ ,\ ]_-$ in end $\mathbb{V}$, (19) defines a homogeneous decomposition $\mathbb{L} = \mathbb{L}_0 \oplus \mathbb{V}$, where now $\mathbb{L}_0$ is the Lie subalgebra of end $\mathbb{V}$ generated by the $C(x,y)$ and the $\Theta_{z_1}\cdots\Theta_{z_i}C(x,y)$. The Lie algebra $\mathbb{L}$ with the Lie bracket (19) is called the *standard embedding* algebra of (C, t, Θ).

To give an action of $\mathbb{G}$ on such homogeneous structures (C,t,Θ), define

$$G\cdot t(x,y) = Gt(G^{-1}x, G^{-1}y) \text{ and } (G\cdot\Theta)_z K = G\Theta_{G^{-1}z}(G^{-1}\cdot K)G^{-1} \tag{22}$$

and $G \cdot C$ in the same way as $G \cdot R$ in Sec. 2. Then $(G\cdot C, G\cdot t, G\cdot\Theta)$ fulfills the seven CT-axioms and the two POH-axioms if (C,t,Θ) does. If $\mathrm{ad}(x,y)$ is given as in (21), then $G\cdot\mathrm{ad}$ is given in the same way by $G\cdot C$ and $G\cdot t$. Hence homogeneous curvature structures lie in *homogeneous* $\mathbb{G}$-orbits.

More general let $(\mathbb{V}', \langle\ ,\ \rangle')$ be a second pseudo-orthogonal vector space and $G: \mathbb{V}' \to \mathbb{V}$ a *linear conformal* mapping, i.e., linear with $\langle Gx', Gy'\rangle = \lambda_G^2 \langle x', y'\rangle'$ and $\lambda_G \neq 0$. Given a homogeneous structure (C', t', Θ') on $\mathbb{V}'$ we get a homogeneous structure $(G\circ C', G\cdot t', G\cdot\Theta')$ on $\mathbb{V}$, the latter being defined as in (22). This makes curv a covariant functor. Any linear conformal mapping G defines an isomorphism $G^*: \mathbb{L}' \to \mathbb{L}$ of the standard embedding algebras by $G^*: K \oplus x \to GKG^{-1} \oplus Gx$ which commutes with the *standard involutions* $K + x \mapsto K - x$. The standard involutions however are automorphisms of the standard embedding algebras only if t and Θ vanish (symmetric case).

The pseudo-orthogonal *reductive* structures (C,t) on $\mathbb{V}$ are characterized by vanishing Θ, see Refs. 13, 6, and 14. The (torsion-free) *pseudo-Riemannian curvature structure* C^∇ on $\mathbb{V}$ was given in Ref. 13 by the following construction: For α in $\mathbb{R}$ define $\nabla_x y = \alpha t(x,y)$ and

$$t^\nabla(x,y) = \nabla_x y - \nabla_y x - t(x,y) = (2\alpha - 1)t(x,y),$$

$$C^\nabla(x,y)z = \nabla_x\nabla_y z - \nabla_y\nabla_x z - \nabla_{t(x,y)}z - C(x,y)z.$$

Then (CT.2) and (CT.6) are valid for (C^∇, t^∇) only if $\alpha = \frac{1}{2}$, and in this case

$$C^\nabla(x,y)z = \tfrac{1}{4}t(x,t(y,z)) - \tfrac{1}{4}t(y,t(x,z)) - \tfrac{1}{2}t(t(x,y),z) - C(x,y)z \tag{23}$$

is a solution of the CT-axioms for $t^\nabla = 0$. In the reductive case again the remaining six CT-axioms and the two POH-axioms hold for any $\mathbb{G}$-transform of (C,t) and $G\cdot(C^\nabla) = (G\cdot C)^\nabla$, hence reductive curvature structures lie in *reductive* $\mathbb{G}$-orbits in $\mathrm{curv}(\mathbb{V}, \langle\ ,\ \rangle)$. It is straightforward to verify that

$$(G\cdot C(z,w))\cdot(G\cdot C) = 0 \text{ and } (G\cdot C(z,w))\cdot(G\cdot t) = 0 \tag{24}$$

for G in $\mathbb{G}$.

A special case of a reductive structure is a *Lie-algebraic* structure where t is a Lie algebra on $\mathbb{V}$. Here (CT.2) splits into the (first) Bianchi identity for C, which implies $C \in \mathrm{curv}(\mathbb{V}, \langle\ ,\ \rangle)$, and the Jacobi identity for t. (POH.2) makes $\langle\ ,\ \rangle$ an invariant form of the Lie algebra. Again if (C,t) is Lie-algebraic, then $(G\cdot C, G\cdot t)$ is as well; hence there are Lie-algebraic orbits among the reductive ones. In general $(\mathbb{V}, t)$ is not a subalgebra or homomorphic image of its standard embedding. An elementary example is given by a semisimple Lie algebra $(\mathbb{V}, [\ ,\ ])$ with Killing form $\langle\ ,\ \rangle$ and $C(x,y) = \mathrm{ad}([x,y])$, $t(x,y) = [x,y]$. Indeed the (split) equations (CT.2), (CT.4), and (CT.6) follow from Jacobi's identity, whereas (CT.3) holds since Lie algebras are Lie triples (see below): the two POH-axioms follow from the invariance of the Killing form. The Ricci form of C is exactly the Killing form.

A special case of a Lie-algebraic structure is a (*pseudo-orthogonal*) *symmetric* (or *Lie triple*-) structure with vanishing t. Here $C(x,y) = \mathrm{ad}(x,y)$ and the associated triple (21) also fulfills the third axiom,

(LT.3) $[u,v,[x,y,z]] - [x,y,[u,v,z]] = [[u,v,x],y,z] + [x,[u,v,y],z]$

of a Lie triple, which in fact is exactly (CT.3) for $K = C(u,v)$. Equivalently this can be written

$$C(u,v)\cdot C = 0 \text{ for } u,v \text{ in } \mathbb{V}. \tag{25}$$

Conversely a Lie triple, i.e., a trilinear composition on $\mathbb{V}$ subject to the three LT-axioms, defines a (symmetric) curvature structure ad if and only if $\mathrm{ad}(x,y)$ is in the pseudo-orthogonal Lie algebra on $(\mathbb{V},\langle\,,\rangle)$ for any x,y in $\mathbb{V}$. In the Lie-algebraic case, C [but not ad in (21)] is already a symmetric curvature structure. The first equation (22) shows that symmetric curvature structures lie in *symmetric* $\mathbb{G}$-orbits in $\mathrm{curv}(\mathbb{V},\langle\,,\rangle)$. Symmetric orbits in $\ker W = \mathrm{Im}\Omega$ can be characterized easily: From (8) and (25) the symmetry of $\Omega(A)=R_0^{A,\mathrm{id}}$ is equivalent to

$$R_0^{A,\mathrm{id}}(u,v)\cdot\Omega(A)=\Omega(R_0^{A,\mathrm{id}}(u,v)\cdot A)=0 \quad \text{for all } u,v, \tag{26}$$

since Ω is injective, hence to $R_0^{A,\mathrm{id}}(u,v)\cdot A = [R_0^{A,\mathrm{id}}(u,v),A]_-=0$ for all u, v, and this is equivalent to $A^2=\tau\mathrm{id}_n$ with τ in $\mathbb{R}$. Especially the (trivial) curvature structures in $\mathbb{R}R_0$, for which $\alpha R_0=\Omega(\alpha\mathrm{id}_n)$, are symmetric. If $\langle\,,\rangle$ is positive definite, the Jordan algebra $J\mathbb{A}(\mathbb{V},\langle\,,\rangle)$ is formal real (or compact) and has no nilpotent elements; whence in this case there are only symmetric $\mathbb{G}$-orbits in $\mathrm{Im}\Omega$ for invertible matrices A. If $\langle\,,\rangle$ is indefinite there are nilpotent (necessarily symmetric) orbits, for instance those through the matrices (18). Clearly besides the orbit $\{0\}$ there are only the two symmetric $\mathbb{G}$-orbits in $\mathbb{R}R_0$ for positive, resp., negative α.

In the same way as in the Lie-algebraic case we get a special type of now symmetric curvature structures by again taking $C(x,y)=\mathrm{ad}([x,y])$ on a Lie algebra $(\mathbb{V},[\,,\,])$ (the first example in Sec. 2, for instance). The axiom (LT.3) is valid for the double Lie bracket $[[\,,\,],]$ as Lie triple composition. They lie in special orbits among the symmetric ones. The specialization $C=0$ in the general Lie-algebraic case gives a third possibility to get a solution of the CT-axioms for a Lie algebra. However these three cases have different standard embedding algebras. For instance, if $C=0$ we have $\mathbb{L}_0=\{0\}$ and the standard embedding becomes $(\mathbb{V},[\,,\,])$ itself.

5. TWO MANIFOLDS CONSTRUCTED IN TERMS OF THE DUFFIN-KEMMER-PETIAU MATRICES

In the following we give a matrix description of two pseudo-Riemannian symmetric spaces whose canonical pseudo-Riemannian curvature structures are R_0, resp. $-R_0$, using the *meson* triples on $(\mathbb{V},\langle\,,\rangle)$, see Ref. 11, more precisely their representation by Duffin–Kemmer–Petiau matrices.

Let $e_1,\ldots,e_n$ be a basis in $\mathbb{V}$ in which x has the components ξ^i, y the components $\eta^i,\ldots$, and in which the matrix of $\langle\,,\rangle$ is I. Consider the real traceless $(n+1)$-square matrices

$$K^{\mp}(x)=\begin{bmatrix}0 & \xi^t I\\ \mp\xi & 0\end{bmatrix}.$$

The n *Duffin–Kemmer–Petiau* matrices are $K^{\mp}(e_i)=\beta_i$. To check their familiar algebraical identities note that

$$K^{\mp}(x)K^{\mp}(y)=\begin{bmatrix}\mp\langle x,y\rangle & 0\\ 0 & \mp\xi\otimes\eta^t I\end{bmatrix}$$

implies $K^{\mp}(x)K^{\mp}(y)K^{\mp}(x)=\mp\langle x,y\rangle K^{\mp}(x)$, from which

$$K^{\mp}(x)K^{\mp}(y)K^{\mp}(z)+K^{\mp}(z)K^{\mp}(y)K^{\mp}(x) = \mp\langle x,y\rangle K^{\mp}(z)\mp\langle z,y\rangle K^{\mp}(x), \tag{27}$$

which is a basis free formulation of the well-known Duffin–Kemmer–Petiau relations $\beta_i\beta_k\beta_l+\beta_l\beta_k\beta_i = I_{ik}\beta_l+I_{kl}\beta_i$ and a representation of the meson Jordan triple. It is easy to verify that the spaces $K^{\mp}(\mathbb{V})$ are closed under double commutation,

$$[[K^{\mp}(x),K^{\mp}(y)]_-,K^{\mp}(z)]_- = \mp\langle y,z\rangle K^{\mp}(x) \pm\langle z,x\rangle K^{\mp}(y), \tag{28}$$

i.e., are Lie triples with curvature structure $\mp R_0$. Using that the trace of $\xi\otimes\eta^t I$ is $\langle x,y\rangle$ we get $\mp\frac{1}{2}\mathrm{tr}K^{\mp}(x)K^{\mp}(y)=\langle x,y\rangle$. Hence the mappings $x\mapsto K^{\mp}(x)$ are isomorphisms of pseudo-orthogonal vector spaces. The pseudo-orthogonal group of $(\mathbb{V},\langle\,,\rangle)$ acts on $K^{\mp}(\mathbb{V})$ by inner automorphisms with $\tilde{G}=\mathrm{diag}(1,G)$, i.e.,

$$J_G: K^{\mp}(x)\mapsto \tilde{G}K^{\mp}(x)\tilde{G}^{-1}=K^{\mp}(Gx),$$

and this establishes a (global) isomorphism $J: G\mapsto J_G$ of the pseudo-orthogonal group of $(\mathbb{V},\langle\,,\rangle)$ onto that of $(K^{\mp}(\mathbb{V}),\ \mathrm{trace})$. This represents the fact that spin zero representations of pseudo-orthogonal groups are ordinary, not only ray representations. Writing $c(x)=\cos\mu\sqrt{\pm\langle x,x\rangle}$ and $s(x)$ for the corresponding sinus series we get

$$\exp(\mu K^{\mp}(x))=\begin{bmatrix}c(x) & \dfrac{s(x)}{\sqrt{\pm\langle x,x\rangle}}\,\xi^t I\\[2ex] \mp\dfrac{s(x)}{\sqrt{\pm\langle x,x\rangle}}\,\xi & \mathrm{id}_n+\dfrac{-1+c(x)}{\langle x,x\rangle}\,\xi\otimes\xi^t I\end{bmatrix}. \tag{29}$$

These exponentials generate two closed n-dimensional symmetric submanifolds $\mathbb{M}^{\pm}$ of $\mathrm{Sl}(n+1,\mathbb{R})$ which are algebraically closed with respect to the symmetric composition $G\square H=GHG^{-1}$ in $\mathrm{Sl}(n+1,\mathbb{R})$. Here "generate" means "generate as a symmetric space," i.e., take all finite $\square$-products of these exponentials.[12] It was shown in Ref. 10 by applying the Loos formulation of symmetric spaces to the de Sitter, resp., the anti-de Sitter, hyperboloids that the $\mathbb{M}^{\mp}$ are two manifolds of $\mathrm{Sl}(n+1,\mathbb{R})$ matrices

$$\begin{bmatrix}1\mp 2\langle z,z\rangle & \mp 2\alpha\zeta^t I\\ 2\alpha\zeta & \mathrm{id}_n\mp 2\zeta\otimes\zeta^t I\end{bmatrix} \text{ with } \pm\alpha^2+\langle z,z\rangle=\pm 1, \tag{30}$$

and α in $\mathbb{R}$, z in $\mathbb{V}$. Note that the matrices (29) are of the form (30) for the choice $z=(+2\langle x,x\rangle)^{-1/2}(-1+\cos\mu\sqrt{\pm\langle x,x\rangle})^{1/2}x$. In the four-dimensional Lorentz case these two manifolds are covered twice by the de Sitter hyperboloids. From the results of Ref. 12 follows that these manifolds are affine and complete with respect to the canonical pseudo-Riemannian $\square$-connection of symmetric spaces such that the geodesics passing through the points id_{n+1} in $\mathbb{M}^{\mp}$ are exactly the curves (29).

Another realization of R_0 and the corresponding pseudo-orthogonal Lie triple can be given in terms of the Clifford algebra $\mathrm{Cl}(\mathbb{V},\langle\,,\rangle)$ over $(\mathbb{V},\langle\,,\rangle)$. The exponential series in $\mathrm{Cl}(\mathbb{V},\langle\,,\rangle)$ guarantees the

existences of an n-dimensional pseudo-Riemannian manifold, $\square$-generated by the $\exp \mu x$ with μ in $\mathbb{R}$ and x in $V \subset \mathrm{Cl}(V, \langle , \rangle)$. Indeed $\mathbb{V}$ embedded in $\mathrm{Cl}(V, \langle , \rangle)$ is the Lie triple defined by R_0 (Ref. 15, p. 232). These manifolds are coverings of the $\mathbb{M}^\dagger$. The same construction starting from (symmetrized) $xy - yx \in \mathrm{Cl}(\mathbb{V}, \langle , \rangle)$, $x, y \in \mathbb{V}$, yields the spin coverings of the pseudo-orthogonal groups. However, a closed form of the points of these coverings corresponding to (30) is not known.

6. REMARKS ON THE CLASSIFICATION OF $\mathbb{G}$-ORBITS IN KERN W

Since dilatations in $\mathbb{V}$ act on $JA(\mathbb{V}, \langle , \rangle)$ only trivially (8) implies that for the classification of $\mathbb{G}$-orbits in kern W it suffices to know the $\mathrm{O}(\mathbb{V}, \langle , \rangle)$-orbits in $JA(\mathbb{V}, \langle , \rangle)$.

The Jordan algebra $JA(\mathbb{V}, \langle , \rangle)$ is simple, hence all derivations are *inner*, i.e., of the form $[L(A), L(B)]_-$ where L is the Jordan left multiplication $2L(A)B = AB + BA$. We have $[L(A), L(B)]_- C = [[A, B]_-, C]_-$ and $[A, B]_- \in \mathrm{so}(\mathbb{V}, \langle , \rangle)$ for all A, B, C in $JA(\mathbb{V}, \langle , \rangle)$. Using a basis, one proves that $\mathrm{so}(\mathbb{V}, \langle , \rangle)$ is spanned by these commutators $[A, B]_-$, whence the identity components of the pseudo-orthogonal group $\mathrm{O}(\mathbb{V}, \langle , \rangle)$ and the automorphism group of the Jordan algebra coincide. Clearly $\mathrm{O}(\mathbb{V}, \langle , \rangle)$ lies entirely in this automorphism group. The proof of the converse is rather lengthy.

$JA(\mathbb{V}, \langle , \rangle)$ can be decompsed into the set $\mathbf{J}$ of invertible and the set $\mathbb{P}$ of noninvertible elements, and invertibility in $JA(\mathbb{V}, \langle , \rangle)$ coincides with that in the matrix sense. This decomposition is invariant under the action of the automorphism group. In $\mathbb{J}$ there are orbits of involutive matrices, i.e., $A^2 = \lambda \mathrm{id}_n$ with a nonzero constant λ in each orbit, in $\mathbb{P}$ there are special orbits of nilpotent matrices [if $JA(\mathbb{V}, \langle , \rangle)$ is not formal-real] and orbits of idempotent matrices. Helwig[16,17] has given a description of idempotent orbits with respect to the identity component of the automorphism group for arbitrary simple real Jordan algebras. Let us specialize this to the Lorentz case with $I = \mathrm{diag}(1, -\mathrm{id}_3)$ being the matrix of $\langle , \rangle$ in a standard basis of $\mathbb{V}$. The maximal formal real subalgebra of $JA(\mathbb{V}, \langle , \rangle)$ is the eigenspace of eigenvalue 1 of the involution automorphism $A \vdash IAI^{-1}$, i.e., the direct sum of the Jordan algebras of symmetric matrices in one and in three dimensions. Since this direct sum is not simple the set of primitive idempotents decomposes into two $\mathrm{SO}_0(\mathbb{V}, \langle , \rangle)$-orbits (Ref. 16, satz 10.5) (primitive means that an idempotent cannot be written as the sum of two other mutually orthogonal idempotents). Let us denote the four standard primitive idempotents in the Lorentz–Jordan algebra with only one entry 1 on the diagonal by $E_0, \ldots, E_3$. Then one proves that there is no Lorentz matrix G such that $GE_0G^{-1} = E_i$ for $i = 1, 2, 3$, but that the E_1, E_2, E_3 can be transformed in this way into each other. Hence these two orbits are even $\mathrm{O}(\mathbb{V}, \langle , \rangle)$-orbits. The degree of primitivity (Ref. 18, p. 78) of the Peirce 1-space (Ref. 18, p. 154) (which is a Jordan subalgebra) of an idempotent is called its *length* (Ref. 17, p. 320). A primitive idempotent has length one, the maximal occuring length is that of the identity element, called the *degree* of the Jordan algebra. In the Lorentz case it is four. The length is invariant under the action of the automorphism group. In Ref. 17 (p. 320) it is shown that $\mathrm{SO}_0(\mathbb{V}, \langle , \rangle)$ acts transitively on each topological component in the sets $\mathbf{J}^r$ of idempotents of length r. Here one has $r = 1, 2, 3, 4$. For $r = 2$ there are two orbits, one containing the standard idempotents diag(1, 1, 0, 0), diag(1, 0, 1, 0), diag(1, 0, 0, 1), which easily can be transformed into each other by a Lorentz matrix, the other containing the Lorentz similar idempotents diag(0, 1, 1, 0), diag(0, 1, 0, 1), diag(0, 0, 1, 1). There is no Lorentz transformation transforming one element of the first set into one of the second set. Similarly the idempotents of length three decompose into two orbits, the first containing diag(1, 1, 1, 0), diag(1, 1, 0, 1), diag(1, 0, 1, 1), and the second containing diag(0, 1, 1, 1). To get the dimensions of these orbits note that Helwig has shown (Ref. 16, p. 23) that the Peirce $\frac{1}{2}$-component of an idempotent is the tangent space to the orbits through this idempotent in this point. A verification shows that the two $r = 1$ orbits have dimensions three, the two $r = 2$ orbits have dimensions four, and the two $r = 3$ orbits have dimensions three. These six orbits exhaust the set of all idempotents different from the trivial ones 0 and id_4. One easily calculates the stability groups in $\mathrm{O}(1, 3; \mathbb{R})$ for the above standard idempotents which gives an idea of the topological structure of the corresponding orbits.

Among the various orbits in $\mathbf{J}$ we mention only one which contains the matrix $\mathrm{diag}(-1, \mathrm{id}_3) = -I$. Writing $x = \xi^0 + x_R$ with ξ^0 in $\mathbb{R}$, x_R in $\mathbb{R}^3$, and $(,)$ for the Euclidean product we get

$$R_0^{-I, \mathrm{id}}(x, y)z = \tfrac{1}{2}(\langle Iy, z \rangle x + \langle y, z \rangle Ix - \langle Ix, z \rangle y - \langle x, z \rangle Iy)$$
$$= (y_R, z_R)x_R - (x_R, z_R)y_R,$$

i.e., $R_0^{-I, \mathrm{id}}(x, y)$ has only a $\mathbb{R}^3$-component which is exactly the trivial curvature structure on $\mathbb{R}^3$ with respect to $(,)$. Since the corresponding Lie triple is the double cross product in Euclidean $\mathbb{R}^3$ this curvature structure lies in a symmetric orbit which is given by a Lie algebra, namely the direct sum of the Lie algebras $\mathbb{R}$ and su(2). The Lorentz manifolds $\mathbb{R} \times \mathbb{S}^3$ and $\mathbb{S}^1 \times \mathbb{S}^3$ have their pseudo-Riemannian curvature structure in this orbit. To show that the dimension of this orbit again is three, note that the stability group of $-I$ in $\mathrm{O}(1, 3; \mathbb{R})$ is given by the matrices $\mathrm{diag}(\pm 1, A)$ with $A^t A = \mathrm{id}_3$.

The *gravitational equation* for $R \in \mathrm{curv}(\mathbb{V}, \langle , \rangle)$ is (without cosmological term)

$$\rho_{\mathcal{L}R}(x, y) - \frac{\mathrm{Sc}(R)}{n}\langle x, y \rangle = \rho_R(x, y) - 2\frac{\mathrm{Sc}(R)}{n}\langle x, y \rangle$$
$$= \kappa \langle T_R x, y \rangle,$$

where the gravitational constant is denoted by κ and the *energy–momentum transformation* is

$$\kappa T_R = L_{\mathcal{L}R} - \frac{\mathrm{Sc}(R)}{n}\mathrm{id}_n$$
$$= L_R - 2\frac{\mathrm{Sc}(R)}{n}\mathrm{id}_n \in JA(\mathbb{V}, \langle , \rangle).$$

If $W R = 0$, then from kern $L = \{0\}$ there is the equivalent form

$$R - 2\frac{\mathrm{Sc}(R)}{n(n-1)} R_0 = \kappa L^{-1}(T_R).$$

Using $\kappa \mathrm{tr} T_R = -\mathrm{Sc}(R)$, the gravitational equations become

$$\rho_R(x,y) = \kappa\langle\left(T_R - 2\frac{\mathrm{tr}T}{n}\mathrm{id}_n\right)x, y\rangle.$$

It is easy to see that $T_{\Omega(A)} = 0$ is equivalent to $A = 0$. The transformation property of T_R under G, resp. $L\mathrm{G}$, is given by (10), i.e., by

$$G \cdot T_R = \lambda_G^2 T_{G \circ R}, \text{ resp. } Q \cdot T_R = T_{Q \circ T} + 2\mu_Q T_R.$$

For a physical identification of the above $\mathbb{G}$-orbits in kern W one has to calculate $T_{\Omega(A)}$. We leave this simple task to the reader.

7. IRREDUCIBLE, SEMISIMPLE, AND FRIEDMANN $\mathbb{G}$-ORBITS

Let us add some remarks on the Ricci form of non-Einsteinian curvature structures. From (4)

$$\rho_{\Omega(A)}(x,y) = \frac{n-2}{2}\langle Ax, y\rangle + \frac{\mathrm{tr}A}{2}\langle x, y\rangle,$$

$$A \in JA(\mathbb{V}, \langle\,,\,\rangle). \tag{31}$$

A verification gives $\rho_{\Omega(A)}(\Omega(A)(x,y)z, w) + \rho_{\Omega(A)}(z, \Omega(A)(x,y)w) = [(n-2)/4]\{\langle A^2x, z\rangle\langle y, w\rangle - \langle x, z\rangle\langle A^2y, w\rangle + \langle A^2x, w\rangle\langle y, z\rangle - \langle x, w\rangle\langle A^2y, z\rangle\}$. Hence for symmetric $\Omega(A)$, i.e., $A^2 = \lambda\,\mathrm{id}_n$, the $\Omega(A)(x,y)$ annihilate the Ricci form. In general, however, the $\Omega(A)(x,y)$ do not exhaust this annihilating Lie subalgebra of $\mathrm{so}(\mathbb{V}, \langle\,,\,\rangle)$; as will be shown elsewhere in the case of the nilpotent matrix (18) these elements span only a two-dimensional subalgebra of the three-dimensional annihilating algebra.

One can increase the structure theory of $\mathbb{G}$-orbits by introducing more orbit properties. One concept is *irreducibility* (Ref. 13, p. 56) in homogeneous orbits which means that the adjoint action of $\mathbb{L}_0$ and $\mathbb{V}$ is irreducible. Another concept is *semisimplicity* which means that the Ricci form is nondegenerate.[19] In ImΩ this is equivalent to the invertibility of the matrix $A + [\mathrm{trace}A/(n-2)]\mathrm{id}_n$. The symmetric (whence also the Lie algebra) semisimple curvature structures are classified, for instance in Ref. 12.

Another kind of non-Einsteinian curvature structure, considered frequently in general relativity are the following: Given u in $\mathbb{V}$ with $\langle u,u\rangle \neq 0$ we get idempotents S_u in $JA(\mathbb{V}, \langle\,,\,\rangle)$ by

$$S_u z = \frac{\langle u, z\rangle}{\langle u, u\rangle} u, \quad z \in \mathbb{V}.$$

The *Friedmann* (or *Robertson—Walker*) curvature structures in ImΩ are given by

$$A_u = \frac{(3n-4)\rho - (n-4)p}{n(n-1)(n-2)}\,\mathrm{id}_n - \frac{2}{n-2}(\rho + p)S_u$$

with ρ the density, p the pressure, and u the velocity of a relativistic perfect fluid, Ricci form and energy—momentum transformation are

$$\rho_{\Omega(A_u)}(x,y) = \frac{2}{n}(\rho - p)\langle x, y\rangle - (\rho + p)\frac{\langle u,x\rangle\langle u,y\rangle}{\langle u,u\rangle},$$

$$T_{\Omega(A_u)} = \frac{4}{n}p\,\mathrm{id}_n - (\rho + p)S_u.$$

One proves that for the "equation of state" $(2+n)p = (2-n)\rho$, the Ricci form becomes $(n+2)^{-1}4\rho\{\langle x,y\rangle - \langle u,u\rangle^{-1}\langle u,x\rangle\langle u,y\rangle\}$ which is degenerate with kernel $\mathbb{R}u$ (the fixed point set of S_u). In the other cases the Ricci form is nondegenerate, hence $\Omega(A_u)$ is semisimple. It is easy to see that

$$G \cdot S_u = S_{Gu}, \quad G \cdot A_u = A_{Gu}, \quad G \in \mathrm{G},$$

which shows that there are *Friedmann* $\mathbb{G}$-orbits in $JA(\mathbb{V}, \langle\,,\,\rangle)$ parametrized by ρ and p, whereas the normalized velocity $\langle u,u\rangle^{-1/2}u$ depends on the curvature structure in such an orbit.

The condition that A_u defines a symmetric $\Omega(A_u)$, i.e., $A_u^2 = \lambda\,\mathrm{id}_n$, gives besides the uninteresting case $p = -\rho$, the same equation $(2+n)p = (2-n)\rho$ as above for the nonsemisimple Friedmann case. Hence

$$A_u = \frac{4\rho}{(n-2)(n+2)}\{\mathrm{id}_n - 2S_u\}$$

describes a symmetric but not semisimple G-orbit in ImΩ. We intend to come back to this case elsewhere.

8. OPEN PROBLEMS

It remains to solve three problems. The first is to complete the classification of $\mathbb{G}$-orbits in $\mathrm{curv}(\mathbb{V}, \langle\,,\,\rangle)$ which is related to the Petrov classification of curvature structures. The second is to show that pseudo-Riemannian manifolds whose Levi-Civita curvature structures lie in the same orbit, i.e., are conformally equivalent, are related to each other exactly by a (pseudo-Riemannian) covering, or at least to generalize this from the symmetric to the homogeneous case: If two symmetric curvature structures are conformally equivalent their Lie triples and consequently their standard embeddings are isomorphic, hence given the exponential mapping the associated Lie groups are related to each other by covering. From the representation of a symmetric space as a homogeneous space of its standard embedding group (group of displacements in Ref. 12) it then follows that the symmetric spaces cover each other.

The third problem is to construct at least one example in each orbit, as done in Sec. 5 for the trivial orbits. An algebraic question to ask is whether there are A-modifications of the Duffin—Kemmer—Petiau representation such that the exponential series generates examples of pseudo-Riemannian manifolds with Levi-Civita connections in other than the trivial $\mathbb{G}$-orbits in ImΩ. If any two examples in one orbit are related to each other by covering and if one succeeds in constructing a simply connected example, then it is straightforward to construct all other possible manifolds in the given orbit from this universal cover. In physics this would be a constructive approach to cosmology. Starting from given curvature structures the constructive approach has an easy solution for the symmetric orbits. In this case the $R(x,y)$ span a Lie

subalgebra of $so(V, \langle , \rangle)$ and it is easy to embed this into its standard embedding algebra. If R is not symmetric, the problem is more involved. For the construction of the standard embedding of a reductive but not symmetric structure for instance, one must know C and t, i.e., one has to solve Eq. (23) for the canonical curvature C and torsion t in terms of the given pseudo-Riemannian torsion-free Levi-Civita curvature $R = C^{\nabla}$. Note that in general C^{∇} is not a symmetric curvature structure, i.e., (25) does not hold, and hence there is no standard embedding Lie algebra constructed from C^{∇} and $t^{\nabla} = 0$.

For the Friedmann orbits in the preceding section the density ρ and the pressure p are constants, whereas in general relativity one considers them to be functions of time. A mathematical formulation of this probably is the concept of deformation of curvature structures in analogy to the well-known deformation theory of algebraic structures. Roughly speaking it consists of in applying a timelike one-parameter subgoup of $\mathbb{G}$ to the curvature structure, deforming it in its $\mathbb{G}$-orbit. It remains to study the resulting deformation of the pseudo-Riemannian manifold.

The classification of $\mathbb{G}$-orbits in $JA(V, \langle , \rangle)$ should be compared with that of the Lorentz orbits in Minkowski space. Minkowki space is a Jordan algebra as well[20] which, however, is formal real (or compact) and hence has no nilpotent elements. To get the Lorentz group there the group must be enlarged from the automorphism group of the Jordan algebra to the so-called structure group, see Ref. 18.

[1]I. M. Singer and J. A. Thorpe, "The Curvature of 4-Dimensional Einstein Spaces," in *Global Analysis*, *Papers in Honor of K. Kodaira*, edited by S. Spencer and S. Iyanaga (Princeton U. P., Princeton, N. J. and University of Tokyo Press, Tokyo, 1968).

[2]A. Gray, "Invariants of Curvature Operators in Four-Dimensional Riemannian Manifolds," in *Proc of the* 13th *Biennial Seminar of the Canadian Mathematical Congress* (Halifax, 1971).

[3]M. Marcus, *Finite-Dimensional Multilinear Algebra* II (Dekker, New York, 1976).

[4]A. Stehney, "Principal Null Directions without Spinors," J. Math. Phys. **17**, 1793–96 (1976).

[5]J. A. Thorpe, "Curvature and the Petrov Canonical Form," J. Math. Phys. **10**, 1–6 (1969).

[6]K. Nomizu, "On the Decomposition of Generalized Curvature Tensor Fields," in *Differential Geometry*, *Papers in Honor of K. Yano* (Kinukuniya, Tokyo, 1972).

[7]O. Kowalski, "Partial Curvature Structures and Conformal Transformations," J. Diff. Geom. **8**, 53–70 (1973).

[8]R. S. Kulkarni, "Curvature and Metric," Ann. Math. **91**, 311–31 (1970).

[9]R. S. Kulkarni, "Curvature Structures and Conformal Transformations," J. Diff. Geom. **4**, 53–70 (1970).

[10]H. Tilgner, "Symmetric Spaces in Cosmology," Rep. Math. Phys. **5**, 51–64 (1974).

[11]N. Jacobson, *Structure and Representations of Jordan Algebras* (American Math. Soc., Providence, Rhode Island, 1968).

[12]O. Loos, *Symmetric Spaces* I (Benjamin, New York, 1969).

[13]K. Nomizu, "Invariant Affine Connections on Homogeneous Spaces," Am. J. Math. **76**, 33–65 (1954).

[14]A. Sagle, "On Anticommutative Algebras and Homogeneous Spaces," J. Math. Mech. **16**, 1381–94 (1967).

[15]N. Jacobson, *Lie Algebras* (Interscience, New York, 1966).

[16]K. H. Helwig, "Halbeinfache reelle Jordan-Algebren," Math. Z. **109**, 1–28 (1969).

[17]K. H. Helwig, "Jordan-Algebren und symmetrische Räume I," Math. Z. **115**, 315–349 (1970).

[18]H. Braun and M. Koecher, *Jordan-Algebren* (Springer, Berlin, 1966).

[19]S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry II* (Interscience, New York, 1969).

[20]H. Tilgner, "Symmetric Spaces in Relativity and Quantum Theories," in *Group Theory in Nonlinear Problems*, edited by A. O. Barut (Reidel, Dordrecht, Holland, 1974).

Nonlinear tunnelling

Alan C. Newell

Department of Mathematics, Clarkson College of Technology, [a] *Potsdam, New York 13676*
(Received 18 October 1977)

The propagation of a soliton pulse towards a finite potential barrier is examined, and it is found that in certain circumstances, depending on the ratio of soliton amplitude to barrier height, the soliton can tunnel through the barrier in a lossless manner. It is suggested that the phenomenon of barrier penetration by nonlinear pulses may have widespread application.

1. INTRODUCTION

The concept of linear tunnelling, of central importance in modern physics, follows naturally from consideration of the linear equation

$$u_{tt} - (c^2(x)u_x)_x + n^2(x)u = 0 \tag{1.1}$$

(subscripts refer to partial derivatives). The equation serves as a model for discussing the propagation of disturbances of infinitesimal amplitude along a Josephson junction, in plasmas or more simply in the mechanical model (due to Scott[1]) for the sine–Gordon equation which consists of a line of pendulae under the influence of gravity strung closely together on a torsion wire. The parameters $c^2(x)$ and $n^2(x)$ may be functions of position. Seeking solutions of the form $u(x,t) = \psi(x)\exp(-i\omega t)$, we obtain the Sturm–Liouville problem,

$$(c^2(x)\psi_x(x))_x + (\omega^2 - n^2(x))\psi = 0, \tag{1.2}$$

which, when $c^2 = 1$, is the Schrödinger equation describing the probability density function $\psi(x)$ for an electron of kinetic energy ω^2 moving in a potential $n^2(x)$. Consider the potential, $n^2 = 0$, $x < 0$, $n^2 = V_1 > \omega^2$, $0 < x < L$, and $n^2 = V_2 < \omega^2$, $x > L$. An incident wave arriving at $x = 0$ is partially reflected and partially transmitted through the shadow zone $0 < x < L$ in which the amplitude of the wave suffers the exponential decay $\exp[-(V_1 - \omega^2)^{1/2}L]$. No free wave $\exp[i(kx - \omega t)]$ can propagate in this barrier region since the wavenumber k given by the dispersion relation $k = k(\omega, x)$,

$$k^2 = \frac{1}{c^2(x)}(\omega^2 - n^2(x)) \tag{1.3}$$

(here we have tacitly assumed that n and c are slowly varying) only permits a real solution for $n^2(x) < \omega^2$. The point x (say $x = 0$) at which $n^2(x) = \omega^2$ is called a caustic and the group velocity of the linear wave is zero there.

On the other hand, a small displacement $u(x,t)$ which obeys the nonlinear equation

$$u_{tt} - (c^2 u_x)_x + n^2 F(u) = 0 \tag{1.4}$$

may admit a periodic solution with an amplitude dependent frequency which is below the cutoff frequency n.[2] For instance, if $F(u) = \sin u \simeq u - \frac{1}{6}u^3 + o(u^3)$, and the disturbance is small, we may show that a solution of the form

$$u(x,t) = \epsilon A(X,T)\exp[i/\epsilon\theta(X,T)]$$

$$+ (*),\quad X = \epsilon x,\quad T = \epsilon t,\quad \theta_t = -\omega,\quad \theta_x = k, \tag{1.5}$$

which assumes $c^2(X)$ and $n^2(X)$ are slowly varying, leads to a nonlinear dispersion relation

$$k^2 = \frac{1}{c^2}(\omega^2 - n^2 + \tfrac{1}{2}n^2\epsilon^2 AA^*), \tag{1.6}$$

which suggests that k may still be real for values of ω^2 below the cutoff frequency n^2. Alternatively, if we imagine $n^2(x)$ to increase through the value ω^2 from below ($n^2(0) = \omega^2$), then the point X_p at which k^2 is zero lies in the shadow region of the linear problem, i.e., X_p is such that $n^2(X_p) = \omega^2(1 - \frac{1}{2}AA^*)^{-1}$ and is thus greater than zero.

Unfortunately, in these cases, the periodic wave is *unstable* in the Whitham[2] and Benjamin–Feir[3] sense and a wavetrain of frequency ω, initiated at a station $X = X_1$, breaks up into separate local pulses soon thereafter. If $u(x,t)$ is sufficiently small, then these pulses are solitons of the nonlinear Schrödinger (NLS) equation

$$q_t - iq_{xx} - 2iq^2q^* = 0 \tag{1.7}$$

although in order to incorporate their genuine nonlinear, as opposed to their quasilinear, tunnelling character, we must use a modified NLS (or MNLS) equation which adds a term $i\delta\epsilon q_{xt}$ to the right-hand side of (1.7).

For certain special choices of $F(u)$, we may waive the small amplitude restriction. One such case which models both the Josephson junction and its mechanical analog referrred to earlier is the sine–Gordon model for which $F(u) = \sin u$. Consider the case where c^2 and n^2 are constant. Then it is known that a pulse of finite duration initiated at X_1 will decompose into (a) a series of solitions (both kinks or 2π pulses and breathers or 0π pulses) and (b) radiation which consists of nonlinear phonons (the linear limit of a phonon is a sinusoidal wave $\exp(ikx - i\omega t)$ with frequency $\omega^2 = c^2k^2 + n^2$ greater than n^2) and near the light cone a component with self similar structure.[4] The soliton comes in two varieties. The kink or 2π pulse is a steady progressing wave

[a] This work was supported by NSF Grant MCS75-07508-A02 and ONR Grant N00014-76-C-0867. Part of it was carried out while the author enjoyed a Guggenheim fellowship visiting Cambridge University.

 0022-2488/78/1905-1126$1.00

$$u(x,t) = 4\tan^{-1}\exp[n/c\gamma(x - Vt)], \quad \gamma = (1 - V^2c^{-2})^{-1/2}, \tag{1.8}$$

which corresponds to a complete twist in the line of pendula. Its energy is γM_0c^2. The second type and the one of central interest for this paper is the breather or 0π pulse [with energy $2\gamma M_0c^2(1 - \Omega^2/n^2)^{1/2}$]

$$u(x,t) = 4\tan^{-1}\frac{2\eta_2 c}{\Omega}\operatorname{sech}(2\eta_2)\frac{\gamma}{\epsilon}(X - VT) \times \cos\left(\frac{1}{\epsilon}\Omega\gamma(T - VXc^{-2})\right), \tag{1.9}$$

and the first feature we point out is that its frequency Ω

$$\Omega = (n^2 - 4\eta_2^2c^2)^{1/2} \tag{1.10}$$

lies below the cutoff frequency n. We have exploited this fact before and shown how a breather can phase-lock onto an applied ac field with a frequency below n.[5] In this paper we suggest that the breather or more generally (although only for small amplitudes) the NLS soliton is the means by which coherent forms of energy can tunnel without loss to the interior of regions where linear theory would allow only edge effects (cf. London penetration depth in superconductivity, see Ref. 6). In a medium (such as a Josephson junction of varying width or along which the temperature varies, or a plasma with increasing density), in which the frequency $n(X)$ increases past the excitation frequency ω, the pulse will penetrate into the medium a distance X_p given by

$$\omega^2 = n^2(X_p) - 4\eta_2^2c^2, \tag{1.11}$$

at which point it will turn around and return unless the potential barrier $n^2(x)$ decreases again.

We distinguish three different types of tunnelling. The first two are lossless and occur because (a) the frequency $\omega + 2\xi\epsilon$ of the oscillatory phase of a particular soliton is greater than the central frequency ω of the excitation (which we might call quasilinear tunnelling or lossless linear tunnelling of a nonlinear pulse) or (b) *more importantly*, the effective frequency (wavenumber) of the pulse is decreased (increased) by an amplitude dependence and pulses of sufficiently large amplitudes may succeed in penetrating the barrier completely (lossless nonlinear tunnelling). The third type of tunnelling is a tunnelling with losses in which even though the soliton itself cannot penetrate all the way through the potential barrier, the influence of its leading edge is felt in the region beyond the barrier where it can create secondary pulses and radiation. However, the exponential decay of the signal from one side of the barrier to the other depends on the pulse amplitude. We call this nonlinear tunnelling with loss or normal nonlinear tunnelling. We remark that it might be more accurate to prefix each definition with "soliton" as here we neglect altogether the propagation of the phonons (radiation).

We begin in Sec. 2 by showing the way in which low amplitude solutions of (1.4) are described by a modified nonlinear Schrödinger equation (MNLS), and in particular relate the parameters natural to (1.7) to those in (1.9). Then, in Secs. 3–6, we describe the propagation of a pulse through a region of increasing plasma frequency $n(x)$, then through the region of the (linear caustic) at $X = 0$, and finally discuss the circumstances under which one can expect it to tunnel without loss through a potential barrier region. The results should have many applications, and some of these are discussed in the conclusion. In the Appendix we write down the principal results (the derivations of these results are to be found in Ref. 5) of perturbation theory which are used.

2. THE NONLINEAR SCHRÖDINGER (NLS) CONNECTION

In the low amplitude limit, $\eta_2 = \epsilon\eta_1$ (1.9) becomes

$$u(X,T) = 4\epsilon\eta_1\frac{c}{\Omega}\operatorname{sech}2\eta_1\gamma(X - VT) \times \exp[-i\epsilon\gamma(T - XVc^{-2})] + (*), \tag{2.1}$$

and the reader may recognize the amplitude to be the envelope soliton of the NLS equation. To see this in another way, substitute

$$u(X,T) = \epsilon A(X,T)\exp[i(kX - \omega T)/\epsilon] + (*), \tag{2.2}$$

in (1.4) with $F(u) = u - \frac{1}{6}u^3$ and find the envelope equation

$$-2i\omega A_T - 2ikc^2A_X + \epsilon A_{TT} - \epsilon A_{XX} = \beta\epsilon n^2A^2A^*, \quad \beta = \tfrac{1}{2}, \tag{2.3}$$

together with the dispersion relation (1.3). The transformation

$$\tau = T - \frac{\omega}{c^2k}X, \quad y = \frac{1}{2}\frac{n^2}{c^4k^3}Y, \quad Y = \epsilon X, \quad A = \left(\frac{2}{\beta}\right)^{1/2}\frac{1}{ck}B \tag{2.4}$$

leads to the exact equation

$$B_y - iB_{\tau\tau} - 2iB^2B^* = -\frac{i\epsilon\omega}{c^2k^2}B_{y\tau} + \frac{i\epsilon^2}{4}\frac{n^2}{c^4k^4}B_{yy}. \tag{2.5}$$

Neglect of the right-hand side leaves the NLS equation whose general solution consists of solitons and radiation. The soliton expression is

$$B = 2\eta\operatorname{sech}2\eta(\tau + 4\xi y - \tau_1) \times \exp[-2i\xi\tau - 4i(\xi^2 - \eta^2)y + i\chi] \tag{2.6}$$

with constant τ_1 and χ, and the corresponding solution $u(X,T)$ may be written,

$$u(X,T) = \epsilon\left(\frac{2}{\beta}\right)^{1/2}\frac{2\eta}{ck}\operatorname{sech}2\eta\left(T - \frac{\omega}{c^2k}X + 2\xi\frac{n^2}{c^4k^3}\epsilon X - \tau_0\right) \times \exp\left[-i(\omega + 2\xi\epsilon)T/\epsilon + i\left(k + \frac{2\xi\epsilon\omega}{kc^2} + 2\eta^2\frac{n^2}{c^4k^3}\epsilon^2\right)X/\epsilon\right], \tag{2.7}$$

or, alternatively,

$$u(X,T)=\epsilon\left(\frac{2}{\beta}\right)^{1/2}\frac{2\eta}{ck}\,\mathrm{sech}2\eta[T-k_1(\omega+2\xi\epsilon)X-\tau_1 + O(\epsilon\xi)^2]\exp\Big[-i(\omega+2\xi\epsilon)T/\epsilon+i\Big(k(\omega+2\xi\epsilon) + 2\eta^2\frac{n^2}{c^4k^3}\epsilon^2\Big)X/\epsilon\Big], \tag{2.8}$$

where $k(\omega)$ is given in (1.3) and $k_1=\partial k/\partial\omega$, the inverse of the linear group velocity. We note that the nonlinear term affects the wavenumber of the oscillation and the shape, but not the phase, of the amplitude (the hyperbolic secant). Pulses of all amplitudes with the same frequency travel at the same speed! This feature often does not faithfully reflect the properties of the original model for which (2.5) with $\epsilon=0$ is an approximation, and one would prefer to see some amplitude dependence in the phase of the amplitude function. This is introduced by including the first term on the right-hand side of (2.5). However, the resulting equation, the modified NLS (MNLS) is no longer exactly integrable (as far as we know). Nevertheless, if we neglect terms of $O(\epsilon\xi^2,\epsilon\xi\eta)$ with respect to the term $O(\epsilon\eta^2)$ (i.e., the amplitude $\epsilon\eta$ is much greater than the frequency shift $\epsilon\xi$), then the principal effect of this term is accounted for if

$$\tau_1=-\frac{4\epsilon\omega\eta^2}{k^2c^2}y = -\frac{2\omega n^2}{c^6k^5}(\epsilon\eta)^2X.$$

Indeed, if $\xi=0$, this is an exact result. The result can be found either by use of the perturbation procedure outlined in Ref. 5 or directly by assuming that only τ_1 varies with respect to y. With this inclusion, the velocity of the pulse

$$V=\left(k_1(\omega+2\xi\epsilon)-\frac{2\omega n^2}{c^6k^5}(\epsilon\eta)^2\right)^{-1} \tag{2.9}$$

is larger for larger amplitudes, reflecting more accurately the characteristics of the breather (1.9).

The connection between the parameters in (2.1) and (2.8) is as follows:

$$2\eta_1\gamma=2\eta\left(k_1(\omega+2\xi\epsilon)-\frac{2\omega n^2}{c^6k^5}(\epsilon\eta)^2\right), \tag{2.10a}$$

$$2\eta_1\gamma V=2\eta, \tag{2.10b}$$

$$\gamma\Omega=\omega+2\xi\epsilon, \tag{2.10c}$$

$$\gamma\Omega\frac{V}{c^2}=k(\omega+2\xi\epsilon)+2\frac{n^2}{c^4k^3}(\epsilon\eta)^2. \tag{2.10d}$$

Using $k^2=(1/c^2)(\omega^2-n^2)$, we find on squaring (2.10d), retaining terms of order $(\epsilon\eta)^2$, and using (2.10c),

$$\Omega^2=n^2-4(n^2/c^2k^2)(\epsilon\eta)^2. \tag{2.11}$$

To leading order $V=1/k_1=ck^2/\omega$, $\gamma=\omega/n$, and $\eta=\eta_1\cdot(ck^2/n)$, whence (2.11) is (1.10).

3. APPROACHING A CAUSTIC

We examine the following situation depicted in Fig. 1. We will specify that all the medium properties change slowly with respect to the soliton length scale $Y=\epsilon X$. Imagine that at $Y=Y_1=\epsilon X_1$ a pulse $u(X,T)$ is initiated with central frequency ω {for examples, $u(X,T)=\epsilon\exp(-i\omega T/\epsilon)$, $T_1<T<T_2$, $u(X_1,T)=\epsilon\exp(-i\omega T/\epsilon)\,\mathrm{sgn}[T-(T_1+T_2)/2]$, $T_1<T<T_2$} and progresses through a region $Y_1<Y<Y_0$ (AB) of constant medium properties. In this interval the pulse will decompose into its soliton components (provided $\int_{-\infty}^{\infty}u(X_1,T)dT>\epsilon(2/\beta)^{1/2}(c_0k_0)^{-1}\pi/2$ and we will seek to follow one of the solitons,

$$u(X,T)=\epsilon\left(\frac{2}{\beta}\right)^{1/2}\cdot\frac{2\eta_0}{c_0k_0}\times\mathrm{sech}2\eta_0\Big[T-\Big(k_1(\omega+2\xi_0\epsilon) - 2\frac{\omega n_0^2}{c_0^6k_0^5}(\epsilon\eta_0)^2\Big)(X-X_0)\Big]\times\exp\Big[-i(\omega+2\xi_0\epsilon)T/\epsilon+i\Big(k(\omega+2\xi_0\epsilon) + \frac{2n_0^2}{c_0^4k_0^3}(\epsilon\eta)^2\Big)(X-X_0)/\epsilon\Big], \tag{3.1}$$

through its subsequent evolution in the region $X_0<X<0$ as it travels towards the linear caustic at $X=0$. Without loss of generality we may take $\xi_0=0$. For, suppose $n^2(X)=(\omega+2\xi_0\epsilon)^2[1+2\sigma\epsilon(X-2\xi_0/\sigma\omega)]$, then we can, neglecting terms of order $\epsilon^2\xi^2$, write this as $n^2(X)=\omega^2(1+2\sigma\epsilon X)$ by redefining ω and applying a simple translation. That is, to say, we expect a pulse of frequency $\omega+2\xi_0\epsilon$ to penetrate a distance $X_p=2\xi_0/\sigma\omega$ past the point at which $n^2=\omega^2$. This is what has already been referred to as the linear tunnelling of a nonlinear pulse or quasilinear tunnelling.

We next seek to find the solution in region II, $Y_0<Y<0$ [$Y_0=O(\sigma^{-1})$, and each of the parameters c^2 and n^2 changes by an order one amount], which matches to

$$u(X_0,T)=\epsilon(2/\beta)^{1/2}\cdot(2\eta_0/c_0k_0)\,\mathrm{sech}2\eta_0T\exp(-i\omega T/\epsilon) \tag{3.2}$$

at $Y=Y_0$. In (1.4), with $F(u)=u-\beta/3u^3+o(u^3)$, set (following Whitham[2])

$$u(X,T)=\epsilon A(X,T)\exp[i/\epsilon\theta(X,T)]+(*), \tag{3.3}$$

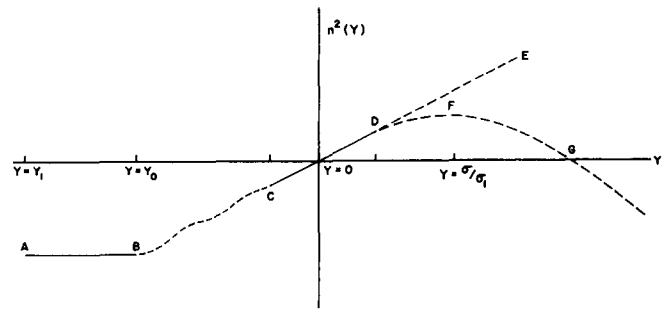

FIG. 1. Graph of $n^2(Y)$ vs Y.

where $\theta_x = \epsilon\theta_X = k$ and $\theta_t = \epsilon\theta_T = -\omega$ are the local wavenumber and frequency. Substitution of (3.3) in (1.4) gives us the local dispersion relation,

$$k^2 = (1/c^2)(\omega^2 - n^2), \tag{3.4}$$

the equation expressing conservation of waves,

$$k_T + \omega_X = 0, \tag{3.5}$$

and finally the amplitude equation

$$-2i\omega A_T - 2ikc^2 A_X - i(kc^2)_X A + \epsilon A_{TT} - \epsilon A_{XX} = \beta n^2 \epsilon A^2 A^*. \tag{3.6}$$

Since $k = k(\omega, X)$, that is, the slowly varying parameters depend only on X, we have from (3.5) $\omega_T + \omega_1\omega_x = 0$, $\omega_1 = \partial\omega/\partial k = k_1^{-1}$. From initial conditions we have $\omega = \text{const}$. Thus k is a function of X only, and its value at any position is given by (3.4). We assume now that c^2 and n^2 depend on σY, $0 < \epsilon \ll \sigma \ll 1$. It is natural to expect that, to first order, the amplitude moves with the local group velocity and thus it is appropriate to introduce the new independent variables

$$\tau = T - \frac{1}{\epsilon}\int_{Y_0}^{Y} k_1(Y)\,dY, \tag{3.6a}$$

$$Y = \epsilon X, \tag{3.6b}$$

whence the amplitude equation becomes

$$-2i\frac{k^3c^4}{n^2}A_y - A_{\tau\tau} - \beta c^2 k^2 A^2 A^* = i\frac{(c^2k)_Y}{(c^2k)}\frac{k^3c^4}{n^2}A - \frac{2c^2\epsilon}{\omega_1}\frac{c^2k}{n^2}A_{Y\tau} + o(\epsilon). \tag{3.7}$$

Notice that if we neglect the dispersion and nonlinear terms [in particular the former since we could account for the nonlinear term with a nonlinear dispersion relation; cf. (1.6)], then the balance of the first terms on both sides of Eq. (3.7) gives us conservation of wave action, i.e., $c^2kAA^* = \text{const}$. As the caustic is approached $k \propto \sqrt{-X}$ and thus $|A| \propto (-X)^{-1/4}$, the usual WKB result. However, the neglected dispersion is $O(k^{-3})$ and clearly cannot be ignored; in fact in our model it dominates the term proportional to $(c^2k)_Y$. Introducing the Y dependent change of amplitude

$$A = \left(\frac{2}{\beta}\right)^{1/2}\cdot\frac{1}{ck}B, \tag{3.8}$$

and rescaling the Y coordinate

$$y = \frac{1}{2}\int_{Y_0}^{Y}\frac{n^2}{c^4k^3}\,dY, \tag{3.9}$$

we find

$$B_y - iB_{\tau\tau} - 2iB^2B^* = \frac{k_y}{2k}B - \frac{i\epsilon}{\omega_1 k}B_{y\tau} + O(\epsilon^2). \tag{3.10}$$

The first term on the right-hand side is of order σ/Y and so the following result is only valid for $\sigma/Y \ll 1$, i.e., $0 < \sigma \ll 1$ and $Y = O(1)$ for example. It is not valid all the way down to $Y = 0$. Previous work[5] has shown that if

$$B = 2\eta\,\text{sech}2\eta(\tau - \tau_0)\exp 4i\int\eta^2 dy, \tag{3.11}$$

then $\eta_y = (k_y/k)\eta$, which integrates to

$$\eta/\eta_0 = k/k_0. \tag{3.12}$$

Also $\tau_0 = -4\epsilon\omega\int_0^y(\eta^2/c^2k^2)\,dy$. This term arises from the $B_{y\tau}$ term in (3.10) and introduces an amplitude dependence into the phase of the amplitude of B. The corresponding $u(X, T)$ is

$$u(X,T) = \epsilon\left(\frac{2}{\beta}\right)^{1/2}\frac{\eta_0}{ck_0}\,\text{sech}2\eta_0\frac{k}{k_0}\left(\tau + 4\epsilon\omega\frac{\eta_0^2}{k_0^2}\int_0^Y\frac{1}{c^2}dy\right) \times\exp\left(-\frac{i\omega T}{\epsilon} + \frac{i}{\epsilon^2}\int_{Y_0}^{Y}k(Y)\,dY + 4i\int_0^y\eta^2\,dy\right). \tag{3.13}$$

Note closely the somewhat unexpected result (the surprise only comes if one believes that the action c^2kAA^* is conserved) that the maximum amplitude of $u(X, T)$ changes from one constant $\epsilon(2/\beta)^{1/2}\eta_0/c_0k_0$ to another $\epsilon(2/\beta)^{1/2}\eta_0/c(0)k_0$ as $Y \to 0$, and does not undergo WKB enhancement!

Since our next task will be to develop a solution in the neighborhood of $Y = 0$ to which we will match (3.13), we will now give the asymptotic expression for the integrals contained in the phases of (3.13). In order to calculate these expressions, we assume that, close to $Y = 0$,

$$n^2 = \omega^2(1 + 2\sigma Y), \quad c^2 = 1,$$
$$\text{and}\quad k = (\omega/c)\sqrt{-2\sigma Y} \approx \omega\sqrt{-2\sigma Y}. \tag{3.14}$$

The way in which n^2 approaches ω^2 is important but only the value c^2 at $Y = 0$ is required. The fact that we have taken $c(0) = 1$ means that its value has been used to provide the original velocity scale in the problem. We find

$$y = \frac{1}{2}\int_{Y_0}^{Y}\frac{n^2}{c^4k^3}dY \sim \frac{1}{2\sigma\omega}\frac{1}{(-2\sigma Y)^{1/2}} + C_1, \tag{3.15a}$$

$$\int_{Y_0}^{Y}k\,dY \sim -\frac{\omega}{3\sigma}(-2\sigma Y)^{3/2} + C_2, \tag{3.15b}$$

$$\int_{Y_0}^{Y}k_1\,dY \sim \frac{1}{\sigma}(-2\sigma Y)^{1/2} + \epsilon\tau_0, \tag{3.15c}$$

$$\frac{1}{2}\int_{Y_0}^{Y}\eta^2\frac{n^2}{c^4k^3}dY \sim -\frac{\omega\eta_0^2}{2\sigma k_0^2}(-2\sigma Y)^{1/2} + C_3, \tag{3.15d}$$

$$\frac{\eta_0^2}{k_0^2}\int_{Y_0}^{Y}\frac{n^2}{c^6k^3}dY \sim \frac{\eta_0^2}{k_0^2}\frac{1}{2\sigma\omega}\frac{1}{(-2\sigma Y)^{1/2}} + C_4. \tag{3.15e}$$

The constants are given by

$$C_1 = -\frac{1}{2\sigma\omega}\frac{1}{(-2\sigma Y_0)^{1/2}} + \frac{1}{2\omega^3} \times\int_{Y_0}^{0}\left(\frac{n^2\omega^3}{c^4k^3} - \frac{\omega^2}{(-2\sigma Y)^{3/2}}\right)dY,$$

$$C_2 = \int_{Y_0}^{0}k\,dY,$$

$$C_3 = \frac{\eta_0^2}{2k_0^2}\int_{Y_0}^{0}\frac{n^2}{c^4k}dY,$$

$$C_4 = \frac{\eta_0^2}{2\omega^3 k_0^2}\int_{Y_0}^{0}\left(\frac{n^2\omega^3}{c^6k^3} - \frac{\omega^2}{(-2\sigma Y)^{3/2}}\right)dY$$

$$- \frac{\eta_0^2}{k_0^2}\frac{1}{2\sigma\omega}\frac{1}{(-2\sigma Y_0)^{1/2}}, \tag{3.16}$$

and

$$\tau_0 = \frac{1}{\epsilon}\int_{Y_0}^{0}\frac{\omega}{c^2k}dY. \tag{3.17}$$

Only τ_0, which measures the approximate time of travel of the pulse between $Y = Y_0$ and $Y = 0$, plays any substantive role in the subsequent analysis.

4. NEAR THE CAUSTIC AT $X = 0$

Here we assume

$$u(X, T) = \epsilon A(X, T)\exp(-i\omega T/\epsilon) + (*) \tag{4.1}$$

and let $c^2 = 1$, $n^2 = \omega^2(1 + 2\sigma\epsilon X)$. The amplitude equation is then

$$-2i\omega A_T + \epsilon A_{TT} - \epsilon A_{XX} = -2\sigma\omega^2 XA + \beta\epsilon n^2 A^2 A^*. \tag{4.2}$$

It turns out, in a construction to be given in some generality in Sec. 5, that if the term A_{TT} can be neglected with respect to the term A_{XX}, (4.2) can be converted to the canonical NLS equation. The condition that A_{TT}/A_{XX} is small is that σY is small. Thus this solution cannot be attached directly to $Y = Y_0 = O(1/\sigma)$ but must be related to the initial pulse (3.1) through the intermediate solution described in Sec. 3. We find

$$u(X, T) = \epsilon\left(\frac{2}{\beta}\right)^{1/2}\frac{2\bar{\eta}}{\omega}\operatorname{sech}2\bar{\eta}\left(\bar{X} - \frac{\sigma\epsilon T^2}{2} - 2\bar{\xi}\frac{\epsilon T}{\omega} - X\right)$$

$$\times\exp\left[-i\omega\frac{T}{\epsilon} - i\sigma\omega X\left(T + \frac{2\bar{\xi}}{\sigma\omega}\right)\right.$$

$$\left. - \frac{i\sigma^2\omega\epsilon}{6}\left(T + \frac{2\bar{\xi}}{\sigma\omega}\right)^3 + 2i\bar{\eta}^2\frac{\epsilon T}{\omega} + i\bar{\phi}\right]. \tag{4.3}$$

We now write $u(X, T)$ as given by (4.3) in terms of the variables of the intermediate solution. Note from (3.6) and (3.15),

$$T = \tau + \tau_0 - \frac{1}{\sigma\epsilon}(-2\sigma Y)^{1/2}, \quad X = \frac{1}{\epsilon}Y. \tag{4.4}$$

Matching amplitudes gives

$$\bar{\eta} = \eta_0\omega/k_0. \tag{4.5}$$

We may write the phase of the hyperbolic secant in (4.3) as $2\bar{\eta}(X_p - \frac{1}{2}\sigma\epsilon(T + 2\bar{\xi}/\sigma\omega)^2 - X)$, where X_p is the *penetration distance*, i.e., the distance into the shadow zone $X > 0$ that the soliton travels before turning around. Using (4.4), this phase is

$$2\eta_0\frac{\omega}{k_0}\left\{X_p - \frac{\sigma\epsilon}{2}\left[\left(\tau_0 + \frac{2\bar{\xi}}{\sigma\omega}\right)^2 + 2\left(\tau_0 + \frac{2\bar{\xi}}{\sigma\omega}\right)\left(\tau - \frac{1}{\sigma\epsilon}(-2\sigma Y)^{1/2}\right)\right.\right.$$

$$\left.\left. + \tau^2 - \frac{2\tau}{\sigma\epsilon}(-2\sigma Y)^{1/2} - \frac{2\sigma Y}{\sigma^2\epsilon^2}\right] - \frac{Y}{\epsilon}\right.$$

$$= 2\eta_0\frac{\omega}{k_0}\left[X_p - \frac{\sigma\epsilon}{2}\left(\tau_0 + \frac{2\bar{\xi}}{\sigma\omega}\right)^2\right.$$

$$\left. - \sigma\epsilon\left(\tau_0 + \frac{2\bar{\xi}}{\sigma\omega}\right)\left(\tau - \frac{1}{\sigma\epsilon}(-2\sigma Y)^{1/2}\right)\right.$$

$$\left. + \tau(-2\sigma Y)^{1/2} - \frac{\sigma\epsilon}{2}\tau^2\right].$$

On the other hand, the phase of the hyperbolic secant in (3.13) is

$$2\eta_0\frac{\omega}{k_0}\left(\tau(-2\sigma Y)^{1/2} + \frac{2\epsilon\eta_0^2}{\sigma k_0^2} + 4\epsilon\omega C_4(-2\sigma Y)^{1/2}\right).$$

Note that the $(-2\sigma Y)^{-1/2}$ singularity in $\int(1/c^2)\,dy$ is cancelled by the decreasing wavenumber $k = (\omega/c)(-2\sigma Y)^{1/2}$. Thus

$$X_p - \frac{\sigma\epsilon}{2}\left(\tau_0 + \frac{2\bar{\xi}}{\sigma\omega}\right)^2 = \frac{2\epsilon\eta_0^2}{\sigma k_0^2}, \tag{4.6a}$$

$$\tau_0 + \frac{2\bar{\xi}}{\sigma\omega} = 4\epsilon\omega C_4 = C, \tag{4.6b}$$

which expressions determine the position and velocity parameters X_p and $(-4\bar{\xi})$ near the caustic. Note that the accumulated speed $(-4\bar{\xi})$ is governed principally by the travel time τ_0. In fact, since τ_0 is $O(1/\epsilon)$,

$$-4\bar{\xi} = 2\sigma\omega\tau_0 + O(1) \tag{4.7}$$

and to leading order

$$X_p = 2\epsilon\eta_0^2/\sigma k_0^2. \tag{4.8}$$

If the original pulse had frequency $\omega + 2\xi_0\epsilon$, then the penetration depth is $X_p = 2\xi_0/\sigma\omega + 2\epsilon\eta_0^2/\sigma k_0^2$.

We have defined the *penetration depth* X_p to be the point of furthest penetration of the amplitude into the medium. It is interesting that it also admits the following interpretation. Define the local frequency $\bar{\omega}(X, T)$ to be the negative of the $t(= T/\epsilon)$ derivative of the oscillatory phase,

$$\bar{\omega}(X, T) = -\epsilon\theta_T = \omega(1 + \sigma\epsilon X) - 2(\epsilon\eta_0)^2\omega/k_0^2 + \omega V^2/2, \tag{4.9}$$

$$= \gamma\Omega = \gamma(n(X) - 2(\epsilon\eta_0)^2\omega/k_0^2), \tag{4.10}$$

to leading order, where $V = -\sigma\epsilon(T + 2\bar{\xi}/\sigma\omega)$ is the pulse velocity, $\gamma = (1 - V^2/c^2)^{-1/2} \simeq 1 + V^2/2$, and Ω is the frequency of the pulse in its own reference frame. Note that (4.10) is simply the square root of (2.11) to the first approximation. In particular, $\bar{\omega}(X_p, T_p) = \omega$, where $T_p = -2\bar{\xi}/\sigma\omega$. Thus X_p, the penetration distance, is that point where the natural frequency in the medium $n(X) - 2\omega/k_0^2(\epsilon\eta_0)^2$ has increased to ω, the external frequency or the frequency seen by an observer in the fixed frame. It is to be expected that if the barrier height $n^2(X)$ should decrease before X_p, then the soliton would propagate through the barrier in a lossless fashion. In the following section, we demonstrate this to be the case for the example $n^2 = \omega^2(1 + 2\sigma Y - \sigma_1 Y^2)$.

Note also that the local wavenumber $K = \epsilon\theta_X$ is given by

$$K = -\epsilon\sigma\omega(T + 2\bar{\xi}/\sigma\omega) = \omega V, \tag{4.11}$$

which changes from plus to minus as $(T + 2\bar{\xi}/\sigma\omega)$ increases through zero. By taking $k = -(\omega/c)(-2\sigma Y)^{1/2}$ in (3.14), we may verify that the reflected pulse is exactly the same as the incoming one with the velocity reversed.

On the other hand, suppose the barrier $n^2(X)$ were to continue to increase along DE (see Fig. 1) after X_p to

some point, L say, and then decrease suddenly to a value n_2^2 below ω^2. Although one would not obtain lossless transmission of the original soliton, there would be energy available in the region $X > L$ for the creation of further wavepackets due to the effect of the leading edge of the soliton penetrating into the free wave zone. Indeed the stimulus at $X = L$ would be given by $u(L, T)$, and would give rise to solitons (under certain conditions) and phonons in the region $X > L$. If L were large with respect to X_p, then the shape of the pulse would be approximately constant in amplitude $\epsilon(2/\beta)^{1/2} \cdot (2\eta_0/k_0) \exp[-2\eta_0(\omega/k_0)(L - X_p)]$ with the phase $\exp[-i\sigma\omega L(T + 2\bar{\xi}/\sigma\omega)]$ for the duration of the turnaround time $T + 2\bar{\xi}/\sigma\omega = [(2/\sigma\epsilon)(L - X_p)]^{1/2}$. Note in particular that the rate of exponential decay in the amplitude of $u(L, T)$ depends on the amplitude of the incoming pulse in contrast to the situation in normal linear tunnelling. Thus for $X > L$ [where $n = n_2$, $c = c_2$, $k_2 = (1/c_2^2)(\omega^2 - n_2^2)^{1/2}$]

$$u(X, T) = \epsilon\left(\frac{2}{\beta}\right)^{1/2} \cdot \frac{1}{c_2 k_2} D\left(T - \frac{\omega}{c_2^2 k_2}(X - L), \frac{1}{2}\frac{n_2^2}{c_2^4 k_2^3}(Y - \epsilon L)\right)$$

and D satisfies $D_y - iD_{\tau\tau} - 2iD^2D^* = 0$. To find the decomposition of the initial pulse $u(L, T)$ in this region, we solve the Zakharov–Shabat eigenvalue problem (see Ref. 4)

$$V_{1\tau} + i\hat{\xi}V_1 = Qe^{-i\kappa\tau}V_2, \quad V_{2\tau} - i\hat{\xi}V_2 = -Qe^{i\kappa\tau}V_1,$$

which is equivalent to ($V_1 = \phi_1 \exp(-i\kappa\tau/2)$, $V_2 = \phi_2 \exp(i\kappa\tau/2)$, $\zeta = \hat{\xi} - \kappa/2$)

$$\phi_{1\tau} + i\zeta\phi_1 = Q\phi_2, \quad \phi_{2\tau} - i\zeta\phi_2 = -Q\phi_1,$$

where $Q = 0$, $|\tau| > \tau_2 = [(2/\sigma\epsilon)(L - X_p)]^{1/2}$, $Q = 2\eta_0 c_2 k_2/c_0 \cdot \exp[-2\eta_0(\omega/k_0)(L - X_p)]$, $|\tau| < [(2/\sigma\epsilon)(L - X_p)]^{1/2}$. The condition for soliton creation is that $Q\tau_2 > \pi/2$.

Finally, we also verify that with the choices (4.6) and $\bar{\phi}$ the oscillatory phases of the solutions (3.13) and (4.3) match. The oscillatory phase of (3.13) is (leaving out the factor i)

$$-\frac{\omega T}{\epsilon} - \frac{1}{\epsilon^2}\frac{\omega}{3\sigma}(-2\sigma Y)^{3/2} - 2\eta_0^2\frac{\omega}{\sigma k_0^2}(-2\sigma Y)^{1/2} + \text{const.} \tag{4.12}$$

Using (4.4), (4.5), and (4.6), the oscillatory phase of (4.3) is

$$-\frac{\omega T}{\epsilon} - \tfrac{1}{6}\sigma^2\omega\epsilon\left(-\frac{1}{\sigma^3\epsilon^3}(-2\sigma Y)^{3/2} + 3(\tau + C)\frac{1}{\sigma^2\epsilon^2}(-2\sigma Y) - 3(\tau + C)^2\frac{1}{\sigma\epsilon}(-2\sigma Y)^{1/2} + (\tau + C)^3\right) - \frac{\sigma\omega Y}{\epsilon}(\tau + C) - \frac{\omega}{2\epsilon^2\sigma}(-2\sigma Y)^{3/2} + 2\eta_0^2\frac{\omega}{k_0^2}\epsilon\left(\tau + \tau_0 - \frac{1}{\sigma\omega}(-2\sigma Y)^{1/2}\right) + \bar{\phi}. \tag{4.13}$$

Add the second and seventh terms and obtain $-(\omega/3\epsilon^2\sigma)(-2\sigma Y)^{3/2}$ which matches with the second term of (4.12). The third and sixth terms in (4.13) cancel. The terms proportional to the amplitudes match to leading order. The constant terms are matched to $\bar{\phi}$. The remaining terms may be neglected to the order of the approximation used.

5. A PARABOLIC BARRIER

We now examine the case (the branch DF, Fig. 1) when

$$n^2 = \omega^2(1 + 2\sigma Y - \sigma_1 Y^2) = \omega^2\left[1 + \frac{\sigma^2}{\sigma_1} - \sigma_1\left(Y - \frac{\sigma}{\sigma_1}\right)^2\right], \tag{5.1}$$

where both σ, σ_1 are small and σ^2/σ_1 is also. In this case the amplitude equation is ($c^2 = 1$)

$$-2i\omega A_T - \epsilon A_{XX} + \epsilon A_{TT} = -2\sigma\omega^2 XA - 2(-\sigma_1\omega^2/2)X^2A + \beta\epsilon\omega^2A^2A^*, \tag{5.2}$$

which under the transformation $S = \epsilon T/2\omega$, $A = [(2/\beta)^{1/2}/\omega]B$ becomes

$$B_S - iB_{XX} - 2iB^2B^* = -2i\alpha XB - 2i\bar{\alpha}X^2B - i(\epsilon/2\omega)^2B_{SS}. \tag{5.3}$$

with $\alpha = \sigma\omega^2/\epsilon$, $\bar{\alpha} = -\sigma_1\omega^2/2$. The following change of variables,

$$x = X + a(S), \quad t = S, \quad B = C(x, t)\exp[iXE + i(aE + F)], \tag{5.4}$$

with a, E, and F functions of S (or t) reduces (5.3) to

$$C_t - iC_{xx} - 2iC^2C^* = -2i\bar{\alpha}x^2C \tag{5.5}$$

[here we have dropped the $O(\epsilon^2)$ terms in (5.3)], provided

$$a' + 2E = 0, \quad (F + aE)' = -E^2 + 2\bar{\alpha}a^2, \quad E' = -2\alpha + 4\bar{\alpha}a. \tag{5.6}$$

Note that if $\bar{\alpha}$ (or σ_1) $= 0$, $E = -2\alpha S = -\sigma\omega T$, $a = 2\alpha S^2 = \sigma\epsilon T^2/2$, and $F + aE = -\frac{4}{3}\alpha^2S^3 = -\frac{1}{6}\sigma^2\omega\epsilon T^3$, and (5.5) is the canonical NLS equation which admits an exact solution. The fact that (5.3) with the gradient term proportional to X (the first term on the right-hand side) is integrable was first noted by Chen and Liu[7] using the inverse scattering transform method directly. The transformation used above is a much neater way of achieving the same result. Now we will treat $\bar{\alpha}$ as small (i.e., σ_1 small) and use the perturbation theory result (A4) given in the Appendix. If

$$C(x, t) = 2\hat{\eta}\,\mathrm{sech}2\hat{\eta}(\hat{x} - x) \times \exp(-2i\hat{\xi}x - 4i\int\hat{\xi}^2\,dt + 4i\hat{\eta}^2t + i\hat{\phi}), \tag{5.7}$$

then $\hat{\eta}_t = 0$, $\hat{x}_t = -4\hat{\xi}$, $\hat{\xi}_t = 2\bar{\alpha}\hat{x}$. Thus

$$\hat{x} = \kappa_1\exp[\epsilon\sigma_1^{1/2}(T - \tau_0)] + \kappa_2\exp[-\epsilon\sigma_1^{1/2}(T - \tau_0)], \tag{5.8}$$

where the time T is measured from τ_0 the approximate

time of arrival of the pulse at $X=0$. But, $x=X+a(S)$ and

$$a(S)=\frac{\alpha}{2\bar{\alpha}}[1-\cosh 2\omega\sigma_1^{1/2}(S-\bar{S})]$$

$$=\frac{\sigma}{\epsilon\sigma_1}[\cosh\sigma_1^{1/2}\epsilon(T-\tau_0)-1] \tag{5.9}$$

(simply note that $a''=-2E'=8\bar{\alpha}a+4\alpha$) which in the small time limit $S-\bar{S}\to 0$ (or for small σ_1) reduces to $2\alpha(S-\bar{S})^2$ which to leading order is the $\frac{1}{2}\sigma\epsilon(T+2\bar{\xi}/\sigma\omega)^2$ term in (4.3). Thus the phase of the hyperbolic secant is

$$2\hat{\eta}(\hat{x}(T)-a(T)-X) \tag{5.10}$$

and the condition that the pulse will tunnel through the barrier is that the coefficient of the exponential with positive argument is positive. This coefficient is $\kappa_1+\alpha/4\bar{\alpha}=\kappa_1-\sigma/2\epsilon\sigma_1$. It remains to determine κ_1 which we do by matching the constant term and term proportional to $T-\tau_0$ in the phase (5.10) to the constant term and term proportional to $(T-\tau_0)$ in the phase X_p $-\frac{1}{2}\sigma\epsilon(T-\tau_0+\tau_0+2\bar{\xi}/\sigma\omega)^2-X$. We find

$$\kappa_1+\kappa_2=X_p,\quad \kappa_1-\kappa_2=(\sigma/\sigma_1^{1/2})C, \tag{5.11}$$

where C is the order one constant of (4.6). However, we have assumed σ^2/σ_1 to be small and so $\kappa_1=\kappa_2=\frac{1}{2}X_p$. Thus the lossless tunnelling condition is that

$$X_p>\sigma/\epsilon\sigma_1$$

or that

$$Y_p=\epsilon X_p>\sigma/\sigma_1$$

namely, the penetration depth $Y_p=2(\epsilon\eta_0)^2/\sigma k_0^2$ is further into the material than the point $Y=\sigma/\sigma_1$ at which the potential barrier is maximum. If $X_p<\sigma/\epsilon\sigma_1$, then we expect the far side of the barrier to be influenced weakly by the leading edge of the soliton.

6. CONCLUSION

Our simple model has demonstrated that nonlinear tunnelling is very different from linear tunnelling. It has a broad range of potential application. For example, the supplementary heating of a Tokamak plasma from an r.f. source requires a fast electromagnetic wave (whistler mode) to propagate across an evanescent zone at the plasma edge and, at least for high temperatures, it may also have to tunnel across a second evanescent zone before it reaches the region of the lower hybrid resonance.[8] It would be of great value if one could design the incident pulse in such a way so as to achieve, if not total, then at least partial penetration of the barrier region. One may be assured that if the pulse is at all nonlinear and confined to a narrow spectrum (and in the plasma heating problem there are solid reasons for both these assumptions), then the NLS equation and its modifications obtain and the analysis of this paper applies. Indeed the canonical character of the NLS equation [it applies to those situations which are (a) weakly nonlinear, (b) almost monochromatic (c) strongly dispersive, (d) one-dimensional] ensures the broad applicability of the nonlinear tunnelling ideas. It is worth stressing that the fact that the NLS equation is almost integrable is important for it allows us to verify that the pulse is only slowly modulated as it traverses the region of gradual field gradients and that it is not totally scattered. The excitation of other modes (for example, radiation) can be computed (see Ref. 5). On the other hand, we do not yet know how much effect the lifting of the "almost integrability" condition might have. There are several examples of solitary pulses which are solutions of nonintegrable systems (such as the ϕ^4 model of field theory) which display a remarkable resilience (scatter only slightly) even in the presence of strong perturbations.

However, there are situations for which the amplitude limitation can be lifted with confidence; for example, the class of problems (see Refs. 1, 9) for which the sine–Gordon equation obtains. It is reasonable to expect that breathers of all amplitudes can tunnel. In the superconductivity context (the Josephson junction), the breather carries units of the first magnetic moment $\int_{-\infty}^{\infty}x\,\mathbf{B}\cdot\hat{y}dx$ (here x is the direction of propagation, z the direction of current flow across the junction, and $\hat{y}$ the unit vector in the perpendicular direction) into the insulator. The breather has zero magnetic charge as $\mathbf{B}\cdot\hat{y}\propto\partial u/\partial x$ and $u(X,T)\to 0$ as $X\to\pm\infty$. Nevertheless, once inside the insulator, the breather can be pumped (see Ref. 5) to the energy required to cause it to split into a fluxon–antifluxon pair. The frequency of the breather decreases to zero (and therefore its tunnelling capability increases) with increasing amplitude (or energy).

There are several features of the analysis which deserve further comment. First is the fact that, once the soliton is formed, its amplitude undergoes only a weak change as the caustic is approached. This would suggest that the groups of "giant waves" sometimes observed in various parts of the ocean are less likely to be the result of adverse currents (which cause a caustic to occur) as suggested by Smith[10] (see also Peregrine and Thomas[11]) and more likely to be simply the result of the focussing of a wavetrain into soliton pulses. (These can occur provided there is some waveguide action to suppress their instability to disturbances which depend on the perpendicular coordinate.) With minor modifications, the previous analysis will also apply to this situation.

A second mathematical feature of considerable interest is the fact that although the NLS equation is canonical it fails, to the order of the approximation involved, to account for amplitude dependence in the velocity of the pulse amplitude and after all, it is the pulse amplitude velocity which is the nonlinear analog of group velocity and therefore the velocity of energy propagation. Because of this, it may be necessary in a variety of situations other than the one considered in this paper to include the $i\delta\epsilon q_{xt}$ term in order to model more faithfully certain propagation characteristics of the system being approximated.

Many open problems remain. In particular it would be useful to analyze the potential tunnelling of nonlinear pulses through sharp or discontinuous gradients in the field parameters. In the absence of analytical machinery, a numerical investigation may prove to be the

most fruitful approach. One possibility is to study the propagation of lattice waves (say for the Toda lattice) through regions of heavy masses. It is conjectured that the fundamental rule expressed by (1.11) holds; namely, if the excitation frequency is greater than the nonlinearly adjusted natural frequency, then the pulse will tunnel in a lossless manner. It is also natural to ask what becomes of the radiation (the phonon modes). We might conjecture that for sufficiently steep field gradients, the WKB enhancement of a wavetrain could trigger the appearance of a secondary soliton pulse (in the language of the inverse scattering transform, a discrete eigenvalue can cross from the lower half complex wavenumber plane where it represents radiation to the upper half where it represents a soliton as the area criterion[4] is satisfied). On the other hand for gradual field gradients, it is suggested that in order to describe the structure of the radiation the complex equation of Painlevé type

$$A_{XX} - (2\sigma\omega^2/\epsilon)XA + \beta n^2 A^2 A^* = 0, \tag{6.1}$$

obtained by ignoring the T dependence in (4.2), may be appropriate. Indeed for sufficiently large amplitudes one can expect to come arbitrarily close to the singular solution of this equation. Also, in parallel with the ideas of Ablowitz and Segur,[12] one can analyze the solutions of (4.2) by converting to the canonical NLS equation, and then analyzing the solutions with the exact theory (the Marcenko equations). The analog to the Airy function (which, using the exact modified Korteweg–de Vries equation theory, generates the solutions of the Painlevé equation of the second kind) is the similarity solution of the NLS equation (see Benney and Newell[13]). Some amplitude enhancement may be expected to occur for the phonon modes.

APPENDIX

These results are derived in Ref. 5. Let the leading order solution of

$$q_t - iq_{xx} - 2iq^2q^* = F, \quad F \text{ small}, \tag{A1}$$

be

$$q(x,t) = 2\eta \operatorname{sech} 2\eta(x - \bar{x}) \exp[-2i\xi x - 4i\int(\xi^2 - \eta^2)\,dt]. \tag{A2}$$

If $F = -\Gamma(t)q$, then

$$\eta_t = -2\Gamma(t)\eta, \quad \xi_t = 0, \bar{x}_t = -4\xi + O(\epsilon). \tag{A3}$$

If

$$F = -2i\bar{\alpha}x^2 q,$$

then

$$\eta_t = 0, \quad \xi_t = 2\bar{\alpha}\bar{x}, \quad \bar{x}_t = -4\xi. \tag{A4}$$

[1] A. Barone, F. Esposito, C. J. Magee, and A. C. Scott, "Theory and Applications of the sine-Gordon equation," Nuovo Cimento 1, 227 (1971).
[2] G. B. Whitham, *Linear and Nonlinear Waves* (Wiley, New York, 1976).
[3] T. B. Benjamin and J. E. Feir, "The Disintegration of Wavetrains in Deep Water Pt. I.," J. Fluid Mech. 27, 417 (1966).
[4] M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, "Method for Solving the sine-Gordon equation," Phys. Rev. Lett. 30, 1262 (1973).
[5] D. J. Kaup and A. C. Newell, "Solitons as Particles and Oscillators," Proc. Soc. Lond. Ser. A (to be published).
[6] L. Solymar, *Superconductive Tunnelling and Applications* (Wiley-Interscience, New York, 1972).
[7] H. H. Chen and C. S. Liu, "Solitons in Nonuniform Media," Phys. Rev. Lett. 37, 693 (1976).
[8] A. Bers, C. Karney, and K. Theilhaber, "Whistler Wave Excitation and its Parametric Down-Conversion to Electrostatic Ion Cyclotron Waves," RLE Progress Report #115 (Mass. Inst. of Technology, January, 1975).
[9] A. C. Scott, "Propagation of Magnetic Flux on a Long Josephson Tunnel Junction," Nuovo Cimento, Ser. X, 69, 241 (1970).
[10] R. Smith, "Giant Waves," J. Fluid Mech. 77, 417 (1976).
[11] D. H. Peregrine and G. P. Thomas, "Finite Amplitude Waves on Nonuniform Currents," Proc. IUTAM Symposium on Surface Gravity Waves on Water of Varying Depth, Canberra, July 1976.
[12] M. J. Ablowitz and H. Segur, "Exact Linearization of Painlevé Transcendents," Phys. Rev. Lett. 38, 1103 (1977).
[13] D. J. Benney and A. C. Newell, "The Propagation of Long Wave Envelopes," J. Math. Phys. (presently Stud. Appl. Math.) 46, 133 (1967).

Cluster expansion for lattice gauge theories with fermions

J. L. Challifour and D. Weingarten[a)]

Physics Department, Indiana University, Bloomington, Indiana 47401

A cluster expansion is constructed for Euclidean lattice gauge theories including fermions. For sufficiently small values of the bare qauge coupling constant g^{-2} and the fermion propagation constant k, we prove the existence of gauge invariant infinite volume vacuum expectation values, convergence of Wilson's strong coupling expansion, and an inequality related to quark confinement.

1. INTRODUCTION

Osterwalder and Seiler[1] have obtained a number of rigorous results for pure gauge theories on a Euclidean lattice[2] by using the cluster expansion[3], now a familiar tool of constructive field theory. In the present article we will extend their results to gauge theories including fermions.

Beginning with an arbitrary compact gauge group G on a finite subset Λ of the sites of a d-dimensional lattice we will prove:

(a) If g^{-2} and k are sufficiently small, the limit $\Lambda \to \infty$ exists (g is the bare gauge coupling constant and k is Wilson's[2] fermion propagation constant).

(b) The limiting theory possesses a unique vacuum and a mass gap. The vacuum is both gauge invariant and invariant under lattice translations.

(c) The vacuum expectation values of the infinite volume theory are analytic in complex g^{-2} and k at the origin. Thus Wilson's strong coupling expansion converges.

(d) A class of states in the infinite volume theory with $G=\mathrm{SU}(N)$ and $d=2$, obey an inequality related to quark confinement.[2]

For simplicity we will consider only one flavor of quark; our results continue to hold, however, for any finite number of flavors.

2. CLUSTER EXPANSION

Let Z^d be an infinite d-dimensional lattice with sites s and a finite subset Λ. $L(\Lambda)$ denotes the oriented nearest neighbor links (bonds) and $P(\Lambda)$ the oriented plaquettes in Λ. R stands for a nontrivial unitary irreduceable representation of the gauge group G. To each link $l=(s_1, s_2)\in L(\Lambda)$ we assign $U(l)\in R$ with $U(s_1, s_2)=U(s_2, s_1)^*$ and to each site s assign a collection $\Psi_{i\alpha}(s)$, $\bar{\Psi}_{j\beta}(s)$ of Grassman variables where i, j are Euclidean spinor indices and α, β indices for R and its complex conjugate, respectively. $\{\gamma_\mu\}$ will be a set of Hermitian γ matrices with $\{\gamma_\mu, \gamma_\nu\}=2\delta_{\mu\nu}$. If $\hat{\mu}$ is a unit lattice vector in the positive μ direction, for the links $(s, s\pm\hat{\mu})$ define $\Gamma(s, s\pm\hat{\mu})=(1\mp\gamma_\mu)/2$. We will use $\int d\mu_F(\cdot)$ for "fermion integration" over all Grassman variables on the sites of Z^d incorporating the term $\exp[-\sum_{s\in Z^d}\bar{\Psi}(s)\Psi(s)]$ in addition to the usual expression.[2,4] The gauge field measure μ_G consists of a product of copies of Haar measure on G for each independent $U(l)$, $l\in L(Z^d)$.

The action $A(\Lambda)$ is given by

$$A(\Lambda)=k\sum_{(s_1,s_2)\in L(\Lambda)}\bar{\Psi}(s_1)\Gamma(s_1,s_2)U(s_1,s_2)\bar{\Psi}(s_2) + g^{-2}\sum_{(s_1,\ldots,s_4)\in P(\Lambda)}\mathrm{Tr}[U(s_1,s_2)U(s_2,s_3)U(s_3,s_4)\times U(s_4,s_1)]. \quad (2.1)$$

For any $\mathcal{J}$ which is a continuous function of a finite number of gauge variables multiplied by a finite product of Grassman variables or a sum of such terms

$$\langle\mathcal{J}\rangle_\Lambda = Z(\Lambda)^{-1}\int d\mu_G\int d\mu_F\,\mathcal{J}\exp A(\Lambda), \quad (2.2)$$

where $Z(\Lambda)$ is the partition function defined by the normalization $\langle 1\rangle_\Lambda=1$. The cluster expansion of $\langle\mathcal{J}\rangle_\Lambda$ is obtained by making replacements in (2.2) in favor of the variables

$$\rho_p=\exp\{g^{-2}\,\mathrm{Tr}[U(s_1,s_2)U(s_2,s_3)U(s_3,s_4)U(s_4,s_1)]\}-1, \quad p\in P(\Lambda), \quad (2.3)$$

$$\rho_m=\exp[k\bar{\Psi}_{i\alpha}\Gamma_{ij}(s_1,s_2)U_{\alpha\beta}(s_1,s_2)\bar{\Psi}_{j\beta}(s_2)]-1, \quad m\in M(\Lambda), \quad (2.4)$$

where $M(\Lambda)$ denotes all $m=(l,\gamma)$ for which $l\in L(\Lambda)$ and γ is a sequence of indices $\gamma=(i,\alpha,j,\beta)$. Combining (2.1)–(2.4) and resuming over sets of links and plaquettes $Q\subset M(\Lambda)\cup P(\Lambda)$ which are not connected to the support of $\mathcal{J}$, $S_{\mathcal{J}}$ leads to[3,1]

$$\langle\mathcal{J}\rangle_\Lambda=\sum_{Q'}\int d\mu_G\int d\mu_F\,\mathcal{J}\prod_{r\in Q'}\rho_r\{Z[\Lambda-(S_{Q'}\cup S_{\mathcal{J}})]/Z(\Lambda)\}. \quad (2.5)$$

The sum is restricted to sets $Q'\subset M(\Lambda)\cup P(\Lambda)$ which are connected to $S_{\mathcal{J}}$ and we use $S_{Q'}$ to denote sites appearing in links or plaquettes of Q'.

Define $[q_F; q_G]$ to be the set of Q' for which the number of elements $|Q'\cap M(\Lambda)|=q_F$, $|Q'\cap P(\Lambda)|=q_G$. Then the proof of (a)–(d) relies on the following bound for the convergence of (2.5) for small $|g^{-2}|$ and $|k|$:

$$\sum_{Q'\in[q_F;\,w_G]}\left|\int d\mu_G\int d\mu_F\,\mathcal{J}\prod_{r\in Q'}\rho_r\{Z[\Lambda-S_{Q'}\cup S_{\mathcal{J}}]/Z(\Lambda)\}\right| \le 2^{c_1|S_{\mathcal{J}}|}\|\mathcal{J}\|\,|c_2k|^{q_F}|c_3g^{-2}|^{q_G}. \quad (2.6)$$

The norm $\|\mathcal{J}\|$ is the sum of the suprema of the gauge parts of $\mathcal{J}$. The constants c_1 and c_2 depend on d, and

a)Work supported in part by the U.S. Department of Energy.

 0022-2488/78/1905-1134$1.00

c_3 depends on d, G and R but none depend on $\mathcal{J}$ or Λ. To obtain this bound notice first that for any finite set M

$$\int d\mu_F \prod_{(s_1,s_2;\, i,\alpha,\, j,\beta)\in M} \Psi_{i\alpha}(s_1)\Psi_{j\beta}(s_2) = \pm 1, 0.$$

In addition, $|\Gamma_{ij}(s_1,s_2)|$, $|U_{\alpha\beta}(s_1,s_2)| \le 1$. When $|g^{-2}|$ is small enough, $|\rho_p| \le b_1|g^{-2}|$, so that for fixed Q'

$$\left|\int d\mu_G \int d\mu_F \mathcal{J} \prod_{r\in Q'} \rho_r\right| \le \|\mathcal{J}\|\, |k|^{q_F} |b_1 g^{-2}|^{q_G}. \tag{2.7}$$

Next, an estimate of the number of Q' for fixed q_F and q_G can be obtained as in Ref. 3. We find

$$|[q_F; q_G]| \le 2^{|T_{\mathcal{J}}|} b_2^{q_F+q_G},$$

where $T_{\mathcal{J}}$ is the set of points corresponding to links or plaquettes bordering S and b_2 is a function of lattice geometry. For another constant b_3 determined by lattice geometry we have $|T_{\mathcal{J}}| \le b_3|S_{\mathcal{J}}|$. Finally, $S_{Q'}$ obeys $|S_{Q'}| \le 2q_F + 4q_G$, and following Ref. 1 we obtain the estimate $2^{-|S|} \le |Z(\Lambda - S)/Z(\Lambda)| \le 2^{|S|}$. These bounds combined with (2.7) and (2.8) lead directly to (2.6).

The methods developed in Refs. 1, 3 and 5 applied to (2.5) and (2.6) yield exponential clustering

$$|\langle \mathcal{J}_v \mathcal{G}\rangle_\Lambda - \langle \mathcal{J}_v\rangle_\Lambda \langle \mathcal{G}\rangle_\Lambda| \le B(\mathcal{J},\mathcal{G}) \exp(-m|v|), \tag{2.9}$$

where $\mathcal{J}_v$ is the displacement of $\mathcal{J}$ by a lattice vector v, and (2.9) then gives results (a)–(c).

If Λ is chosen to be $\{x \mid |x_i| \le M_i\}$, cyclic boundary conditions in Ref. 6 yield an expectation $\langle\cdots\rangle'_\Lambda$ slightly different from the one discussed so far. By a simple adaptation of the procedure used in Refs. 1 and 3 to prove the existence of the infinite volume limit of $\langle\cdots\rangle_\Lambda$, however, one can also show that, as $\Lambda \to Z^d$, $\langle\mathcal{J}\rangle'_\Lambda \to \langle\mathcal{J}\rangle_\Lambda$ uniformly in complex k and g^{-2} in a neighborhood of the origin. Thus $\lim \langle\mathcal{J}\rangle'_\Lambda$ also exists and is identical to $\lim\langle\mathcal{J}\rangle_\Lambda$.

3. QUARK CONFINEMENT

For the expectation with cyclic boundary conditions $\langle\cdots\rangle'_\Lambda$ with $d=4$ and $G=$SU(N), properties of the physical Hilbert space and transfer matrix have been obtained by Lüscher[6]. An obvious extension of Lüscher's work can be carried through for arbitrary d. Using the results in Ref. 6, combined with our propositions (a) and (b), we can define, in the standard way, a physical Hilbert space H and transfer matrix T for the infinite-volume theory. A larger space $\hat{H} \supset H$ will also be needed. In the gauge with $U(s, s\pm\hat{\mu}) = 1$, where μ is the time direction, a basis for H can be represented by gauge invariant polynomials in the variables $\Psi(s)$, $\bar{\Psi}(s)$, $U(s,t)$ with time components s^μ, $t^\mu \ge 0$ while a basis for $\hat{H}$ can be represented by polynomials in nonnegative time variables without the restriction of gauge invariance. T is interpreted as $\exp(-H)$ for the lattice Hamiltonian H.

Quarks are confined if none of the states in H can be interpreted as including one or more quarks moving freely separated by a large distance from all other physical particles in the state. We will now prove an inequality which, for $d=2$, suggests that quarks are confined but is weaker than a full proof of confinement. For $d>2$, the inequality continues to hold but is not directly related to confinement. The proof will be given for arbitrary d.

Let $\mathcal{J}$ be a polynomial in zero-time variables representing a nonnull vector in H. Choose a fermion in $\mathcal{J}$ and move it progressively farther away from the remaining sites in $\mathcal{J}$ introducing in the process additional $U(s,t)$ needed to maintain gauge invariance. In particular, suppose $\mathcal{J}$ has the form $\sum_\alpha \bar{\Psi}_{i\alpha}(s)\mathcal{J}_\alpha$, and let $\mathcal{J}(n)$ be

$$\mathcal{J}(n) = \sum_{\alpha_1,\cdots,\alpha_n} \bar{\Psi}_{i\alpha_n}(s_n) \times U_{\alpha_n\alpha_{n-1}}(s_n, s_{n-1})\cdots U_{\alpha_1\alpha}(s_1, s)\mathcal{J}_\alpha, \tag{3.1}$$

with all the sites $s_1, \cdots, s_n$ distinct from each other and from the sites in $\mathcal{J}_\alpha$ and all at time zero. An expression similar to $\mathcal{J}(n)$ with a $\Psi_{i\alpha}$ instead of $\bar{\Psi}_{i\alpha}$ moved away from $\mathcal{J}_\alpha$ can, of course, also be treated by the methods used here, as can more general connections between s_n and s. We require $\sum_\alpha(\mathcal{J}_\alpha, \mathcal{J}_\alpha) > 0$, where $(\cdots,\cdots)$ is the inner product of H. Let $P(E)$ be the projection onto the subspace with energy less than E. We will prove

$$(\mathcal{J}(n), P(E)\mathcal{J}(n)) \le C(\mathcal{J}(n), \mathcal{J}(n)) \exp(-\lambda n) \tag{3.2}$$

for a constant C independent of n and a constant λ independent of $\mathcal{J}$ and E for values of k and g^{-2} in the strong coupling regime. In other words we will show that the total probability that a normalized state proportional to $\mathcal{J}(n)$ happens to have energy less than E falls at least exponentially with n.

For $d=2$, there is only one possible $\mathcal{J}(n)$ in (3.1) with $s_1 \cdots s_n$ all distinct from each other, and using the condition $U(s,s') = U(s',s)^*$ it is not hard to show that this $\mathcal{J}(n)$ incorporates the only possible gauge field configuration combining $\bar{\Psi}_{i\beta}(s_n)$ and $\mathcal{J}_\alpha$ into a gauge invariant state. For $d>2$, on the other hand, an infinite number of different states can be found in (3.1) and quark confinement requires at least a proof of (3.2) for a state in (3.1) and quark confinement requires at least a proof of (3.2) for a state given by an arbitrary superposition of such $\mathcal{J}(n)$. It is precisely by forming such superpositions in an unconfined theory with $d>2$ that one would expect to be able to separate out a quark without expending large amounts of energy.

To obtain (3.2) it is sufficient to show

$$(\mathcal{J}(n), T\mathcal{J}(n)) \le C'(\mathcal{J}(n), \mathcal{J}(n)) \exp(-\lambda n). \tag{3.3}$$

By definition of T and the physical inner product

$$(\mathcal{J}(n), T\mathcal{J}(n)) = \langle \mathcal{J}(n)^c \mathcal{J}_1(n)\rangle, \tag{3.4}$$

where $\mathcal{J}_1(n)$ is $\mathcal{J}(n)$ displaced from $s^\mu = 0$ to $s^\mu = 1$ and $(\cdots)^c$ denotes a conjugation[6]. An upper bound on (3.4) results by applying the cluster expansion and the Peter–Weyl theorem[7] in the form $\int d\mu_G U_{\alpha\beta}(s,t) = 0$ for all $\alpha, \beta, (s,t)$ and noting that only those Q' arise in (2.5) for which $q_F + q_G \ge n - c$. Using (2.6) with $\|\mathcal{J}^c(n)\mathcal{J}_1(n)\| \le \|\mathcal{J}\|^2$ leads to (3.3) without a factor of $(\mathcal{J}(n), \mathcal{J}(n))$ on the right side.

To complete (3.3) a lower bound on $(\mathcal{J}(n), \mathcal{J}(n))$ is needed. Let $\mathcal{J}_0(n)$ be the polynomial obtained from $\mathcal{J}(n)$ by replacing fermion variables $\psi(s)$, $\bar{\psi}(s)$ by Fock space operators $\hat{\Psi}(s)$, $\hat{\Psi}^+(s)\gamma_0$ with the anticommutation relations

$$\{\Psi_{i\alpha}(s),\ \hat{\Psi}_{j\beta}(t)\}=\{\hat{\Psi}^{\dagger}_{i\alpha},\ \hat{\Psi}_{j\beta}(t)\}=0,$$
$$\{\Psi_{i\alpha}(s),\ \Psi^{\dagger}_{j\beta}(t)\}=\delta_{ij}(B^{-1})_{s\alpha t\beta}, \tag{3.5}$$

where

$$B_{s\alpha t\beta}=\delta_{st}\delta_{\alpha\beta}-\tfrac{1}{2}kU(s,t)_{\alpha\beta}\sum_{\mu}(\delta_{s+\hat{\mu},t}+\delta_{s-\hat{\mu},t}). \tag{3.6}$$

Then Lüscher[6] gives

$$(\mathcal{J}(n),\mathcal{J}(n))_{\Lambda}$$
$$=\mathrm{Tr}\{T_{\Lambda}^{2M+1}N[\mathcal{J}_0(n)]^{*}N[\mathcal{J}_0(n)]\}/\mathrm{Tr}T_{\Lambda}^{2M+1}, \tag{3.7}$$

where $(\ldots,\ldots)_{\Lambda}$ is the inner product and T_{Λ} the transfer matrix for the theory with cyclic boundary conditions in the volume Λ having $2M+1$ sites in the time direction. $N(\ldots)$ is a normal ordering. Proposition (2) of Ref. 6 implies

$$(\mathcal{J}(n),\mathcal{J}(n))_{\Lambda}=\sum_{\text{contractions}}\langle\mathcal{J}(\overbrace{n)^{c}\mathcal{J}(n}^{})\rangle_{\Lambda}, \tag{3.8}$$

where fermions in $\mathcal{J}(n)^{c}$ are contracted with fermions in $\mathcal{J}(n)$ according to Wick's theorem.

The desired lower bound for (3.8) can be found by considering first those terms in which fermions of $\mathcal{J}^{c}_{\alpha}$ or $\mathcal{J}_{\alpha}$ are not contracted with any at s_n. For each of these terms, carry out a cluster expansion and select Q' for which there exists a site s_j in (3.1) such that Q' does not connect any s_k, $k\geq j$, to any s_l, $l<j$, or to a site in the support of $\mathcal{J}_{\alpha}$. For a Q' of this sort, sum all the terms in (2.5) differing from the original only by the choice of gauge indices in fermion members ρ_m of Q'. The result has the form

$$X=Y\sum_{\beta\beta'}A_{\beta\beta'}B_{\beta\beta'}, \tag{3.9}$$

where $A_{\beta\beta'}$ arises from components of Q' connected to s_k, $k\geq j$, $B_{\beta\beta'}$ from components connected to s_l, $l<j$, or to $\mathcal{J}_{\alpha}$, Y is a ratio of partition functions and β,β' are gauge indices across site s_j. Since R is irreducible with dimension D and both $A_{\beta\beta'}$ and $B_{\beta\beta'}$ are gauge invariant, we have $A_{\beta\beta'}=A\delta_{\beta\beta'}$ and $B_{\beta\beta'}=B\delta_{\beta\beta'}$. If these expressions are placed in (3.9), then since each $U(s_m,s_{m+1})$ is unitary X becomes a set of terms contributing to a modified cluster expansion of

$$\sum_{\text{contractions}}D^{-1}\langle\sum_{\alpha,\beta,j}\gamma_{0ij}\Psi_{j\alpha}(\overbrace{s_n)\bar{\Psi}_{i\alpha}(s_n}^{})\overbrace{\mathcal{J}^{c}_{\beta}\mathcal{J}_{\beta}}^{}\rangle_{\Lambda}, \tag{3.10}$$

where the Q' in the cluster sum are connected either to the support of $\mathcal{J}_{\beta}$ or to the sites s_k in (3.1). Let us compare (3.8) and (3.10). The modified cluster expansion of (3.10) duplicates the cluster expansion of (3.8) for all Q' which contain a dividing site except those with contraction between site s_n and $\mathcal{J}_{\beta}$ or $\mathcal{J}^{c}_{\beta}$. For small k an expansion of B^{-1} in (3.5) and (3.6) shows the missing contractions are of $O(k^n)$ while the common Q' include all terms with $q_F+q_G\leq n-1$. Moreover $\|\mathcal{J}^{c}(n)\mathcal{J}(n)\|\leq\|\mathcal{J}\|^2$ and

$$\|\sum_{\alpha,\beta,j}\gamma_{0ij}\psi_{j\alpha}(s_n)\bar{\psi}_{i\alpha}(s_n)\mathcal{J}^{c}_{\beta}\mathcal{J}_{\beta}\|\leq D\|\sum_{\beta}\mathcal{J}^{c}_{\beta}\mathcal{J}_{\beta}\|.$$

Applying (2.6) we obtain a bound of $O[(k+g^{-2})^n]$, uniform in Λ, on the difference between (3.10) and (3.8). Then, using exponential clustering, (2.9), we can approximate (3.10)

$$\sum_{\text{contractions}}D^{-1}\langle\sum_{\alpha j}\gamma_{0ij}\Psi_{j\alpha}\overbrace{(s_n)\bar{\Psi}_{i\alpha}(s_n)}^{}\rangle_{\Lambda}\langle\sum_{\beta}\overbrace{\mathcal{J}^{c}_{\beta}\mathcal{J}_{\beta}}^{}\rangle_{\Lambda}. \tag{3.11}$$

Lüscher's proposition(2)converts (3.11) to $D^{-1}\sum_{\alpha}(\Psi_{i\alpha}(s_n),\bar{\Psi}_{i\alpha}(s_n))\sum_{\beta}(\mathcal{J}_{\beta},\mathcal{J}_{\beta})$ as $\Lambda\to Z^{d}$. Using translational invariance of the infinite-volume theory we obtain the following result. If k and g^{-2} are sufficiently small, for all zero-time polynomials $\mathcal{J}_{\beta}$ in (3.1) and every $\epsilon>0$ there exists an $n(\epsilon)$ such that when $n>n(\epsilon)$

$$\left|(\mathcal{J}(n),\mathcal{J}(n))-D^{-1}\sum_{\alpha}(\Psi_{i\alpha}(s),\Psi_{i\alpha}(s))\sum_{\beta}(\mathcal{J}_{\beta},\mathcal{J}_{\beta})\right|<\epsilon.$$

If $\sum_{\alpha}(\Psi_{i\alpha}(s),\ \bar{\Psi}_{i\alpha}(s))\sum_{\beta}(\mathcal{J}_{\beta},\mathcal{J}_{\beta})>0$ we can conclude (3.3) holds. The first of these factors can always be made greater than zero by choosing k and g^{-2} sufficiently small but greater than zero so that $\sum_{\alpha}(\Psi_{i\alpha}(s_n),\bar{\Psi}_{i\alpha}(s))$ is nearly given by the leading nonzero term in its cluster expansion. The second factor is greater than zero by assumption. This completes the proof of (3.3).

4. CONCLUSION

A number of remarks might be useful concerning the limitations on the class of $\mathcal{J}(n)$ to which (3.2) applies. First of all, it seems likely that the set of vectors represented by polynomials in zero-time variables spans all of $\mathcal{H}$. A similar result has been claimed for $(\Phi^4)_2$[8], while for the finite volume lattice gauge theory it follows from Ref. 6 that the zero-time polynomials span $\mathcal{H}$. Moreover, if we do interpret the zero-time $\Psi(s)$ and $\bar{\Psi}(s)$ as quark field variables, it is not consistent to apply this same interpretation to $\Psi(s)$ and $\bar{\Psi}(s)$ with $s^{\mu}>0$. If $\mathcal{G}$ is a polynomial in $s^{\mu}=n>0$ fields we have $\mathcal{G}=T^{n}\mathcal{G}_0$ for a corresponding $s^{\mu}=0$ polynomial $\mathcal{G}_0$. Then using the existence of a mass gap and the uniqueness of the infinite volume vacuum, it is easily shown that as $n\to\infty$, $T^{n}\mathcal{G}_0$ becomes a multiple of the vacuum.

The restriction $\sum_{\beta}(\mathcal{J}_{\beta},\mathcal{J}_{\beta})>0$ we would expect is fulfilled by all $\mathcal{J}_{\beta}$ arising from nonnull $\mathcal{J}$. If the fermion field $\bar{\Psi}_{i\alpha}(s)$ pulled out of $\mathcal{J}$ is a field variable normal ordered to the left according to Ref. 6, then $(\mathcal{J},\mathcal{J})>0$ implies $\sum_{\beta}(\mathcal{J}_{\beta},\mathcal{J}_{\beta})>0$ by a simple application of the Cauchy–Schwarz inequality. If $\bar{\Psi}_{i\alpha}(s)$ is normal ordered to the right, this argument fails, however for any particular $\mathcal{J}_{\beta}$, $\sum_{\beta}(\mathcal{J}_{\beta})>0$ can still be arranged if $(\mathcal{J},\mathcal{J})>0$ by choosing k and g^{-2} sufficiently close to zero that the leading nonzero term in the cluster expansion of $\sum_{\beta}(\mathcal{J}_{\beta},\mathcal{J}_{\beta})$ dominates.

ACKNOWLEDGMENTS

One of us (D.W) would like to thank L. Susskind for a valuable discussion. While this paper was in press a preprint by K. Gawedzki appeared covering similar results. We thank E. Seiler for bringing this work to our attention.

[1]K. Osterwalder and E. Seiler, Harvard preprint (1977).
[2]K. G. Wilson, Phys. Rev. D 10, 2445 (1974).
[3]J. Glimm, A. Jaffe, and T. Spencer, in *Constructive Quantum Field Theory*, Springer Lecture Notes in Physics, edited by G. Velo and A. S. Wightman (Springer-Verlag, Berlin, 1973), Vol. 25.
[4]F. Berezin, *The Method of Second Quantization* (Academic, New York, 1966).
[5]J. Ginibre, Commun. Math. Phys. 16, 310 (1970).
[6]M. Lüscher, Commun. Math. Phys. 54, 283 (1977); Osterwalder and Seiler, Ref. 1.
[7]E. Hewitt and K.A. Ross, *Abstract Harmonic Analysis* (Springer, New York, 1970), Vol. 2, Sec. 27.40.
[8]S. Albeverio, Bielefield preprint, 1977.

Uniform approximations to integral and integrodifferential equations

Rina Ling

Department of Mathematics, California State University, Los Angeles, California 90032
(Received 3 November 1977)

Integral and intergrodifferential equations containing a small parameter are studied. An approximate solution is obtained and proved to be uniformly valid in t as the parameter tends to zero.

1. INTRODUCTION

Equations of integral and integrodifferential types, which contain a small or large parameter, occur in many physical contexts; see, for example, Refs. 1–4. In most cases, the exact solutions cannot be obtained or are in a complicated form; therefore we must resort to a form of approximation which is uniformly valid in the interval of interest.

In this work, uniformly valid approximations are obtained for certain integral and integrodifferential equations containing a small parameter. The equations contain monotonic increasing or decreasing kernels. References 2, 3 in the past have obtained approximations to equations with particular kernels, and the integral equations studied in Ref. 5 contain a different class of kernels.

In the course of this investigation, theorems from Refs. 5–8 are applied; Refs. 7, 8 are on qualitative behavior of Volterra integral equations.

2. INTEGRAL EQUATIONS

In this section, equations of the form

$$f(t;\epsilon)=1-\int_0^t K(\epsilon(t-\tau))f(\tau;\epsilon)\,d\tau, \tag{2.1}$$

where ϵ is a small positive parameter, are studied. (For simplicity, the argument ϵ will be dropped from the various functions from now on. Also $K*f$ will be used to denote the convolution of K and f.) Let $g(t)$ be the solution to the equation with the kernel pulled out,

$$g(t)=1-K(\epsilon t)\int_0^t g(\tau)\,d\tau. \tag{2.2}$$

It can be shown, under certain conditions on the kernel, that the error $f(t)-g(t)$ tends to zero as $\epsilon\to 0$, uniformly on $[0, T]$, and $f(t)-g(t)\to 0$ as $t\to\infty$. It would be assumed that all involved derivatives of $K(t)$ are continuous.

First note that $g(t)$ can be solved. If, in (2.2), we let

$$G(t)=\int_0^t g(\tau)\,d\tau,$$

then

$$G'(t)+K(\epsilon t)G(t)=1$$

and so

$$G(t)=\int_0^t E(\tau)\,d\tau/E(t), \tag{2.3}$$

where

$$E(t)=\exp[\int_0^t K(\epsilon\tau)\,d\tau];$$

therefore

$$g(t)=1-K(\epsilon t)G(t).$$

Properties of $g(t)$ can be found under certain conditions on the kernel. The analysis used for studying the properties of $g(t)$ is similar to that in Refs. 3, 5.

In the following theorem, uniformity of the approximation $g(t)$ is proved for equations with positive decreasing kernels.

Theorem 2.1: If (1) $K(t)>0$, (2) $K'(t)<0$, (3) $K(\infty)=A\neq 0$, and (4) $K'(t)/K(t)$ is nondecreasing [i.e., $\ln K(t)$ is convex], then (1) $f-g=o(1)$ as $\epsilon\to 0$, uniformly on $[0, T]$ and (2) $f(t)-g(t)\to 0$ as $t\to\infty$ fixed, for ϵ fixed.

Proof: Note first the following properties of $g(t)$.

(1) The fact that $K'(t)<0$ implies that $g(t)$ is always positive.

This can be seen by differentiating (2.2) to obtain

$$g'(t)=-K(\epsilon t)g(t)-\epsilon K'(\epsilon t)G(t);$$

therefore at any zero of g, g' is positive, but $g(0)=1$, so $g(t)>0$ for all t.

(2) $g(t)$ is absolutely integrable. From (2.3),

$$G(\infty)=\lim_{t\to\infty}[1/K(\alpha t)]=1/A.$$

Therefore,

$$\int_0^\infty |g(t)|\,dt=\int_0^\infty g(t)\,dt$$
$$=\int_0^\infty G'(t)\,dt=1/A.$$

The error $f-g$ satisfies an integral equation with the same kernel as that for $f(t)$, but with source term

$$\phi(t)=K(\epsilon t)\int_0^t g(\tau)\,d\tau-\int_0^t K(\epsilon\tau)g(t-\tau)\,d\tau; \tag{2.4}$$

therefore by convolution theorem[6]

$$f(t)-g(t)=\phi(t)+\phi*f'. \tag{2.5}$$

From (2.4),

$$\phi(t)=\int_0^t [K(\epsilon t)-K(\epsilon\tau)]g(t-\tau)\,d\tau,$$

so

$$|\phi(t)|=\int_0^t [K(\epsilon\tau)-K(\epsilon t)]g(t-\tau)\,d\tau$$
$$\leq [K(0)-K(\epsilon t)]\int_0^t g(\tau)\,d\tau$$
$$\leq [K(0)-K(\epsilon T)]\int_0^\infty g(\tau)\,d\tau$$
$$=[K(0)-K(\epsilon T)]\cdot(1/A)$$
$$=o(1)\quad \text{as } \epsilon\to 0, \text{ uniformly on } [0, T].$$

 0022-2488/78/1905-1137$1.00

By Theorem 2 in Ref. 7, $f'(t) \leq 0$ and $0 < f(t) \leq 1$. So, from (2.5),

$$\begin{aligned} |f-g| &\leq o(1)(1+\int_0^t |f'(\tau)\,d\tau) \\ &= o(1)(1+f(0)-f(t)) \\ &\leq o(1)\cdot 2 \\ &= o(1) \quad \text{as } \epsilon \to 0, \text{ uniformly on } [0,T]. \end{aligned}$$

It can be shown that both $f(t)$ and $g(t)$ tend to zero as $t \to \infty$. Since $f(t)$ is monotonic and bounded, $f(\infty)$ exists. By Equivalence Theorem 1.1.1 in Ref. 8, $f(t)$ also satisfies

$$f(t) = \exp(-\gamma t) - L * f,$$

where

$$L(t) = (a-\gamma)\exp(-\gamma t) + \epsilon \int_0^t \exp[-\gamma(t-\tau)]K'(\epsilon\tau)\,d\tau,$$

$a = K(0)$, γ any constant.

Since

$$\int_0^\infty L(t)\,dt = -1 + K(\infty)/\gamma = -1 + A/\gamma,$$

$$f(\infty) = -f(\infty)(-1 + A/\gamma),$$

so that $f(\infty) = 0$. For $g(t)$, the equation is

$$g(t) = 1 - K(\epsilon t)G(t),$$

since

$$\lim_{t\to\infty} K(\epsilon t)G(t) = 1, \quad g(\infty) = 0.$$

Therefore, $f(t) - g(t) \to 0$ as $t \to \infty$, for ϵ fixed.

The next theorem is on equations with positive increasing kernels.

Theorem 2.2: If (1) $K(t) > 0$, (2) $K'(t) > 0$, (3) $K(0) \neq 0$, (4) $K''(t) < 0$ and (5) $K''(t)/K'(t)$ is nondecreasing [$\ln K'(t)$ is convex], then (1) $f - g = o(1)$ as $\epsilon \to 0$, uniformly on $[0, T]$ and (2) $f(t) - g(t) \to 0$ as $t \to \infty$, for ϵ fixed.

Proof: As before, note first some properties of $g(t)$.

(1) $g(t)$ has exactly one zero. The fact that $K(t)$ is positive and increasing implies that $G(\infty) = 0$, and since $G(0) = 0$, there exists t_0 such that $0 < t_0 < \infty$ and $G'(t_0) = 0$, that is, $g(t_0) = 0$. The uniqueness comes from the fact that K' is positive and the equation

$$g'(t) = -K(\epsilon t)g(t) - \epsilon K'(\epsilon t)G(t).$$

(2) For g positive, $g(t)$ is bounded by one, since $g(0) = 1$.

(3) For g negative, a bound on g can be obtained as follows. If $t > t_0$, then

$$g(t) = 1 - \frac{K(\epsilon t)}{E(t)} \int_0^{t_0} E(\tau)\,d\tau + \int_{t_0}^t E(\tau)\,d\tau,$$

since

$$\int_{t_0}^t E(\tau)\,d\tau = \frac{E(t)}{K(\epsilon t)} - \frac{E(t_0)}{K(\epsilon t_0)} + \int_{t_0}^t E(\tau)\frac{\epsilon K'(\epsilon\tau)}{K^2(\epsilon\tau)}\,d\tau$$

and

$$\int_0^{t_0} E(\tau)\,d\tau = \frac{E(t_0)}{K(\epsilon t_0)},$$

so

$$g(t) = -\frac{K(\epsilon t)}{E(t)} \int_{t_0}^t E(\tau)\frac{\epsilon K'(\epsilon\tau)}{K^2(\epsilon\tau)}\,d\tau. \tag{2.6}$$

A second integration by parts yields

$$\begin{aligned} &\int_{t_0}^t E(\tau)\frac{\epsilon K'(\epsilon\tau)}{K^2(\epsilon\tau)}\,d\tau \\ &= \frac{E(t)\epsilon K'(\epsilon t)}{K^3(\epsilon t)} - \frac{E(t_0)\epsilon K'(\epsilon t_0)}{K^3(\epsilon t_0)} \\ &\quad - \int_{t_0}^t \left(\frac{\epsilon^2 K''(\epsilon\tau)}{K^3(\epsilon\tau)} - \frac{3K^2(\epsilon\tau)\epsilon^2 K'^2(\epsilon\tau)}{K^6(\epsilon\tau)}\right) E(\tau)\,d\tau; \end{aligned}$$

therefore, from (2.6),

$$\begin{aligned} g(t) = &-\frac{\epsilon K'(\epsilon t)}{K^2(\epsilon t)} + \frac{E(t_0)\epsilon K'(\epsilon t_0)K(\epsilon t)}{E(t)K^3(\epsilon t_0)} \\ &- \frac{K(\epsilon t)}{E(t)} \int_{t_0}^t \epsilon^2 \left(\frac{3K'^2(\epsilon\tau) - K(\epsilon\tau)K''(\epsilon\tau)}{K^4(\epsilon\tau)}\right) E(\tau)\,d\tau. \end{aligned} \tag{2.7}$$

Since $-K''/K'$ is nonincreasing,

$$\begin{aligned} &\int_{t_0}^t \epsilon \left(\frac{3K'^2(\epsilon\tau) - K(\epsilon\tau)K''(\epsilon\tau)}{K^2(\epsilon\tau)K'(\epsilon\tau)}\right) \cdot \frac{E(\tau)\epsilon K'(\epsilon\tau)}{K^3(\epsilon\tau)}\,d\tau \\ &\leq \epsilon \left(\frac{3K'^2(\epsilon t_0) - K(\epsilon t_0)K''(\epsilon t_0)}{K^2(\epsilon t_0)K'(\epsilon t_0)}\right) \int_{t_0}^t E(\tau)\frac{\epsilon K'(\epsilon\tau)}{K^2(\epsilon\tau)}\,d\tau, \end{aligned}$$

so from (2.6), (2.7),

$$\begin{aligned} |g(t)| &\leq \frac{\epsilon K'(\epsilon t)}{K^2(\epsilon t)} + \epsilon \left(\frac{3K'^2(\epsilon t_0) - K(\epsilon t_0)K''(\epsilon t_0)}{K^2(\epsilon t_0)K'(\epsilon t_0)}\right) |g(t)| \\ &\leq \frac{\epsilon K'(0)}{K^2(0)} + \epsilon \left(\frac{3K'(0)}{K^2(0)} - \frac{K''(0)}{K'(0)}\frac{1}{K(0)}\right) |g(t)|, \end{aligned}$$

$$(1-\delta)|g(t)| \leq \frac{\epsilon K'(0)}{K^2(0)}, \quad \text{where } 0 < \delta < 1 \text{ for } \epsilon \text{ small},$$

$$|g(t)| \leq \frac{\epsilon}{1-\delta}\frac{K'(0)}{K^2(0)} = o(1) \quad \text{as } \epsilon \to 0, \text{ uniformly for } t_0 < t.$$

As in Theorem 2.1, the error $f - g$ satisfies

$$f(t) - g(t) = \phi(t) + \phi * f',$$

or

$$f(t) - g(t) = \phi' * f, \tag{2.8}$$

where, from (2.4),

$$\begin{aligned} \phi'(t) = &\epsilon K'(\epsilon t)\int_0^t g(\tau)\,d\tau + K(\epsilon t)g(t) - K(0)g(t) \\ &- \epsilon \int_0^t K'(\epsilon\tau)g(t-\tau)\,d\tau. \end{aligned}$$

For $0 \leq t \leq t_0$,

$$\begin{aligned} |\phi'(t)| &\leq \epsilon K'(0)T + [K(\epsilon T) - K(0)] + \epsilon K'(0)T \\ &= o(1) \quad \text{as } \epsilon \to 0. \end{aligned}$$

For $t_0 < t \leq T$,

$$\begin{aligned} \phi'(t) = &\epsilon \int_0^{t_0} [K'(\epsilon t) - K'(\epsilon\tau)]g(t-\tau)\,d\tau \\ &+ \epsilon \int_{t_0}^t [K'(\epsilon t) - K'(\epsilon\tau)]g(t-\tau)\,d\tau \\ &+ [K(\epsilon t) - K(0)]g(t), \end{aligned}$$

$$|\phi'(t)| \leq \epsilon K'(0)T + \epsilon K'(0)T\frac{\epsilon}{1-\delta}\frac{K'(0)}{K^2(0)}$$

$$+[K(\epsilon T)-K(0)]\frac{\epsilon}{1-\delta}\frac{K'(0)}{K^2(0)}$$

$$=o(1) \quad \text{as } \epsilon \to 0.$$

By remark after Theorem 1.2.3. in Ref. 8, $|f(t)| \leq 1$ for ϵ sufficiently small and so, from (2.8),

$$|f-g| \leq o(1)\cdot T$$

$$=o(1) \quad \text{as } \epsilon \to 0, \text{ uniformly on } [0, T].$$

It can be shown that $f(t)-g(t) \to 0$ as $t \to \infty$. By Theorem 1.1.1 in Ref. 8, $f(t)$ also satisfies

$$f(t)=\exp(-\gamma t)-L*f,$$

where

$$L(t)=(a-\gamma)\exp(-\gamma t)+\epsilon\int_0^t \exp[-\gamma(t-\tau)]K'(\epsilon\tau)\,d\tau,$$

$a=K(0)$, γ any constant.

Let $u(t)=1-L*u$; then

$$f(t)=u(t)-\gamma\exp(-\gamma t)*u. \tag{2.9}$$

By an appropriate choice of γ (see the proof of Theorem 1.2.4 in Ref. 8), we have $L(t)>0$, $L'(t)<0$, $\ln L(t)$ is convex, so $u' \leq 0$ and $0<u(t)\leq 1$. Since $u(t)$ is monotone and bounded, $u(\infty)$ exists. Therefore, from (2.9), $f(\infty)=0$. As is shown in Theorem 2.1, $g(\infty)=0$ and so $f(t)-g(t) \to 0$ as $t \to \infty$, for ϵ fixed.

3. INTEGRODIFFERENTIAL EQUATIONS

Integrodifferential equations of the form

$$f'(t)=-mf(t)-\epsilon\int_0^t k(\epsilon(t-\tau))f(\tau)\,d\tau, \quad f(0)=1, \tag{3.1}$$

where ϵ is a small positive parameter can be transformed into the integral equations studied in Sec. 2. The function $g(t)$ will represent the solution to the equivalent integral equation with the kernel pulled out. In the following theorem, uniformity of the approximation is proved for equations of the form (3.1) containing negative kernels $k(t)$.

Theorem 3.1: If (1) $k(t)<0$, (2) $m>-\int_0^\infty k(t)\,dt$ and (3) $k(t)/(m+\int_0^t k(\tau)\,d\tau)$ is nondecreasing, then (1) $f-g=o(1)$ as $\epsilon \to 0$, uniformly on $[0, T]$ and (2) $f(t)-g(t) \to 0$ as $t \to \infty$, for ϵ fixed.

Proof: (3.1) can be written in the form

$$f(t)=1-\int_0^t K(\epsilon(t-\tau))f(\tau)\,d\tau,$$

where

$$K(t)=m+\int_0^t k(\tau)\,d\tau.$$

The above kernel $K(t)$ satisfies all the hypotheses of Theorem 2.1, $K(t)>0$, $K'(t)<0$, $K(\infty)\neq 0$, and K'/K is nondecreasing. The conclusion now follows.

Some examples of kernels $k(t)$ that satisfy the hypotheses of Theorem 3.1 are $k(t)=-e^{-t}$ and $k(t)=-1/(1+t)^2$, for any $m>1$.

In the next theorem, Eqs. (3.1) with positive decreasing kernels $k(t)$ are studied.

Theorem 3.2: If (1) $k(t)>0$, (2) $k'(t)<0$, (3) $m>0$ and (4) k'/k is nondecreasing, then (1) $f-g=o(1)$ as $\epsilon \to 0$, uniformly on $[0, T]$ and (2) $f(t)-g(t) \to 0$ as $t \to \infty$, for ϵ fixed.

Proof: As in Theorem 3.1, (3.1) is written in the form

$$f(t)=1-\int_0^t K(\epsilon(t-\tau))f(\tau)\,d\tau,$$

where

$$K(t)=m+\int_0^t k(\tau)\,d\tau.$$

The fact that the above kernel $K(t)$ satisfies all the hypotheses of Theorem 2.2 leads to the conclusion of this theorem.

Some examples of kernels $k(t)$ that satisfy all hypotheses of Theorem 3.2 are $k(t)=e^{-t}$ and $k(t)=1/(1+t)$.

In the remaining part of this section, we consider another class of integrodifferential equations, of the form

$$F'(t)=-\sigma_0-\epsilon\int_0^t k(t-\tau)F(\tau)\,d\tau, \quad F(0)=1, \tag{3.2}$$

where σ_0 is a constant source term and ϵ is a small positive parameter. Let $f(t)$ be the solution to the corresponding homogeneous equation,

$$f'(t)=-\epsilon\int_0^t k(t-\tau)f(\tau)\,d\tau, \quad f(0)=1. \tag{3.3}$$

When $k(t)=1/(1+t)$, (3.2) arises in cosmic ray transport.[2] And (3.3), with $k(t)=1/(1+t)$, has been studied in Ref. 3.

The next theorem in this work is on (3.3) for a certain class of kernels $k(t)$. Use will be made of the theorem in Ref. 5, in which the integral equation studied is of the form

$$f(t)=1-\epsilon\int_0^t K(t-\tau)f(\tau)\,d\tau$$

and $g(t)$ is the solution to the equation

$$g(t)=1-\epsilon K(t)\int_0^t g(\tau)\,d\tau.$$

It is now stated here.

Theorem: (1) If $K(t)>0$, $t>0$, (2) $a/(t+b)\leq K'(t)\leq c/(t+d)$, where a, b, c, and d are positive constants, (3) $-a/(t+b)^2\leq K''(t)<0$ and (4) $K''(t)/K'(t)$ is nondecreasing, then $f-g=o(1)$ as $\epsilon \to 0$, uniformly on $[0, \infty)$.

So for (3.3), we have the following theorem. The function $g(t)$ will represent the solution to the equivalent integral equation with the kernel pulled out.

Theorem 3.3: (1) If $a/(t+b)\leq k(t)\leq c/(t+d)$, where a, b, c, and d are positive constants, (2) $-a/(t+b)^2\leq k'(t)<0$ and (3) $k'(t)/k(t)$ is nondecreasing, then $f-g=o(1)$ as $\epsilon \to 0$, uniformly on $[0, \infty)$.

Proof: (3.3) can be written in the form

$$f(t)=1-\epsilon\int_0^t K(t-\tau)f(\tau)\,d\tau,$$

where

$$K(t)=\int_0^t k(\tau)\,d\tau.$$

The conclusion follows from an application of the theorem in Ref. 5.

Returning to the inhomogeneous problem (3.2), let $h(t)=g(t)-\sigma_0*g$, where the function $g(t)$ is the one corresponding to Theorem 3.3.

Theorem 3.4: If the kernel $k(t)$ in (3.2) satisfies all hypotheses of Theorem 3.3, then $F-h=o(1)$ as $\epsilon \to 0$, uniformly on $[0, T]$.

Proof: By quadrature[3] the inhomogeneous solution $F(t)$ can be expressed in terms of the homogeneous solution $f(t)$, namely,

$$F(t)=f(t)-\sigma_0 * f. \tag{3.4}$$

From (3.4), the error $F-h$ satisfies

$$F(t)-h(t)=f(t)-g(t)-\sigma_0 * (f-g).$$

By theorem 3.3, $f-g=o(1)$ as $\epsilon \to 0$, uniformly on $[0, \infty)$; therefore,

$$|F-h| \le o(1)+|\sigma_0| To(1), \quad \text{for } 0 \le t \le T$$
$$=o(1) \quad \text{as } \epsilon \to 0, \text{ uniformly on } [0, T].$$

[1]G. J. Habetler and B. J. Matkowsky, "Uniform asymptotic expansions in transport theory with small mean free paths, and the diffusion approximation," J. Math. Phys. **16**, 846—54 (1975).

[2]F. B. Hansion, A. Klimas, G. V. Ramanathan, and G. Sandri, "Analysis of a model for transport of charged particles in a random magnetic field," J. Math. Anal. Appl. **44**, 786—98 (1973).

[3]F. B. Hansion, A. Klimas, G. V. Ramanathan, and G. Sandri, "Uniformly valid asymptotic solution to a Volterra equation on an infinite interval," J. Math. Phys. **14**, 1592—1600 (1973).

[4]E. W. Larsen, "Solutions of the steady, one-speed neutron transport equation for small mean free paths," J. Math. Phys. **15**, 299—305 (1974).

[5]R. Ling, "Uniformly valid solutions to Volterra integral equations," J. Math. Phys. **18**, 2019—25 (1977).

[6]R. Bellman and K. L. Cooke, *Differential-Difference Equations* (Academic, New York, 1963).

[7]A. Friedman "On integral equations of Volterra type," J. d'analyse Math. **XI**, 381—413 (1963).

[8]R. Ling, "Integral equations of Volterra type," to be published in J. Math. Anal. Appl.

On the calculation of the probability density at the origin for an *s*-state

Nanny Fröman

Institute of Theoretical Physics, University of Uppsala, Uppsala, Sweden
(Received 11 November 1977)

Phase-integral formulas for the probability density at the origin of an *s*-state are derived, on various assumptions as to the behavior of the potential in the neighborhood of the origin. In particular, the analog of the Fermi–Segrè formula is given for the case that the potential is regular at the origin.

1. INTRODUCTION

In a previous paper[1] a simple phase-integral formula for calculating quantal expectation values without the use of wavefunctions, for the situation referred to in Fig. 1(a) of the present paper, was derived. In a later paper[2] the formula was given in a more general form, covering all the three cases illustrated in Figs. 1(a), (b), and (c). Though the wavefunction does not appear in the final expectation value formula, the derivation is based on the use of the phase-integral approximations described in Refs. 3 and 4, possibly modified (cf. Ref. 5 and pp. 126–31 in Ref. 6). Expressions for the normalization factors of the phase-integral approximations representing bound state wavefunctions in the cases of Figs. 1(a) and 1(b) were given in Refs. 7, 8, and 9, respectively, and a general formula covering all the three cases, i.e., also case (*c*), was given in Ref. 2.

In the present paper we shall use the above-mentioned results for deriving expressions for the probability density at the origin for a state with zero angular momentum, on the various assumptions concerning the qualitative behavior of the potential illustrated in Figs. 1(a), (b), and (c) and explained in the caption to Fig. 1.

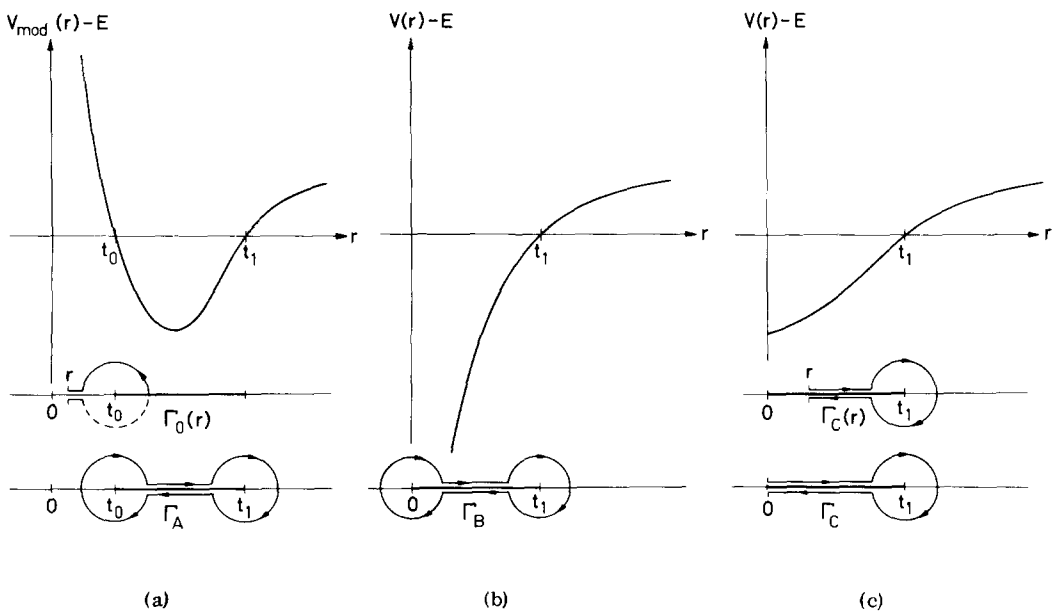

FIG. 1. Figure 1(a) shows the qualitative behavior of the modified potential, $V_{mod}(r)$, as assumed in the case treated in Sec. 2 A. The curve is drawn as in a typical radial problem, but $V_{mod}(r)$ may, alternatively, tend to infinity when $r \to +\infty$. The points t_0 and t_1 are classical turning points in the generalized sense, i.e., real zeros of $Q^2_{mod}(r)$. Figures 1(b) and 1(c) show the qualitative behavior of the physical potential $V(r)$ for the cases treated in Secs. 2 B and 2 C, respectively. There is in these cases only one classical turning point, t_1. In case (c) the derivative $dV(r)/dr$ may in particular be equal to zero at the origin, or may be negative. The potential $V(r)$ may alternatively tend to infinity when $r \to +\infty$. The cuts in the complex r-plane are indicated by bold lines. The contours of integration occurring in the formulas pertaining to each case [(a), (b), or (c)] are also depicted. In case (a), the contour $\Gamma_0(r)$ is a nonclosed contour encircling the generalized classical turning point t_0, the broken part of this contour lying on the second Riemann sheet. The contour Γ_A is a closed loop encircling both of the generalized classical turning points t_0 and t_1. In case (b) the contour Γ_B is a closed loop encircling the origin and the classical turning point t_1. In case (c) the contour $\Gamma_C(r)$ is a nonclosed loop emerging from the real value $r+i0$ on the upper lip of the cut, encircling the classical turning point t_1, and ending at $r-i0$ on the lower lip of the cut. The particular contour $\Gamma_C(0)$ is also denoted by Γ_C. The phase of $Q_{mod}(r)$ in case (a) and of $Q(r)$ in cases (b) and (c) is chosen such that $Q_{mod}=|Q_{mod}|$ and $Q=|Q|$, respectively, on the upper lip of the cut in the classically allowed region, which means that the contour integrals of $q(r)$ over Γ_A, Γ_B, and Γ_C are positive.

 0022-2488/78/1905-1141$1.00

For the case that the potential is regular at the origin, as depicted in Fig. 1(c), the analog of the Fermi–Segrè formula [the latter referring to the case in Fig. 1(b)] is derived.

We shall now quote some general formulas in order to make the present paper, at least to some extent, self contained. For further details we refer to Refs. 1 and 2.

Let the radial Schrödinger equation be

$$\frac{d^2u}{dr^2} + Q^2(r)u = 0, \tag{1}$$

where, since l is assumed to be equal to zero, we have, with conventional notation,

$$Q^2(r) = \frac{2m}{\hbar^2}[E - V(r)], \tag{2}$$

the physical potential $V(r)$ being assumed to be at the most as singular as $1/r$ at the origin. If we write a solution of (1) as

$$u = q^{-1/2}(r)\exp[\pm i\int^r q(r)\,dr], \tag{3}$$

the phase-integral approximations[3–5] (cf. also pp. 126–31 in Ref. 6) on which the present treatment is based, are obtained by replacing the exact function $q(r)$ in (3) by a truncated series, which for the $(2N+1)$th-order phase-integral approximation is written

$$q(r) = Q_{\text{mod}}(r)\sum_{n=0}^{N} Y_{2n}. \tag{4}$$

Expressions for $Y_0\ (=1)$, Y_2, and Y_4 can be found in Ref. 3, for Y_6 and Y_8 in Refs. 7 and 10, and for all the functions Y_{2n} up to Y_{20} in Ref. 11. By analogy with (2) we write (cf. Ref. 5 and pp. 126–31 in Ref. 6)

$$Q^2_{\text{mod}}(r) = \frac{2m}{\hbar^2}[E - V_{\text{mod}}(r)]. \tag{5}$$

In the present paper we shall, in case (a), use modified approximations assuming $Q^2_{\text{mod}}(r)$ to be chosen such that

$$\lim_{r\to 0} r^2[Q^2_{\text{mod}}(r) - Q^2(r)] = -\tfrac{1}{4}. \tag{6a}$$

In cases (b) and (c) we shall use unmodified phase-integral approximations which means that

$$Q^2_{\text{mod}}(r) = Q^2(r). \tag{6b,c}$$

The phase-integral quantization condition covering all the three cases in Figs. 1(a), (b), (c) is given by Eq. (19) with (20a, b, c,) in Ref. 2.

2. FORMULAS FOR THE PROBABILITY DENSITY AT THE ORIGIN IN THE THREE CASES (a), (b), (c)

If we normalize $u_n(r)$ such that

$$\int_0^\infty u_n^2(r)\,dr = 1, \tag{7}$$

the normalized three-dimensional probability density, $\psi_n^2(0)$, at the origin for an s-state is

$$\psi_n^2(0) = \frac{1}{4\pi}\lim_{r\to 0}\left(\frac{u_n(r)}{r}\right)^2 = \frac{1}{4\pi}[u_n'(0)]^2, \tag{8}$$

where the prime denotes the derivative with respect to r. This probability density can be exactly expressed in terms of the expectation value of $dV(r)/dr$ by the formula [see, e.g., Eq. (16a) in Ref. 12 or Eq. (40a) in Ref. 2].

$$\psi_n^2(0) = \frac{m}{2\pi\hbar^2}\int_0^\infty \frac{dV(r)}{dr}u_n^2(r)\,dr = \frac{m}{2\pi\hbar^2}\left\langle\frac{dV(r)}{dr}\right\rangle. \tag{9}$$

[Note that in Ref. 2 the symbol ψ_n was used to denote a one-dimensional or a radial wavefunction, i.e., ψ_n in Ref. 2 corresponds to u_n in the present paper.]

In the following subsections we shall discuss the calculation of $\psi_n^2(0)$ by means of the phase-integral approximations mentioned in the previous section.

A. The potential $V(r)$ is either regular or Coulomb-like when $r\to 0$. A modification fulfilling (6a) is used

According to the assumptions made in connection with the differential equation (1), the physical potential $V(r)$ is at the most as singular as $1/r$ at the origin. In the present subsection we adopt no further restriction on the behavior of the potential in the neighborhood of the origin, and thus we assume that $V(r)$ is either regular or Coulomb-like at the origin. We shall use modified phase-integral approximations with $Q^2_{\text{mod}}(r)$ chosen according to (6a), which means that $V_{\text{mod}}(r)$ will qualitatively have the behavior shown in Fig. 1(a) both when $V(r)$ is regular and when $V(r)$ is Coulomb-like at the origin. Due to the form of $V_{\text{mod}}(r)$, there is a classically forbidden region (in the generalized sense) in the neighborhood of the origin. The function $Q^2_{\text{mod}}(r)$ has thus two real zeros (generalized classical turning points) between which there is a classically allowed region (in the generalized sense) as illustrated in Fig. 1(a).

Since the contour Γ_A in Fig. 1(a) does not enclose the origin, the function $dV(r)/dr$ is regular within Γ_A, even when the potential $V(r)$ is Coulomb-like. We can therefore use formula (26a) in Ref. 2 for the expectation value of $dV(r)/dr$ in the right-hand member of (9), obtaining

$$\psi_n^2(0) = \frac{m}{2\pi\hbar^2}\left(\int_{\Gamma_A}\frac{dV(r)}{dr}\frac{dr}{q(r,E_n)}\Big/\int_{\Gamma_A}\frac{dr}{q(r,E_n)}\right), \tag{10}$$

where Γ_A is the closed contour of integration depicted in Fig. 1(a). We recall that the contour Γ_A shall enclose those zeros of $q(r)$ which exist in the neighborhood of the generalized classical turning points, if the order of the phase-integral approximation used, i.e., $2N+1$, is larger than unity (see the comments at the end of Sec. 2 in Ref. 2).

For the sake of completeness we also give the expression for $\psi_n^2(0)$ obtained by evaluating (8) directly by means of phase-integral wavefunctions modified according to (6a), instead of using (9). Except for a constant normalization factor, the wavefunction $u_n(r)$ is, to the left of t_0 and even in the neighborhood of the origin, given approximately by the expression (3) with

the sign in the exponent chosen appropriately such that the wavefunction decreases when r decreases and tends to zero. The normalized wavefunction is given by (29a) in Ref. 8, but can also be obtained by using the expression for the normalization factor given by Eqs. (23) and (25a) in Ref. 2. Inserting the normalized expression for $u_n(r)$ into (8), we obtain

$$\psi_n^2(0)=\frac{1}{4\pi}\lim_{r\to 0}\frac{\exp(-2|\tfrac{1}{2}\int_{\Gamma_0(r)}q(r)\,dr|)}{r^2|q(r)|\int_{\Gamma_A}dr/q(r)}, \tag{11}$$

where $\Gamma_0(r)$ is the nonclosed contour of integration depicted in Fig. 1(a).

Formula (11), involving a limiting procedure, looks more complicated than (10). Furthermore, formula (10) is based on the formula (26a) in Ref. 2 for calculating expectation values, and this formula in turn, as derived in Ref. 1, is based on the use of the quantization condition and not on the wavefunction itself. For reasons mentioned in Sec. V in Ref. 2, we may therefore expect formula (10) to be more accurate than (11). In addition, when the modification (6a) is used, the case $l=0$ is unfavorable, as concerns the accuracy of the phase-integral *wavefunctions*. [Cf. the comment below Eq. (11.28) on p. 120 in Ref. 13; see also the illustration of the accuracy of phase-integral wavefunctions in Ref. 14.]

To illustrate what we have just said about the accuracy of (10) and (11) we shall evaluate $\psi_n^2(0)$ for the Coulomb potential $V(r)=-Ze^2/r$, choosing [cf. (6a)] $Q_{\text{mod}}^2(r)-Q^2(r)=-1/(4r^2)$. Formula (10) then yields already in the first-order approximation the exact result [cf. Eq. (32) in Ref. 2], the higher-order contributions being zero. Applying formula (11) to the same potential and writing the result as follows,

$$\psi_n^2(0)=k_\nu[\psi_n^2(0)]_{\text{exact}}, \tag{12}$$

where $\nu=1, 3, 5$ refers to the first-, third-, and fifth-order phase-integral approximation, respectively, we obtain

$$k_1\approx\frac{\tilde{e}}{2\pi}\left(1-\frac{1}{2n}\right)^{1/2-n}\left(1+\frac{1}{2n}\right)^{1/2+n}, \tag{13a}$$

where $\tilde{e}\approx 2.7183\ldots$ is Neper's number,

$$k_3=k_1\exp\left(-\frac{2n^2-1}{3(4n^2-1)}\right), \tag{13b}$$

$$k_5=k_3\exp\left(\frac{8n^6-6n^4-9n^2-1}{45(4n^2-1)^3}\right). \tag{13c}$$

The numerical results displayed in Table I show that, for the potential $V(r)=-Ze^2/r$ under consideration, formula (11) is not particularly good, although an essential improvement is obtained, when one proceeds from the first to the third order. We also notice that, for a fixed order of approximation, the error is of the same order of magnitude for small as for large quantum numbers. According to what has been said above, we can expect formula (10) to be much more satisfactory than (11), in general, i.e., not only accidentally for the special case of the pure Coulomb-potential.

In the present context, we also mention a formula for the probability density at the origin derived by Young and Uhlenbeck [see Eq. (8) in Ref. 15] with the aid of the first-order JWKB-approximation. Their formula differs from our formula (11), taken in the first-order approximation, because they determine the normalization factor directly by integrating the square of the JWKB wavefunction over the appropriate interval. This procedure leads to a formula in which certain unnecessary terms appear, which cause the formula to be less accurate than the first-order version of (11) for small quantum numbers, while when $n\to\infty$ the additional terms disappear and the formula becomes of the same accuracy as (11), specialized to the first order. According to the figures given in Ref. 15 for the case of hydrogenic atoms, the relative errors for $n=1$ and $n=2$ are 0.30 and 0.20, respectively, to be compared with our results 0.12 and 0.16, respectively, given in Table I. (For further comments on the convenience of using more sophisticated formulas, not involving wavefunctions, we refer the reader to Ref. 2.)

TABLE I. Relative error, $k_\nu-1$, of formula (11) for $\psi_n^2(0)$ in the case of s-states in the potential $V(r)=-Ze^2/r$, calculated in the first-, third-, and fifth-order phase-integral approximation by means of (13a), (13b), and (13c).

	Relative error of formula (11) for $\psi_n^2(0)$		
n	First order	Third order	Fifth order
1	0.12	6×10^{-3}	-0.8×10^{-3}
2	0.16	-4.0×10^{-3}	-1.6×10^{-3}
3	0.171	-4.4×10^{-3}	-1.72×10^{-3}
4	0.173	-4.50×10^{-3}	-1.75×10^{-3}
5	0.174	-4.52×10^{-3}	-1.759×10^{-3}
∞	0.176	-4.53×10^{-3}	-1.764×10^{-3}

B. The potential $V(r)$ is Coulomb-like and attractive close to the origin. The unmodified expression for $q(r)$ is used

In the neighborhood of the origin the physical potential $V(r)$ is assumed to be attractive and to correspond to the potential of a point charge with charge number Z situated at the origin. Thus, as r tends to zero, the potential $V(r)$ behaves as $-Ze^2/r$, where e is the electron charge, and hence $Q^2(r)$ behaves as $2mZe^2/(\hbar^2 r)=2Z/(a_0 r)$, where $a_0=\hbar^2/(me^2)$ is the Bohr radius. The qualitative behavior of $V(r)$ is depicted in Fig. 1(b).

In the treatment of the analogous situation in Ref. 2, we used the modification $Q_{\text{mod}}^2(z)=Q^2(z)+l(l+1)/z^2$, which, in the particular case of $l=0$ under consideration here, corresponds to the use of unmodified phase-integral approximations.

It should first be noted that, on the assumptions just mentioned, we cannot obtain a phase-integral expression for $\psi_n^2(0)$ by using the relation (9) for the following reason. Since $dV(r)/dr$ behaves as Ze^2/r^2 at the origin, the auxiliary potential [see Eq. (13) in Ref. 1], which would be used in the derivation of a phase-integral formula for $\langle dV(r)/dr\rangle$, would be such that the quantization condition (19) with (20b) in Ref. 2 would no longer be valid if the actual physical potential were replaced by the auxiliary potential in question. Thus,

the procedure devised in Ref. 1 does not apply to the derivation of a phase-integral formula for $\langle dV(r)/dr\rangle$ in the present situation. Neither can we, since we use unmodified phase-integral approximations, obtain $\psi_n^2(0)$ directly on the basis of (8) in a similar way as is possible in case (a) (cf. the previous subsection) and in case (c) (cf. the following subsection), because, as is well known, the unmodified phase-integral approximations are not valid at the origin, when the potential has a Coulomb singularity there.

In the case to be treated in the present subsection we start from a formula for the probability density at the origin, which was given in Ref. 9 as an intermediate step in the derivation in that paper of the Fermi–Segrè formula. Thus, according to Eq. (16) with (14) in Ref. 9, we have

$$\psi_n^2(0)=\frac{Z}{a_0^2e^2}\left(\frac{\partial}{\partial E}\,\frac{1}{2}\int_{\Gamma_B}q(r)\,dr\right)^{-1}_{E=E_n}. \tag{14}$$

We remark, in passing, that by utilizing formula (21) in Ref. 2 for the level density, we obtain from (14) the Fermi–Segrè formula [cf. Eq. (18) in Ref. 9]

$$\psi_n^2(0)=\frac{Z}{\pi a_0^2e^2}\,\frac{dE_n}{dn}, \tag{15}$$

where dE_n/dn can be obtained from spectroscopic data.

Using the relation (25a) in Ref. 2 for the derivative with respect to the energy E in the right-hand member of (14), we obtain

$$\psi_n^2(0)=\frac{2Z}{a_0}\left(\int_{\Gamma_B}\frac{dr}{q(r,E_n)}\right)^{-1}. \tag{16}$$

If dE_n/dn is known from experiment, one can obtain the probability density at the origin from (15), without knowing the analytic form of $V(r)$, except that $V(r)$ shall behave as $-Ze^2/r$ close to the origin. On the other hand, if we know the analytic form of $V(r)$ completely, i.e., not only its behavior at the origin, we can evaluate the right-hand member of (16), obtaining the probability density at the origin without recourse to experimental data on dE_n/dn. For such an evaluation the simplification achieved by the use of Eq. (25a) in Ref. 2 in the step from (14) to (16) is essential.

For the pure Coulomb potential $V(r)=-Ze^2/r$ as well as for the potential $V(r)=-\gamma Ze^2/(e^{\gamma r}-1)$, where γ is a positive constant, both formulas (15) and (16) yield exact results already in the first-order approximation, the higher-order contributions being zero.[9]

For the important case that $V(r)$ is attractive and Coulomb-like in the immediate neighborhood of the origin, both formula (10), which is based on the use of the modification (6a), and formula (16), which is based on the use of unmodified phase-integral approximations, are applicable, and, as already mentioned, both yield exact results for the particular case of the pure Coulomb potential $V(r)=-Ze^2/r$, if in (10) one uses the modification $Q^2_{\rm mod}(r)=Q^2(r)-1/(4r^2)$. Formula (16) is simpler than (10) and has the advantage of being directly related to the Fermi–Segrè formula.

C. The potential $V(r)$ is regular at the origin and $Q^2(0)>0$. The unmodified expression for $q(r)$ is used

If the potential $V(r)$ is regular and $Q^2(r)$ is positive at $r=0$, there is, when $l=0$, a classically allowed region from the origin to the classical turning point t_1 as depicted in Fig. 1(c). This figure is drawn with $dV(r)/dr>0$ at $r=0$, but $dV(r)/dr$ may also be zero or negative at the origin. When $dV(r)/dr$ is negative at the origin, we must add the requirement that the energy shall be well above $V(0)$, in order that our single-well quantization condition be adequate.

Let us now derive an expression for $\psi_n^2(0)$ on the basis of (8). Solving the radial Schrödinger equation (1) with the aid of phase-integral approximations, and using the connection formula (21) in Ref. 4 to trace the bound-state solution from the classically forbidden region to the classically allowed region, and utilizing expression (23) in Ref. 2 for the normalization factor, we obtain the following normalized, approximate solution in the classically allowed region on the upper edge of the cut

$$u_n(r)=2C_nq^{-1/2}(r)\cos\left(\tfrac{1}{2}\int_{\Gamma_{C(r)}}q(r)dr-\frac{\pi}{4}\right), \tag{17}$$

the normalization factor, with the phase chosen conveniently, being given by

$$C_n=\left(\frac{2\hbar^2}{m}\,\frac{\partial}{\partial E}\,\frac{1}{2}\int_{\Gamma_C}q(r)\,dr\right)^{-1/2}_{E=E_n}, \tag{17a}$$

where the contours $\Gamma_C(r)$ and Γ_C [$=\Gamma_C(0)$] are those depicted in Fig. 1(c). When we impose the condition $u_n(0)=0$ on the solution (17), we obtain the quantization condition [see Eqs. (19) and (20c) in Ref. 2]

$$\tfrac{1}{2}\int_{\Gamma_C}q(r)\,dr=(n+\tfrac{3}{4})\pi. \tag{18}$$

Differentiating (17) with respect to r and using the quantization condition (18), we obtain, from (8) and (17a), the formula

$$\begin{aligned}\psi_n^2(0)&=C_n^2q(0)/\pi\\&=\frac{mq(0)}{2\pi\hbar^2}\left(\frac{\partial}{\partial E}\,\frac{1}{2}\int_{\Gamma_C}q(r)\,dr\right)^{-1}_{E=E_n},\end{aligned} \tag{19}$$

where, here as well as in the following formulas, $q(0)$ means $q(0+i0)$.

Using formula (21) in Ref. 2, we obtain from our Eq. (19) the formula

$$\psi_n^2(0)=\frac{mq(0)}{2\pi^2\hbar^2}\,\frac{dE_n}{dn}, \tag{20}$$

which is analogous to the Fermi–Segrè formula (15), the latter formula applying to the case that the potential $V(r)$ is Coulomb-like at the origin. Thus, according to (20), if dE_n/dn is known from experiment, it suffices, in the present case of a potential that is regular at the origin, to know the analytic form of $V(r)$ at the origin, such that $q(0)$ can be obtained from (4), in order to calculate $\psi_n^2(0)$. The requirement that $q(0)$ must be known in the present case corresponds to the requirement in case (b) that the factor Ze^2, determining the strength of the Coulomb singularity at the origin,

must be known in order that one shall be able to obtain $\psi_n^2(0)$ from the Fermi–Segrè formula (15).

On the other hand, if we use Eq. (25b) in Ref. 2 for the derivative with respect to the energy in the right-hand member of (19), we obtain the counterpart of (16) in the preceding subsection, namely

$$\psi_n^2(0)=\frac{q(0)}{\pi}\left\{\int_{\Gamma_C}\frac{dr}{q(r,E_n)}+(1-\delta_{N,0})\times\left[\sum_{j=0}^{N-1}(-1)^j\left(\frac{1}{2q}\frac{d}{dr}\right)^{2j+1}\frac{1}{q^2(r,E_n)}\right]_{r=0+i0}\right\}^{-1}. \tag{21}$$

As regards the first, third, and fifth order, more explicit formulas can be obtained from (19) in the present paper and Eqs. (25b′)–(25b‴) in Ref. 2.

In the special case that Q^2, and hence also q^2, are even functions of r, the differential operators in the right-hand member of (21), operating on $1/q^2$, yield an expression which is zero at the origin. Instead of verifying this directly, we can realize it by noting that when $Q^2(-r)=Q^2(r)$ we may regard $u_n(r)$ in (17) as an odd state wavefunction belonging to the potential well obtained by extending $V(r)$ to negative r. Determining C_n in (17) by considering this extended potential for which the normalization factor, given by (23) and (25a) in Ref. 2, involves a closed contour integral (encircling the turning points $\pm t_1$), being twice the corresponding integral along Γ_C, we find that $\psi_n^2(0)$ will agree with the expression which (21) reduces to when the sum inside the square brackets is zero. Thus,

$$\psi_n^2(0)=\frac{q(0)}{\pi}\left(\int_{\Gamma_C}\frac{dr}{q(r,E_n)}\right)^{-1},\quad \text{if } Q^2(-r)=Q^2(r). \tag{21′}$$

Formulas (19)–(21′) were derived on the basis of (8). Let us now consider the possibility of obtaining a formula for $\psi_n^2(0)$ on the basis of (9). Since $dV(r)/dr$ is regular within Γ_C, we can use Eq. (28) in Ref. 2, which was derived from Eq. (26b) in Ref. 2, to obtain a phase-integral expression for $\langle dV(r)/dr\rangle$, and hence we can, alternatively, obtain $\psi_n^2(0)$ on the basis of (9). Inserting that expression for $\langle dV(r)/dr\rangle$ into (9), we actually obtain a final formula which is identical to (21). Hence, in case (c), the accuracy of $\psi_n^2(0)$, even when derived from (9), is determined by the accuracy of the function $q(r)$, as given by (4), at the origin. This is to be expected, since the quantization condition in case (c), on which the expectation value formula pertinent to that case is based, was derived by imposing the condition that the wavefunction be zero at the origin. We had a different situation in Sec. 2 A, where the formula (10) for $\psi_n^2(0)$, derived on the basis of (9), was simpler and more accurate than the formula (11), derived on the basis of (8).

Referring to the comments on accuracy made in Sec. V of Ref. 2, we draw attention to the fact that, in formulas (10) and (16), which are capable of yielding exact results for the test cases considered, we have closed contour integrals, and the accuracy of $\psi_n^2(0)$ is not directly dependent on the accuracy or validity of the phase-integral wavefunction in the neighborhood of the origin, as is the case with formula (21).

As a convenient test case for the formulas (20) and (21′) for $\psi_n^2(0)$ obtained in the present subsection we consider the s-states of the three-dimensional harmonic oscillator. As is well known, the energy eigenvalues are obtained exactly from the first-order JWKB-quantization condition, i.e., the condition (18) specialized to the first-order approximation. They are identical with the odd-state energies of the linear oscillator and are thus given by $E_n=(2n+\frac{3}{2})\hbar\omega$, $n=0, 1, 2, \cdots$, where ω is the angular frequency of the oscillator. The exact expression for $\psi_n^2(0)$ can be obtained by utilizing well-known properties of the Hermite polynomials and evaluating the exact wavefunction at $r=0$. The exact expression obtained in this way reads

$$[\psi_n^2(0)]_{\text{exact}}=\left(\frac{m\omega}{\pi\hbar}\right)^{3/2}\frac{(2n+1)!}{2^{2n}(n!)^2},\quad n=0, 1, 2, \cdots. \tag{22}$$

Since the phase-integral quantization condition is exact for the harmonic oscillator, the approximate nature of (20) and (21′) derives from the factor $q(0)$, and both formulas yield the same expression, namely

$$\psi_n^2(0)=\left(\frac{m\omega}{\pi\hbar}\right)^{3/2}\left(\frac{4n+3}{\pi}\right)^{1/2}\times\left[1+\frac{1}{4(4n+3)^2}-\frac{19}{32(4n+3)^4}+\cdots\right], \tag{23}$$
$$n=0, 1, 2, \cdots$$

where the three terms within the square brackets correspond to the first-, third-, and fifth-order phase-integral contributions, respectively. Formula (23) is identical to the result which is obtained by using Stirling's asymptotic expansion for the factorials occurring in the exact formula (22), and expanding the resulting formula in powers of $(4n+3)^{-2}$. By comparison with the exact formula (22) the relative errors of $\psi_n^2(0)$ obtained according to formula (23) have been calculated for the eleven lowest eigenstates and are given in Table II.

For a situation when $(dV(r)/dr)_{r=0}\neq 0$, formula (21) has been checked by application to a linear potential,

TABLE II. Relative error of $\psi_n^2(0)$, i.e., $[\psi_n^2(0)-\psi_n^2(0)_{\text{exact}}]/\psi_n^2(0)_{\text{exact}}$, for s-states of a three-dimensional harmonic oscillator, calculated by means of (22) and (23) in the first-, third-, and fifth-order phase-integral approximation.

	Relative error of formula (21′) for $\psi_n^2(0)$		
n	First order	Third order	Fifth order
0	-2.3×10^{-2}	4.4×10^{-3}	-2.8×10^{-3}
1	-4.9×10^{-3}	2.1×10^{-4}	-3.2×10^{-5}
2	-2.0×10^{-3}	3.8×10^{-5}	-2.5×10^{-6}
3	-1.1×10^{-3}	1.1×10^{-5}	-4.0×10^{-7}
4	-6.9×10^{-4}	4.5×10^{-6}	-1.0×10^{-7}
5	-4.7×10^{-4}	2.0×10^{-6}	-3.2×10^{-8}
6	-3.4×10^{-4}	1.1×10^{-6}	-1.2×10^{-8}
7	-2.6×10^{-4}	6.4×10^{-7}	-5.5×10^{-9}
8	-2.0×10^{-4}	3.9×10^{-7}	-2.6×10^{-9}
9	-1.6×10^{-4}	2.6×10^{-7}	-1.4×10^{-8}
10	-1.4×10^{-4}	1.7×10^{-7}	-7.7×10^{-10}

for which the radial Schrödinger equation can be solved exactly in terms of an Airy function.[16] The agreement between exact results and results obtained from (21) is excellent, the relative error in the fifth-order approximation being of the order 10^{-4} for the ground state, i.e., for $n=0$, of the order 10^{-6} for $n=1$, and further diminishing with increasing values of the quantum number n.

ACKNOWLEDGMENTS

I wish to thank fil. lic. Örjan Dammert for checking the manuscript carefully and making valuable comments. The assistance of fil. mag. Anders Hökback and fil.mag. Bo Thidé in evaluating the numerical results in Table I and Table II is gratefully acknowledged. The research work reported here was supported in part by the Swedish Natural Science Research Council.

[1]N. Fröman, Phys. Lett. A 48, 137 (1974).
[2]N. Fröman, Phys. Rev. A 17, 493 (1978).
[3]N. Fröman, Ark. Fys. 32, 541 (1966).
[4]N. Fröman, Ann. Phys. (N.Y.) 61, 451 (1970).
[5]N. Fröman and P.O. Fröman, Ann. Phys. (N.Y.) 83, 103 (1974).
[6]N. Fröman and P.O. Fröman, Nuovo Cimento B 20, 121 (1974).
[7]S. Yngve, J. Math. Phys. 12, 114 (1971).
[8]P.O. Fröman, Ann. Phys. (N.Y.) 88, 621 (1974).
[9]N. Fröman and P.O. Fröman, Phys. Rev. A 6, 2064 (1972).
[10]N. Fröman and P.O. Fröman, Nucl. Phys. A 147, 606 (1970).
[11]J.A. Campbell, J. Comp. Phys. 10, 308 (1972).
[12]D.M. Fradkin and F. Calogero, Nucl. Phys. 75, 475 (1966).
[13]N. Fröman and P.O. Fröman, *JWKB-Approximation, Contributions to the Theory* (North-Holland, Amsterdam, 1965) (Russian translation: MIR, Moscow, 1967).
[14]N. Fröman and W. Mrazek, J. Phys. A 10, 1287 (1977).
[15]L.A. Young and G.E. Uhlenbeck, Phys. Rev. 36, 1154 (1930).
[16]B. Thidé, to be published.

Variational method and Mathieu equation

D. Y. Hsieh

Division of Applied Mathematics, Brown University, Providence, Rhode Island 02912
(Received 27 October 1977)

A variational method is developed to study the linear and nonlinear Mathieu equations. For the linear Mathieu equation, various modes of the Mathieu functions, the characteristic curves, and the stability regions are found, which agree with the established results. The variational method is then applied to the nonlinear Mathieu equation. Approximate periodic solutions of various modes are found in a similar manner, and their stability also investigated.

I. INTRODUCTION

In a previous paper,[1] it was shown that the recursion relations to find the Mathieu functions can be obtained from the Mathieu equation by a direct variational method. Take the Mathieu equation

$$\frac{d^2x}{dt^2} + (\alpha + \beta\cos 2t)x = 0. \tag{1}$$

The solutions of the last equation are those which would make the functional

$$J = \int_0^t \left[\left(\frac{dx}{dt}\right)^2 - \alpha x^2 - \beta x^2\cos 2t\right] dt, \tag{2}$$

an extremum. If we take the asymptotic trial solution as

$$x = \frac{A_0}{2} + \sum_{k=1} A_k \cos kt + \sum_{k=1} B_k \sin kt, \tag{3}$$

then the recursion relations for the Mathieu functions[2] are obtained when we assume that the A_k's and B_k's are constants.

In this paper, we shall explore the potentials of the variational method further by treating the A_k's and B_k's as slowly varying functions of time. It may then be shown that the well-established stability criterions can also be obtained by the variational analysis. The same procedure is then applied to the nonlinear Mathieu equation

$$\frac{d^2x}{dt} + (\alpha + \beta\cos 2t)x + \gamma x^3 = 0. \tag{4}$$

Various modes of the periodic solutions and their stability are studied by the direct variational method. The results may shed light on the role played by the nonlinear effects in problems with parameteric resonance.

II. LINEAR MATHIEU EQUATION

In Eq. (3), let us take the A_k's and B_k's as slowly varying functions of time. Then, using the procedure developed previously,[1] we obtain the approximate expression of the functional J for large t,

$$J \approx \int_0^t dt \Bigg\{ -\frac{1}{4}\left[\left(\frac{dA_0}{dt}\right)^2 + \alpha A_0^2\right] - \tfrac{1}{4}\beta A_1^2 + \tfrac{1}{2}\sum_{m=1}\left[\left(mA_m - \frac{dB_m}{dt}\right)^2 - \alpha A_m^2\right] - \tfrac{1}{2}\beta\sum_{m=0} A_m A_{m+2} + \tfrac{1}{4}\beta B_1^2 + \tfrac{1}{2}\sum_{m=1}\left[\left(mB_m + \frac{dA_m}{dt}\right)^2 - \alpha B_m^2\right] - \tfrac{1}{2}\beta\sum_{m=1} B_m B_{m+2}\Bigg\}. \tag{5}$$

The variation with respect to A_m and B_m thus leads to:

$$\frac{d^2A_0}{dt^2} + \alpha A_0 + \beta A_2 = 0, \tag{6}$$

$$-\frac{d}{dt}\left(\frac{dA_1}{dt} + B_1\right) + \left(A_1 - \frac{dB_1}{dt}\right) + \left(-\alpha A_1 - \frac{\beta}{2}(A_1 + A_3)\right) = 0, \tag{7}$$

$$-\frac{d}{dt}\left(\frac{dA_m}{dt} + mB_m\right) + m\left(mA_m - \frac{dB_m}{dt}\right) + \left(-\alpha A_m - \frac{\beta}{2}(A_{m-2} + A_{m+2})\right) = 0, \quad m \geq 2, \tag{8}$$

$$-\frac{d}{dt}\left(\frac{dB_1}{dt} - A_1\right) + \left(\frac{dA_1}{dt} + B_1\right) + \left(-\alpha B_1 + \frac{\beta}{2}(B_1 - B_3)\right) = 0, \tag{9}$$

and

$$-\frac{d}{dt}\left(\frac{dB_m}{dt} - mA_m\right) + m\left(mB_m + \frac{dA_m}{dt}\right) + \left(-\alpha B_m - \frac{\beta}{2}(B_{m-2} + B_{m+2})\right) = 0, \quad m \geq 2. \tag{10}$$

When the A_m's and B_m's are constants, we recover the usual recursion relations. In the following, we shall study the case that β is small and concentrate on some specific modes for a more detailed study. In particular, we shall study the $\{ce_1, se_1\}$ and $\{ce_2, se_2\}$ modes for illustration.

A. The mode with $m = 1$

In Eq. (3), if we simply takes $x = A_1\cos t + B_1\sin t$, then we obtain only Eqs. (7) and (9), with A_3 and B_3 missing. Since A_1 and B_1 are assumed to be slowly varying functions of time, let us neglect terms with second derivatives with respect to time. Then we obtain

$$-2\frac{dB_1}{dt} + [(1 - \alpha) - \beta/2]A_1 = 0 \tag{11}$$

and

$$2\frac{dA_1}{dt} + [(1 - \alpha) + \beta/2]B_1 = 0. \tag{12}$$

From Eqs. (11) and (12), it is clear that the system has the following equilibrium points:

$$\text{(0): } A_1 = B_1 = 0, \text{ i.e., the null solution;}$$

$$\text{(S1): } A_1 = 0, \ B_1 \text{ arbitrary, and } \alpha = 1 + \beta/2. \tag{13}$$

This corresponds to the mode se_1, and (13) is the equation of the characteristic curve $\alpha(\beta)$ for this

 0022-2488/78/1905-1147$1.00

mode up to $O(\beta)$.

(C1): $B_1 = 0$, A_1 arbitrary, and

$$\alpha = 1 - \beta/2. \tag{14}$$

This corresponds to the mode ce_1, and (14) is the equation of the characteristic curve $\alpha(\beta)$ for this mode up to $O(\beta)$. If we want to compute the relation $\alpha(\beta)$ to the next higher order of β, then we should include A_3 and B_3 in our trial solution and hence also the equations with $m = 3$. The higher the orders of β we want to include, the more the equations we need to consider.

The stability of the equilibrium points can be investigated readily from Eqs. (11) and (12). Let a_1 and b_1 be the perturbed quantities from the equilibrium values of A_1 and B_1 respectively. Then for each separate case we have:

$$(0): \quad -2\frac{db_1}{dt} + [(1-\alpha) - \beta/2]a_1 = 0,$$

$$2\frac{da_1}{dt} + [(1-\alpha) + \beta/2]b_1 = 0,$$

or

$$4\frac{d^2a_1}{dt^2} + [(1-\alpha)^2 - (\beta/2)^2]a_1 = 0. \tag{15}$$

Thus the null solution is $\begin{cases}\text{stable}\\ \text{unstable}\end{cases}$ if

$$(1-\alpha)^2 \gtrless \left(\frac{\beta}{2}\right)^2. \tag{16}$$

The stability criterion again agrees with the established result.[3]

$$(S1): \quad -2\frac{db_1}{dt} + 2(1-\alpha)^2 a_1 = 0$$

$$2\frac{da_1}{dt} = 0.$$

Thus $a_1 = \text{const}$, and $b_1 \propto t$. The equilibrium point (S1) is unstable. This is again consistent with the established result.

Similarly, the equilibrium point (C1) is also found to be unstable.

The characteristic curves for the modes se_1 and ce_1 as well as the stability region for the null solution are shown in Fig. 1.

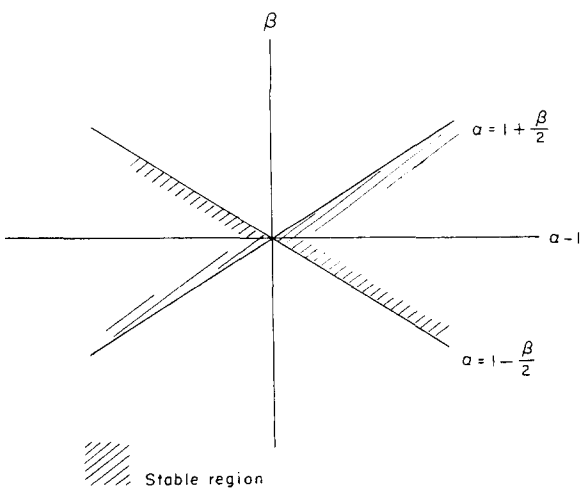

FIG. 1. Characteristic curves and stability region for the null solution of the $m = 1$ mode.

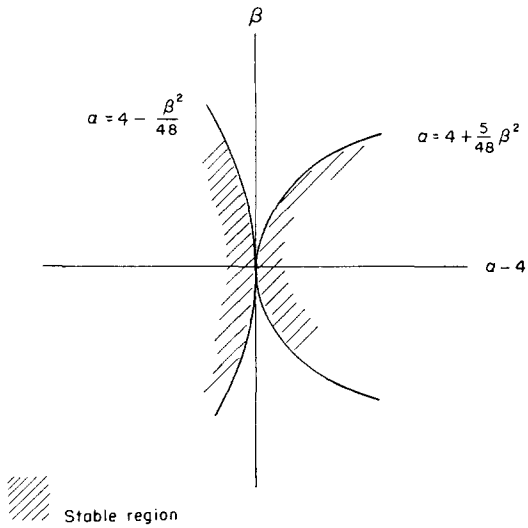

FIG. 2. Characteristic curves and stability region for the null solution of the $m = 2$ mode.

B. The mode with $m = 2$

For this mode, let us try to obtain results which are valid up to $O(\beta^2)$. Thus, we shall take the trial solution as

$$x = \frac{A_0}{2} + A_2\cos 2t + A_4\cos 4t + B_2\sin 2t + B_4\sin 4t.$$

Then we obtain Eqs. (6) and the equations corresponding to $m = 2$ and $m = 4$ from (8) and (10), with A_6 and B_6 missing. If terms with second derivatives with respect to time are neglected, we have:

$$\alpha A_0 + \beta A_2 = 0, \tag{17}$$

$$-4\frac{dB_2}{dt} + \left((4-\alpha)A_2 - \frac{\beta}{2}(A_0 + A_4)\right) = 0, \tag{18}$$

$$-8\frac{dB_4}{dt} + \left((32-\alpha)A_4 - \frac{\beta}{2}A_2\right) = 0, \tag{19}$$

$$4\frac{dA_2}{dt} + \left((4-\alpha)B_2 - \frac{\beta}{2}B_4\right) = 0, \tag{20}$$

$$8\frac{dA_4}{dt} + \left((32-\alpha)B_4 - \frac{\beta}{2}B_2\right) = 0. \tag{21}$$

We consider the following equilibrium points:

(0): $A_0 = A_2 = A_4 = B_2 = B_4 = 0$, i.e., the null solution;

$$(S2): \quad A_0 = A_2 = 0, \quad B_4 = \frac{\beta}{32 - 2\alpha}B_2, \quad \text{and} \quad \alpha = 4 - \frac{\beta^2}{48} + O(\beta^4). \tag{22}$$

This solution corresponds to the mode se_2, and (22) is the equation of the characteristic curve up to $O(\beta^2)$.

$$(C2): \quad B_2 = B_4 = 0, \quad A_0 = -\frac{\beta}{\alpha}A_2, \quad A_4 = \frac{\beta}{32 - 2\alpha}A_2, \quad \text{and}$$

$$\alpha = 4 + \tfrac{5}{48}\beta^2 + O(\beta^4). \tag{23}$$

This solution corresponds to the mode ce_2, and (23) is the equation of the characteristic curve up to $O(\beta^2)$.

The stability of these equilibrium points can be in investigated in the similar manner as the previous

case. It is easy to see that the equilibrium points (S2) and (C2) are not stable. For the equilibrium point (0), let us take the perturbed quantities to be proportional to $e^{\nu t}$, then we can obtain the characteristic equation as follows,

$$c_4\nu^4 + c_2\nu^2 + c_0 = 0,$$

where

$$c_4 = (128)^2,$$

$$c_2 = 64\left[(32-2\alpha)^2 + 4(8-2\alpha)\left(8-2\alpha+\frac{\beta^2}{\alpha}\right) + 4\beta^2\right],$$

$$c_0 = \left[\left(8-2\alpha+\frac{\beta^2}{\alpha}\right)(32-2\alpha) - \beta^2\right]\left[(32-2\alpha)(8-2\alpha) - \beta^2\right].$$

Thus the solution is stable if and only if c_4, c_2, and c_0 have the same sign. For small β and in the neighborhood of $\alpha = 4$, we have $c_2 \approx 64\times(24)^2$, and

$$c_0 \approx (24)^2[\alpha - 4 - \tfrac{5}{48}\beta^2 + O(\beta^4)][\alpha - 4 + \beta^2/48 + O(\beta^4)].$$

Thus the solution is stable if

$$\alpha > 4 + \tfrac{5}{48}\beta^2 \quad \text{and} \quad \alpha > 4 - \frac{\beta^2}{48}$$

or

$$\alpha < 4 + \tfrac{5}{48}\beta^2 \quad \text{and} \quad \alpha < 4 - \frac{\beta^2}{48}.$$

The characteristic curves for the modes se_2 and ce_2 as well as the stability region of the null solution are shown in Fig. 2. They are consistent with the established result.[3]

III. NONLINEAR MATHIEU EQUATION

The solutions of the nonlinear Mathieu equation (4) are those which would make the functional

$$I = \int_0^t \left[\left(\frac{dx}{dt}\right)^2 - \alpha x^2 - \beta x^2\cos 2t - \tfrac{1}{2}rx^4\right] dt \tag{24}$$

an extremum.

For the linear Mathieu equation, it has been established that the null solution is stable only for certain regions in the (α, β) parameter plane. Athough periodic solutions can be found on certain characterisitc curves in the (α, β) plane, yet these solutions are not stable. Whether the growth of such unstable solutions could be arrested by the nonlinear interactions is still an unanswered question. We hope the following investigation can shed some light on this difficult problem. The programs follow similar procedures as presented in the last section. Thus, we shall look for some approximate asymptotic periodic solution and then investigate the stability of these solutions. Again we shall study the case that β is small, and illustrate our procedure on particular modes corresponding to $m = 1$ and $m = 2$ as in the previous section.

A. The mode with $m = 1$

Let us take the asymptotic trial solution

$$x = A_1\cos t + B_1\sin t, \tag{25}$$

and substitute directly in (24). Since A_1 and B_1 are assumed to be slowly varying functions of time, we obtain an approximate expression of I:

$$I \tilde{=} \tfrac{1}{2}\int_0^t dt\left[\left(\frac{dA_1}{dt} + B_1\right)^2 + \left(\frac{dB_1}{dt} - A_1\right)^2 - \alpha(A_1^2 + B_1^2) - \frac{\beta}{2}(A_1^2 - B_1^2) - \tfrac{3}{8}r(A_1^4 + 2A_1^2B_1^2 + B_1^4)\right].$$

The variation of I with respect to A_1 and B_1, together with the neglect of the second derivatives leads to

$$-2\frac{dB_1}{dt} + \left((1-\alpha) - \frac{\beta}{2} - \frac{3}{4}r(A_1^2 + B_1^2)\right)A_1 = 0 \tag{26}$$

and

$$2\frac{dA_1}{dt} + \left((1-\alpha) + \frac{\beta}{2} - \frac{3}{4}r(A_1^2 + B_1^2)\right)B_1 = 0. \tag{27}$$

We shall investigate the following equilibrium points:

(0): $A_1 = B_1 = 0$, i.e., the null solution. The stability criterion turns out to be the same as the corresponding case for the linear Mathieu equation, which is given by (15) and (16), and depicted in Fig. 1. This is as expected, since the nonlinear terms do not affect the null solution.

(S1): $A_1 = 0$, $B_1^2 = \frac{4}{3r}\left((1-\alpha) + \frac{\beta}{2}\right)$. (28)

Let the perturbed quantities be a_1 and b_1, then the stability equations are

$$-2\frac{db_1}{dt} - \beta a_1 = 0$$

and

$$2\frac{da_1}{dt} - 2\left((1-\alpha) + \frac{\beta}{2}\right)b_1 = 0,$$

or

$$\frac{d^2a_1}{dt^2} - \frac{\beta}{2}\left((\alpha - 1) - \frac{\beta}{2}\right)a_1 = 0.$$

Thus the solution is stable if and only if

$$\beta(\alpha - 1 - \beta/2) < 0.$$

Since B_1^2 is positive, the solution is permissible only if the parameter r and $(1 + \beta/2 - \alpha)$ is of the same sign. The permissible and stable region in the (α, β) plane for this solution is shown in Fig. 3.

(C1): $B_1 = 0$, $A_1^2 = \frac{4}{3r}[(1-\alpha) - \beta/2]$. (29)

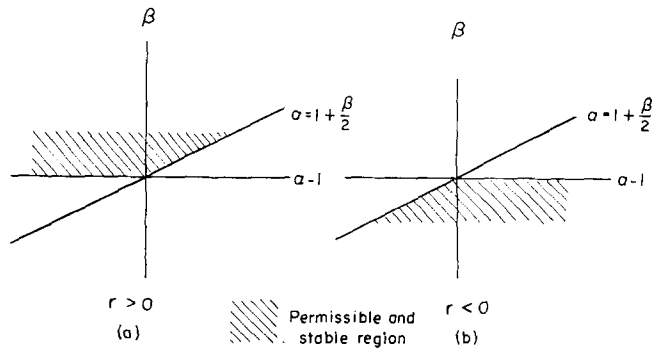

FIG. 3. Permissible and stable region of the (S1) mode of the nonlinear Mathieu equation: (a) for $r > 0$, (b) for $r < 0$.

Then the solution is stable if and only if

$$\beta(\alpha - 1 + \beta/2) > 0.$$

The solution (29) is premissible only if r and $(1 - \beta/2 - \alpha)$ are of the same sign. The permissible and stable region for this solution is shown is Fig. 4.

(CS1): If $A_1 \neq 0$ and $B_1 \neq 0$, then the equilibrium solution exists only if $\beta = 0$. Thus the Mathieu equation reduces to the Duffing equation. For this particular case, we have

$$A_1^2 + B_1^2 = \frac{4}{3r}(1 - \alpha).$$

Thus these equilibrium solutions are permissible only if r and $(1 - \alpha)$ are of the same sign. It is readily shown that these solutions are stable.

B. The mode with $m = 2$

Let us now take the asymptotic trial solution

$$x = \frac{A_0}{2} + A_2 \cos 2t + A_4 \cos 4t + B_2 \sin 2t + B_4 \sin 4t, \tag{30}$$

and substitute in (24). Then an approximate expression of I can be obtained as follows:

$$I \cong \tfrac{1}{2}\int_0^t dt\left[\frac{1}{2}\left(\frac{dA_0}{dt}\right)^2 - \frac{\alpha}{2}A_0^2 + \left(\frac{dA_2}{dt} + 2B_2\right)^2 + \left(\frac{dA_4}{dt} + 4B_4\right)^2 + \left(\frac{dB_2}{dt} - 2A_2\right)^2 + \left(\frac{dB_4}{dt} - 4A_4\right)^2 - \alpha(A_2^2 + A_4^2 + B_2^2 + B_4^2) - \beta(A_0A_2 + A_2A_4 + B_2B_4) - r\left(\frac{A_0^4}{16} + \frac{3}{4}A_0^2(A_2^2 + A_4^2 + B_2^2 + B_4^2) + \tfrac{3}{2}A_0A_4(A_2^2 - B_2^2) + \tfrac{3}{8}(A_2^4 + A_4^4 + B_2^4 + B_4^4) + \tfrac{3}{4}(A_2^2B_2^2 + A_4^2B_4^2) + \tfrac{3}{2}(A_2^2A_4^2 + A_2^2B_4^2 + B_2^2B_4^2 + B_2^2A_4^2)\right)\right]. \tag{31}$$

The variation of I with respect to A_0, A_2, A_4, B_2, and B_4, together with the neglect of the second derivatives leads to

$$-\alpha A_0 - \beta A_2 - r\left(\frac{A_0^3}{4} + \frac{3}{2}A_0(A_2^2 + A_4^2 + B_2^2 + B_4^2) + \tfrac{3}{2}A_4(A_2^2 - B_2^2)\right) = 0. \tag{32}$$

$$-8\frac{dB_2}{dt} + (8 - 2\alpha)A_2 - \beta(A_0 + A_4) - r[\tfrac{3}{2}A_2(A_0^2 + A_2^2 + B_2^2 + 2A_4^2 + 2B_4^2 + 2A_0A_4)] = 0, \tag{33}$$

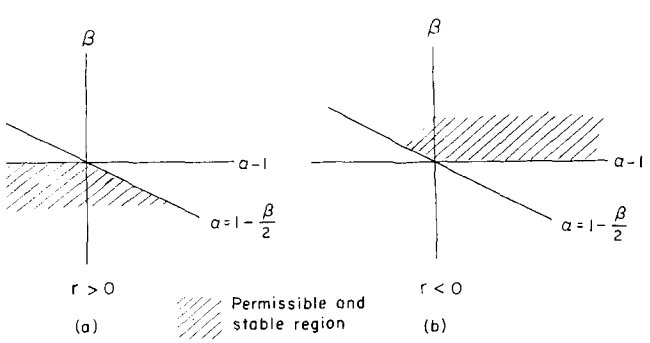

FIG. 4. Permissible and stable region of the (C1) mode of the nonlinear Mathieu equation: (a) for $r > 0$, (b) for $r < 0$.

$$-16\frac{dB_4}{dt} + (32 - 2\alpha)A_4 - \beta A_2 - r[\tfrac{3}{2}A_4(A_0^2 + 2A_2^2 + 2B_2^2 + A_4^2 + B_4^2) + \tfrac{3}{2}A_0(A_2^2 - B_2^2)] = 0, \tag{34}$$

$$8\frac{dA_2}{dt} + (8 - 2\alpha)B_2 - \beta B_4 - r[\tfrac{3}{2}B_2(A_0^2 + A_2^2 + B_2^2 + 2A_4^2 + 2B_4^2 - 2A_0A_4)] = 0, \tag{35}$$

$$16\frac{dA_4}{dt} + (32 - 2\alpha)B_4 - \beta B_2 - r[\tfrac{3}{2}B_4(A_0^2 + 2A_2^2 + 2B_2^2 + A_4^2 + B_4^2)] = 0. \tag{36}$$

We shall investigate the following equilibrium points:

(0): $A_0 = A_2 = A_4 = B_2 = B_4 = 0$, i.e., the null solution. The stability criterion for this solution is the same as the corresponding case for the linear Mathieu equation, which is depicted in Fig. 2.

(S2): $A_0 = A_2 = A_4 = 0$. Let us consider the solution such that $B_2 = O(1)$, and $B_4 = O(\beta)$ for small β. Then, up to $O(\beta^2)$, we have

$$B_2^2 = \frac{1}{3r}\left(4(4 - \alpha) - \frac{16 - \alpha}{(8 + \alpha)^2}\beta^2\right), \tag{37}$$

$$B_4 = \left(\frac{4 - \alpha}{3r}\right)^{1/2} \cdot \frac{\beta}{8 + \alpha}. \tag{38}$$

To investigate the stability of this mode of solution, we can again write down the stability equation for the perturbed quantities from (32)–(36). Take the perturbed quantities to be proportional to $e^{\nu t}$. Then we can obtain the following characteristic equation after straightforward calculations,

$$D_4\nu^4 + D_2\nu^2 + D_0 = 0, \tag{39}$$

where

$$D_4 = (128)^2,$$

$$D_2 = 64(16 + 2\alpha)^2 + O(\beta^2),$$

$$D_0 = -\left[2(8 + \alpha)\left(\frac{1}{8 - \alpha} + \frac{16 - \alpha}{2(8 + \alpha)^2}\right) - \left(\frac{16 - 3\alpha}{8 - \alpha}\right) + O(\beta^2)\right] \times \left[8(8 + \alpha)(4 - \alpha) - \left(\frac{3(16 - \alpha)}{8 + \alpha} - 1\right)\beta^2 + O[(4 - \alpha)\beta^2; \beta^4]\right]\beta^2.$$

Thus if $4 - \alpha = O(\beta^2)$, we have

$$D_0 \cong -576\beta^2\left(4 - \alpha - \frac{\beta^2}{48}\right) + O(\beta^6). \tag{40}$$

The solution is stable if and only if D_4, D_2, and D_0 are of the same sign. Since D_4 and D_2 are positive, the solution is stable if and only if D_0 is positive. For $4 - \alpha = O(\beta^2)$, the solution is thus stable if

$$\alpha > 4 - \beta^2/48. \tag{41}$$

Now for $4 - \alpha = O(\beta^2)$, the solution is permissible if

$$\frac{4}{3r}\left(4 - \alpha - \frac{\beta^2}{48}\right) > 0,$$

as may be seen from Eq. (37). Thus the permissible (S2) mode is always stable if $r < 0$, and always unstable if $r > 0$. The permissible and stable region is shown in Fig. 5.

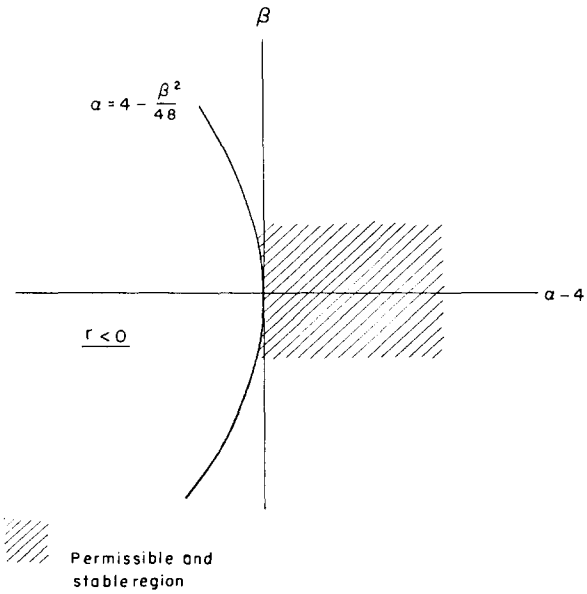

FIG. 5. Permissible and stable region of the (S2) mode of the nonlinear Mathieu equation.

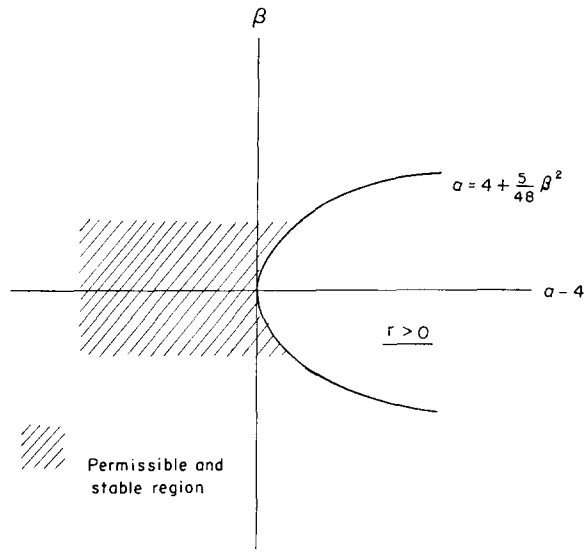

FIG. 6. Permissible and stable region of the (C2) mode of the nonlinear Mathieu equation.

(C2): $B_2 = B_4 = 0$. Again consider the mode that $A_2 = O(1)$ and $A_0 = O(\beta)$, $A_4 = O(\beta)$ for small β. Then up to $O(\beta^2)$, we have

$$A_2^2 = \frac{1}{3r}\{4(4-\alpha) - 2[(n+p) + 2(4-\alpha)(n^2 + 2np + 2p^2)]\beta^2\}, \tag{42}$$

$$A_0 = n\left(\frac{4(4-\alpha)}{3r}\right)^{1/2}\beta, \tag{43}$$

$$A_4 = p\left(\frac{4(4-\alpha)}{3r}\right)^{1/2}\beta, \tag{44}$$

where

$$n = \frac{1}{(\alpha - 8) + (\alpha/12)(4-\alpha)},$$

$$p = -\frac{\alpha}{24}n.$$

The stability of the (C2) mode can be investigated in a similar manner as before. Let us take the perturbed quantities to be proportional to $e^{\nu t}$, then we can obtain the following characteristic equation from the stability equation,

$$E_4\nu^4 + E_2\nu^2 + E_0 = 0, \tag{45}$$

where

$$E_4 = (128)^2,$$

$$E_2 = 64\times(16 + 2\alpha)^2 + O(\beta^2),$$

and

$$E_0 = [1 - (16+2\alpha)(n+p)]\left\{8(8+\alpha)(4-\alpha) + \left[1 + (16+2\alpha)\left(\frac{1}{\alpha - 8} - 3[n+p]\right)\right]\beta^2\right\}\beta^2 + O[\beta^6; \beta^4(\alpha-4);\ \beta^2(\alpha-4)^2].$$

Thus if $\alpha - 4 = O(\beta^2)$, we have $n + p \cong -\frac{5}{24}$, thus

$$E_0 = 576\beta^2(4 - \alpha + \tfrac{5}{48}\beta^2) + O(\beta^6). \tag{46}$$

Therefore, for $\alpha - 4 = O(\beta^2)$, the solution is stable if and only if

$$E_0 > 0,$$

or

$$\alpha < 4 + \tfrac{5}{48}\beta^2. \tag{47}$$

Now from (42), the solution is permissible for $\alpha - 4 = O(\beta^2)$, if

$$\frac{4}{3r}(4 - \alpha + \tfrac{5}{48}\beta^2) > 0. \tag{48}$$

Thus the permissible (C2) mode is always stable if $r > 0$ and always unstable if $r < 0$. The permissible and stable region is shown in Fig. 6.

From the above analysis, we found that for $r > 0$, only the (C2) mode can be present asymptotically and the (S2) mode is unstable, while for $r < 0$, only the (S2) mode can stably exist.

The existence of approximate asymptotic solutions of the nonlinear Mathieu equation does not answer the question whether and how an initial disturbance will evolve into these asymptotic solutions. However it does show that in the linearly unstable region of the (α, β) parameter space, stable solutions can be found. The approach we presented above should also be applicable to other types of nonlinear problems in finding asymptotic periodic solutions.

ACKNOWLEDGMENT

This study was supported by ERDA and NSF.

[1] D. Y. Hsieh, J. Math. Phys. **16**, 275—80 (1975).
[2] J. Mathews and R. L. Walker, *Mathematical Methods of Physics* (Benjamin, New York, 1970).
[3] N. W. McLachlan, *Theory and Application of Mathieu Functions* (Dover, New York, 1964).

On the positivity of energy in general relativity[a)]

Pong Soo Jang

Department of Physics, Syracuse University, Syracuse, New York 13210
(Received 8 December 1977)

We show that the positive energy argument of Geroch for time-symmetric initial data sets can be generalized to general initial data sets.

I. INTRODUCTION

With any asymptotically flat initial data set of Einstein's equation, there is associated a certain number E whose physical interpretation is "the total energy of the system, including contributions from both matter and the gravitational field." This number E is called the Arnowitt–Deser–Misner (ADM) energy.[1] The positive energy conjecture states that if the local energy condition for matter is satisfied this total energy E is also nonnegative and vanishes when and only when the data set is that for Minkowski space.[2]

The ADM energy E is obtained by performing a flux integral over a topological 2-sphere in the asymptotic region of the initial surface.[1] Although there are various forms of this integral, they all give the same value for an asymptotically flat initial data set. Several years ago, Geroch proposed[3] a particularly convenient flux form for the positive energy conjecture. Using his energy expression, he has given an argument to establish the validity of the positive energy conjecture for *time-symmetric* initial data sets.[3] In this paper, we shall generalize his argumant to general initial data sets.

In Sec. II, we shall briefly review the positive energy conjecture. In Sec. III, we shall present a characterization of initial data sets for Minkowski space, which turns out to be very convenient for the positive energy conjecture. Using results of Sec. III we generalize Geroch's argument to the full conjecture in Sec. IV. In Sec. V, we shall discuss the remaining problems and other related topics.

II. THE POSITIVE ENERGY CONJECTURE

To fix the notations, we shall briefly review the positive energy conjecture. An initial data set for a space–time consists of an R^3 manifold S on which there are given a positive-definite metric q_{ab}, a symmetric tensor field p^{ab} (the extrinsic curvature), a local mass density μ, and a local current density J^a. These fields on S must obey the constraint equations

$$R - p^{ab}p_{ab} + p^2 = 2\mu, \tag{1}$$

$$D_a(p^{ab} - pq^{ab}) = J^b, \tag{2}$$

where R is the Ricci scalar of the metric q_{ab}, p is the trace of p^{ab}, and D_a is the covariant derivative operator defined by q_{ab}. Furthermore, μ and J^a must obey the local energy condition,

$$\mu \geq (J^a J_a)^{1/2}. \tag{3}$$

An initial data set is said to be asymptotically flat if in the asymptotic region the metric q_{ab} approaches the Euclidean metric not slower than $1/r$ and $p^{ab}p_{ab}$ approaches zero not slower than $1/r^4$, where r is any typical radial distance.[4] From now on an initial data set is always assumed to be asymptotically flat. With any (asymptotically flat) initial data set one can define the total energy, called ADM energy,[1] of the system as a flux integral over a topological 2-sphere in the asymptotically distant region. Although there are various forms of this flux integral, they all give the same value for an asymptotically flat initial data set. The following form, proposed by Geroch,[3,5] turns out to be most convenient for the positive energy conjecture,

$$E = c^{-1}A^{1/2}\int(2\Omega - \pi^2)\,dA, \tag{4}$$

where c is a constant which turns out to be $c = 64\pi^{3/2}$, A is the area of the integration surface which asymptotically approaches a metric sphere in the asymptotic region, Ω is the intrinsic scalar curvature of the integration surface, and π is the trace of the extrinsic curvature of the integration surface as a submanifold of S.

The positive energy conjecture states the following. For a nonsingular asymptotically flat initial data set, the ADM energy E is nonnegative and vanishes if and only if

$$R^{ab} = p^{am}p^b_m - p\,p^{ab}, \tag{5}$$

$$D^{[a}(p^{b]c} - pq^{b]c}) = 0 \tag{6}$$

(i.e., the initial data set is that for Minkowski space). Note that $\mu = 0$ and $J^a = 0$ follows from (5) and (6).

III. DATA FOR MINKSOWSKI SPACE AND *W*-TRANSFORMATION

An initial data set with flat metric and vanishing p^{ab} is obviously a data set for Minkowski space. There are, however, many other initial data sets for Minkowski space which are not so easy to identify. These data sets can be characterized by the fact that they satisfy Eqs. (5) and (6). Even though this is a legitimate characterization, Eqs. (5) and (6) are cumbersome conditions to be identified as consequences of vanishing E. In view of this difficulty, we shall introduce another characterization of the initial data sets for Minkowski space in the following theorem.[6]

Theorem I: An initial data set, i.e., an R^3-manifold with positive definite metric q_{ab} and tensor field p^{ab}, is

[a)]Supported by NSF Grant MPS74-15246.

 0022-2488/78/1905-1152$1.00

that for Minkowski space if and only if there exist a scalar field f and a flat metric k_{ab} as follows:

$$p^{ab} = D^a D^b f/(1 + D^n f D_n f)^{1/2}, \tag{7}$$

$$q_{ab} = k_{ab} - D_a f D_b f. \tag{8}$$

In the Appendix we shall give a proof of this theorem.

This theorem suggests to us the following approach to the positive energy conjecture. Given an initial data set, choose a scalar field w on S such that w will reduce to f of the Theorem I when the given set is that for Minkowski space. Next, introduce new quantities $\bar{p}^{ab}$ and $\bar{q}_{ab}$ such that

$$\bar{p}^{ab} = p^{ab} - D^a D^b w/(1 + D^n w D_n w)^{1/2}, \tag{9}$$

$$\bar{q}_{ab} = q_{ab} + D_a w D_b w. \tag{10}$$

Rewrite the constraint equations in terms of the barred quantities. Then, what we need to show is that E is nonnegative and vanishes if and only if $\bar{p}^{ab} = 0$ and $\bar{q}^{ab}$ is flat.

Following this program, we shall first specify how to choose w for a given data set. Let p^{ab} and q_{ab} be an asymptotically flat initial data set for Minkowski space. Then there exists a scalar field f on S which satisfies (7) and (8). Therefore, f also satisfies the following quasilinear elliptic differential equation

$$\left[p^{ab} - \frac{D^a D^b f}{(1 + D^n f D_n f)^{1/2}}\right]\left(q_{ab} - \frac{D_a f D_b f}{1 + D^n f D_n f}\right) = 0$$

with the asymptotic condition $|D_a f| = O(1/r)$. With this observation, we proceed as follows. For a given asymptotically flat initial data set, let w be a solution of the following quasilinear elliptic partial differential equation

$$\left[p^{ab} - \frac{D^a D^b w}{(1 + D^n w D_n w)^{1/2}}\right]\left(q_{ab} - \frac{D_a w D_b w}{1 + D^n w D_n w}\right) = 0 \tag{11}$$

with the asymptotic condition $|D_a w| = O(1/r)$. Here we assume such a solution exists. We postpone to the conclusion further discussion of this issue.

Next, we introduce $\bar{p}^{ab}$ and $\bar{q}_{ab}$ as in Eqs. (9) and (10) with this w. Note that the difference between q_{ab} and $\bar{q}_{ab}$ is the order of $1/r^2$ in the asymptotic region. Hence the total energy defined in terms of q_{ab} is the same as that in terms of $\bar{q}_{ab}$. That is to say,

$$\begin{aligned} E &= c^{-1} A^{1/2} \int (2\Omega - \pi^2)\, dA \\ &= c^{-1} \bar{A}^{1/2} \int (2\bar{\Omega} - \bar{\pi}^2)\, d\bar{A}, \end{aligned} \tag{12}$$

where the barred quantities $\bar{A}$, $\bar{\Omega}$, and $\bar{\pi}$ represent the corresponding quantities with respect to $\bar{q}_{ab}$ on S. Hence $\bar{A}$ is the area of the integration surface with respect to the metric $\bar{q}_{ab}$ on S. Considering the integration 2-surface as a submanifold of S, $\bar{\Omega}$, and $\bar{\pi}$ are the scalar curvature of the induced metric of the 2-surface from $\bar{q}_{ab}$ and the trace of the extrinsic curvature of the 2-surface determined from $\bar{q}_{ab}$, respectively.

We now rewrite the constraint equations replacing the unbarred quantities (e.g., q_{ab} and p^{ab}) by the barred quantities (e.g., $\bar{q}_{ab}$ and $\bar{p}^{ab}$). Let $\bar{D}_a$ be the derivative operator defined by $\bar{q}_{ab}$, that is to say, the derivative operator which satisfied $\bar{D}_a \bar{q}_{bc} = 0$. The relationship between D_a and $\bar{D}_a$ is the following:

$$\begin{aligned} \bar{D}_a T^{b\cdots c}{}_{d\cdots e} = {} & D_a T^{b\cdots c}{}_{d\cdots e} + C^b{}_{am} T^{m\cdots c}{}_{d\cdots e} + \\ & + \cdots + C^c{}_{am} T^{b\cdots m}{}_{d\cdots e} - C^m{}_{ad} T^{b\cdots c}{}_{m\cdots e} - \\ & - \cdots - C^m{}_{ae} T^{b\cdots c}{}_{d\cdots m} \end{aligned} \tag{13}$$

for arbitrary tensor field $T^{b\cdots c}{}_{d\cdots e}$ on S, where

$$C^b{}_{am} = (1 + D^n w D_n w)^{-1} D^b w D_a D_m w. \tag{14}$$

Since curvature tensor is obtained from the derivative operator, and since we know the relationship between D_a and $\bar{D}_a$, we can find the relationship between R_{ab} and $\bar{R}_{ab}$, where $\bar{R}_{ab}$ is the Ricci tensor obtained from $\bar{D}_a$. We have

$$\bar{R}_{ab} = R_{ab} + D_m C^m{}_{ab} - D_a C^m{}_{bm} + C^m{}_{ab} C^n{}_{mn} - C^m{}_{an} C^n{}_{bm}. \tag{15}$$

Using these facts, we can rewrite the constraint equations as follows:

$$\begin{aligned} J^a = {} & \bar{D}_b \bar{p}^{ab} - q^{ac} D_c (\bar{p}^{db} q_{db}) + K^{1/2} \bar{R}^{ab} \bar{D}_b w \\ & + K^{3/2} \bar{D}^a w (\bar{R}^{mn} \bar{D}_m w \bar{D}_n w) - K(\bar{D}^a w \bar{p}^{mn} \bar{D}_m \bar{D}_n w \\ & + \bar{p}^{am} \bar{D}^n w \bar{D}_m \bar{D}_n w), \end{aligned} \tag{16}$$

$$\begin{aligned} 2\mu = {} & \bar{R} - \bar{p}^{ab} \bar{p}^{cd} q_{ac} q_{bd} - 2K^{1/2} \bar{p}^{ab} \bar{D}_a \bar{D}_b w \\ & + 2K \bar{R}^{ab} \bar{D}_a w \bar{D}_b w + (\bar{p}^{ab} q_{ab})^2 \\ & + 2K^{-1/2} (\bar{p}^{ab} q_{ab}) q^{cd} D_c D_d w, \end{aligned} \tag{17}$$

where $K = 1 + D_a w D_b w q^{ab}$, $\bar{D}_a w = D_a w$, and $\bar{D}^a w = \bar{q}^{ab} \bar{D}_b w$. Eliminating $\bar{R}^{ab} \bar{D}_a w \bar{D}_b w$ from (17) using (16), we obtain

$$\begin{aligned} 2[\mu - K^{-1/2} D_a w J^a] = {} & \bar{R} - \bar{p}^{ab} \bar{p}^{cd} q_{ac} q_{bd} + (\bar{p}^{ab} q_{ab})^2 \\ & - 2\bar{D}_a [K^{-1/2} \{\bar{p}^{ab} - (\bar{p}^{mn} q_{mn}) q^{ab}\} D_b w]. \end{aligned} \tag{18}$$

This equation is the only information from constraint equations which we shall need. Note that the left side of Eq. (18) is nonnegative because of the local energy condition (3).

IV. A POSITIVE ENERGY ARGUMENT

Introduce a function t on S such that the two-dimensional surfaces $t = \text{const}$ in S are nested topological 2-spheres with the innermost surface reducing to a point. Define a scalar field $\bar{\phi}$ on S by

$$\bar{\phi} \bar{r}^a D_a t = 1, \tag{19}$$

where $\bar{r}^a$ denotes the unit outward vector field normal to the 2-surfaces with respect to $\bar{q}_{ab}$. Relative to this $\bar{r}^a$ we decompose the vector which appears in the last term of Eq. (18) as follows:

$$K^{-1/2} \{\bar{p}^{ab} - (\bar{p}^{mn} q_{mn}) q^{ab}\} D_b w = v \bar{r}^a + T^a, \tag{20}$$

where $T^a \bar{r}^b \bar{q}_{ab} = 0$. Note $v = O(1/r^3)$ asymptotically since $p^{ab} = O(1/r^2)$ and $|D_a w D_b w q^{ab}| = O(1/r^2)$ asymptotically.

For each value of t, set

$$f(t) = \int [2\bar{\Omega} - (\bar{\pi} + v)^2]\, d\bar{A}, \tag{21}$$

where the integral extends over the surface $t = \text{const}$. Since $v = O(1/r^3)$ and $\bar{\pi} = O(1/r)$ asymptotically, this $f(t)$, for $t \to \infty$, reduces to the integral in the energy expression (12). Therefore, once we show that $f(t) \geq 0$ as $t \to \infty$ we can conclude that $E \geq 0$.

Let ξ_{ab} be the induced metric on the $t = \text{const}$ surfaces

from $\bar{q}_{ab}$. Let $\bar{\pi}^{ab}$ be the extrinsic curvature of the $t=$ const surface with respect to $\bar{q}_{ab}$. Then, the rate of change of ξ_{ab} and $\bar{\pi}=\bar{\pi}^{ab}\bar{q}_{ab}$ with respect to t (i.e., their Lie derivative by $\bar{\phi}\bar{r}^a$) are given by

$$\mathcal{L}_{\bar{\phi}\bar{r}}\bar{\xi}_{ab}=2\bar{\phi}\bar{\pi}_{ab}, \tag{22}$$

$$\mathcal{L}_{\bar{\phi}\bar{r}}\bar{\pi}=-\tfrac{1}{2}\bar{\phi}\bar{\pi}^{ab}\bar{\pi}_{ab}+\tfrac{1}{2}\bar{\phi}\bar{\Omega}-\tfrac{1}{2}\bar{\phi}\bar{R}-\tfrac{1}{2}\bar{\phi}\bar{\pi}^2-\bar{\partial}^a\bar{\partial}_a\bar{\phi}, \tag{23}$$

where, as the notation suggests, $\bar{\pi}_{ab}$ is defined by $\bar{\pi}_{ab}=\bar{\pi}^{mn}\bar{q}_{ma}\bar{q}_{nb}$, and where $\bar{\partial}_a$ denotes the derivative operator on $t=$ const surfaces defined by the induced metric $\bar{\xi}_{ab}$ from $\bar{q}_{ab}$.

Substituting (20) into (18), we obtain

$$2[\mu-K^{-1/2}D_a wJ^a]=\bar{R}-\bar{p}^{ab}\bar{p}^{cd}q_{ac}q_{bd}+(p^{ab}q_{ab})^2 -2\bar{D}_a(v\bar{r}^a)-2\bar{\partial}_a T^a-2\bar{\phi}^{-1}T^a\bar{\partial}_a\bar{\phi}. \tag{24}$$

This equation tells us the rate of change of v with respect to t.

Using the Gauss–Bonnet theorem and Eqs. (22), (23), and (24) the rate of change of $f(t)$ with respect to t is obtained:

$$\begin{aligned}\frac{df(t)}{dt}&=-\int\bar{\phi}D_a[(\pi+v)^2\bar{r}^a]\,d\bar{A}\\&=-\tfrac{1}{2}\int\bar{\phi}(\bar{\pi}+v)[2\bar{\Omega}-(\bar{\pi}+v)^2]\,d\bar{A}\\&\quad+2\int\bar{\phi}(\bar{\pi}+v)[(\mu-K^{-1/2}J^a w_a)+\tfrac{1}{2}(\bar{\pi}^{ab}\bar{\pi}_{ab}-\tfrac{1}{2}\bar{\pi}^2)\\&\quad+\partial_a T^a+\bar{\phi}^{-1}\bar{\partial}^a\bar{\partial}_a\bar{\phi}+\bar{\phi}^{-1}T^a\bar{\partial}_a\bar{\phi}\\&\quad+\bar{\pi}^{ab}\bar{\pi}^{cd}q_{ac}q_{bd}-(\bar{\pi}^{ab}q_{ab})^2-\tfrac{1}{2}v^2]\,d\bar{A}.\end{aligned} \tag{25}$$

Analysis up to now works for any one-parameter family of 2-surfaces. We now choose the 2-surfaces such that

$$\bar{\phi}(\bar{\pi}+v)=1. \tag{26}$$

Here we assume such a choice can be made. We postpone to the conclusion further discussion of this issue.

Then, Eq. (25) becomes

$$\begin{aligned}\frac{df(t)}{dt}&=-\tfrac{1}{2}f(t)+2\int[(\mu-K^{-1/2}J^aD_a w)\\&\quad+\tfrac{1}{2}(\bar{\pi}^{ab}\bar{\pi}_{ab}-\tfrac{1}{2}\bar{\pi}^2)+\bar{\partial}_a T^a+\bar{\partial}_a(\bar{\phi}^{-1}\bar{\partial}^a\bar{\phi})]\,d\bar{A}\\&\quad+\int[\bar{\pi}^{ab}\bar{\pi}^{cd}q_{ac}q_{bd}-(\bar{\pi}^{ab}q_{ab})^2-\tfrac{1}{2}v^2\\&\quad+2\bar{\phi}^{-1}T^a\bar{\partial}_a\phi+2\bar{\phi}^{-2}\bar{\partial}_a\bar{\phi}\bar{\partial}^a\bar{\phi}]\,d\bar{A}\\&\geq-\tfrac{1}{2}f(t)+\int[\bar{\pi}^{ab}\bar{\pi}^{cd}q_{ac}q_{bd}-(\bar{\pi}^{ab}q_{ab})^2-\tfrac{1}{2}v^2\\&\quad+2\bar{\phi}^{-1}T^a\bar{\partial}_a\phi+2\bar{\phi}^{-2}\bar{\partial}_a\phi\,\bar{\partial}^a\phi]\,d\bar{A}.\end{aligned} \tag{27}$$

The above inequality is obtained by using the local energy condition, the fact $(\bar{\pi}^{ab}\bar{\pi}_{ab}-\tfrac{1}{2}\bar{\pi}^2)\geq 0$ and that an integration of a divergence of a vector field on a 2-sphere vanishes.

The last integral of Eq. (27) is also nonnegative because of the defining equation of w, (11). This can be shown as follows. Let $n^a=(D_m wD_n wq^{mn})^{-1/2}D^a w$ and decompose $\bar{p}^{ab}$ with respect to n^a,

$$\bar{p}^{ab}=\alpha n^a n^b+n^{(a}\beta^{b)}+\gamma^{ab},$$

where $\beta^a n^b q_{ab}=0$, $\gamma^{ab}=\gamma^{(ab)}$, and $\gamma^{ab}n^c q_{ac}=0$. Then, Eq. (20) can be written as

$$v\bar{r}^a+T^a=K^{-1/2}(D_m wD_n wq^{mn})^{1/2}(\tfrac{1}{2}\beta^a-\gamma n^a).$$

Taking the norm of $(v\bar{r}^a+T^a)$ with respect to $\bar{q}_{ab}$, we obtain

$$v^2+T^aT^b\bar{q}_{ab}=K^{-1}(D_m wD_n wq^{mn})(\tfrac{1}{4}\beta^a\beta_a+K\gamma^2). \tag{28}$$

The defining equation (11) of w can now be rewritten in terms of α and γ as follows:

$$K^{-1}\alpha+\gamma=0. \tag{29}$$

Using the above two equations, (28) and (29), the last integral of Eq. (26) can be written as

$$\begin{aligned}&\int[\bar{\pi}^{ab}\bar{\pi}^{cd}q_{ac}q_{bd}-(\bar{\pi}^{ab}q_{ab})^2-\tfrac{1}{2}v^2\\&\quad+2\bar{\phi}^{-1}T^a\bar{\partial}_a\bar{\phi}+2\bar{\phi}^{-2}\bar{\partial}_a\bar{\phi}\bar{\partial}^a\bar{\phi}]\,d\bar{A}\\&=\int[\beta^a\beta_a(\tfrac{1}{2}-\tfrac{1}{8}K^{-1}D_n wD_m wq^{mn})+(\gamma^{ab}\gamma_{ab}-\tfrac{1}{2}\gamma^2)\\&\quad+\tfrac{1}{2}(T^a+2\bar{\phi}^{-1}\bar{\partial}^a\phi)(T^b+2\bar{\phi}^{-1}\bar{\partial}^b\bar{\phi})\bar{q}_{ab}+\tfrac{3}{2}K^{-1}\alpha^2]\,d\bar{A}\geq 0.\end{aligned} \tag{30}$$

Therefore, we obtain finally the following inequality:

$$\frac{df(t)}{dt}\geq-\tfrac{1}{2}f(t) \tag{31}$$

As we move outward from one $t=$ const surface to the next, t will increase or decrease depending on the sign of the evolution function $\bar{\phi}$. The sign of $\bar{\phi}$, in turn, is determined by the sign of $(\bar{\pi}+v)$ through the equation $(\bar{\pi}+v)\phi=1$. First, suppose t always increases as we move outward. Then, since $f(t)\to 0$ as the surfaces reduce to a point, Eq. (31) implies $f(t)\geq 0$ for every t. Second, suppose in some region of S, t decreases as we move outward. However, that region should be bounded since $\bar{\pi}$ is positive and is order of $1/r$ and v is order of $1/r^2$ in the asymptotic region. Now, let $t=\underline{t}$ be the boundary of the region which includes the asymptotic region and in which t increases as we move outward. Hence, across $t=\underline{t}$ surface $(\bar{\pi}+v)$ changes sign, in other words, $(\bar{\pi}+v)$ vanishes on the $t=\underline{t}$ surface. Then, we observe that $f(\underline{t})=16\pi$ because of the Gauss–Bonnet theorem. This positivity of $f(\underline{t})$ combined with the validity of Eq. (31) from there on implies that $f(t)\geq 0$ for surfaces outside $t=\underline{t}$ surface. Hence, we conclude

$$\lim_{t\to\infty}f(t)\geq 0$$

for both of the above two cases. It now follows from the remarks about Eq. (21) that E is nonnegative. Furthermore, E can vanish only when $f\to 0$ at $t\to\infty$, and this is possible only when $f(t)$ vanishes for every t because of Eq. (31). The $f(t)$ can be zero for every t only when $\bar{p}^{ab}=0$ and $\bar{q}_{ab}$ is flat everywhere because of Eqs. (27), (30), and (31). In other words, E vanishes only when p^{ab} and q_{ab} are initial data for Minkowski space.

V. CONCLUSION

The argument presented in the previous two chapters is complete except that we assumed the existence of scalar field w which satisfies the quasilinear elliptic partial differential equation (11) with the asymptotic condition $|D_a w|=O(1/r)$ and that we assumed the existence of one parameter t family of smooth 2-surfaces which can be obtained from one point by the evolution equation $\phi=(\bar{\pi}+v)^{-1}$. Some intuitive arguments suggest that these two field t and w exist for smooth metric and

the extrinsic curvature on a smooth R^3-manifold S. However, as far as the author knows, the existence of such l and w have not yet been proved rigorously. Hence, a precise mathematical result of our argument would be that the positive energy conjecture is true for initial data sets which admits two scalar fields l and w satisfying the above conditions.

An interesting feature of Geroch's argument for the time-symmetric initial data set is that a slight modification of it rules out a possible class of counterexamples to the cosmic censor hypothesis proposed by Penrose in that case.[5] Since we now have generalized Geroch's argument to the full energy conjecture, it would be interesting to see if one can modify this full argument to rule out the counterexamples to the cosmic censor hypothesis in general.

Finally, it might be appropriate to mention that any further argument for the existence of l and w should distinguish the initial data sets with an apparent horizon from the others. When there exists an apparent horizon H, on S, difficulties with the w equation and with the surface evolution of l could occur across H. In this case case, one can hope only that w exists outside H with the boundary condition that $\nabla_a w$ is orthogonal to H and only that l exists outside H with the property that the 2-surface $l=0$ is H. Once this can be shown rigorously, the validity of $df/dl \geq -f/2$ outside H, and $f \to 16\pi$ as $l \to 0^+$ would lead to the conclusion that $E \geq 0$.

ACKNOWLEDGMENTS

I wish to thank Joshua Goldberg, Peter Bergmann, Edward Flaherty, and especially Robert Geroch for many interesting discussions.

APPENDIX: PROOF OF THEOREM I

Let a three-dimensional surface S with a positive definite metric q_{ab} and an extrinsic curvature p^{ab} be an initial data set for Minkowski space with metric g_{ab}. We shall show that there exists f and k_{ab} on S which satisfy Eqs. (7) and (8). Let $\hat{S}$ be a three-dimensional Euclidean plane in Minkowski space. Then these two manifolds S and $\hat{S}$ are diffeomorphic to each other by the following mapping ψ. Let p be a point of $\hat{S}$. We draw a timelike geodesic through p, which is perpendicular to the surface $\hat{S}$. This geodesic meets S at one and only one point p. The ψ maps $\hat{p}$ to p. On $\hat{S}$ we have a flat metric $\hat{k}_{ab}$. By the diffeomorphism ψ, $\hat{k}_{ab}$ induces a flat metric k_{ab} on S. Let the length of the geodesic between p and $\hat{p}$ be f. This f can be considered as a scalar field on S. We shall show that the fields, f and k_{ab}, on S satisfies Eqs. (7) and (8).

Let $\hat{\xi}^a$ be a constant vector field which is tangential to $\hat{S}$. Let $\hat{n}^a$ be a unit timelike constant field which is orthogonal to $\hat{S}$. Let ξ^a denote the induced vector field on S from $\hat{\xi}^a$ on $\hat{S}$ by the diffeomorphism ψ. Then ξ^a and $\hat{\xi}^a$ on S are related by the following equation:

$$\xi^a = \hat{\xi}^a + \hat{n}^a(\xi^b D_b f),$$

where D_a is the derivative operator on S defined by q_{ab}. Using the definitions of q_{ab} and $\hat{k}_{ab}$, we obtain

$$\begin{aligned}\xi^a \xi\, q_{ab} = \xi^a \xi^b g_{ab} &= \hat{\xi}^a \hat{\xi}^b \hat{k}_{ab} - \xi^a \xi^b D_a f D_b f \\ &= \xi^a \xi^b k_{ab} - \xi^a \xi^b D_a f D_b f.\end{aligned}$$

This equation holds for arbitrary vector ξ^a at p. Therefore, we have

$$q_{ab} = k_{ab} - D_a f D_b f.$$

The unit normal vector field n^a to S can be expressed as follows:

$$n^a = (\hat{n}^a + D^a f)/(1 + D^n f D_n f)^{1/2}.$$

Now, the extrinsic curvature p^{ab} of S can be calculated

$$\begin{aligned}p^{ab} = q^{am} q^{bn} \nabla_m n_n &= q^{am} q^{bn} \frac{\nabla_m D_n f}{(1 + D^n f D_n f)^{1/2}} \\ &= \frac{D^a D^b f}{(1 + D^n f D_n f)^{1/2}},\end{aligned}$$

where ∇_m is the covariant derivative operator on Minkowski space. This completes the necessary part of the proof.

To show the sufficient part, we reverse the above construction. Suppose q_{ab}, p^{ab}, k_{ab} and f on a R^3-manifold S satisfy Eqs. (7) and (8). First, choose a Euclidean plane $\hat{S}$ with the induced metric $\hat{k}_{ab}$ in Minkowski space. The S with k_{ab} is isometric to $\hat{S}$ with $\hat{k}_{ab}$, because both are R^3 manifold with flat metric. From the isometry we have an induced scalar field f on $\hat{S}$. Using this f as the length of the timelike geodesics from $\hat{S}$ to S', we draw the surface S' in Minkowski space. Then it is immediate that the surface S' with the induced metric q'_{ab} from the Minkowski metric is the immersion of the starting manifold S with q_{ab} such that the extrinsic curvature p'^{ab} of S' coincides with the given tensor p^{ab} of S (more precisely, p'^{ab} of S' coincides with the induced tensor on S' from the given tensor p^{ab} on S).

[1] R. Arnowitt, S. Deser and C.W. Misner, Phys. Rev. **118**, 1100 (1960); **121**, 1556 (1961); **122**, 997 (1961); R. Geroch, J. Math. Phys. **13**, 956 (1972).

[2] D. Brill and S. Deser, Ann. Phys. **50**, 548 (1968); D. Brill, S. Deser, and L.D. Fadeev, Phys. Lett. **26A**, 538 (1968); P.S. Jang, J. Math. Phys. **17**, 141 (1976); N.S. Murchadha and J.W. York, Phys. D **10**, 2345 (1974); Y. Choquet and J. Marsden, C.R. Acad. Sci. **282**, 609 (1976); R. Geroch, Proc. Sym. Pure Math. Vol. XXVII, 401 (1973).

[3] R. Geroch, Ann. N.Y. Acad. Sci. **224**, 108 (1973).

[4] For a more precise definition of asymptotic flatness, see R. Geroch, J. Math. Phys. **13**, 956 (1972).

[5] P.S. Jang and R. Wald, J. Math. Phys. **18**, 41 (1977).

[6] This theorem is an unpublished result of Geroch.

A simple model of the integrable Hamiltonian equation[a)]

Franco Magri

Istituto di Matematica del Politecnico, Piazza Leonardo da Vinci 32, 21033 Milano, Italy
(Received 18 April 1977)

A method of analysis of the infinite-dimensional Hamiltonian equations which avoids the introduction of the Bäcklund transformation or the use of the Lax equation is suggested. This analysis is based on the possibility of connecting in several ways the conservation laws of special Hamiltonian equations with their symmetries by using symplectic operators. It leads to a simple and sufficiently general model of integrable Hamiltonian equation, of which the Korteweg–de Vries equation, the modified Korteweg–de Vries equation, the nonlinear Schrödinger equation and the so-called Harry Dym equation turn out to be particular examples.

INTRODUCTION

The aim of this paper is to suggest a constructive approach to the infinite-dimensional integrable Hamiltonian equations, i.e., to the evolution equations possessing an infinite sequence of independent integrals which are in involution. The present analysis is based on the study of the connection between the symmetries and the conservation laws of the evolution equations. The main result is in showing a simple model of integrable Hamiltonian equation, of which the Korteweg–de Vries equation, the modified Korteweg–de Vries equation, the nonlinear Schrödinger equation, and the so-called Harry Dym equation turn out to be particular examples.

The analysis proceeds as follows. In Sec. 1 it is shown that any conservation law of an infinite-dimensional Hamiltonian equation is connected with a symmetry transformation. The study of the connection between the symmetries and the conservation laws of a given evolution equation is thus reduced to the study of its Hamiltonian structures. In Sec. 2 it is shown by an example that a given evolution equation may be endowed with different Hamiltonian structures. Each of them provides a way of connecting the conservation laws with the symmetries. Let us then consider an equation endowed with two of such connections, and let us use the former to associate the conservation laws with the symmetries and the latter to conversely associate the symmetries with the conservation laws. One is thus able to obtain a new conservation law from a given one. In Sec. 3 it is shown that highly ordered chains of integrals which are in involution can be constructed in this way for special twofold Hamiltonian equations. Such equations provide a simple model of integrable Hamiltonian equation. The examples of Sec. 5 seem to suggest that this model is not only conceptually simple but also effective in the applications.

1. SYMMETRIES AND CONSERVATION LAWS OF HAMILTONIAN EVOLUTION EQUATIONS

In this section an operator approach to the symmetries and to the conservation laws of any system of evolution equations

$$\partial_t u^A(\underline{x}, t) = k^A(u^B, u^B_j, u^B_{jl}, \cdots), \tag{1.1}$$

is suggested. The field functions $u^A(\underline{x}, t)$ are supposed to be defined, at any instant of time, in a fixed region Ω of $\mathbb{R}^3$ and to vanish on the boundary of this region; the subscripts denote the partial derivatives of these functions with respect to the space coordinates x^j.

We set up the study of Eq. (1.1) into the linear space U of the field functions regarded as functions of the space coordinates only. Consequently, any n-tuple $u^A(\underline{x}, t_0)$ will be simply denoted by $u(t_0)$ and will be referred to as a point of this space. The given evolution equations will be synthetized into the single operator equation

$$\partial_t u = \mathbf{K}(u), \tag{1.2}$$

where $\mathbf{K}$ is the formal differential operator defined by the functions $k^A(u^B, u^B_j, \cdots)$. The space U will be called the configuration space associated with the abstract evolution equation (1.2). The purpose of this operator approach is to suggest a simple way of extending to infinite-dimensional systems the geometric analysis developed for the classical Hamiltonian mechanics in the phase space.[1]

A. Symmetries

The object of the theory of the symmetries is the study of the manifold of the solutions of Eq. (1.2) in the configuration space U. We shall limit ourselves to a local study of such a manifold and so we only consider the infinitesimal symmetry transformations. They are the infinitesimal point transformations

$$\bar{u} = u + \epsilon \mathbf{S}(u), \tag{1.3}$$

of the configuration space into itself which map every solution again into a solution.[2] The operator $\mathbf{S}$ is called the generator of the symmetry mapping and is regarded as defining a "contravariant" vector field on the space U. The lines of this vector field are the orbits of the symmetry mapping.

The symmetry condition is readily obtained if one observes that for any solution $u(t)$ it is

$$\partial_t \bar{u} - \mathbf{K}(\bar{u})$$

$$\overset{(A2)}{=} \partial_t \cancel{u} + \epsilon \partial_t \mathbf{S}(u) - \mathbf{K}(\cancel{u}) - \epsilon \mathbf{K}'_u \mathbf{S}(u)$$

a)This work has been sponsored by the Consiglio Nazionale delle Ricerche, Gruppo per la Fisica-Matematica.

 0022-2488/78/1905-1156$1.00

$$\overset{(A6)}{=} \epsilon[\mathbf{S}'_u \partial_t u - \mathbf{K}'_u \mathbf{S}(u)] \tag{1.4}$$

$$\overset{(1.2)}{=} \epsilon[\mathbf{S}'_u \mathbf{K}(u) - \mathbf{K}'_u \mathbf{S}(u)],$$

where $\mathbf{S}'_u$ is the Gateaux derivative of the operator $\mathbf{S}$, which is supposed to not explicitly depend on the time (see Appendix A). Hence, the *symmetry condition* is

$$\mathbf{S}'_u \mathbf{K}(u) - \mathbf{K}'_u \mathbf{S}(u) \triangleq 0, \tag{1.5}$$

where the symbol $\triangleq$ means that the equality is required to hold only for the solutions. For simplicity, however, in this paper we shall only consider the symmetry generators for which condition (1.5) is identically verified (the condition being *a fortiori* verified on the manifold of the solutions).

Equation (1.5) expresses the structural relation which connects the given equation to its symmetries, independently of the specific form either of the equation or of the symmetry mappings. It is a commutation relation, the left side being the commutator of the two nonlinear operators $\mathbf{S}$ and $\mathbf{K}$.[3] Therefore, the set of the generators of the symmetry mappings constitutes a Lie algebra. This means that if two of such generators $\mathbf{S}_j$ and $\mathbf{S}_k$ are composed according to the formula

$$[\mathbf{S}_j, \mathbf{S}_k](u) \equiv \mathbf{S}'_{ju} \mathbf{S}_k(u) - \mathbf{S}'_{ku} \mathbf{S}_j(u), \tag{1.6}$$

a third generator is obtained again.

B. Conservation laws

The study of the manifold of the solutions is the object also of the theory of the conservation laws, but the standpoint is different and, so to speak, dual to that of the theory of the symmetries.

Besides the configuration space U, one considers a second space V, put in duality with U by a convenient bilinear form $\langle v, u \rangle$,[4] and then one considers the operators $\mathbf{Q}: U \to V$ (see Fig. 1). Such operators may be regarded as defining the "covariant" vector fields on U. For such fields it is possible to introduce the concept of elementary circulation

$$\delta C \equiv \langle \mathbf{Q}(u), \delta u \rangle, \tag{1.7}$$

and so it is possible to consider the *conservative* covariant vector fields, for which the circulation does not depend on the line but only on the endpoints. As is known,[4] in order that the field be conservative it is

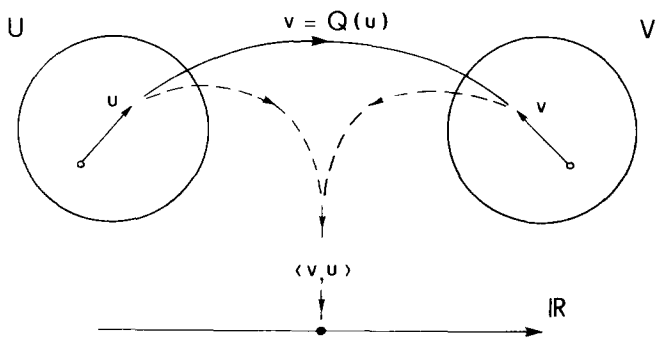

FIG. 1.

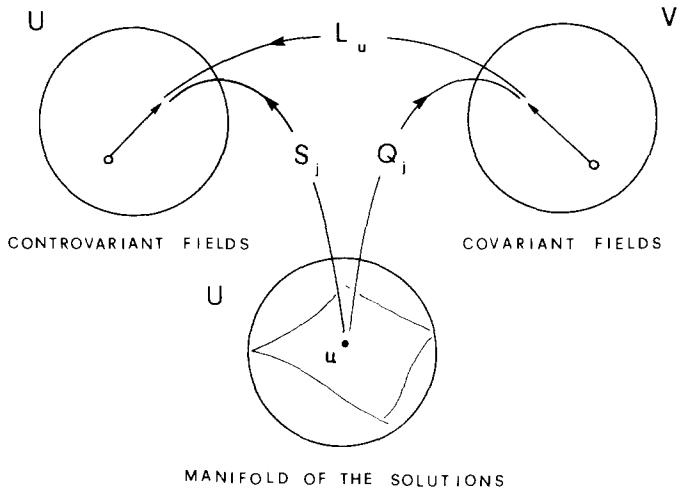

FIG. 2.

necessary and sufficient that

$$\langle \mathbf{Q}'_u du, \delta u \rangle = \langle \mathbf{Q}'_u \delta u, du \rangle, \tag{1.8}$$

for any pair of variations du and δu of the field functions. The operators $\mathbf{Q}$ verifying this condition are called *potential operators*.[4] For such operators the circulation from a fixed point u_0 to any point u defines a functional $F[u]$, so that the elementary circulation is given by

$$\delta F[u] = \langle \mathbf{Q}(u), \delta u \rangle. \tag{1.9}$$

For this reason, the operator $\mathbf{Q}$ is also called the *gradient* of the functional F.

The theory of the conservation laws associates a special set of conservative covariant vector fields with the given equation by the requirement that the corresponding functionals keep their value $F[u(t)]$ independent of t for any solution $u(t)$. These functionals are called *integrals*[5] of the given equation and the corresponding potential operators may be called *"integrating"* operators.[6] Therefore, the theory of the symmetries studies the manifold of the solutions by using contravariant vector fields while the theory of the conservation laws studies the same manifold by using covariant vector fields. In this sense the two formalisms are dual.

The following condition

$$\langle \mathbf{Q}(u), \mathbf{K}(u) \rangle \triangleq 0, \tag{1.10}$$

on the integrating operators $\mathbf{Q}$ is readily obtained if one observes that it is

$$\partial_t F[u(t)] \overset{(1.9)}{=} \langle \mathbf{Q}(u), \partial_t u \rangle$$

$$\overset{(1.2)}{=} \langle \mathbf{Q}(u), \mathbf{K}(u) \rangle, \tag{1.11}$$

for any solution $u(t)$. As in the case of the symmetry generators, however, we shall only consider the integrating operators for which the condition (1.10) is identically verified.

C. Connecting the conservation laws with the symmetries

The problem of connecting the conservation laws with

the symmetries requires the introduction of a metric operator, which associates the covariant with the contravariant vector fields as is usual in Riemannian geometry. By a *metric operator* it is meant a linear operator $\mathbf{L}_u : V \to U$ (which may nonlinearly depend on the point u) mapping the covariant vector fields $\mathbf{Q}_j$ into the contravariant vector fields $\mathbf{S}_j$ according to the relation

$$\mathbf{S}_j(u) = \mathbf{L}_u \mathbf{Q}_j(u) \tag{1.12}$$

(examples will be given in Secs. 2 and 5, see Fig. 2).

It is the purpose of this subsection to study the special class of the metric operators verifying the following two conditions:

$$\langle dv, \mathbf{L}_u \delta v\rangle = -\langle \delta v, \mathbf{L}_u\, dv\rangle, \tag{1.13}$$

$$\langle dv, \mathbf{L}'_u(\delta v; \mathbf{L}_u \Delta v)\rangle + \langle \delta v, \mathbf{L}'_u(\Delta v; \mathbf{L}_u\, dv)\rangle + \langle \Delta v, \mathbf{L}'_u(dv; \mathbf{L}_u \delta v)\rangle = 0, \tag{1.14}$$

where $\mathbf{L}'_u$ is the Gateaux derivative of $\mathbf{L}_u$ with respect to u (see Appendix B), and to show that they allow to connect the integrating operators with the symmetry generators of the evolution equations. Such metric operators will be called *symplectic operators* with respect to the prefixed bilinear form $\langle v, u\rangle$.

To this end, consider any operator $\mathbf{S}_j$ associated with a *potential* operator $\mathbf{Q}_j$ by means of a symplectic operator $\mathbf{L}_u$. It is called a *Hamiltonian operator* and it verifies the following condition,

$$\langle dv, \mathbf{S}'_{ju} \mathbf{L}_u \delta v\rangle - \langle \delta v, \mathbf{S}'_{ju} \mathbf{L}_u\, dv\rangle = \langle dv, \mathbf{L}'_u(\delta v; \mathbf{S}_j(u))\rangle, \tag{1.15}$$

(see Appendix B). This condition implies that the commutator $[\mathbf{S}_j, \mathbf{S}_k]$ of any pair of Hamiltonian operators verifies the relation

$$\langle dv, [\mathbf{S}_j, \mathbf{S}_k](u)\rangle \equiv \langle dv, \mathbf{S}'_{ju}\mathbf{S}_k(u) - \mathbf{S}'_{ku}\mathbf{S}_j(u)\rangle \overset{(B4)}{=} \langle dv, \mathbf{L}_u \mathbf{Q}'_{ju}\mathbf{S}_k(u) + \mathbf{L}'_u(\mathbf{Q}_j(u); \mathbf{S}_k(u)) - \mathbf{S}'_{ku}\mathbf{L}_u \mathbf{Q}_j(u)\rangle$$

$$\overset{(1.15)}{=} \langle dv, \mathbf{L}_u \mathbf{Q}'_{ju}\mathbf{S}_k(u)\rangle + \langle dv, \mathbf{L}'_u(\mathbf{Q}_j(u); \mathbf{S}_k(u))\rangle - \langle \mathbf{Q}_j(u), \mathbf{S}'_{ku}\mathbf{L}_u\, dv\rangle - \langle dv, \mathbf{L}'_u(\mathbf{Q}_j(u); \mathbf{S}_k(u))\rangle$$

$$\overset{(1.13)}{=} \langle dv, \mathbf{L}_u(\mathbf{Q}'_{ju}\mathbf{S}_k(u) + \tilde{\mathbf{S}}'_{ku}\mathbf{Q}_j(u))\rangle, \tag{1.16}$$

which shows that

$$[\mathbf{S}_j, \mathbf{S}_k](u) = \mathbf{L}_u(\mathbf{Q}'_{ju}\mathbf{S}_k(u) + \tilde{\mathbf{S}}'_{ku}\mathbf{Q}_j(u)), \tag{1.17}$$

where $\tilde{\mathbf{S}}'_{ku}$ is the adjoint of $\mathbf{S}'_{ku}$ with respect to the prefixed bilinear form $\langle v, u\rangle$ (see Appendix A). Since the operator

$$\mathbf{Q}_{jk}(u) \equiv \mathbf{Q}'_{ju}\mathbf{S}_k(u) + \tilde{\mathbf{S}}'_{ku}\mathbf{Q}_j(u), \tag{1.18}$$

is the gradient of the functional

$$F_{jk}[u] \equiv \langle \mathbf{Q}_j(u), \mathbf{S}_k(u)\rangle, \tag{1.19}$$

as is proved by

$$\delta F_{jk}[u] = \langle \mathbf{Q}'_{ju}\delta u, \mathbf{S}_k(u)\rangle + \langle \mathbf{Q}_j(u), \mathbf{S}'_{ku}\delta u\rangle \overset{(1.8)}{=} \langle \mathbf{Q}'_{ju}\mathbf{S}_k(u), \delta u\rangle + \langle \tilde{\mathbf{S}}'_{ku}\mathbf{Q}_j(u), \delta u\rangle = \langle \mathbf{Q}'_{ju}\mathbf{S}_k(u) + \mathbf{S}'_{ku}\mathbf{Q}_j(u), \delta u\rangle, \tag{1.20}$$

relation (1.17) shows that $[\mathbf{S}_j, \mathbf{S}_k]$ is again a Hamiltonian operator relative to $\mathbf{L}_u$. Therefore, the operators $\mathbf{S}_j$ make a Lie algebra, and this Lie algebra structure induces a corresponding structure on the operators $\mathbf{Q}_j$ and on the functionals F_j according to the scheme of Fig. 3. The functional F_{jk} is the *Poisson bracket* of the functionals F_j and F_k associated with the Hamiltonian operators $\mathbf{S}_j$ and $\mathbf{S}_k$ (see Ref. 3, Sec. 5).

A simple property of this algebraic structure is that the condition

$$F_{jk}[u] \equiv \langle \mathbf{Q}_j(u), \mathbf{S}_k(u)\rangle = 0 \tag{1.21}$$

implies

$$[\mathbf{S}_j, \mathbf{S}_k] = 0. \tag{1.22}$$

From the point of view of the theory of the symmetries and of the conservation laws, conditions (1.21) and (1.22) mean that $\mathbf{Q}_j$ and $\mathbf{S}_j$ are respectively an integrating operator and a symmetry generator of the evolution equation

$$\partial_t u = \mathbf{S}_k(u) \equiv \mathbf{L}_u \mathbf{Q}_k(u). \tag{1.23}$$

Therefore, the symplectic operator $\mathbf{L}_u$, associating $\mathbf{Q}_j$ with $\mathbf{S}_j$ according to (1.12), connects the integrating operators of Eq. (1.23) with its symmetry generators. This property explains the importance of the symplectic operators in the study of the evolution equations.

The problem of connecting the conservation laws with the symmetries of the given equation (1.2) is thus reduced to that of recasting this equation into the Hamiltonian form (1.23), by decomposing the operator $\mathbf{K}$ as follows,

$$\mathbf{K}(u) = \mathbf{L}_u \mathbf{Q}(u), \tag{1.24}$$

where $\mathbf{L}_u$ is a suitable symplectic operator and $\mathbf{Q}$ is a potential operator. To find such operators, if any, it is useful to observe that, according to (1.15), $\mathbf{L}_u$ must be

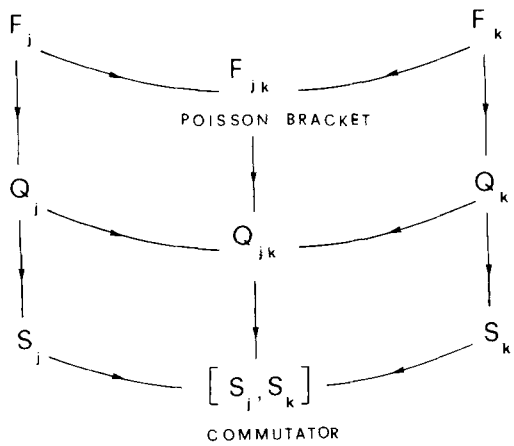

FIG. 3

coupled to $\mathbf{K}$ by the condition

$$\langle dv, \mathbf{K}'_u \mathbf{L}_u \delta v\rangle - \langle \delta v, \mathbf{K}'_u \mathbf{L}_u dv\rangle = \langle dv, \mathbf{L}'_u(\delta v; \mathbf{K}(u))\rangle, \tag{1.25}$$

and that $\mathbf{Q}$ must be an integrating operator of the given equation, as is proved by

$$\langle \mathbf{Q}(u), \mathbf{K}(u)\rangle = \langle \mathbf{Q}(u), \mathbf{L}_u \mathbf{Q}(u)\rangle \overset{(1,13)}{=} 0 \tag{1.26}$$

(examples will be given in Sec. 5). When condition (1.25) is fulfilled, one says that the symplectic operator $\mathbf{L}_u$ makes the given equation Hamiltonian. It can thus be stated that *every* symplectic operator $\mathbf{L}_u$ making the given equation Hamiltonian maps its integrating operators $\mathbf{Q}_j$ into its symmetry generators $\mathbf{S}_j$ according to relation (1.12). This is the main result on which the following analysis of the integrable Hamiltonian equations rests upon.

2. AN EXAMPLE: THE KORTEWEG-de VRIES EQUATION

It is the purpose of this section to show by an example how the integrable Hamiltonian equations may be analyzed by using *only* the connection between the symmetries and the integrating operators previously pointed out.

Consider the KdV equation

$$u_t + auu_x + u_{xxx} = 0 \tag{2.1}$$

and observe that it admits the following *two* Hamiltonian decompositions:

$$u_t + \partial_x\left(\frac{a}{2}u^2 + u_{xx}\right) = 0, \tag{2.2}$$

$$u_t + \left(\partial_{xxx} + \tfrac{2}{3}au\partial_x + \frac{a}{3}u_x I\right)u = 0, \tag{2.3}$$

where the operators

$$\mathbf{L}_u\varphi \equiv \varphi_x, \tag{2.4a}$$

$$\mathbf{M}_u\varphi \equiv \varphi_{xxx} + \tfrac{2}{3}au\varphi_x + \frac{a}{3}u_x\varphi, \tag{2.4b}$$

and

$$\mathbf{Q}_1(u) \equiv u,$$

$$\mathbf{Q}_2(u) \equiv \frac{a}{2}u^2 + u_{xx}, \tag{2.5}$$

are respectively two symplectic operators and two potential operators with respect to the bilinear form

$$\langle v, u\rangle \equiv \int_\Omega v(x,t)u(x,t)\,dx. \tag{2.6}$$

The former decomposition is well known[7]; the latter seems not to have been previously reported.

It follows that two symplectic operators are at our disposal to pass from the integrating operators of the KdV equation to its symmetry generators. They can be used to recover the infinite sequence of conservation laws associated with this equation, as follows. Consider the integrating operator $\mathbf{Q}_1$ and associate with it the symmetry generator

$$\mathbf{S}_2(u) \equiv \mathbf{M}_u\mathbf{Q}_1(u) = auu_x + u_{xxx} = \partial_x\mathbf{Q}_2(u), \tag{2.7}$$

by means of the higher-order symplectic operator $\mathbf{M}_u$. The inverse operator of the second symplectic operator ∂_x then allows to obtain from $\mathbf{S}_2$ the new integrating operator $\mathbf{Q}_2$. By iterating this process, first obtain from $\mathbf{Q}_2$

$$\mathbf{S}_3(u) \equiv \mathbf{M}_u\mathbf{Q}_2(u) = \partial_x\left(\tfrac{5}{16}a^2u^3 + \tfrac{5}{3}auu_{xx} + \tfrac{5}{6}au_x^2 + u_{xxxx}\right), \tag{2.8}$$

and then

$$\mathbf{Q}_3(u) = \tfrac{5}{16}a^2u^3 + \tfrac{5}{3}auu_{xx} + \tfrac{5}{6}au_x^2 + u_{xxxx}, \tag{2.9}$$

and so on. In this way an infinite sequence of independent integrating operators of the KdV equation can be constructed according to the recursion formula

$$\partial_x\mathbf{Q}_{j+1}(u) = \left(\partial_{xxx} + \tfrac{2}{3}au\partial_x + \frac{a}{3}u_x I\right)\mathbf{Q}_j(u). \tag{2.10}$$

The functionals F_j associated with the potential operators $\mathbf{Q}_j$ constitute the infinite sequence of integrals of the KdV equation.

The previous result, of course, is well known[8]; but the present analysis, which is based only on the study of the given equation and which avoids the introduction of the Bäcklund transformation or the use of the Lax equation,[9] provides a different point of view. Above all, this analysis emphasizes the role of the pairs of suitably coupled symplectic operators in the study of the integrable Hamiltonian equations. This point of view will be systematically developed in the next section.

3. INTEGRABLE HAMILTONIAN EQUATIONS

The example of the KdV equation suggests the study of the *twofold* Hamiltonian equations. This requires that we have to first discuss under which conditions two symplectic operators $\mathbf{L}_u$ and $\mathbf{M}_u$ have Hamiltonian operators in common. In this section we prove that if $\mathbf{L}_u$ and $\mathbf{M}_u$ have at least *one* Hamiltonian operator in common and they verify the *coupling condition*[10]

$$\begin{aligned}&\langle dv, \mathbf{L}'_u(\delta v; \mathbf{M}_u\Delta v)\rangle + \langle \delta v, \mathbf{L}'_u(\Delta v; \mathbf{M}_u dv)\rangle + \langle \Delta v, \mathbf{L}'_u(dv; \mathbf{M}_u\delta v)\rangle \\ &\quad = -\left[\langle dv, \mathbf{M}'_u(\delta v; \mathbf{L}_u\Delta v)\rangle + \langle \delta v, \mathbf{M}'_u(\Delta v; \mathbf{L}_u dv)\rangle\right. \\ &\qquad \left. + \langle \Delta v; \mathbf{M}'_u(dv; \mathbf{L}_u\delta v)\rangle\right],\end{aligned} \tag{3.1}$$

and if for one of them, say $\mathbf{L}_u$, condition (1.15) on the Hamiltonian operators is sufficient as well,[11] then they have a possibly infinite sequence of commuting Hamiltonian operators in common.

Assume that

$$\mathbf{S}_j(u) \equiv \mathbf{L}_u\mathbf{Q}_j(u) \tag{3.2}$$

is any Hamiltonian operator common both to $\mathbf{L}_u$ and $\mathbf{M}_u$ and construct

$$\mathbf{S}_{j+1}(u) \equiv \mathbf{M}_u\mathbf{Q}_j(u) \tag{3.3}$$

(see Fig. 4). Our aim is to show that $\mathbf{S}_{j+1}$ is a common Hamiltonian operator again. For this purpose, it suffices to prove that $\mathbf{M}_u\mathbf{Q}_j$ satisfies condition (1.15) on the operators which are Hamiltonian with respect to $\mathbf{L}_u$. On account of (B4), this condition is explicitly given by

$$\begin{aligned}&\langle dv, \mathbf{M}_u\mathbf{Q}'_{ju}\mathbf{L}_u\delta v + \mathbf{M}'_u(\mathbf{Q}_j(u); \mathbf{L}_u\delta v)\rangle \\ &\quad - \langle \delta v, \mathbf{M}_u\mathbf{Q}'_{ju}\mathbf{L}_u dv + \mathbf{M}'_u(\mathbf{Q}_j(u); \mathbf{L}_u dv)\rangle \\ &\quad = \langle dv, \mathbf{L}'_u(\delta v; \mathbf{M}_u\mathbf{Q}_j(u))\rangle.\end{aligned} \tag{3.4}$$

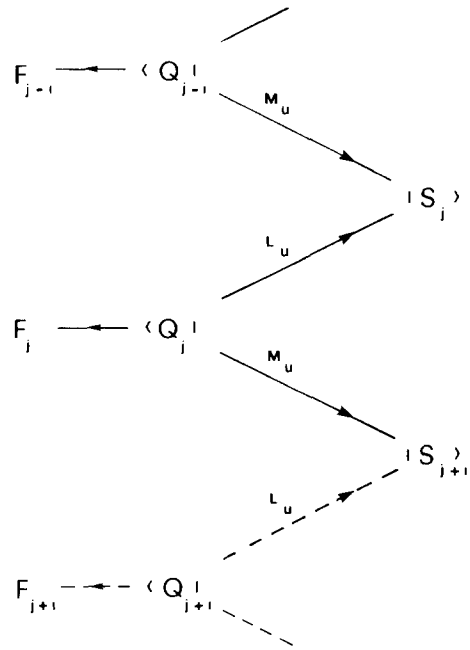

FIG. 4.

Now, this condition readily follows from the coupling condition and from the fact that $\mathbf{S}_j$ is a common Hamiltonian operator. In fact, the assumption on $\mathbf{S}_j$ means that $\mathbf{L}_u\mathbf{Q}_j$ obeys the following condition,

$$\langle dv,\ \mathbf{L}_u\mathbf{Q}'_{ju}\mathbf{M}_u\delta v + \mathbf{L}'_u(\mathbf{Q}_j(u);\mathbf{M}_u\delta v)\rangle - \langle \delta v,\ \mathbf{L}_u\mathbf{Q}'_{ju}\mathbf{M}_u dv + \mathbf{L}'_u(\mathbf{Q}_j(u);\mathbf{M}_u dv)\rangle = \langle dv,\ \mathbf{M}'_u(\delta v;\mathbf{L}_u\mathbf{Q}_j(u))\rangle, \tag{3.5}$$

and the coupling condition (3.1) implies that (3.4) and (3.5) coincide for any $\mathbf{Q}_j$.[12] This proves the statement.

It follows that, once we know one common Hamiltonian operator, we are able to construct successively a possibly infinite sequence of such operators according to the recursion formula

$$\mathbf{L}_u\mathbf{Q}_{j+1}(u) = \mathbf{M}_u\mathbf{Q}_j(u). \tag{3.6}$$

All the properties of this sequence stem from relation (3.6). So, for example, consider the Poisson bracket of any pair of functionals associated with the sequence (see Fig. 4). The following recursion formula is induced by (3.6),

$$\begin{aligned} F_{jk}[u] &\overset{(1.19)}{=} \langle \mathbf{Q}_j(u),\ \mathbf{L}_u\mathbf{Q}_k(u)\rangle \\ &\overset{(3.6)}{=} -\langle \mathbf{Q}_j(u),\ \mathbf{M}_u\mathbf{Q}_{k-1}(u)\rangle \\ &\overset{(1.13)}{=} -\langle \mathbf{Q}_{k-1}(u),\ \mathbf{M}_u\mathbf{Q}_j(u)\rangle \\ &\overset{(3.6)}{=} \langle \mathbf{Q}_{k-1}(u),\ \mathbf{L}_u\mathbf{Q}_{j+1}(u)\rangle \\ &\overset{(1.13)}{=} \langle \mathbf{Q}_{j+1}(u),\ \mathbf{L}_u\mathbf{Q}_{k-1}(u)\rangle \\ &= F_{j+1,k-1}[u]. \end{aligned} \tag{3.7}$$

By iteration, one finds (by assuming $j < k$)

$$F_{jk}[u] = F_{j+1,k-1}[u] = F_{j+2,k-2}[u] = \cdots = F_{kj}[u], \tag{3.8}$$

and then

$$F_{jk}[u] \equiv \langle \mathbf{Q}_j(u), \mathbf{S}_k(u)\rangle = 0, \tag{3.9}$$

on account of the skew symmetry of the Poisson bracket. *All the functionals F_j are thus in involution.* From this property the corresponding relation

$$[\mathbf{S}_j, \mathbf{S}_k] = 0, \tag{3.10}$$

readily follows (recall Fig. 3), showing that *the operators $\mathbf{S}_j$ constitute a set of commuting operators.*

Hence we have constructed a sequence of twofold Hamiltonian equations

$$\partial_t u = \mathbf{S}_k(u), \tag{3.11}$$

which have remarkable properties. The conditions (3.9) and (3.10) mean that each of such equations has a possibly infinite sequence of symmetry generators $\mathbf{S}_j$, of integrating operators $\mathbf{Q}_j$, and of integrals F_j which are in involution. Often this sequence is indeed infinite (see the examples of Sec. 5). In this case, *any Eq. (3.11) is an infinite-dimensional integrable Hamiltonian equation.*

4. SUMMARY

The present paper has mainly dealt with the following two results: The symmetry generators and the integrating operators of any Hamiltonian equation are connected in pairs, and such pairs may be connected into a highly ordered chain for special twofold Hamiltonian equations. A characteristic property of this chain is that the symmetries are automatically in involution, so that the chain may actually define a whole hierarchy of integrable Hamiltonian equations.

These results suggest either a sufficiently general procedure of constructing the integrable Hamiltonian equations or a systematic way of analyzing the Hamiltonian structure of a given evolution equation. Dealing with the first problem, we consider a suitable pair of symplectic operators coupled according to (3.1), and we look for their Hamiltonian operators by solving condition (1.25), where $\mathbf{K}$ is regarded as the unknown operator. If we find *one* Hamiltonian operator common both to $\mathbf{L}_u$ and $\mathbf{M}_u$, we can construct an infinite hierarchy of integrable Hamiltonian equations. Dealing with the second problem, we look for the symplectic operators which make the given equation Hamiltonian by solving condition (1.25) with respect to the unknown operator $\mathbf{L}_u$. If we find *two* solutions of this condition which are coupled according to (3.1) and for which the sequence defined by the recursion formula (3.6) is infinite, the given equation turns out to be an infinite-dimensional integrable Hamiltonian equation. At the same time, the recursion formula (3.6) directly defines the sequence of conservation laws associated with this equation. Examples of this procedure will be given in the next section.

5. APPLICATIONS

As a first example, consider the so-called *Harry Dym equation*[13]

$$\partial_t u = \partial_{xxx}(u^{-1/2}). \tag{5.1}$$

It is manifestly a Hamiltonian equation, since the operators

$$\mathbf{M}_u\varphi \equiv \varphi_{xxx}, \tag{5.2}$$

and

$$\mathbf{Q}_1(u) \equiv u^{-1/2}, \tag{5.3}$$

are respectively a symplectic and a potential operator with respect to the bilinear form (2.6).

According to the method of analysis previously worked out, we look for a second symplectic operator making Eq. (5.1) Hamiltonian. To this end, let us consider the operators

$$\mathbf{L}_u\varphi \equiv 2a(u)\varphi_x + \mathring{a}(u)u_x\varphi, \tag{5.4}$$

which are symplectic with respect to (2.6) for any choice of the function $a(u)$ ($\mathring{a} = da/du$). By trying to fulfill condition (1.25) by means of (5.4), the function

$$a(u) = u, \tag{5.5}$$

is obtained. Since the two symplectic operators (5.2) and (5.4) [with $a(u)$ given by (5.5)] verify the coupling condition (3.1), it turns out that the Harry Dym equation is an integrable Hamiltonian equation, and that the infinite sequence of its conservation laws is defined by the recursion formula

$$(2u\partial_x + u_x I)\mathbf{Q}_{j+1}(u) = \partial_{xxx}\mathbf{Q}_j(u), \tag{5.6}$$

$\mathbf{Q}_1$ being given by (5.3). The search for the symplectic operators making the Harry Dym equation Hamiltonian has thus led to a simple analysis of this equation.

The previous procedure seems to require some preliminary guess of the form of the second symplectic operator in order to be effective. This difficulty can be bypassed as follows. Consider, as a second example, the *nonlinear Schrödinger equation*

$$i\psi_t + \psi_{xx} + 2\psi^2\bar{\psi} = 0, \tag{5.7}$$

and write it in the form

$$\psi_t = i(\psi_{xx} + 2\psi^2\bar{\psi}). \tag{5.8}$$

Since the operators

$$\mathbf{L}\varphi = i\varphi, \tag{5.9}$$

and

$$\mathbf{Q}(\psi) = \psi_{xx} + 2\psi^2\bar{\psi}, \tag{5.10}$$

are respectively a symplectic and a potential operator with respect to the bilinear form

$$\langle \psi, \varphi \rangle = \int_\Omega (\psi\bar{\varphi} + \bar{\psi}\varphi)\, dx, \tag{5.11}$$

(5.8) is a first (well known) Hamiltonian decomposition of the nonlinear Schrödinger equation.

To find a second symplectic operator $\mathbf{M}_\psi$ making the Eq. (5.7) Hamiltonian, let us look for the simplest integrating operators of this equation. The following four integrating operators:

$$\begin{aligned} &\mathbf{Q}_1(\psi) = \psi, && \mathbf{S}_1(\psi) = i\psi; \\ &\mathbf{Q}_2(\psi) = -i\psi_x, && \mathbf{S}_2(\psi) = \psi_x; \\ &\mathbf{Q}_3(\psi) = -(\psi_{xx} + 2\psi^2\bar{\psi}), && \mathbf{S}_3(\psi) = -i(\psi_{xx} + 2\psi^2\bar{\psi}); \\ &\mathbf{Q}_4(\psi) = i(\psi_{xxx} + 6\psi\bar{\psi}\psi_x), && \mathbf{S}_4(\psi) = -(\psi_{xxx} + 6\psi\bar{\psi}\psi_x) \end{aligned} \tag{5.12}$$

can be readily obtained by using the condition (1.10). The operators $\mathbf{S}_j$ are the symmetry generators associated with them by the symplectic operator (5.9). According to (3.3), our problem is to find a symplectic operator $\mathbf{M}_\psi$ fulfilling the relations ($j = 1, 2, 3$)

$$\mathbf{S}_{j+1}(\psi) = \mathbf{M}_\psi\mathbf{Q}_j(\psi). \tag{5.13}$$

The inspection of these relations leads to the following integrodifferential operator

$$\mathbf{M}_\psi\varphi = \varphi_x + 2\psi\int_a^x (\bar{\psi}\varphi - \psi\bar{\varphi})\, d\xi. \tag{5.14}$$

This operator, however, is not skew symmetric with respect to the bilinear form (4.11), because of the integral. Let us then write it in the equivalent form

$$\begin{aligned} \mathbf{M}_\psi\varphi = \varphi_x &+ \psi\int_a^x (\bar{\psi}\varphi - \psi\bar{\varphi})\, d\xi \\ &+ \psi\int_b^x (\bar{\psi}\varphi - \psi\bar{\varphi})\, d\xi, \end{aligned} \tag{5.15}$$

where a and b denote the endpoints of the interval Ω of definition of the field functions. The operator (4.15) obeys the conditions on the symplectic operators with respect to the bilinear form (4.11) and the coupling condition with (4.9). Therefore, the results obtained in Sec. 3 can be applied to the nonlinear Schrödinger equation. In particular, the following recursion formula for its conservation laws,

$$\begin{aligned} i\mathbf{Q}_{j+1}(\psi) = \partial_x\mathbf{Q}_j &+ \psi\int_a^x [\bar{\psi}\mathbf{Q}_j(\psi) - \psi\bar{\mathbf{Q}}_j(\psi)]\, d\xi \\ &+ \psi\int_b^x [\bar{\psi}\mathbf{Q}_j(\psi) - \psi\bar{\mathbf{Q}}_j(\psi)]\, d\xi, \end{aligned} \tag{5.16}$$

is obtained.

The same analysis, finally, can be repeated for the *modified KdV equation*

$$u_t + au^2u_x + u_{xxx} = 0. \tag{5.17}$$

One obtains the following two Hamiltonian decompositions:

$$u_t + \partial_x\left(\frac{a}{3}u^3 + u_{xx}\right) = 0, \tag{5.18a}$$

$$\begin{aligned} u_t + \Big(&\partial_{xxx} + \tfrac{2}{3}av^2\partial_x + \tfrac{2}{3}vv_xI \\ &- \frac{a}{3}v_x\int_a^x v_\xi I\, d\xi - \frac{a}{3}v_x\int_b^x v_\xi I\, d\xi\Big)v = 0, \end{aligned} \tag{5.18b}$$

whose symplectic operators again verify the coupling condition (3.1) with respect to the bilinear form (2.6). The modified KdV equation is thus another example of the special twofold Hamiltonian equations considered in this paper. These examples show that the method of analysis based on the search for the symplectic operators making the given equation Hamiltonian is not only conceptually simple but also effective in the applications. They point out, moreover, that some of the more

interesting evolution equations considered in the literature have a *common structure*, which is well described by the model of integrable Hamiltonian equation developed in Sec. 3.

APPENDIX A

The Gateaux derivative of an operator $\mathbf{S}: U \to U$ may be denoted by $\mathbf{S}'_u$ and is defined by

$$\mathbf{S}'_u\varphi \equiv \frac{d}{d\epsilon}\mathbf{S}(u+\epsilon\varphi)\big|_{\epsilon=0}, \tag{A1}$$

so that to the first order in ϵ it is

$$\mathbf{S}(u+\epsilon\varphi) = \mathbf{S}(u) + \epsilon\mathbf{S}'_u\varphi. \tag{A2}$$

Its adjoint operator $\tilde{\mathbf{S}}'_u$, relative to the prefixed bilinear form $\langle v, u\rangle$, is defined by

$$\langle dv, \mathbf{S}'_u du\rangle = \langle \tilde{\mathbf{S}}'_u dv, du\rangle. \tag{A3}$$

It is a linear mapping of the dual space V into itself.

If $\mathbf{S}$ is given by

$$\mathbf{S}(u) = s(u, u_x, u_{xx}, \cdots), \tag{A4}$$

it is

$$\mathbf{S}'_u\varphi = \frac{\partial s}{\partial u}\varphi + \frac{\partial s}{\partial u_x}\varphi_x + \cdots, \tag{A5}$$

and so the following identity

$$\begin{aligned}\partial_t\mathbf{S}(u) &= \frac{\partial s}{\partial u}\partial_t u + \frac{\partial s}{\partial u_x}\partial_t u_x + \cdots \\ &= \frac{\partial s}{\partial u}u_t + \frac{\partial s}{\partial u_x}\partial_x u_t + \cdots \\ &= \mathbf{S}'_u\partial_t u,\end{aligned} \tag{A6}$$

can be readily verified.

APPENDIX B

The Gateaux derivative of the metric operator $\mathbf{L}_u$ is defined by

$$\mathbf{L}'_u(\varphi;\psi) \equiv \frac{d}{d\epsilon}\mathbf{L}_{u+\epsilon\psi}\varphi\big|_{\epsilon=0}. \tag{B1}$$

Consequently, to the first order in ϵ it is

$$\mathbf{L}_{u+\epsilon\psi}\varphi = \mathbf{L}_u\varphi + \epsilon\mathbf{L}'_u(\varphi;\psi), \tag{B2}$$

and the Gateaux derivative of the operator

$$\mathbf{S}(u) \equiv \mathbf{L}_u\mathbf{Q}(u) \tag{B3}$$

is given by

$$\begin{aligned}\mathbf{S}'_u\psi &= \frac{d}{d\epsilon}\mathbf{L}_{u+\epsilon\psi}\mathbf{Q}(u+\epsilon\psi)\big|_{\epsilon=0} \\ &= \frac{d}{d\epsilon}[\mathbf{L}_u\mathbf{Q}(u) + \epsilon\mathbf{L}_u\mathbf{Q}'_u\psi \\ &\quad + \epsilon\mathbf{L}'_u(\mathbf{Q}(u);\psi) + \cdots]_{\epsilon=0} \\ &= L_u\mathbf{Q}'_u\psi + \mathbf{L}'_u(\mathbf{Q}(u);\psi).\end{aligned} \tag{B4}$$

To prove condition (1.15), observe that every symplectic operator satisfies the relation

$$\langle \delta v, \mathbf{L}'_u(dv;du)\rangle = -\langle dv, \mathbf{L}'_u(\delta v;du)\rangle, \tag{B5}$$

which is obtained by differentiating (1.13) with respect to u. Therefore,

$$\begin{aligned}&\langle dv, \mathbf{S}'_u\mathbf{L}_u\delta v\rangle - \langle \delta v, \mathbf{S}'_u\mathbf{L}_u dv\rangle \\ &\overset{(B4)}{=} \langle dv, \mathbf{L}_u\mathbf{Q}'_u\mathbf{L}_u\delta v + \mathbf{L}'_u(\mathbf{Q}(u);\mathbf{L}_u\delta v)\rangle \\ &\quad - \langle \delta v, \mathbf{L}_u\mathbf{Q}'_u\mathbf{L}_u dv + \mathbf{L}'_u(\mathbf{Q}(u);\mathbf{L}_u dv)\rangle \\ &\overset{(B5)}{=} -\langle \mathbf{Q}(u), \mathbf{L}'_u(dv;\mathbf{L}_u\delta v)\rangle - \langle \delta v, \mathbf{L}'_u(\mathbf{Q}(u);\mathbf{L}_u dv)\rangle \\ &\overset{(1.14)}{=} \langle dv, \mathbf{L}'_u(\delta v;\mathbf{L}_u\mathbf{Q}(u))\rangle \\ &\quad = \langle dv, \mathbf{L}'_u(\delta v;\mathbf{S}(u))\rangle.\end{aligned} \tag{B6}$$

This proves that condition (1.15) is necessary. As regards the problem if this condition is sufficient as well, we can only remark that there exist symplectic operators for which this is true. This can be verified, for example, for the symplectic operators (2.4a), (5.4), (5.9) and (5.18a) considered in this paper. This fact justifies us to assume, in Sec. 3, that for the symplectic operator $\mathbf{L}_u$, condition (1.15) is sufficient as well.

[1] R. Abraham, *Foundations of Mechanics* (Benjamin, New York, 1967); E. C. G. Sudarshan and N. Mukunda, *Classical Dynamics: A Modern Perspective* (Wiley, New York, 1974); V. Arnold, *Les Méthodes Mathématiques de la Mécanique Classique* (Mir, Moscow, 1976).

[2] F. Magri, Nuovo Cimento B **34**, 334 (1976).

[3] F. Magri, Ann. Phys. **99**, 196 (1976), Sec. 4.

[4] For a more detailed discussion of the theory of the potential operators see Ref. 3, Sec. 2, and the references given there.

[5] J. Rzewuski, *Field Theory* (Polish Sci. Publ., Warsaw, 1964), part 1°, p. 101.

[6] The integrating operators may be regarded as an extension to general evolution equations of the well-known integrating factors of the ordinary differential equations of the first order.

[7] C. S. Gardner, J. Math. Phys. **12**, 1548 (1971).

[8] C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, Commun. Pure Appl. Math. **27**, 114 (1974), formula (3.20).

[9] G. L. Lamb, Jr., J. Math. Phys. **15**, 2157 (1974); P. D. Lax, Commun. Pure Appl. Math. **21**, 467 (1968).

[10] The coupling condition (3.1) has a simple algebraic meaning, namely that the sum of the two symplectic operators $\mathbf{L}_u$ and $\mathbf{M}_u$ is itself a symplectic operator. I thank Professor C. Cercignani for this observation.

[11] See the remark at the end of Appendix B.

[12] Take into account property (B5), which is verified by every symplectic operator.

[13] M. D. Kruskal, in *Dynamical Systems, Theory and Applications*, edited by J. Moser (Springer, Berlin, 1975), p. 313.

Tensor harmonics on the 3-sphere[a)]

Robert T. Jantzen

Department of Physics, University of California, Berkeley, California 94720
(Received 21 June 1977)

The scalar, vector, and tensor harmonics on the 3-sphere are developed by its identification with SU(2), enabling familiar angular momentum techniques to be employed. The application to spatially homogeneous cosmology is discussed. In this context the classic work of Lifshitz and the recent approach of Hu are bridged. Finally spinor harmonics are introduced.

1. INTRODUCTION

We consider three representations of the Lie group SU(2), whose underlying manifold is the 3-sphere $\mathbf{S}^3$, by dragging along of the tensor algebra over the group by left translation, inverse right translation and conjugation. This leads to a simple way of dealing with the tensor harmonics on $\mathbf{S}^3$ which we identify with SU(2). Specifically, define $\hat{\mathbf{e}}_a = -i\sigma_a$ and $\hat{\mathbf{e}}_4 = 1_2$, where $\{\sigma_a\}$ are the Pauli matrices and 1_2 the two-dimensional identity matrix, and let $\{\hat{e}_\alpha\}$ be the standard basis of R^4 considered as a vector space.[1] We then identity $z^\alpha \hat{e}_\alpha \in \mathbf{S}^3 = \{y^\alpha \hat{e}_\alpha \in R^4 \mid \delta_{\alpha\beta} y^\alpha y^\beta = 1\}$ with $z^\alpha \hat{\mathbf{e}}_\alpha \in \mathrm{SU}(2)$. (Alternatively one may identify $\mathbf{S}^3$ with the unit quaternions of the quaternion algebra induced on R^4 by the identification of $\{\hat{e}_\alpha\}$ and $\{\hat{\mathbf{e}}_\alpha\}$.) As a reminder of this identification we will use the symbol $\mathbf{G}$ to denote $\mathrm{SU}(2) \sim \mathbf{S}^3$ in what follows.

2. STRUCTURE AND GEOMETRY OF G

Let $\mathfrak{g}$ be the three-dimensional vector space of left invariant vector fields on $\mathbf{G}$ (the Lie algebra of $\mathbf{G}$) and $\tilde{\mathfrak{g}}$ its right invariant counterpart. Introduce also the dual vector spaces $\mathfrak{g}^*$ and $\tilde{\mathfrak{g}}^*$ of respectively left and right invariant 1-forms. "~" may be interpreted as a map which associates to each left invariant tensor field on $\mathbf{G}$ the right invariant field whose value at the identity coincides with that of the original field at the identity.

Let $\{y_\alpha\}$ be standard Cartesian coordinates on R^4 and $\{\partial_\alpha = \partial/\partial y^\alpha\}$ the coordinate frame. The identity $\hat{\mathbf{e}}_4$ of SU(2) corresponds to the north pole of $\mathbf{S}^3$ and has coordinates $\{0, 0, 0, 1\}$. The vector fields $\mathbf{L}_{\alpha\beta} = y^\alpha \partial_\beta - y^\beta \partial_\alpha$ generate the rotations of R^4 about the origin [the identity representation of SO(4, R)] and restrict naturally to vector fields on the orbit $\mathbf{G}$. Let $2e_a = \mathbf{L}_{4a} - \mathbf{L}_{bc}$ and $2\tilde{e}_a = \mathbf{L}_{4a} + \mathbf{L}_{bc}$ be defined on $\mathbf{G}$, where (a, b, c) is a cyclic permutation of $(1, 2, 3)$. $\{e_a\}$ and $\{\tilde{e}_a\}$ are the canonical bases of $\mathfrak{g}$ and $\tilde{\mathfrak{g}}$ which agree with the Cartesian derivatives $\{\frac{1}{2}\partial_a\}$ at the identity. [They correspond to the basis $\{\frac{1}{2}\hat{e}_a\}$ of the matrix Lie algebra of SU(2).] They satisfy

$$[e_a, e_b] = C^c{}_{ab} e_c, \quad [\tilde{e}_a, \tilde{e}_b] = -C^c{}_{ab} \tilde{e}_c, \quad [e_a, \tilde{e}_b] = 0, \tag{2.1}$$

with $C^c{}_{ab} = \epsilon_{cab}$. Each basis is a global (analytic) frame on $\mathbf{G}$, having dual frames $\{\omega^a\}$ and $\{\tilde{\omega}^a\}$, respectively, which are the corresponding dual bases of $\mathfrak{g}^*$ and $\tilde{\mathfrak{g}}^*$. The 1-forms $\frac{1}{2}\omega^a$ and $\frac{1}{2}\tilde{\omega}^a$ result from restriction to $\mathbf{G}$ of the 1-forms on R^4 whose components in the Cartesian frame are the same as those of the vector fields $2e_a$ and $2\tilde{e}_a$ on R^4, respectively.

The Euclidean metric $\delta_{\alpha\beta} dy^\alpha \otimes dy^\beta$ restricts to the following bi-invariant metric on $\mathbf{G}$ which is a constant multiple of the Killing metric and whose six-dimensional Killing Lie algebra is $\mathfrak{g} \oplus \tilde{\mathfrak{g}}$:

$$\begin{aligned} &\mathbf{g} = g_{ab} \omega^a \otimes \omega^b = g_{ab} \tilde{\omega}^a \otimes \tilde{\omega}^b, \\ &g_{ab} = \mathbf{g}(e_a, e_b) = \mathbf{g}(\tilde{e}_a, \tilde{e}_b) = \tfrac{1}{4}\delta_{ab}, \\ &\mathbf{g}^{-1} = g^{ab} e_a \otimes e_b = g^{ab} \tilde{e}_a \otimes \tilde{e}_b, \\ &g^{ab} = \mathbf{g}^{-1}(\omega^a, \omega^b) = \mathbf{g}^{-1}(\tilde{\omega}^a, \tilde{\omega}^b) = 4\delta^{ab}. \end{aligned} \tag{2.2}$$

The volume element of the metric is

$$\begin{aligned} \eta &= g^{1/2} \omega^1 \otimes \omega^2 \otimes \omega^3 = g^{1/2} \tilde{\omega}^1 \otimes \tilde{\omega}^2 \otimes \tilde{\omega}^3 \\ &= (3!)^{-1} \eta_{abc} \omega^a \otimes \omega^b \otimes \omega^c, \end{aligned} \tag{2.3}$$

where $g = (4)^{-3}$ is the determinant of the matrix $\mathbf{g}$ whose entries are the components g_{ab}. It is also convenient to have a normalized volume element $\hat{\eta}$ whose integral over $\mathbf{G}$ is unity, obtained from η by dividing out the volume of $\mathbf{G}$. (To each of these volume elements corresponds a bi-invariant Lebesque measure on $\mathbf{G}$.)

Suppose we consider $\binom{p}{q}$-tensor fields on $\mathbf{G}$ with complex valued components in any (analytic) frame which are analytic in the real sense.[2] This vector space $CT^{p,q}(\mathbf{G})$ has a natural Hermitian inner product induced by the metric. If S and T are two such tensor fields with components $S^{a\cdots}{}_{b\cdots}$ and $T^{a\cdots}{}_{b\cdots}$ in the frame $\{e_a\}$, their inner product is

$$\langle S, T \rangle = \int_G \hat{\eta}(S, T), \quad (S, T) = \overline{S^{a\cdots}{}_{b\cdots}} T_{a\cdots}{}^{b\cdots}, \tag{2.4}$$

where the indices are raised and lowered with the metric as usual:

$$T_{a\cdots}{}^{b\cdots} = g_{ac} \cdots g^{bd} \cdots T^{c\cdots}{}_{d\cdots}.$$

Since $(e_a, e_b) = \frac{1}{4}\delta_{ab}$, factors of 4 will often appear. This is so because the natural metric on $\mathbf{G}$ is $4\mathbf{g}$ rather than the induced metric $\mathbf{g}$ and with respect to which $\{e_a\}$ is an orthonormal frame.

Standard formulas[3] may be used to evaluate the components of the metric connection and its Riemann and Ricci tensors in the frame $\{e_a\}$:

a)Work supported by NSF Grant MCS-7621525.

 0022-2488/78/1905-1163$1.00

$$\nabla_{e_a} e_b = \Gamma^c{}_{ab} e_c, \quad \Gamma^c{}_{ab} = \tfrac{1}{2} C^c{}_{ab}, \qquad (2.5)$$

$$R^{ab}{}_{cd} = \delta^{ab}_{cd}, \quad R^a{}_b = 2\delta^a{}_b.$$

If T is a $\binom{p}{q}$-tensor field, the components of its covariant derivative ∇T in this frame are given by

$$T^{a\cdots}{}_{b\cdots;c} = e_c T^{a\cdots}{}_{b\cdots} + \Gamma^a{}_{cd} T^{d\cdots}{}_{b\cdots} + \cdots - \Gamma^d{}_{cb} T^{a\cdots}{}_{d\cdots} - \cdots . \qquad (2.6)$$

Replacing $\Gamma^a{}_{bc}$ by $2\Gamma^a{}_{bc} = C^a{}_{bc}$ yields the formula for $(\pounds_{e_c} T)^{a\cdots}{}_{b\cdots}$, where $\pounds_{e_c}$ is the Lie derivative with respect to e_c. These latter components may also be interpreted as the components of the covariant derivative of T with respect to a flat connection ${}^R\nabla$ (with torsion) whose components are ${}^R\Gamma^a{}_{bc} = C^a{}_{bc}$ in the frame $\{e_a\}$ and whose global parallel transport is right translation. $\mathbf{g}$ is covariant constant with respect to this connection. The same formula with $\Gamma^a{}_{bc}$ replaced by 0 holds for $(\pounds_{\mathcal{E}_c} T)^{a\cdots}{}_{b\cdots}$ and one may introduce a left connection ${}^L\nabla$ in a similar way.

We use the notation ΔT for the ordinary Laplacian of T:

$$(\Delta T)^{a\cdots}{}_{b\cdots} = - T^{a\cdots}{}_{b\cdots;c}{}^{;c}. \qquad (2.7)$$

From the identity

$$T^{c\cdots}{}_{d\cdots;a;b} = (\nabla_{e_b} \nabla_{e_a} T)^{c\cdots}{}_{d\cdots} - T^{c\cdots}{}_{d\cdots;e} \Gamma^e{}_{ba}$$

and the fact that $\Gamma^e{}_{ba} g^{ba} = 0$, it follows that

$$\Delta T = - g^{ab} \nabla_{e_a} \nabla_{e_b} T. \qquad (2.8)$$

Similarly, if Φ is a function,

$$\Delta \Phi = - g^{ab} e_a e_b \Phi. \qquad (2.9)$$

By "index lowering" any $\binom{p}{q}$-tensor field $\hat{T}$ with $r = p + q$ is equivalent to a $\binom{0}{r}$-tensor field T with components $T_{a_1 \cdots a_r}$. Following Lichnerowicz,[5] we define the divergence δT and the DeRham Laplacian $\Delta_{DR} T$ of T (and hence of T by index raising) by the formulas

$$(\delta T)_{a_1 \cdots a_{r-1}} = - T^b{}_{a_1 \cdots a_{r-1};b}$$

$$(\Delta_{DR} T)_{a_1 \cdots a_r} = (\Delta T)_{a_1 \cdots a_r} + \sum_s T_{a_1 \cdots b \cdots a_r} R^b{}_{a_s} - \sum_{s \neq t} T_{a_1 \cdots c \cdots d \cdots a_r} R^c{}_{a_s}{}^d{}_{a_t}. \qquad (2.10)$$

When acting on differential forms this reduces to the usual DeRham Laplacian $\Delta_{DR} = d\delta + \delta d$. (For a function Φ, $\delta \Phi = 0$, and $\Delta_{DR} \Phi = \Delta \Phi$.)

3. DRAGGING ACTION

Let $\mathbf{L}_u$, $\mathbf{R}_u$, and $\mathbf{AD}_u = \mathbf{L}_u \circ \mathbf{R}_{u^{-1}}$ denote left translation, right translation, and conjugation by $u \in \mathbf{G}$. (Each of these is an isometry of $\mathbf{g}$ and left and right translations commute.) Use the same symbols to denote the operators which drag along tensor fields by these diffeomorphisms and hence induce linear transformations on the vector spaces $CT^{p,q}(\mathbf{G})$. For example, if Φ, X, σ are a function, vector field and one-form on $\mathbf{G}$ and h a diffeomorphism of $\mathbf{G}$ into itself, the dragged along fields are[6]

$$h\Phi = \Phi \circ h^{-1},$$

$$(hX)(u) = dh(h^{-1}(u)) X(h^{-1}(u)),$$

$$(h\sigma)(u) = dh^{-1}(u)^* \sigma(h^{-1}(u)).$$

For higher rank tensor fields these apply to each factor in a tensor product:

$$h(\Phi X \otimes \cdots \otimes \sigma \cdots) = h\Phi\, hX \otimes \cdots \otimes h\sigma \cdots .$$

A tensor field T satisfying $hT = T$ is called h-invariant. On a Lie group left (right) invariant tensor fields are invariant under dragging by all left (right) translations and have constant components in a left (right) invariant frame, while bi-invariant tensor fields are invariant under dragging by both left and right translations.

The requirement for a map ρ from a group $\mathbf{G}$ into the group $\mathbf{GL}(V)$ of invertible linear transformations of a vector space V to be a homomorphism and therefore a representation of $\mathbf{G}$ is $\rho_{u_1 u_2} = \rho_{u_1} \circ \rho_{u_2}$, where ρ_u is the value of ρ at u and $\circ$ indicates composition of the linear transformations. This is satisfied by each of the dragging maps $\mathbf{L}$, $\mathbf{R}^{-1}$ and $\mathbf{AD}$ which therefore determine representations of $\mathbf{G}$ on the space $CT^{p,q}(\mathbf{G})$ called, respectively, the left, right, and adjoint $\binom{p}{q}$-tensor dragging representations. [By $\mathbf{R}_u^{-1}$ we mean $(\mathbf{R}_u)^{-1} = \mathbf{R}_{u^{-1}}$. Note that the dragging operator $\mathbf{AD}_u$ equals $\mathbf{L}_u \mathbf{R}_{u^{-1}} = \mathbf{R}_{u^{-1}} \mathbf{L}_u$.] These are unitary representations with respect to the inner product (2.4) since $\hat{\eta}$ is bi-invariant.

Denote the one-parameter group of diffeomorphisms generated by a complete analytic vector field X by $\{X_t \mid t \in R\}$ (the flow of X). The dragging operator X_t, when acting on an analytic tensor field T, has the Lie derivative expansion:

$$X_t T = e^{-t \pounds_X} T, \quad \pounds_X T = - (d/dt) \big|_0 X_t T. \qquad (3.1)$$

Tensor fields invariant under the flow of X ($X_t T = T$ for all $t \in R$) have vanishing Lie derivative with respect to X.

From Lie group theory it is well known that for $X \in \mathfrak{g}$:

$$X_t = \mathbf{R}_{\exp(tX)}, \quad \tilde{X}_t = \mathbf{L}_{\exp(tX)},$$

$$(\tilde{X} - X)_t = \mathbf{AD}_{\exp(tX)}, \qquad (3.2)$$

where the exponential map $\exp: \mathfrak{g} \to \mathbf{G}$ may be defined by $\exp(X) = X_1(u_0) = \tilde{X}_1(u_0)$. u_0 denotes the identity of $\mathbf{G}$. The corresponding dragging operators therefore have the expansions:

$$\mathbf{R}_{\exp(tX)} = \exp(- t \pounds_X), \quad \mathbf{L}_{\exp(tX)} = \exp(- t \pounds_{\tilde{X}}),$$

$$\mathbf{AD}_{\exp(tX)} = \exp[- t \pounds_{(\tilde{X} - X)}], \qquad (3.3)$$

valid when acting on analytic tensor fields. As a consequence the action of X and $\tilde{X}$ on a function Φ is given by

$$X\Phi = (d/dt) \big|_0 \Phi \circ \mathbf{R}_{\exp(tX)}, \quad \tilde{X}\Phi = (d/dt) \big|_0 \Phi \circ \mathbf{L}_{\exp(tX)}. \qquad (3.4)$$

$\mathfrak{g}$ is said to generate the right translations and $\tilde{\mathfrak{g}}$ the left translations. The adjoint diffeomorphisms or conjugations are generated by the Lie algebra with basis $\{\mathbf{L}_{bc} = \tilde{e}_a - e_a\}$, where (a, b, c) is a cyclic permutation of $(1, 2, 3)$. These are the rotations of $\mathbf{G}$ which leave the identity (north pole) fixed. [The adjoint group acting on SU(2) corresponds to the action on $\mathbf{S}^3$ of the SO(3, R) subgroup of SO(4, R) which leaves $\hat{e}_4$ fixed.] The orbits of this action are 2-spheres of constant y^4 which degenerate to points at the poles. Standard spherical coordinates $\{\chi, \theta, \phi\}$ on $\mathbf{G}$ are conveniently adapted to these orbits; $\{\theta, \phi\}$ are standard spherical coordinates on the 2-sphere of radius $\sin\chi$ for which $y^4 = \cos\chi$.

The (linear) adjoint group is the subgroup of $\mathbf{GL}(\mathfrak{g})$ induced by the dragging action of the adjoint group on $\mathfrak{g}$:

$$\mathbf{A}d(u)X = \mathbf{AD}_u X, \quad X \in \mathfrak{g}\ .$$

Its matrix representation with respect to the basis $\{e_a\}$ of $\mathfrak{g}$ is SO(3, R) and provides the standard link between SU(2) and SO(3, R):

$$\mathbf{A}d(u)e_a = e_b \mathcal{R}^b{}_a(u).$$

$\mathcal{R}$ is an SO(3, R)-valued function on $\mathbf{G}$ which also provides the transformation between the canonical left and right invariant frames:

$$\tilde{e}_a = e_b \mathcal{R}^{-1b}{}_a, \quad \tilde{\omega}^a = \mathcal{R}^a{}_b \omega^b,$$

$$\mathcal{R}^c{}_a \delta_{cd} \mathcal{R}^d{}_b = \delta_{ab}, \quad \mathcal{R}^a{}_c \delta^{cd} \mathcal{R}^b{}_d = \delta^{ab}. \tag{3.5}$$

If $\mathbf{K}_a$ is the matrix whose components are $K_a{}^b{}_c = C^b{}_{ac}$ and if $\delta_{ab} n^a n^b = 1$, then $\mathcal{R}(\exp\theta n^a \mathbf{K}_a)$ is the matrix of the rotation of R^3 about the origin by an angle θ about the direction specified by the unit vector with Cartesian components n^a. This equality is established by means of the following identities for $X \in \mathfrak{g}$:

$$\pounds_X e_a = [X, e_a] = X^c C^b{}_{ca} e_b\ ,$$

$$\mathbf{AD}_{\exp X} e_a = \mathbf{R}_{\exp(-X)} \mathbf{L}_{\exp(X)} e_a = \mathbf{R}_{\exp(-X)} e_a = e^{tX} e_a.$$

Let us introduce the notation $\mathbf{L}_a = ie_a$, $\tilde{\mathbf{L}}_a = -i\tilde{e}_a$, $\mathbf{J}_a = \mathbf{L}_a + \tilde{\mathbf{L}}_a$ and use the same symbols for the corresponding Lie derivatives,[7] i.e., $\mathbf{L}_a T = \pounds_{\mathbf{L}_a} T = i\pounds_{e_a} T$, etc. Then the first two sets mutually commute and all three satisfy the standard angular momentum commutation relations, i.e., $[\mathbf{L}_a, \mathbf{L}_b] = i\epsilon_{abc}\mathbf{L}_c$. Define also $\mathbf{L}^2 = \delta^{ab}\mathbf{L}_a\mathbf{L}_b$, $\tilde{\mathbf{L}}^2 = \delta^{ab}\tilde{\mathbf{L}}_a\tilde{\mathbf{L}}_b$, $\mathbf{J}^2 = \delta^{ab}\mathbf{J}_a\mathbf{J}_b$, each of which commutes with its corresponding set and with each other, and finally introduce the raising and lowering operators for each set, i.e., $\mathbf{L}_\pm = \mathbf{L}_1 \pm i\mathbf{L}_2$. Combining the new notation with our previous formulas we may write:

$$\mathbf{R}^{-1}{}_{\exp(\theta n^a e_a)} = \exp(-i\theta n^a \mathbf{L}_a),$$

$$\mathbf{L}_{\exp(\theta n^a e_a)} = \exp(-i\theta n^a \tilde{\mathbf{L}}_a),$$

$$\mathbf{AD}_{\exp(\theta n^a e_a)} = \exp(-i\theta n^a \mathbf{J}_a). \tag{3.6}$$

We will refer to $\tilde{\mathbf{L}}_a$, $\mathbf{L}_a$, and $\mathbf{J}_a$ as left, right, and total angular momentum. They generate the left, right, and adjoint dragging representations. We will refer to $\mathbf{L}^2$, $\tilde{\mathbf{L}}^2$, and $\mathbf{J}^2$ as their squares. [In fact by the remarks following (2.6), $\mathbf{L}_a = i^R\nabla_{e_a}$ and $\tilde{\mathbf{L}}_a = -i^L\nabla_{\tilde{e}_a}$, and $\tilde{\mathbf{L}}^2$ and $\mathbf{L}^2$ are the Laplacians of the torsion geometries $(\mathbf{G}, 4g, {}^L\nabla)$ and $(\mathbf{G}, 4g, {}^R\nabla)$.] Since $\Delta\Phi = -g^{ab}e_a e_b \Phi = -g^{ab}\tilde{e}_a\tilde{e}_b\Phi$ if Φ is a function,

$$\Delta\Phi = 4\mathbf{L}^2\Phi = 4\tilde{\mathbf{L}}^2\Phi, \tag{3.7}$$

so $\mathbf{L}^2$ and $\tilde{\mathbf{L}}^2$ coincide when acting on functions.

4. REPRESENTATION FUNCTIONS AND SCALAR HARMONICS

Consider the irreducible unitary representations $\{D^J \mid J = 0, \frac{1}{2}, 1, \cdots\}$ of SU(2). Each D^J is a linear transformation-valued function $\mathbf{G}$ satisfying

$$D^J(u_1, u_2) = D^J(u_1) D^J(u_2). \tag{4.1}$$

The basis $\{\mathcal{J}_a^J\}$ of the Lie algebra of the representation D^J determined by the "basis" $\{\mathbf{L}_a\}$ of $\mathfrak{g}$ has the same commutation relations and is defined by

$$\mathcal{J}_a^J = i(d/dt)\big|_0 D^J(\exp te_a),$$
$$D^J(\exp\theta n^a e_a) = \exp(-i\theta n^a \mathcal{J}_a^J). \tag{4.2}$$

Using (3.4), (4.1), and (4.2) it is easy to compute the following derivatives of the function D^J:

$$\mathbf{L}_a D^J = D^J \mathcal{J}_a^J, \quad -\tilde{\mathbf{L}}_a D^J = \mathcal{J}_a^J D^J,$$
$$\mathbf{L}^2 D^J = \tilde{\mathbf{L}}^2 D^J = (\mathcal{J}^J)^2 D^J. \tag{4.3}$$

Let $\{|\, JM\rangle \mid M = -J, \ldots, J\}$ be the standard orthonormal basis of the $(2J+1)$-dimensional space carrying the representation D^J:

$$\langle JM \mid JN\rangle = \delta_{MN}, \quad \epsilon JM = [J(J+1) - M(M+1)]^{1/2},$$
$$\mathcal{J}_3^J \mid JM\rangle = M \mid JM\rangle, \quad \mathcal{J}_+^J \mid JM\rangle = \epsilon_{JM} \mid J\, M+1\rangle,$$
$$(\mathcal{J}^J)^2 \mid JM\rangle = J(J+1) \mid JM\rangle, \quad \mathcal{J}_-^J \mid JM\rangle = \epsilon_{JM-1} \mid J\, M-1\rangle,$$
$$D^J \mid JM\rangle = \mid JN\rangle D^J{}_{NM}. \tag{4.4}$$

The matrices of D^J and $\mathcal{J}_a^J$ in this basis are defined by

$$D^J{}_{MN} = \langle JM \mid D^J \mid JN\rangle, \quad (\mathcal{J}_a^J)^M{}_N = \langle JM \mid \mathcal{J}_a^J \mid JN\rangle.$$

For example, $(\mathcal{J}_+^J)^M{}_N = \delta_{M,N+1}\epsilon_{JN}$. Taking components of (4.3), one finds

$$\mathbf{L}^2 D^{JM}{}_N = J(J+1) D^{JM}{}_N = \tilde{\mathbf{L}}^2 D^{JM}{}_N,$$
$$\mathbf{L}_3 D^{JM}{}_N = N D^{JM}{}_N, \quad -\tilde{\mathbf{L}}_3 D^{JM}{}_N = M D^{JM}{}_N,$$
$$\mathbf{L}_+ D^{JM}{}_N = \epsilon_{JN} D^J{}_{N+1}, \quad -\tilde{\mathbf{L}}_+ D^{JM}{}_N = \epsilon_{JM-1} D^{JM-1}{}_N,$$
$$\mathbf{L}_- D^{JM}{}_N = \epsilon_{JN-1} D^{JM}{}_{N-1}, \quad -\tilde{\mathbf{L}}_- D^{JM}{}_N = \epsilon_{JM} D^{JM+1}{}_N. \tag{4.5}$$

By defining $|\, JM, JM'\rangle = Q^{JMM'} = (2J+1)^{1/2}(-1)^M \times D^{J-M}{}_{M'}$, one finds that left angular momentum acts on the index M of $Q^{JMM'}$ exactly as right angular momentum acts on the index M' of both $Q^{JMM'}$ and $D^J{}_{M'}$, namely in the standard fashion.

In other words the weight J component occurs in

the decomposition of the left scalar dragging representation $2J+1$ times, once for each value of M', and $|JM,JM'\rangle$ is a standard basis for the subspace of a given M'. An analogous statement holds for the right scalar dragging representation. The behavior of $|JM,JM'\rangle$ under left and right dragging is therefore standard:

$$\mathbf{L}_u\big|JM,JM'\rangle = \big|JN,JM'\rangle D^J{}_{NM}(u),$$

$$\mathbf{R}_u^{-1}\big|JM,JM'\rangle = \big|JM,JN\rangle D^J{}_{NM'}(u). \tag{4.6}$$

Furthermore, from the well-known orthogonality and completeness properties of the representation functions[8] if follows that $\{|JM,JM'\rangle\}$ is an orthonormal basis of $CT^{0,0}(\mathbf{G})$:

$$\langle J_1M_1,J_1M_1'\big|J_2M_2,J_2M_2'\rangle = \langle Q^{J_1M_1M_1'},Q^{J_2M_2M_2'}\rangle$$

$$= \delta_{J_1J_2}\delta_{M_1M_2}\delta_{M_1'M_2'}. \tag{4.7}$$

The diagonalization of $\mathbf{J}^2,\mathbf{J}_3,\mathbf{L}^2=\tilde{\mathbf{L}}^2$ leads to a standard basis of the subspaces of the weight l component of the adjoint scalar dragging representation. This new orthonormal basis $\{Q^J{}_{lm}\}$, where J, l, m refer to $\mathbf{L}^2$, $\mathbf{J}^2$, $\mathbf{J}_3$ and $l=0,\ldots,2J$, is obtained by familiar angular momentum addition using Clebsch–Gordon coefficients[9]:

$$Q^J{}_{lm} = C_{JJ}(lm;MM')Q^{JMM'}$$

$$\langle Q^{J_1}{}_{l_1m_1},Q^{J_2}{}_{l_2m_2}\rangle = \delta_{J_1J_2}\delta_{l_1l_2}\delta_{m_1m_2},$$

$$\mathbf{AD}_uQ^J{}_{lm} = Q^J{}_{ln}D^J{}_{nm}(u). \tag{4.8}$$

This new basis is orthonormal because of the unitary character of the Clebsch–Gordan transformation. Either basis might be called the scalar harmonics on $\mathbf{G}$. When a distinction is required, we will refer to $\{Q^{JMM'}\}$ as the left–right harmonics and $\{Q^J{}_{lm}\}$ as the adjoint harmonics.

By (3.7) the functions $\{Q^{JMM'}\}$ or $\{Q^J{}_{lm}\}$ for fixed J are n_J^2 eigenvectors of the Laplacian Δ with eigenvalue $\Delta^J = 4J(J+1) = n_J^2 - 1$, where $n_J = 2J+1$. The usual approach[11] is to obtain such eigenvectors by separation of variables in standard spherical coordinates $\{\chi,\theta,\phi\}$ on $\mathbf{G}$, in terms of which the expression for the Laplacian is well known:

$$\Delta = -\sin^{-2}\chi\,\partial/\partial\chi(\sin^2\chi\,\partial/\partial\chi) + \sin^{-2}\chi J^2,$$

$$-J^2 = \sin^{-1}\theta\,\partial/\partial\theta(\sin\theta\,\partial/\partial\theta) + \partial^2/\partial\phi^2. \tag{4.9}$$

Here J^2 is the square of the total angular momentum and it is easy to see that $J_3 = -i\partial/\partial\phi$. [The adjoint group acts on the 2-spheres of constant χ exactly as $SO(3,R)$ acts on the 2-spheres of R^3 centered at the origin.] Since $J^2\mathbf{Y}_{lm} = l(l+1)\mathbf{Y}_{lm}$, where $\mathbf{Y}_{lm}(\theta,\phi)$ are the standard spherical harmonics on the 2-sphere, separation of variables with $\mathcal{Y}^n{}_{lm}(\chi,\theta,\phi) = \Pi_{nl}(\chi)\mathbf{Y}_{lm}(\theta,\phi)$ yields the following χ equation and solution[12]:

$$\{\sin^{-2}\chi\,\partial/\partial\chi(\sin^2\chi\,\partial/\partial\chi) + n^2 - 1$$

$$-\sin^{-2}\chi\,l(l+1)\}\Pi_{nl}(\chi) = 0,$$

$$\Pi_{nl}(\chi) = \sin^l\chi C^{l+1}_{n-(l+1)}(\cos\chi),$$

where $C^l{}_n$ is the Gegenbauer polynomial. $\mathcal{Y}^{nJ}_{lm}$ equals Q^J_{lm} up to a constant factor.

5. VECTOR AND TENSOR HARMONICS

Suppose we denote by $|JM,J'M',s,\binom{p}{q}\rangle$ $\binom{p}{q}$-tensor fields of integral spin s (to be explained shortly) which are simultaneous eigenvectors of $\tilde{\mathbf{L}}^2,\tilde{\mathbf{L}}_3,\mathbf{L}^2,\mathbf{L}_3$ with eigenvalues specified by J,M,J',M', respectively. Simultaneous eigenvectors of $\mathbf{J}^2,\mathbf{J}_3,\tilde{\mathbf{L}}^2,\mathbf{L}^2$ are obtained from these by simple angular momentum addition:

$$\Big|lm,JJ',s\binom{p}{q}\rangle_{\mathrm{AD}} = C_{JJ'}(lm,MM')\Big|JM,J'M',s,\binom{p}{q}\Big\rangle \tag{5.1}$$

We shall define the first basis (left–right harmonics) to be a standard orthonormal[13] basis for the left and right $\binom{p}{q}$-tensor dragging representations. The second basis (adjoint harmonics) will then be a standard orthonormal basis for the adjoint $\binom{p}{q}$-tensor dragging representation. It is sufficient, however, to consider only $\binom{0}{p+q}$-tensor fields because of the natural correspondence determined by the bi-invariant metric. (Lie derivatives with respect to $\mathfrak{g}$ and $\tilde{\mathfrak{g}}$ commute with "index raising and lowering" since $\mathbf{g}$ is bi-invariant: $\mathbf{L}_a\mathbf{g} = \tilde{\mathbf{L}}_a\mathbf{g} = 0$.) Furthermore, we are only interested in the 1-form case ("vector harmonics") and in the symmetric $\binom{0}{2}$-tensor field case ("tensor harmonics"). We will use the alternative notation $\mathbf{X}^{JM,J'M'}$ and $\mathbf{X}^{JJ'}_{lm}$ for the vector harmonics ($s=1$) and $\mathbf{T}_s^{JM,J'M'}$ and $\mathbf{T}^{JJ's}_{lm}$ with $s=0,2$ for the tensor harmonics. The antisymmetric $\binom{0}{2}$-tensor harmonics ($s=1$) may be obtained from the vector harmonics by the Hodge star duality operation (which commutes with the DeRham Laplacian). Let $\hat{\sigma}$ denote the vector field associated with the one-form σ (obtained by contracting σ with the contravariant metric $\mathbf{g}^{-1}$ or "raising its index"); for example, $\hat{\omega}^a = 4e_a$. We will have occasion to refer to the vector field harmonics $\hat{\mathbf{X}}^{JM,J'M'}$.

The notion of spin arises in decomposing the adjoint dragging representation in the subspace of either left or right invariant $\binom{0}{r}$-tensor fields. The $\binom{0}{r}$-tensor harmonics are then obtained by coupling the scalar harmonics to spin eigenvectors. The $\binom{0}{r}$-tensor field spin eigenvectors are themselves obtained by decomposing the tensor products of the one-form eigenvectors. Suppose for example we choose the left invariant fields and introduce the spherical basis $\{\omega^A\}$ corresponding to the "Cartesian" basis $\{\omega^a\}$ of $\mathfrak{g}^*$:

$$\omega^{\pm1} = \mp 2^{-1/2}(\omega^1 \pm i\omega^2),\quad \omega^0 = \omega^3,\quad (\omega^A,\omega^B) = 4\delta_{AB}. \tag{5.2}$$

From the formula $\pounds_{e_a}\omega^b = -C^b{}_{ac}\omega^c$ and from the fact that left angular momentum annihilates ω^b so that $\mathbf{J}_a\omega^A = \mathbf{L}_a\omega^A$, it follows that $\{\omega^A\}$ is a standard spin one basis:

$$\mathbf{J}^2\omega^A = 2\omega^A, \quad \mathbf{J}_+\omega^A = \epsilon_{1A}\omega^{A+1},$$
$$\mathbf{J}_3\omega^A = A\omega^A, \quad \mathbf{J}_-\omega^A = \epsilon_{1A-1}\omega^{A-1}. \tag{5.3}$$

(The same is true of $\{\tilde{\omega}^A\}$ for which $\mathbf{J}_a\tilde{\omega}^A = \tilde{\mathbf{L}}_a\tilde{\omega}^A$.) We therefore define $\mathbf{X}^{00,1A} = \omega^A$ and obtain all the vector harmonics by right angular momentum coupling of the scalar harmonics to the left invariant spin one basis $\{\omega^A\}$:

$$\mathbf{X}^{JM,J'M'} = C_{J1}(J'M', NA)\,|\,JM, JN\rangle\,\omega^A,$$
$$J' = J+1, J, J-1. \tag{5.4}$$

From the 1-form basis $\{\omega^A\}$ we may construct the standard second rank spherical basis in the usual fashion yielding spin components with $s = 0, 1, 2$:

$$\omega^{sm} = C_{11}(sm, AB)\omega^A \otimes \omega^B,$$
$$\mathbf{J}^2\omega^{sm} = s(s+1)\omega^{sm}, \quad J_3\omega^{sm} = m\omega^{sm},$$
$$(\omega^{sm}, \omega^{s'm'}) = 4^2\delta_{ss'}\,\delta_{mm'}. \tag{5.5}$$

[By successive coupling one may obtain standard left invariant spin bases for $\binom{0}{r}$-tensor fields with $r > 2$.] The explicit expressions are

$$\omega^{22} = \omega^{+1}\otimes\omega^{+1},$$
$$\omega^{21} = 2^{-1/2}(\omega^{+1}\otimes\omega^0 + \omega^0\otimes\omega^{+1}),$$
$$\omega^{20} = 6^{-1/2}(\omega^{+1}\otimes\omega^{-1} + 2\omega^0\otimes\omega^0 + \omega^{-1}\otimes\omega^{+1}),$$
$$\omega^{2-1} = 2^{-1/2}(\omega^{-1}\otimes\omega^0 + \omega^0\otimes\omega^{-1}),$$
$$\omega^{2-2} = \omega^{-1}\otimes\omega^{-1},$$
$$\omega^{11} = 2^{-1/2}(\omega^{+1}\otimes\omega^0 - \omega^0\otimes\omega^{+1}), \tag{5.6}$$
$$\omega^{10} = 2^{-1/2}(\omega^{+1}\otimes\omega^{-1} - \omega^{-1}\otimes\omega^{+1}),$$
$$\omega^{1-1} = 2^{-1/2}(\omega^0\otimes\omega^{-1} - \omega^{-1}\otimes\omega^0),$$
$$\omega^{00} = 3^{-1/2}(\omega^{+1}\otimes\omega^{-1} - \omega^0\otimes\omega^0 + \omega^{-1}\otimes\omega^{+1}),$$
$$= \mathbf{T}_0{}^{00,00} = -4(3)^{-1/2}g.$$

They correspond to the symmetric traceless ($s = 2$), antisymmetric ($s = 1$) and pure trace ($s = 0$) parts of a second rank covariant tensor field.

By right angular momentum coupling we therefore obtain the tensor harmonics:

$$\mathbf{T}_s^{JM,J'M'} = C_{Js}(J'M', Nm)\,|\,JM, JN\rangle\,\omega^{sm},$$
$$J' = J+s, \ldots, J-s. \tag{5.7}$$

For $s = 0$ this is trivial:

$$\mathbf{T}_0{}^{JM,J'M'} = Q^{JMM'}\omega^{00} = -4(3)^{-1/2}Q^{JMM'}g. \tag{5.8}$$

Both $\{\mathbf{X}^{JM,J'M'}\}$ and $\{\mathbf{T}_s{}^{JM,J'M'}\}$ are orthonormal apart from factors of 4:

$$\langle\mathbf{X}^{J_1M_1,J'_1M'_1}, \mathbf{X}^{J_2M_2,J'_2M'_2}\rangle = 4\delta_{J_1J_2}\delta_{M_1M_2}\delta_{J'_1J'_2}\delta_{M'_1M'_2}$$
$$\langle\mathbf{T}_{s_1}{}^{J_1M_1,J'_1M'_1}, \mathbf{T}_{s_2}{}^{J_2M_2,J'_2M'_2}\rangle = 4^2\delta_{J_1J_2}\delta_{M_1M_2}\delta_{J'_1J'_2}\delta_{M'_1M'_2}$$
$$\times\,\delta_{s_1s_2}. \tag{5.9}$$

However, we could have chosen right invariant spin eigenvectors and left angular momentum coupling. The two approaches are connected by the following result:

$$\mathbf{X}^{1A,00} = C_{11}(00, BC)\,|\,1A, 1B\rangle\,\omega^C = -\tilde{\omega}^A. \tag{5.10}$$

This is not surprising since $D^{1A}{}_B$ are the components of $\mathcal{R}$ in a spherical basis; the manipulation of (3.5) establishes this result. The minus sign appears since an additional minus sign is required in defining left angular momentum relative to right angular momentum. Using the properties of the Clebsch–Gordan coefficients and the representation functions,[10] one may show that the harmonics constructed by left angular momentum coupling of the scalar harmonics to the right invariant spherical spin bases built from $\{-\tilde{\omega}^a\}$ coincide exactly with the harmonics already defined:

$$\mathbf{X}^{JM,J'M'} = C_{J'1}(JM, NA)\,|\,J'N, J'M'\rangle\ (-\tilde{\omega}^A),$$
$$\mathbf{T}_s{}^{JM,J'M'} = C_{J's}(JM, Nm)\,|\,J'N, J'M'\rangle\,\tilde{\omega}^{sm}. \tag{5.11}$$

The vector and tensor harmonics are eigenvectors of the ordinary and DeRham Laplacians. To evaluate the eigenvalues, it is convenient to work in the left invariant frame $\{e_a\}$ and decompose right angular momentum into orbital and spin parts:

$$\mathbf{L}_a = \mathbf{L}_a^{\rm orb} + \mathbf{S}_a$$
$$(\mathbf{L}_a^{\rm orb}T)^{b\cdots}{}_{c\cdots} = L_aT^{b\cdots}{}_{c\cdots}$$
$$(\mathbf{S}_a\mathbf{T})^{b\cdots}{}_{c\cdots} = iC^b{}_{ad}T^{d\cdots}{}_{c\cdots} + \cdots - iT^{b\cdots}{}_{d\cdots}C^d{}_{ac} - \cdots. \tag{5.12}$$

Here $\mathbf{T}^{b\cdots}{}_{c\cdots}$ are the components of a $\binom{p}{q}$-tensor field $\mathbf{T}$ in this frame. The square of the right orbital angular momentum is just $\tilde{\mathbf{L}}^2$:

$$(\tilde{\mathbf{L}}T)^{a\cdots}{}_{b\cdots} = \tilde{\mathbf{L}}^2\mathbf{T}^{a\cdots}{}_{b\cdots} = \mathbf{L}^2\mathbf{T}^{a\cdots}{}_{b\cdots}$$
$$= ((\mathbf{L}^{\rm orb})^2\mathbf{T})^{a\cdots}{}_{b\cdots}, \tag{5.13}$$

while the spin–orbit operator is

$$\mathbf{L}^2 = (\mathbf{L}^{\rm orb})^2 + \mathbf{S}^2 + 2\mathbf{S}_a\mathbf{L}_a^{\rm orb}$$
$$2\mathbf{S}_a\mathbf{L}_a^{\rm orb} = \mathbf{L}^2 - \tilde{\mathbf{L}}^2 - \mathbf{S}^2. \tag{5.14}$$

From (2.5), (2.10), and (5.12) one may verify that the square of the right spin angular momentum is the difference between the ordinary and DeRham Laplacians:

$$\Delta_{\rm DR} - \Delta = \mathbf{S}^2. \tag{5.15}$$

A similar decomposition $\tilde{\mathbf{L}}_a = \tilde{\mathbf{L}}_a^{\rm orb} + \tilde{\mathbf{S}}_a$ of left angular momentum using the right invariant frame $\{\tilde{e}_a\}$ leads to a left spin angular momentum which is easily seen to be related to the right spin by $\tilde{\mathbf{S}}_a = \mathbf{S}_b\mathcal{R}^{-1b}{}_a$ which in turn implies $\tilde{\mathbf{S}}^2 = \mathbf{S}^2$.[14] Since $\mathbf{S}_a\mathbf{T}$

$=\mathbf{L}_a\mathbf{T}=\mathbf{J}_a\mathbf{T}$ if $\mathbf{T}$ is left invariant and $\tilde{\mathbf{S}}_a\mathbf{T}=\tilde{\mathbf{L}}_a\mathbf{T}=\mathbf{J}_a\mathbf{T}$ if $\mathbf{T}$ is right invariant, $\mathbf{S}^2=\tilde{\mathbf{S}}^2$ coincides with $\mathbf{J}^2$ on the invariant tensor fields and is therefore the spin introduced above whose eigenvalues were labeled by s. Functions have spin $s=0$ and vector fields $s=1$ while second rank tensor fields decompose into spins $s=0, 1, 2$.

The covariant derivative and Laplacian may be expressed in terms of these operators using (2.6) and (2.8):

$$
\begin{aligned}
&i\nabla_{e_a}=\mathbf{L}_a^{\mathrm{orb}}+\tfrac{1}{2}\mathbf{S}_a,\\
&\Delta=-4\delta^{ab}\nabla_{e_a}\nabla_{e_a}=4(\mathbf{L}^{\mathrm{orb}})^2+\mathbf{S}^2+4\mathbf{S}_a\mathbf{L}_a^{\mathrm{orb}}\\
&\quad=2(\mathbf{L}^2+\tilde{\mathbf{L}}^2)-\mathbf{S}^2,\\
&\Delta_{\mathrm{DR}}=2(\mathbf{L}^2+\tilde{\mathbf{L}}^2).
\end{aligned}
\tag{5.16}
$$

A spin-s tensor field of left and right angular momentum J and J' respectively is therefore an eigenvector of both Laplacians. Denote the corresponding eigenvalues by $\Delta^{JJ's}$ and $\Delta_{\mathrm{DR}}^{JJ'}$ and define $n_{JJ'}=J+J'+1$, so that $n_{JJ}=n_J$:

$$
\begin{aligned}
&\Delta^{JJ's}=\Delta_{\mathrm{DR}}^{JJ'}-s(s+1),\\
&\Delta_{\mathrm{DR}}^{JJ'}=2\{J(J+1)+J'(J'+1)\}\\
&\quad=n_{JJ'}^2-1+(J-J')^2.
\end{aligned}
\tag{5.17}
$$

6. DIFFERENTIAL PROPERTIES

It is worthwhile knowing how the harmonics behave under the operations of taking divergences, exterior derivatives (when appropriate), and symmetrized covariant derivatives. These operations connect the scalar, vector, and tensor harmonics of fixed left and right angular momentum.

Consider the exterior derivative of the scalar harmonics and let $\mathbf{L}_A$ stand for the spherical components of $\mathbf{L}_a$:

$$
\mathbf{L}_{\pm1}=\mp2^{-1/2}L_{\pm},\quad \mathbf{L}_0=\mathbf{L}_3,
$$

$$
id\Phi=\omega^a\mathbf{L}_a\Phi=(-1)^A\omega^{-A}\mathbf{L}_A\Phi. \tag{6.1}
$$

By the definition of the coefficients $C_{J1}(JM, M+A-A)$[15]:

$$
\begin{aligned}
\mathbf{L}_AQ^{JMM'}&=(-1)^A[J(J+1)]^{1/2}\\
&\quad\times C_{J1}(JM', M'+A-A)Q^{JM,M'+A},\\
id Q^{JMM'}&=[J(J+1)]^{1/2}C_{J1}(JM', NA)Q^{JMN}\omega^A\\
&=[J(J+1)]^{1/2}\mathbf{X}^{JM,JM'}.
\end{aligned}
\tag{6.2}
$$

This is exactly analogous to a similar situation occurring with the scalar and vector harmonics on R^3.

To compute the exterior derivative of the vector harmonics we use the following formulas:

$$
\begin{aligned}
(d\mathbf{A})_{bc}&=2e_{[b}\mathbf{A}_{c]}-\mathbf{A}_dC^d{}_{bc}\\
(*d\mathbf{A})_a&=\tfrac{1}{2}\eta_a{}^{bc}(d\mathbf{A})_{bc}=\epsilon_{abc}(d\mathbf{A})_{bc}=-2i\epsilon_{abc}\mathbf{L}_b\mathbf{A}_c-2\mathbf{A}_a\\
&=((-2\mathbf{S}_b\mathbf{L}_b^{\mathrm{orb}}-\mathbf{S}^2)\mathbf{A})_a=((\tilde{\mathbf{L}}^2-\mathbf{L}^2)\mathbf{A})_a\\
*d\mathbf{A}&=(\tilde{\mathbf{L}}^2-\mathbf{L}^2)\mathbf{A}.
\end{aligned}
\tag{6.3}
$$

Applying this to the left–right vector harmonics, we obtain

$$
\begin{aligned}
\delta*\mathbf{X}^{JM,J'M'}&=*d\,\hat{\mathbf{X}}^{JM,J'M'}=[J(J+1)-J'(J'+1)]\mathbf{X}^{JM,J'M}\\
&=(J-J')n_{JJ'}\mathbf{X}^{JM,J'M'}.
\end{aligned}
\tag{6.4}
$$

The same formula holds for the adjoint vector harmonics.

The divergence of $\mathbf{X}^{JM,J'M'}$ is easy to compute using (6.2) and the formula $\Delta\Phi=(\delta d+d\delta)\Phi=\delta\, d\Phi$:

$$
\begin{aligned}
\delta\mathbf{X}^{JM,J'M'}&=i[J(J+1)]^{-1/2}\delta\, dQ^{JMM'}\\
&=4i[J(J+1)]^{-1/2}Q^{JMM'}.
\end{aligned}
\tag{6.5}
$$

To evaluate $\delta\mathbf{X}^{JM,J\pm1M'}$ we need the divergence formula:

$$
\delta\mathbf{A}=-g^{ab}e_a\mathbf{A}_b=4i\mathbf{L}_a\mathbf{A}_a=4i(-1)^A\mathbf{L}_A\mathbf{A}_{-A},
$$

where $\mathbf{A}_A$ are the spherical components of $\mathbf{A}_a$ defined as in (5.2):

$$
\mathbf{A}_a\omega^a=(-1)^A\mathbf{A}_{-A}\omega^A.
$$

δ commutes with all three angular momenta (since they generate isometries) so that all harmonics obtained from a divergenceless harmonic by the left and right raising and lowering operators are also divergenceless. The explicit calculation:

$$
\mathbf{X}^{JM,J+1\,J+1}=Q^{JMJ}\omega^{+1},
$$

$$
\delta\mathbf{X}^{JM,J+1\,J+1}=-4i(2)^{-1/2}\mathbf{L}_+Q^{JMJ}=0
$$

then establishes $\delta\mathbf{X}^{JM,J+1M'}=0$. The corresponding relation $\delta\mathbf{X}^{JM,J-1,M'}=0$ follows from left–right symmetry [since a similar calculation using (5.11) would show $\delta\mathbf{X}^{J+1M,JM'}=0$]. Thus the vector harmonics are transverse (i.e., divergenceless) for $J\neq J'$ and exact for $J=J'$. This reflects the Hodge decomposition[16] of the space of 1-forms which in our case contains no harmonic elements ($\Delta_{\mathrm{DR}}A=0$).

The divergence of a symmetric second rank covariant tensor field h is given by

$$
\begin{aligned}
(\delta h)_a&=-g^{bc}e_bh_{ca}=4i\mathbf{L}_bh_{ba}\\
(\delta h)_A&=4i(-1)^B\mathbf{L}_Bh_{-BA}.
\end{aligned}
$$

Since $\mathbf{T}_2^{JM,J+2\,J+2}=Q^{JMJ}\omega^{+1}\otimes\omega^{+1}$, it follows exactly as in the vector case that $\delta\mathbf{T}_2^{JM,J\pm2\,M'}=0$, so $\{\mathbf{T}_2^{JM,J\pm2\,M'}\}$ are transverse tracless harmonics. From (6.2) and the formula $\delta(\Phi\mathbf{g})=-d\Phi$, it follows that

$$
\delta\mathbf{T}_0^{JM,JM'}=-4i(3)^{-1/2}[J(J+1)]^{1/2}\mathbf{X}^{JM,JM'}. \tag{6.6}
$$

The remaining spin-2 tensor harmonics may be obtained from vector field harmonics by Lie derivation. To simplify notation, let i stand for the four

indices $JMJ'M'$, δ_{ij} for the product of the four individual Kronecker deltas, n_i for $n_{JJ'}$, and Δ^i_{DR} for $\Delta^{JJ'}_{DR}$. The vector field harmonics $\hat{\mathbf{X}}^i$ are obtained from the one-form harmonics $\mathbf{X}^i$ by replacing ω^A by $4e_A$ in (5.4). Consider the tensor fields $\mathbf{S}^i = \pounds_{\hat{\mathbf{X}}^i}\mathbf{g}$ with components $2\mathbf{X}^i{}_{(a;b)}$ and let $\mathbf{S}^i_2$ and $\mathbf{S}^i_0$ be the traceless and pure trace parts of $\mathbf{S}^i$. A short calculation using $\mathbf{L}_a\mathbf{g} = 0 = \tilde{\mathbf{L}}_a\mathbf{g}$ shows that

$$\mathbf{L}_a\mathbf{S}^i = \pounds_{\mathbf{L}_a\hat{\mathbf{X}}^i}\mathbf{g}, \quad \tilde{\mathbf{L}}_a\mathbf{S}^i = \pounds_{\tilde{\mathbf{L}}_a\hat{\mathbf{X}}^i}\mathbf{g},$$

so $\mathbf{S}^i$ inherits the angular momentum properties of $\hat{\mathbf{X}}^i$ which in turn are identical with those of $\mathbf{X}^i$. It then follows that $\mathbf{S}^i_0$ and $\mathbf{S}^i_2$ are respectively proportional to $\mathbf{T}^i_0$ and $\mathbf{T}^i_2$. The proportionality factors may be determined up to phase by comparing norms. (Once they are obtained for one i, application of the raising and lowering operators shows that the values are independent of M and M'.)

Since $\mathbf{S}^{ia}{}_{;a} = 2\mathbf{X}^{ia}{}_{;a} = -2\delta\mathbf{X}^i$, $\mathbf{S}^i_0$ vanishes when $J' = J \pm 1$ and so $\mathbf{S}^i$ is itself proportional to $\mathbf{T}^i_2$. When $J' = J$:

$$\begin{aligned}\mathbf{S}^i_0 &= -\tfrac{2}{3}\mathbf{g}\delta\mathbf{X}^i \\ &= (2i/3)[J(J+1)]^{1/2}Q^i\mathbf{g} = 2i(3)^{-1/2}[J(J+1)]^{1/2}\mathbf{T}^i_2.\end{aligned} \tag{6.7}$$

In this case since $d\mathbf{X}^i = 0$:

$$\mathbf{S}^i{}_{ab} = 2\mathbf{X}^i{}_{a;b} = 2i[J(J+1)]^{-1/2}Q^i{}_{;a;b}. \tag{6.8}$$

The divergence of $\mathbf{S}^i$ may be computed with the help of the Ricci identity:

$$\begin{aligned}&\mathbf{X}^i{}_{b;a}{}^{;b} = \mathbf{X}^i{}_{b;}{}^{;b}{}_{;a} + \mathbf{X}^i d\mathbf{R}^d{}_{ba}{}^b = ((2-d\delta)\mathbf{X}^i)_a, \\ &\mathbf{S}^i{}_{ab}{}^{;b} = \mathbf{X}^i{}_{a;b}{}^{;b} + \mathbf{X}^i{}_{b;a}{}^{;b}, \\ &\delta\mathbf{S}^i = (\Delta - 2 + d\delta)\mathbf{X}^i.\end{aligned} \tag{6.9}$$

When $J' = J \pm 1$, $\delta\mathbf{X}^i = 0$ so $\delta\mathbf{S}^i = (\Delta - 2)\mathbf{X}^i = (n_i^2 - 4)\mathbf{X}^i$, while if $J = J'$ then $d\delta\mathbf{X}^i = \Delta_{DR}\mathbf{X}^i$ and $\delta\mathbf{S}^i = 2\Delta\mathbf{X}^i = 2(n_i^2 - 3)\mathbf{X}^i$. Let $\delta\mathbf{S}^i = c^i\mathbf{X}^i$. Then $\{\mathbf{S}^i\}$ are orthogonal

$$\begin{aligned}\langle\mathbf{S}^i, \mathbf{S}^j\rangle &= 2\int^{\mathbf{G}}\hat{\eta}\mathbf{S}^{iab}\mathbf{X}^j{}_{a;b} \\ &= 2\langle\delta\mathbf{S}^i, \mathbf{X}^j\rangle = 2c^i\langle\mathbf{X}^i, \mathbf{X}^j\rangle \\ &= 4^2\delta_{ij}c^i/2.\end{aligned} \tag{6.10}$$

When $J' = J + 1$ it then follows that

$$i\mathbf{S}^i = [(n_i^2 - 4)/2]^{1/2}\mathbf{T}^i_2. \tag{6.11}$$

The phase in (6.11) and (6.12) has been determined by evaluating $\mathbf{S}^{JM,J+1\,J+1}$ and $\mathbf{S}^{JM,JJ}$ explicitly using Clebsch–Gordan coefficient formulas which may be found in Ref. 17. When $J' = J$, knowledge of $\langle\mathbf{S}^i_0, \mathbf{S}^i_0\rangle = 4^2\delta_{ij}[(n_i^2 - 1)/3]^{1/2}$ from (6.7) is sufficient to evaluate $\langle\mathbf{S}^i_2, \mathbf{S}^i_2\rangle$ by orthogonality:

$$\begin{aligned}&\langle\mathbf{S}^i_2, \mathbf{S}^j_2\rangle = \langle\mathbf{S}^i, \mathbf{S}^j\rangle - \langle\mathbf{S}^i_0, \mathbf{S}^j_0\rangle = 4^2\delta_{ij}[2(n_i^2 - 4)/3], \\ &i\mathbf{S}^i_2 = [2(n_i^2 - 1)/3]^{1/2}\mathbf{T}^i_2.\end{aligned} \tag{6.12}$$

The Ricci identity may also be used to evaluate other derivatives. For example

$$\begin{aligned}&\mathbf{S}^i{}_{ac;b}{}^{;c} = \mathbf{S}^i{}_{ac}{}^{;c}{}_{;b} + 3\mathbf{S}^i_{2ab}, \\ &-\mathbf{S}^i{}_{ac;b}{}^{;c} - \mathbf{S}^i{}_{bc;a}{}^{;c} = c^i\mathbf{S}^i{}_{ab} - 6\mathbf{S}^i_{2ab}.\end{aligned} \tag{6.13}$$

Let us adopt the convention that by omitting certain right angular momentum indices on the harmonics we mean the linear span of the harmonics with the remaining indices fixed and the omitted ones assuming all possible values. For example $\mathbf{X}^{JM}$ is the space spanned by $\mathbf{X}^{JM,J'M'}$ with JM fixed. Also by $\mathbf{T}^{JM}$ and $\mathbf{T}^{JM,J'}$ we mean the direct sum of the spaces with $s = 0$ and $s = 2$ only. The Hodge decomposition is then reflected in the orthogonal direct sum:

$$\mathbf{X}^{JM} = \mathbf{X}^{JM,J\pm1} \oplus \mathbf{X}^{JM,J}.$$

$(\mathbf{X}^{JM,J\pm1} = \mathbf{X}^{JM,J+1} \oplus \mathbf{X}^{JM,J-1}$, etc.).

Consider $\mathbf{T}^{JM} = \mathbf{T}_2{}^{JM,J\pm2} \oplus \mathbf{T}_2{}^{JM,J\pm1} \oplus \mathbf{T}^{JM,J}$. The first summand contains the transverse traceless harmonics and the second those which result from Lie derivation of $\mathbf{g}$ by transverse vector field harmonics. The last summand is in an obvious notation $\pounds_{\hat{\mathbf{X}}^{JM,J}}g \oplus Q^{JM}\mathbf{g}$ but the direct sum is not orthogonal. If we decompose $\mathbf{T}^{JM,J}$ into an orthogonal direct sum with $\pounds_{\hat{\mathbf{X}}^{JM,J}}\mathbf{g}$ as the first summand, the second summand must be transverse:

$$\begin{aligned}0 = \langle\mathbf{S}^i, P\rangle &= 2\int^{\mathbf{G}}\hat{\eta}\mathbf{X}^{ia;b}P_{ab} = -2\int^{\mathbf{G}}\hat{\eta}\mathbf{X}^{ia}P_{ab}{}^{jb} \\ &= 2\langle\mathbf{X}^i, \delta P\rangle.\end{aligned}$$

Since $P \in \mathbf{T}^{JM,J}$ implies $\delta P \in \mathbf{X}^{JM,J}$, orthogonality to $\pounds_{\hat{\mathbf{X}}^{JM,J}}\mathbf{g}$ requires that $\delta P = 0$. Given a function Φ, the Ricci identity shows immediately that the following combination of Φ and its second derivatives is divergenceless:

$$(\Phi - \tfrac{1}{2}\Delta\Phi)\mathbf{g} - \tfrac{1}{2}\nabla\nabla\Phi.$$

It is exactly these combinations of the scalar harmonics which span the transverse subspace of $\mathbf{T}^{JM,J}$:

$$\begin{aligned}&p^i = (Q^i - \tfrac{1}{2}\Delta Q^i)\mathbf{g} - \tfrac{1}{2}\nabla\nabla Q^i, \\ &P^{ia}{}_a = (3 - \Delta)Q^i = (4 - n_i^2)Q^i, \\ &\langle P^i, P^j\rangle = 4^2\delta_{ij}[(n_i^2 - 3)(n_i^2 - 4)/32].\end{aligned} \tag{6.14}$$

$\{\mathbf{S}^i, \mathbf{P}^i\}$ is an orthogonal basis of $\mathbf{T}^{JM,J}$ and the following is an orthogonal direct sum reflecting the canonical decomposition of symmetric second rank tensor fields on a Riemannian manifold of constant positive curvature[18]:

$$\mathbf{T}^{JM} = \mathbf{T}_2{}^{JM,J\pm2} \oplus \pounds_{\hat{\mathbf{X}}^{JM}}\mathbf{g} \oplus P^{JM}.$$

7. REALITY PROPERTIES AND PARITY

The reality properties of the harmonics are important if one is interested in real tensor fields. Let $\mathbf{X}^{JM}{}_{NA} = Q^{JMN}\omega^A$ and $\mathbf{T}_s{}^{JM}{}_{Nm} = Q^{JMN}\omega^{sm}$. These are standard orthonormal bases (modulo factors of 4) of the left dragging representations but not the right. The left–right harmonics are related to these left harmonics by the real unitary Clebsch–Gordan transformation which decomposes the right angular

momentum tensor product representation into irreducible components. The adjoint harmonics are obtained from the latter harmonics by another real Clebsch–Gordan transformation.

The representation functions[10] and therefore our left–right scalar harmonics satisfy

$$\overline{D^{J\,M}}_{M'} = (-1)^{M-M'} D^{J\,-M}{}_{-M'}, \quad \overline{Q^{JMM'}} = (-1)^{M+M'} Q^{J\,-M\,-N},$$

while the left invariant spherical spin bases satisfy

$$\overline{\omega^A} = (-1)^A \omega^{-A}, \quad \overline{\omega^{sm}} = (-1)^m \omega^{s\,-m}.$$

Together these imply

$$\overline{X^{J\,M}}_{NA} = (-1)^{M+N+A} \mathbf{X}^{J\,-M}{}_{-N-A},$$

$$\overline{\mathbf{T}_s{}^{J\,M}}_{Nm} = (-1)^{M+N+A} \mathbf{T}_s{}^{J\,-M}{}_{-N-m}.$$

The sign is always $(-1)^J$[3]. The reality properties of the left–right and adjoint harmonics are therefore

$$\overline{\mathbf{X}^{JM,J'M'}} = (-1)^{M+M'} \mathbf{X}^{J\,-M,J'\,-M'},$$

$$\overline{\mathbf{X}^{JJ'}_{lm}} = (-1)^m \mathbf{X}^{JJ'}_{l\,-m}$$

$$\overline{\mathbf{T}_s{}^{JM,J'M'}} = (-1)^{M+M'} \mathbf{T}_s{}^{J\,-M,J'\,-M'},$$

$$\overline{\mathbf{T}^{JJ's}_{lm}} = (-1)^m \mathbf{T}^{JJ's}_{l\,-m}.$$

One may also introduce the notion of inversion and parity. Let P denote the inverse diffeomorphism and its corresponding dragging operator. P is a discrete isometry of $\mathbf{g}$ satisfying $P^2 = Id$ (the identity diffeomorphism and dragging operator). It inverts $\mathbf{G}$ about the north pole. In Cartesian coordinates restricted to $\mathbf{G}$, $P\{y^a, y^4\} = \{-y^a, y^4\}$ and hence in terms of spherical coordinates, χ remains unchanged while the two-spheres of constant χ behave exactly like the spheres centered at the origin of R^3 under inversion. Q^J_{lm} therefore has the same parity as $\mathbf{Y}_{lm}$.

Let $X \in \mathfrak{g}$ and Φ be a function:

$$(PX)(u)\Phi = X(u^{-1})\Phi \circ P = (d/dt)\big|_0 \Phi \circ P(u^{-1} \exp tX)$$

$$= (d/dt)\big|_0 \Phi((\exp - tX)u) = -\tilde{X}(u)\Phi.$$

Then it follows that $Pe_a = -\tilde{e}_a$ and $P\omega^a = -\tilde{\omega}^a$. Since $PD^{J\,M}{}_{M'} = (-1)^{M-M'} D^{J\,-M}{}_{-M'}$,[10] the scalar harmonics satisfy:

$$PQ^{J\,MM'} = (-1)^{-2M'} Q^{J\,M'\,M} = (-1)^{2J} Q^{J\,M'\,M}.$$

With the help of (5.11) the inversion properties of the left–right harmonics are easily deduced:

$$P\mathbf{X}^{JM,J'M'} = (-1)^{2J} C_{J1}(J'M', NA) Q^{J\,NM'}(-\tilde{\omega}^A)$$

$$= (-1)^{2J} \mathbf{X}^{J'M'JM},$$

$$P\mathbf{T}_s{}^{JM,J'M'} = (-1)^{2J} \mathbf{T}_s{}^{J'M',JM}.$$

A Clebsch–Gordan coefficient symmetry then yields the inversion properties of the adjoint harmonics:

$$P\mathbf{X}^{JJ'}_{lm} = C_{JJ'}(lm, MM') P\mathbf{X}^{JM,J'M'}$$

$$= (-1)^{2J} C_{JJ'}(lm, MM') \mathbf{X}^{J'M',JM}$$

$$= (-1)^{l+J'-J} C_{JJ'}(lm, M', M) \mathbf{X}^{J'M',JM}$$

$$= (-1)^{l+J'-J} \mathbf{X}^{J'J}_{lm}.$$

The same manipulation shows

$$P\mathbf{T}^{JJ's}_{lm} = (-1)^{l+J'-J} \mathbf{T}^{J'Js}_{lm}, \quad PQ^J{}_{lm} = (-1)^l Q^J_{lm}.$$

For $J < J'$ define the following parity eigenvectors with eigenvalues:

$p(-1)^{l+J'-J}$:

$$\mathbf{X}^{JJ'p}_{lm} = (2)^{-1/2}(\mathbf{X}^{JJ'}_{lm} + p\mathbf{X}^{JJ'}_{lm}),$$

$$\mathbf{T}^{JJ'sp}_{lm} = (2)^{-1/2}(\mathbf{T}^{JJ's}_{lm} + p\mathbf{T}^{JJ's}_{lm}). \tag{7.1}$$

These bases are also orthonormal modulo factors of 4. Also note that (6.4) implies

$$*d\mathbf{X}^J{}^{J+1p}_{lm} = -n_{J\,J+1} \mathbf{X}^J{}^{J+1-p}_{lm}. \tag{7.2}$$

By stereographic projection from the south pole onto the hyperplane $y^4 = 1$, one may associate a point of R^3 (identified with that hyperplane in the natural way) with each point of $\mathbf{G}$ except the south pole. This is a conformal map. Cartesian or spherical coordinates on R^3 induce coordinates on $\mathbf{G} - \{-\hat{e}_4\}$ which might be called conformal Cartesian and spherical coordinates. The latter coordinates $\{r, \theta, \phi\}$ are related to spherical coordinates by $r = 2\tan(\chi/2)$. P and $\mathbf{J}_a$ coincide exactly with the usual parity and total angular momentum on R^3. The metric is explicitly

$$\mathbf{g} = d\chi^2 + \sin^2\chi(d\theta^2 + \sin^2\theta\, d\phi^2)$$

$$= (1 + r^2/4)^{-1}[dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)].$$

Eigenvectors of the vector DeRham Laplacian on $\mathbf{G}$ may be found by separation of variables using the vector spherical harmonics on R^3. This leads to radial equations involving Gegenbauer polynomials.[12] Requiring the eigenvectors to have definite parity produces the adjoint vector harmonics of definite parity. A similar statement holds for the tensor case.[19]

8. WAVE EQUATIONS ON POSITIVE CURVATURE FRIEDMANN SPACE-TIMES

Consider the manifold $\mathbf{M} = R \times \mathbf{G}$ with the following Lorentz metric:

$${}^4\mathbf{g} = a^2(-dt \otimes dt + \mathbf{g}), \tag{8.1}$$

where a is a function of the "time" t. Since we are not here concerned with a specific function a, it suffices to say that a is determined by the Einstein field equations for ${}^4\mathbf{g}$ with a perfect fluid source having the same symmetry group as ${}^4\mathbf{g}$. $(\mathbf{M}, {}^4\mathbf{g})$ is then called a Friedmann space–time.

The vector and tensor harmonics on $\mathbf{S}^3$ were originally developed by Lifshitz in terms of harmonic polynomials on R^4, an approach which does not easily lend itself to explicit calculation. His motivation was the treatment of the perturbations of the Friedmann spacetimes, an elegant and rather classic

application of the machinery of this paper and to which the reader is referred.[11] More recently Hu and Regge have introduced the idea of exploiting the group properties of $\mathbf{S}^3$ to study perturbations of spatially homogeneous space—times $(\mathbf{M}, {}^4\mathbf{g})$ more general than Friedmann.[20] Their technique essentially amounts to the use of left harmonic expansions; however, the special case of Friedmann was never treated explicitly.

It is also of interest to consider wave equations satisfied by test fields on a Friedmann space—time. A harmonic expansion of these fields then allows the wave equations to be reduced to uncoupled ordinary differential equations for the time dependent expansion coefficients. Let $e_0 = \partial/\partial t$ and $\omega^0 = dt$. Then $\{e_\alpha\}$ is a frame on $\mathbf{M}$ with dual frame $\{\omega^\alpha\}$ and structure functions $C^\alpha{}_{\beta\gamma} = \delta^\alpha_a C^a{}_{bc}\delta^b_\beta\delta^c_\gamma$. The spacetime metric is ${}^4g_{\alpha\beta}\omega^\alpha \otimes \omega^\beta$ with ${}^4g_{0\alpha} = a^2\delta_{0\alpha}$ and ${}^4g_{ab} = a^2 g_{ab}$. It is now a straightforward exercise to evaluate the connection and curvature components and the Laplacians using the formulas of the second section.

The scalar Laplacian is

$$ {}^4\Delta\Phi = a^{-4}(a^2\dot{\Phi})^{\cdot} + a^{-2}\Delta\Phi, \tag{8.2} $$

where $\dot{\Phi} = e_0\Phi$. An expansion of Φ in terms of scalar harmonics reduces the Klein—Gordon equation, for example, to the following ordinary differential equation for the expansion coefficients:

$$ \begin{aligned} &\Phi = \sum \Phi^i Q^i, \\ &0 = [({}^4\Delta + \mu^2)\Phi]^i \\ &\quad = a^{-4}(a^2\dot{\Phi}^i)^{\cdot} + [a^{-2}(n_i^2 - 1) + \mu^2]\Phi^i. \end{aligned} \tag{8.3} $$

The sourceless Maxwell equations may be solved explicitly on a Friedmann spacetime with no knowledge of the function a since they are invariant under conformal scaling of the metric.[21] Let ${}^4A = A_0\omega^0 + A$ with $A = A_a\omega^a$ be the vector potential. In Lorentz gauge:

$$ 0 = {}^4\delta {}^4A = a^{-4}(a^2 A_0)^{\cdot} + a^{-2}\delta A, \tag{8.4} $$

the vector potential satisfies the wave equation ${}^4\Delta_{\mathrm{DR}}{}^4A = 0$.[22] The Latin components of this equation are:

$$ 0 = ({}^4\Delta_{\mathrm{DR}}{}^4A)_a = a^{-2}(\ddot{A}_a + (\Delta A)_a). \tag{8.5} $$

By introducing the harmonic expansions:

$$ \begin{aligned} &A_0 = \sum A_0^i Q^i, \\ &A = \sum (A^{JM,J'M'}\mathbf{X}^{JM,J'M'} + A^{JM,J\pm1\,M'}\mathbf{X}^{JM,J\pm1\,M'}), \end{aligned} \tag{8.6} $$

one obtains the following equations for the coefficients:

$$ \begin{aligned} &0 = a^{-2}(a^2 A_0^i)^{\cdot} + i(n_i^2 - 1)A^i \\ &\ddot{A}^i + \Delta^i A^i = 0, \end{aligned} \tag{8.7} $$

where $\Delta^i = n_i^2 - 1$ if $J' = J$ and $\Delta^i = n_i^2$ if $J' = J \pm 1$. Let $\omega^i = (\Delta^i)^{1/2}$. The latter equation has the solution $A^i = \mathcal{A}^i e^{i\omega^i t}$, where $\mathcal{A}^i$ is a constant.

A_0 (which incidentally satisfies the scalar wave equation) and the longitudinal $(J' = J)$ part of A are affected by gauge transformations:

$$ A_0 \to A_0 + \dot{\lambda}, \quad A \to A + d\lambda . $$

To preserve Lorentz gauge λ must satisfy ${}^4\Delta\lambda = 0$. In fact, the gauge dependent parts of ${}^4A$ are removable by a Lorentz gauge preserving transformation with $\dot{\lambda} = -A_0$, exactly as in flat spacetime. In the new gauge A_0 and $A^{JM,JM'}$ vanish. [$A_0 = 0$ forces $\delta A = 0$ by (8.4).]

Assuming this gauge, the solution may also be expressed in terms of the adjoint harmonics of definite parity:

$$ A = \sum A^{JJ+1p}_{lm}\mathbf{X}^{JJ+1p}_{lm}, \quad A^{JJ+1p}_{lm} = \mathcal{A}^{JJ+1p}_{lm}\, e^{in_{JJ+1}t}. $$

The covariant components of the electric field and the contravariant components of the magnetic field density are given by

$$ E_a = F_{a0} = -\dot{A}_a, \quad \beta^a = \tfrac{1}{2}\epsilon^{abc}(dA)_{bc} = 4(*dA)_a, $$

where $*$ is the star operation of $\mathbf{g}$ so that (7.2) may be used. The corresponding expansion coefficients

$$ E = \sum E^{JJ+1}_{lm}\mathbf{X}^{JJ+1}_{lm}, \quad \beta = \sum \beta^{JJ+1}_{lm}\hat{\mathbf{X}}^{JJ+1}_{lm}, $$

are easily evaluated:

$$ E^{JJ+1p}_{lm} = -in_{JJ+1}A^{JJ+1p}_{lm}, \quad \beta^{JJ+1p}_{lm} = -n_{JJ+1}A^{JJ+1-p}_{lm}. $$

The parities of the electric field and magnetic field density of the vector potential $A^{JJ+1p}_{lm}\mathbf{X}^{JJ+1p}_{lm}$ are $p(-1)^{l+1}$ and $p(-1)^l$, respectively. The $p = 1$ and $p = -1$ solutions therefore correspond to the electric and magnetic l-pole radiation of Mashoon.[12]

9. SPINOR HARMONICS

The orthonormal frame $\{\epsilon_\alpha\} = \{a^{-1}e_0,\ 2a^{-1}e_a\}$ may be used to introduce ordinary or Dirac spinor algebras over M. Although spinor fields cannot be decomposed into collections of induced spinor fields on the natural slicing of $M = R \times G$ as can be done with tensor fields, it does make sense to consider a restricted spin connection on G in order to reduce spinor equations on M to ordinary differential equations. We briefly sketch how this may be accomplished.

Let $\{E_q\} = \{E_{1/2}, E_{-1/2}\}$ be the natural basis of C^2 considered as a left invariant spin frame on G associated with the orthonormal frame $\{2e_a\}$.[23] One may extend the action of angular momenta to E_q and hence to any spinor field $\psi = \psi^q E_q$ by

$$ \begin{aligned} &\tilde{L}_a E_q = 0 = L^{\mathrm{orb}}_a E_q , \\ &L_a E_q = S_a E_q = \tfrac{1}{2}\sigma_a E_q. \end{aligned} \tag{9.1} $$

By (3.6) the natural extensions of the dragging actions are then:

$$ L_u E_q = E_q, \quad R_u^{-1}E_q = AD_u E_q = \mathrm{u}E_q, \tag{9.2} $$

where $\mathrm{u} \in \mathrm{Su}(2)$ is the actual matrix of $u \in G$.

$\{E_q\}$ is a standard spin one-half basis satisfying the $s=\frac{1}{2}$ version of (5.3) with $J_3 E_q = qE_q$. This explains our choice of the indices $\{\frac{1}{2}, -\frac{1}{2}\}$ rather than the more conventional indices $\{1,2\}$. We may therefore introduce the spinor harmonics by right angular momentum coupling to the scalar harmonics:

$$\mathcal{Y}^{JM,J'M'} = C_{J(1/2)}(J'M', Nq) Q^{JMN} E_q, \quad J' = J \pm \tfrac{1}{2} \tag{9.3}$$

$$\mathcal{Y}^{JJ'}_{lm} = C_{JJ'}(lm, MM') \mathcal{Y}^{JM,J'M'}$$

$$\mathcal{Y}^{JJ+p/2}_{lm} = 2^{-1/2}(\mathcal{Y}^{JJ+1/2}_{lm} + p\mathcal{Y}^{J+1/2\,J}_{lm})$$

By considerations similar to (6.3), (6.4), and (7.2) one finds

$$2(\tilde{L}^2 - L^2)\mathcal{Y}^{JJ+1/2\,p}_{lm} = -n_{JJ+(1/2)}\mathcal{Y}^{JJ+1/2-p}_{lm}. \tag{9.4}$$

The covariant derivative and ordinary Laplacian may be extended to spinor fields $\psi = \psi^q E_q$ by (5.16), but define a DeRham Laplacian by[5]

$$\Delta_{DR}\psi = (\Delta + \tfrac{1}{4}R)\psi = (\Delta + 2S)\psi = [2(\tilde{L}^2 + L^2) + S^2]\psi, \tag{9.5}$$

where we have used $R=6$ and $S^2\psi = \frac{3}{4}\psi$. In fact, Δ_{DR} is the square of $\sigma_a P_a$, where $P_a = -i\nabla_{2e_a}$:

$$\begin{aligned}
&\sigma_a P_a \psi = -4iS_a(L^{orb}_a + \tfrac{1}{2}S_a)\psi = 2(\tilde{L}^2 - L^2)\psi,\\
&\sigma_a P_a \mathcal{Y}^{JJ+1/2\,p}_{lm} = -n_{JJ+(1/2)}\mathcal{Y}^{JJ+1/2-p}_{lm},\\
&\Delta_{DR}\mathcal{Y}^{JJ+1/2\,p}_{lm} = (\sigma_a P_a)^2 \mathcal{Y}^{JJ+1/2\,p}_{lm}\\
&\qquad = n^2_{JJ+(1/2)}\mathcal{Y}^{JJ+1/2\,p}_{lm}.
\end{aligned} \tag{9.6}$$

Taub has essentially shown[24] that the Dirac equation:

$$(i\gamma^\alpha \nabla_{e\alpha} + m)\Psi = 0$$

may be reduced to the form:

$$(i\partial/\partial t - am\beta)\Phi = \alpha_a P_a \Phi, \tag{9.7}$$

where γ^α, $\beta = \rho_3$, and $\alpha_a = \rho_1\sigma_a$ are the standard Dirac matrices in the notation of Messiah[10] and

$$\Phi = \begin{bmatrix}\phi^1\\ \phi^2\end{bmatrix} = a^{3/2}\Psi$$

is a pair of spinor fields. ϕ^1 and ϕ^2 may each be expanded in terms of the spinor harmonics $\mathcal{Y}^{JJ+1/2\,p}_{lm}$ with time dependent expansion coefficients $\phi^{1JJ+1/2\,p}_{lm}$ and $\phi^{2JJ+1/2\,p}_{lm}$, respectively. Suppressing all indices except p and inserting the expansions into (9.7), one obtains the ordinary differential equations:

$$\begin{aligned}
&(id/dt - am)\phi^{1+} = -n\phi^{2-},\\
&(id/dt - am)\phi^{1-} = -n\phi^{2+},\\
&(id/dt + am)\phi^{2-} = -n\phi^{1+},\\
&(id/dt + am)\phi^{2+} = -n\phi^{1-}.
\end{aligned} \tag{9.8}$$

For example, in an Einstein spacetime in which a is a constant, one obtains solutions with exponential time dependence and frequencies $\omega = \pm (n^2 + a^2m^2)^{1/2}$.

ACKNOWLEDGMENTS

The author wishes to thank A. H. Taub for reading the manuscript and John Uyemura for typing it.

[1] Latin indices $a, b, c, d \ldots$ assume the values 1, 2, 3 while Greek indices assume the values 1, 2, 3, 4 when referrring to R^4 and 0, 1, 2, 3 when referring to space—time.

[2] B. L. Beers and R. S. Millman, J. Math. Phys. **16**, 11 (1975).

[3] We observe the conventions of Ref. 4 except for the connection components. If $\mathbf{T}^a{}_{bc} = \omega^a(\mathbf{T}(e_b, e_c))$ are the components of the torsion tensor $\mathbf{T}$ of a connection ∇ in the frame $\{e_a\}$ with dual frame $\{\omega^a\}$ and structure functions $C^a{}_{bc} = \omega^a([e_b, e_c])$ and if $\mathbf{g} = g_{ab}\,\omega^a \otimes \omega^b$ is a covariant constant metric $\nabla\mathbf{g} = 0$, then the components of the connection in this frame may be obtained from the formula:

$$2\Gamma_{abc} = g_{\{ab,c\}} + C_{\{abc\}} + \mathbf{T}_{\{abc\}},$$

$$\mathbf{S}_{\{abc\}} = \mathbf{S}_{abc} - \mathbf{S}_{bca} + \mathbf{S}_{cab}, \quad g_{ab,c} = e_c\, g_{ab}.$$

The usual index raising and lowering conventions are understood. The components of the Riemann tensor of ∇ in this frame are

$$R^a{}_{bcd} = e_c\Gamma^a{}_{db} - e_d\Gamma^a{}_{cb} - \Gamma^a{}_{eb}\,C^e{}_{cd} + \Gamma^a{}_{ce}\Gamma^e{}_{db} - \Gamma^a{}_{de}\Gamma^e{}_{cb}.$$

The metric connection of $\mathbf{g}$ has $\mathbf{T} = 0$; the components of its Ricci tensor are defined by $\mathbf{R}_{bd} = \mathbf{R}^a{}_{bad}$.

[4] C. W. Misner, K. S. Thorne, J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).

[5] A. Lichnerowicz, "Propagateurs, commutateurs, et anticommutateurs en relativité générale, in *Relativity, Groups, and Topology*, edited by C. Dewitt and B. S. Dewitt (Gordon and Breach, New York, 1964).

[6] The differential $dh(u)$: $\mathbf{TG}_u \to \mathbf{TG}_{h(u)}$ and its transpose $dh(u)^*$: $\mathbf{TG}^*_{h(u)} \to \mathbf{TG}^*_u$ of h at u are defined by

$$[dh(u)X(u)]\,\Phi = X(u)\Phi \circ h$$

$$[dh(u)^*\sigma(h(u))]\;(X(u)) = \sigma(h(u))\;(dh(u)X(u))$$

[7] Recall the identity $[\pounds_X, \pounds_Y] = \pounds_{[X,Y]}$.

[8] N. J. Vilenkin, *Special Functions and the Theory of Group Representations* (American Mathematical Society, Providence, R. I., 1968).

[9] The notation $Cj_1j_2(j_3m_3, m_1m_2)$ of Blatt and Weisskopf is more convenient here that $\langle j_1j_2m_1m_2 \mid j_3m_3\rangle$ as used by Messiah.[10]

[10] A. Messiah, *Quantum Mechanics* (Wiley, New York, 1958) Vol. II.

[11] E. M. Lifshitz and I. M. Khalatnikov, Adv. Phys. **12**, 185 (1963).

[12] B. Mashoon, Phys. Rev. D **8**, 4297 (1973).

[13] Actually these bases will be orthonormal only up to a factor of 4^{p-q}.

[14] To do this one uses the adjoint invariance of the structure constant tensor: $R^a{}_d C^d{}_{fg} R^{-1}{}^s{}_b R^{-1}{}^g{}_c = C^a{}_{bc}$.

[15] A. E. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton University, Princeton, N.J., 1957).

[16] F. W. Warner, *Foundations of Differentiable Manifolds and Lie Groups* (Scott, Foresman and Co., Glenview, Ill. 1971).

[17] L. I. Schiff, *Quantum Mechanics* (McGraw-Hill, New York, 1968).

[18] A. E. Fisher and J. E. Marsden, "Topics in the Dynamics of General Relativity," Enrico Fermi International School of Physics, Varenna, Italy, 1976.

[19] Let $\mathbf{g} \to \rho^2\mathbf{g}$, where $\rho^2 = 1$. Analytic continuation of ρ and χ to pure imaginary values ($\rho \to i, \chi \to i\chi$) yields the Riemannian metric of a geometry of constant negative curvature on $\mathbf{S}^3 - \{-\hat{e}_4\} \sim R^3$. The adjoint harmonics of definite parity may also be analytically continued to harmonics of the new geometry.[11]

[20] B. L. Hu, J. Math. Phys. **15**, 1748 (1974).

[21] L. Infeld and A. E. Schild, Phys. Rev. **70**, 410 (1946).

[22] By introducing a vector potential $F = {}^4d\,{}^4A$, Maxwell's equations ${}^4dF = {}^4\delta F = 0$ are reduced to $0 = {}^4\delta\,{}^4d\,{}^4A = {}^4\Delta_{DR}\,{}^4A - {}^4d\,{}^4\delta\,{}^4A$.

[23] J. S. Dowker, Ann. Phys. **71**, 577 (1972).

[24] A. H. Taub, Phys. Rev. **51**, 512 (1937).

Raising and lowering operators for the orthogonal groups

Adam M. Bincer

Physics Department, University of Wisconsin, Madison, Wisconsin 53706
(Received 28 September 1977)

A formalism for obtaining shift operators for all classical groups in the form of tensor operators is described in application to the orthogonal groups. This formalism is also applicable to the irreps in the discrete series of the noncompact versions of the classical groups.

INTRODUCTION

This is the second in a series of papers dealing with the shift operators for the classical groups. In the first paper[1] of this series we stated without proof the results for the lowering operators of the unitary groups, $U(n)$. In the present paper we develop the formalism for the orthogonal groups, $O(n)$, and give explicit results for the raising, lowering, and weight operators (collectively referred to as shift operators) for $O(n)$. In the paper[2] immediately following we obtain the normalization coefficients for the shift operators of $O(n)$. The corresponding results for the symplectic groups, $Sp(n)$, will be given in future publications.

The importance of the concept of the lowering operator in the study of representations of groups is clear from its definition: The state of any weight in an irreducible representation is obtained from the state of highest weight by application of the appropriate lowering operator. The concept of shift operators for the $U(n)$ groups was first explicitly introduced by Nagel and Moshinsky[3] and by Zhelobenko.[4] For the $O(n)$ groups the shift operators were given by Wong.[5] To the best of our knowledge shift operators for the $Sp(n)$ groups do not appear in the literature.

One of the problems that arise in work in this field is that of devising a suitable notation. Thus, whereas the results of Nagel and Moshinsky are fairly transparent, those of Wong are virtually unusable because of their complexity. The method for obtaining $O(n)$ shift operators to be described in the present work is applicable, with minor modifications, to all classical groups. Even in the case of $U(n)$ our results[1] are simpler than the elegant formulas of Nagel and Moshinsky[3]; in the case of $O(n)$ the simplification is greater yet.

In an attempt to make this paper reasonably self-contained we discuss in Sec. 1 the generators and weights of $O(2\nu+1)$ and in Sec. 2 we describe the canonical chain $O(2\nu+1)\supset O(2\nu)\supset\cdots O(3)\supset O(2)$ and obtain a (modified) Gel'fand–Zetlin pattern[6] completely characterizing an $O(n)$ basis vector. The slight modification of the standard Gel'fand–Zetlin pattern is needed to obtain a more transparent scheme for classifying generators and tensors as raising, weight, and lowering generators and tensors. In Sec. 3 we introduce $O(n)$ tensor operators following the ideas of Louck and Biedenharn,[7] Okubo,[8] and Nwachuku and Rashid.[9] We obtain the shift operators for $O(2k+1)$ in Sec. 4, and for $O(2k)$ in Sec. 5. Section 6 contains some concluding remarks. It is perhaps worth remarking that our shift operators are tensor operators whereas those of Wong[5] are not. In fact the corresponding operators are not equal except when applied to the appropriate state. This is a reflection of the nonuniqueness of the shift operators already noted by Nagel and Moshinsky.[3]

1. GENERATORS AND WEIGHTS OF $O(2\nu+1)$

We denote the generators of $O(2\nu+1)$ by C^a_b with the indices ranging from $-\nu$ to ν. In the Racah basis[10] their commutation relations are

$$[C^a_b, C^c_d] = \delta^c_b C^a_d - \delta^a_d C^c_b + \delta^{\bar{b}}_d C^c_{\bar{a}} - \delta^c_{\bar{a}} C^{\bar{b}}_d, \tag{1.1}$$

where

$$\bar{a} \equiv -a. \tag{1.2}$$

These C's obey

$$C^a_b = -C^{\bar{b}}_{\bar{a}} \tag{1.3}$$

and therefore the number of independent generators (order of the group) is $\nu(2\nu+1)$. Moreover, in unitary representations we demand

$$C^{a\dagger}_b = C^b_a. \tag{1.4}$$

It is clear that the ν generators

$$C^a_a, \quad 1 \leq a \leq \nu, \tag{1.5}$$

may be taken simultaneously diagonal—they form a so-called Cartan subalgebra; we see that the rank of $O(2\nu+1)$ is ν. Let $|w\rangle$ denote a simultaneous eigenstate of the C^a_a:

$$C^a_a|w\rangle = w_a|w\rangle, \quad \bar{\nu} \leq a \leq \nu, \tag{1.6}$$

where

$$w \equiv (w_\nu, w_{\nu-1}, \ldots, w_1, w_0, w_{\bar{1}}, \ldots, w_{\bar{\nu}}). \tag{1.7}$$

w is called the weight of the state $|w\rangle$ and the w_a, $\bar{\nu} \leq a \leq \nu$, are the components of the weight. It follows from Eq. (1.3) that

$$w_{\bar{a}} = -w_a \tag{1.8}$$

and therefore the last $\nu+1$ entries in Eq. (1.7) are redundant and will be usually omitted. Two weights will be called equal if all components are equal; for unequal weights we introduce an ordering as follows:

$$w > w' \quad \text{if} \quad w_a > w'_a, \tag{1.9}$$

where

$$\begin{aligned} w &= (w_\nu, w_{\nu-1}, \ldots, w_{a+1}, w_a, \ldots, w_1), \\ w' &= (w_\nu, w_{\nu-1}, \ldots, w_{a+1}, w'_a, \ldots, w'_1). \end{aligned} \tag{1.10}$$

 0022-2488/78/1905-1173$1.00

It follows from Eq. (1.1) that

$$C_c^c\{C_b^a|w\rangle\} = (w_c + \delta_c^a - \delta_b^c + \delta_c^{\bar{b}} - \delta_{\bar{a}}^c)C_b^a|w\rangle, \tag{1.11}$$

so that we may write

$$C_b^a|w\rangle \propto |w'\rangle, \tag{1.12}$$

$$w'_c = w_c + \delta_c^a - \delta_b^c + \delta_c^{\bar{b}} - \delta_{\bar{a}}^c. \tag{1.13}$$

Consequently

$$w' \gtreqless w \quad \text{if} \quad a \gtreqless b \tag{1.14}$$

and therefore we may classify generators as raising, weight, and lowering generators according to the following scheme:

$$C_b^a \text{ is a } \begin{matrix}\text{raising} \\ \text{weight} \\ \text{lowering}\end{matrix} \text{ generator if } \begin{matrix}a > b \\ a = b. \\ a < b\end{matrix} \tag{1.15}$$

An irreducible representation (irrep) of $O(2\nu+1)$ may be specified by the weight of the state of highest weight. Let $|h\rangle$ be the state of highest weight, i.e.,

$$C_b^a|h\rangle = \delta_b^a h_a|h\rangle \quad \text{for} \quad \bar{\nu} \leq b \leq a \leq \nu, \tag{1.16}$$

where

$$h \equiv (h_\nu, h_{\nu-1}, \ldots, h_1) \tag{1.17}$$

is the weight. It follows from our ordering that

$$h_\nu \geq h_{\nu-1} \geq \cdots \geq h_1. \tag{1.18}$$

The labeling of states within the irrep specified by h makes use of the canonical chain of subgroups

$$O(2\nu+1) \supset O(2\nu) \supset O(2\nu-1) \supset \cdots O(3) \supset O(2), \tag{1.19}$$

and we discuss this next.

2. THE $O(2\nu+1) \supset O(2\nu) \supset \cdots \supset O(3) \supset O(2)$ canonical chain

This chain may be formed in a number of different ways. We find it convenient to proceed as follows: We take for the $O(2\nu)$ subgroup of $O(2\nu+1)$ the group generated by the C_b^a of Sec. 1 with the value 0 *excluded* from the range of the indices. With this one modification all the concepts discussed in Sec. 1 for $O(2\nu+1)$ carry over to $O(2\nu)$. Thus the order of $O(2\nu)$ is $\nu(2\nu-1)$, the rank is ν, and an irrep may be specified by the weight of the state of highest weight.

Let $|q\rangle$ be the state of highest weight, i.e.,

$$C_b^a|q\rangle = \delta_b^a q_a|q\rangle \quad \text{for} \quad \bar{\nu} \leq b \leq a \leq \nu \quad (0 \text{ excluded}), \tag{2.1}$$

where

$$q \equiv (q_\nu, q_{\nu-1}, \ldots, q_1), \tag{2.2}$$

and

$$q_\nu \geq q_{\nu-1} \geq \cdots \geq q_1. \tag{2.3}$$

Now consider an irrep of $O(2\nu+1)$ specified by h as in Sec. 1. Such an irrep is necessarily a representation (possibly reducible) of any subgroup of $O(2\nu+1)$, in particular of $O(2\nu)$. Therefore, for appropriate values of h and q, we must be able to form from the states in the irrep of $O(2\nu+1)$ specified by h, the state $|q\rangle$ defined by Eq. (2.1), which we shall now denote as

$$\left|\begin{matrix}h \\ q\end{matrix}\right\rangle \equiv \left|\begin{matrix}h_\nu, h_{\nu-1}, \ldots, h_1 \\ q_\nu, q_{\nu-1}, \ldots, q_1\end{matrix}\right\rangle. \tag{2.4}$$

The labels in the top row, h, specify the irrep of $O(2\nu+1)$. The labels in the bottom row, q, specify the irrep of $O(2\nu)$, contained in this particular irrep of $O(2\nu+1)$, and at the same time q gives the weight of this state of $O(2\nu+1)$. Since q is the weight and h is the highest weight for this irrep of $O(2\nu+1)$, we must have

$$h \geq q, \tag{2.5}$$

hence in particular

$$h_\nu \geq q_\nu. \tag{2.6}$$

Moreover, the state $|h\rangle$ defined by Eq. (1.16) is obviously given in the notation of Eq. (2.4) by

$$|h\rangle = \left|\begin{matrix}h \\ h\end{matrix}\right\rangle. \tag{2.7}$$

Next we take for the $O(2\nu-1)$ subgroup of $O(2\nu)$ the group generated by the C_b^a of $O(2\nu)$ with the values 1 and -1 excluded from the range of the indices, together with additional generators $\tilde{C}_a^0$, $\tilde{C}_0^a$ defined by

$$\tilde{C}_a^0 = (C_a^1 + C_a^{\bar{1}})/\sqrt{2}, \quad \tilde{C}_0^a = (C_1^a + C_1^{\bar{a}})/\sqrt{2}, \quad 2 \leq |a| \leq \nu. \tag{2.8}$$

It follows from Eq. (1.1) that the commutation relations of these $(\nu-1)(2\nu-1)$ generators are precisely of the same form as Eq. (1.1) with the values 1 and -1 omitted from the range of indices and with C_a^0, C_0^a of Eq. (1.1) replaced by $\tilde{C}_a^0$, $\tilde{C}_0^a$. This proves that this set of generators generates $O(2\nu-1)$. The tilde in $\tilde{C}_a^0$, $\tilde{C}_0^a$ will be omitted whenever there is no risk of confusion.

It is seen that the rank of $O(2\nu-1)$ is $\nu-1$ with the weight generators the same as in the original $O(2\nu+1)$, except that C_1^1 is missing. As before we have that an irrep of $O(2\nu-1)$ may be specified by the weight of the state of highest weight. Let $|r\rangle$ be the state of highest weight, i.e.,

$$C_b^a|r\rangle = \delta_b^a r_a|r\rangle, \quad \bar{\nu} \leq b \leq a \leq \nu \quad (\pm 1 \text{ omitted}), \tag{2.9}$$

where

$$r \equiv (r_\nu, r_{\nu-1}, \ldots, r_2) \tag{2.10}$$

and

$$r_\nu \geq r_{\nu-1} \geq \cdots \geq r_2. \tag{2.11}$$

By repeating the argument used for $O(2\nu+1) \supset O(2\nu)$ but applied to $O(2\nu) \supset O(2\nu-1)$ we have that for appropriate values of q, r the state $|r\rangle$, which we now denote as

$$\left|\begin{matrix}q \\ r\end{matrix}\right\rangle \equiv \left|\begin{matrix}q_\nu, q_{\nu-1}, \ldots, q_2, q_1 \\ r_\nu, r_{\nu-1}, \ldots, r_2\end{matrix}\right\rangle, \tag{2.12}$$

is a state in the irrep of $O(2\nu)$ specified by q, with the labels r specifying the irrep of $O(2\nu-1)$ contained in this irrep of $O(2\nu)$. We note that this state is no longer necessarily an eigenstate of C_1^1, however we still have corresponding to Eq. (2.6) that

$$q_\nu \geq r_\nu. \tag{2.13}$$

Moreover, the state $|q\rangle$ defined by Eq. (2.1) is obviously given in the notation of Eq. (2.12) by

$$\left|\begin{matrix} q \\ q \end{matrix}\right\rangle = \left|\begin{matrix} q_\nu, \ldots, q_2, q_1 \\ q_\nu, \ldots, q_2 \end{matrix}\right\rangle \tag{2.14}$$

and it is an eigenstate of C_1^1 to eigenvalue q_1.

By combining the arguments that led to Eqs. (2.4) and (2.12) we also have that $|\begin{matrix} q \\ r \end{matrix}\rangle$, which we now denote as

$$\left|\begin{matrix} h \\ q \\ r \end{matrix}\right\rangle \equiv \left|\begin{matrix} h_\nu, h_{\nu-1}, \ldots, h_2, h_1 \\ q_\nu, q_{\nu-1}, \ldots, q_2, q_1 \\ r_\nu, r_{\nu-1}, \ldots, r_2 \end{matrix}\right\rangle, \tag{2.15}$$

is a state in the irrep of $O(2\nu+1)$ specified by h. Also, the state $|h\rangle$ defined by Eq. (1.16) is given in the notation of Eq. (2.15) by

$$\left|\begin{matrix} h \\ h \\ h \end{matrix}\right\rangle \equiv \left|\begin{matrix} h_\nu, \ldots, h_2, h_1 \\ h_\nu, \ldots, h_2, h_1 \\ h_\nu, \ldots, h_2 \end{matrix}\right\rangle. \tag{2.16}$$

The pattern for the chain

$$O(2\nu+1) \supset O(2\nu) \supset \cdots \supset O(2k) \supset O(2k-1) \supset \cdots \supset O(2)$$

should now be clear. Thus the $O(2k)$ subgroup is generated by the C_b^a of $O(2\nu+1)$ with the values $\bar{\nu}+k$, $\bar{\nu}+k+1, \ldots, \bar{1}, 0, 1, \ldots, \nu-k-1, \nu-k$ *excluded* from the range of the indices; the $O(2k-1)$ subgroup is generated by the C_b^a of the above $O(2k)$ subgroup with the values $\bar{\nu}+k-1$ and $\nu-k+1$ excluded, together with the additional generators $\tilde{C}_a^0$, $\tilde{C}_0^a$ defined by

$$\begin{aligned} &\tilde{C}_a^0 = (C_a^{\bar{\nu}+k-1} + C_a^{\nu-k+1})/\sqrt{2}, \\ &\tilde{C}_0^a = (C_{\bar{\nu}+k-1}^a + C_{\nu-k+1}^a)/\sqrt{2}, \\ &\qquad \text{for} \quad \nu-k+2 \leq |a| \leq \nu. \end{aligned} \tag{2.17}$$

Changing our notation so that

$$h_a \equiv m_a^{2\nu+1}, \quad q_a \equiv m_a^{2\nu}, \quad r_a \equiv m_a^{2\nu-1}, \tag{2.18}$$

it is clear that according to the indicated chain a labeling of an $O(2\nu+1)$ state vector is given as follows,

$$\left|\begin{matrix} m^{2\nu+1} \\ m^{2\nu} \\ m^{2\nu-1} \\ \cdot \\ \cdot \\ \cdot \\ m^{2k+1} \\ m^{2k} \\ m^{2k-1} \\ \cdot \\ \cdot \\ \cdot \\ m^3 \\ m^2 \end{matrix}\right\rangle \tag{2.19}$$

where

$$m^{2k+1} = (m_\nu^{2k+1}, m_{\nu-1}^{2k+1}, \ldots, m_{\nu-k+1}^{2k+1}), \quad 1 \leq k \leq \nu, \tag{2.20}$$

$$m^{2k} = (m_\nu^{2k}, m_{\nu-1}^{2k}, \ldots, m_{\nu-k+1}^{2k}), \quad 1 \leq k \leq \nu, \tag{2.21}$$

with the labels subject to the following inequalities,

$$m_\nu^i \geq m_\nu^{i-1} \quad \text{and} \quad m_j^i \geq m_{j-1}^i. \tag{2.22}$$

In fact, in a unitary representation the labels m have additional restrictions placed upon them—this will become clear when we construct the normalized shift operators. The pattern (2.19), somewhat modified to account for different conventions in the ordering and labeling of weights, is known as the Gel'fand–Zetlin[9] pattern for $O(2\nu+1)$. The corresponding pattern for $O(2\nu)$ is obtained by simply omitting the topmost row in (2.19). It is seen that the number of labels in the pattern is $\nu(\nu+1)$ for $O(2\nu+1)$ and ν^2 for $O(2\nu)$, which in each case is half the sum of order and rank of the group. Consequently we have precisely the right number of labels for a complete labeling of an $O(n)$ state vector, which justifies the use of the term "canonical," as defined by Baird and Biedenharn,[11] for this chain of subgroups.

We shall denote the pattern (2.19) by

$$\left|\begin{matrix} m^{2\nu+1} \\ m^{2\nu} \\ \cdot \\ \cdot \\ \cdot \\ m^n \end{matrix}\right\rangle \tag{2.23}$$

in the case that

$$m_j^i = m_j^n \tag{2.24}$$

for all j, and all $i \leq n$. Obviously it describes the state of highest weight of $O(n)$, that weight being m^n.

3. O(n) TENSOR OPERATORS

A set of operators T_d^c, with the indices in the range appropriate to $O(n)$, will be called an $O(n)$ two-tensor operator if it satisfies

$$[C_b^a, T_d^c] = \delta_b^c T_d^a - \delta_d^a T_b^c + \delta_d^{\bar{b}} T_{\bar{a}}^c - \delta_{\bar{a}}^c T_d^{\bar{b}}. \tag{3.1}$$

It is obvious that a subset of the T_d^c with the indices restricted to the range appropriate to a subgroup $O(n')$ of $O(n)$ is an $O(n')$ two-tensor operator.

More precisely, T_d^c is an $O(2\nu+1)$ two-tensor operator if the range of the indices is

$$0 \leq |c|, |d| \leq \nu; \tag{3.2}$$

it is an $O(2k)$ two-tensor operator if the range of the indices is

$$\nu-k+1 \leq |c|, |d| \leq \nu; \tag{3.3}$$

it is an $O(2k-1)$ two-tensor operator if the range of the indices is

$$\begin{aligned} &\nu-k+2 \leq |c|, |d| \leq \nu, \\ &\text{and} \\ &c, d = 0, \end{aligned} \tag{3.4}$$

where

$$\begin{aligned} &T_a^0 = (T_a^{\bar{\nu}+k-1} + T_a^{\nu-k+1})/\sqrt{2}, \\ &T_0^a = (T_{\bar{\nu}+k-1}^a + T_{\nu-k+1}^a)/\sqrt{2}, \\ &\qquad \text{for} \quad \nu-k+2 \leq |a| \leq \nu, \end{aligned} \tag{3.5}$$

and

$$T^0_0 = \tfrac{1}{2}(T^{\nu-k+1}_{\nu-k+1} + T^{\bar{\nu}+k-1}_{\nu+k-1} + T^{\nu-k+1}_{\nu+k-1} + T^{\bar{\nu}+k-1}_{\nu-k+1}). \tag{3.6}$$

In the above, the subgroups $O(2k)$, $O(2k-1)$ are those from the canonical chain defined in Sec. 2.

It follows from Eq. (3.1) that

$$T^a_b \text{ is a } \begin{matrix}\text{raising}\\ \text{weight}\\ \text{lowering}\end{matrix} \text{ two-tensor operator if } \begin{matrix}a>b\\ a=b.\\ a<b\end{matrix} \tag{3.7}$$

The generators are an example of a two-tensor operator but there are others. In particular if T^a_b and W^a_b are $O(n)$ two-tensor operators, then so are $(TW)^a_b$ and $(T\circ W)^a_b$, where

$$(TW)^a_b \equiv \sum_c T^a_c W^c_b, \tag{3.8}$$

$$(T\circ W)^a_b \equiv \sum_c T^c_b W^a_c, \tag{3.9}$$

with c ranging over the range appropriate to $O(n)$.

A set of operators V_d will be called an $O(2\nu+1)$ one-tensor operator if it satisfies

$$[C^a_b, V_d] = -\delta^a_d V_b + \delta^{\bar{b}}_d V_{\bar{a}} \quad \text{for } \bar{\nu} \le a, b, d \le \nu. \tag{3.10}$$

Again, the subset of V_d obtained by restricting the indices to the range

$$\nu - k + 1 \le |d| \le \nu \tag{3.11}$$

yields an $O(2k)$ one-tensor operator; the subset obtained by restricting the indices to

$$\nu - k + 2 \le |d| \le \nu \quad \text{and} \quad d = 0 \tag{3.12}$$

yields an $O(2k-1)$ one-tensor operator. Here

$$V_0 \equiv (V_{\nu-k+1} + V_{\bar{\nu}+k-1})/\sqrt{2}. \tag{3.13}$$

It follows that

$$V_d \text{ is a } \begin{matrix}\text{raising}\\ \text{weight}\\ \text{lowering}\end{matrix} \text{ one-tensor operator if } \begin{matrix}d<0\\ d=0.\\ d>0\end{matrix} \tag{3.14}$$

More precisely, the $O(2k)$ one-tensor operator V_d when applied to a state of given $O(2k)$ weight produces a state with the weight component $w_{\bar{d}}$ raised by one unit if $\bar{\nu} \le d \le \bar{\nu} + k - 1$, or it produces a state with the component w_d lowered by one unit if $\nu - k + 1 \le d \le \nu$. Similarly the $O(2k-1)$ one-tensor operator V_d, when applied to a state of given $O(2k-1)$ weight, produces a state with the component $w_{|d|}$ raised by one unit, left unchanged, or lowered by one unit according as $\bar{\nu} \le d \le \bar{\nu} + k - 2$, $d = 0$, $\nu - k + 2 \le d \le \nu$.

An example of an $O(n)$ one-tensor operator is the set of generators C^p_d, where d ranges over the values appropriate to $O(n)$ and p is fixed and outside the range appropriate to $O(n)$. There are, however, other one-tensor operators. In particular, by contraction of an $O(n)$ one-tensor and an $O(n)$ two-tensor we may form another $O(n)$ one-tensor:

$$(VT)_a \equiv \sum_c V_c T^c_a, \tag{3.15}$$

$$(T\circ V)_a \equiv \sum_c T^c_a V_c, \tag{3.16}$$

with all indices in the range appropriate to $O(n)$.

It is clear that one can define in a similar fashion one-tensor operators with a superscript instead of a subscript,

$$[C^a_b, V^c] = \delta^c_b V^a - \delta^a_{\bar{b}} V^{\bar{b}}; \tag{3.17}$$

however for the application to the problem of shift operators this turns out to be unnecessary.

4. O(2k + 1) SHIFT OPERATORS

The $O(2k+1)$ shift operators will be denoted by ${}^{2k+1}S_\mu$ with the index μ in the $O(2k)$ range, i.e.,

$$\nu - k + 1 \le |\mu| \le \nu. \tag{4.1}$$

Throughout this section all indices are in the $O(2k)$ range. The shift operators are defined by

$${}^{2k+1}S_\mu \left|\begin{matrix} m^{2\nu+1} \\ \cdot \\ \cdot \\ \cdot \\ m^{2k+1} \\ m^{2k}\end{matrix}\right\rangle \propto \left|\begin{matrix} m^{2\nu+1} \\ \cdot \\ \cdot \\ \cdot \\ m^{2k+1} \\ \tilde{m}^{2k}\end{matrix}\right\rangle, \tag{4.2}$$

where

$$m^{2k} \equiv (m^{2k}_\nu, m^{2k}_{\nu-1}, \ldots, m^{2k}_{|\mu|}, \ldots, m^{2k}_{\nu-k+1}) \tag{4.3}$$

and

$$\tilde{m}^{2k} \equiv (m^{2k}_\nu, m^{2k}_{\nu-1}, \ldots, m^{2k}_{|\mu|} - \mu/|\mu|, \ldots, m^{2k}_{\nu-k+1}), \tag{4.4}$$

so that we have a lowering operator for $\mu > 0$ and a raising operator for $\mu < 0$.

Since the patterns on both sides of (4.2) describe $O(2k)$ states of highest weight we have [compare Eq. (2.1)]

$$C^a_b \left|\begin{matrix} m^{2\nu+1} \\ \cdot \\ \cdot \\ \cdot \\ m^{2k+1} \\ m^{2k}\end{matrix}\right\rangle = \delta^a_b m^{2k}_a \left|\begin{matrix} m^{2\nu+1} \\ \cdot \\ \cdot \\ \cdot \\ m^{2k+1} \\ m^{2k}\end{matrix}\right\rangle, \quad a \ge b, \tag{4.5}$$

and similarly for the pattern involving $\tilde{m}^{2k}$. It then follows that we must have

$$[C^a_a, {}^{2k+1}S_\mu] \left|\begin{matrix} m^{2\nu+1} \\ \cdot \\ \cdot \\ \cdot \\ m^{2k+1} \\ m^{2k}\end{matrix}\right\rangle = (\delta^\mu_a - \delta^\mu_{\bar{a}})\, {}^{2k+1}S_\mu \left|\begin{matrix} m^{2\nu+1} \\ \cdot \\ \cdot \\ \cdot \\ m^{2k+1} \\ m^{2k}\end{matrix}\right\rangle \tag{4.6}$$

and

$$[C^a_b, {}^{2k+1}S_\mu] \left|\begin{matrix} m^{2\nu+1} \\ \cdot \\ \cdot \\ \cdot \\ m^{2k+1} \\ m^{2k}\end{matrix}\right\rangle = 0 \quad \text{for } a > b. \tag{4.7}$$

We may view Eqs. (4.6) and (4.7) as the definition of the shift operators.

Now let $V(\mu)_d$ be, for fixed value of μ, an O($2k$) one-tensor operator. The μ is the argument serves to identify the particular operator—we shall actually construct $2k$ such operators. It follows from its definition, Eqs. (3.10) and (3.11), that if we set

$$ {}^{2k+1}S_\mu = V(\mu)_\mu, \tag{4.8} $$

then Eq. (4.6) is satisfied and Eq. (4.7) becomes

$$ [\delta^{\bar{b}}_\mu V(\mu)_{\bar{a}} - \delta^a_\mu V(\mu)_b] \left| \begin{matrix} m^{2\nu+1} \\ \cdot \\ \cdot \\ \cdot \\ m^{2k+1} \\ m^{2k} \end{matrix} \right\rangle = 0, \quad a > b. \tag{4.9} $$

Now we note that it is sufficient to require

$$ V(\mu)_d \left| \begin{matrix} m^{2\nu+1} \\ \cdot \\ \cdot \\ \cdot \\ m^{2k+1} \\ m^{2k} \end{matrix} \right\rangle = 0 \quad \text{for } \mu > d \tag{4.10} $$

to ensure that Eq. (4.9) is satisfied.

Thus the problem of determining O($2k+1$) shift operators is reduced to the problem of finding O($2k$) one-tensor operators satisfying Eq. (4.10). This problem is now solved recursively as follows. Let $V(\mu)_d$ be an O($2k$) one-tensor operator that satisfies Eq. (4.10) and consider $V(\mu+1)_d$ defined by

$$ V(\mu+1)_d \equiv \{V(\mu)(C - c^{2k}_\mu \mathbb{1})\}_d, \tag{4.11} $$

where c^{2k}_μ is a number to be determined below. First, it is obvious [see Eq. (3.15)] that $V(\mu+1)_d$ is an O($2k$) one-tensor operator. Second, we have [where the prime on $\sum$ serves as a reminder that all indices are in the O($2k$) range]

$$ \begin{aligned} V(\mu+1)_d &= \sum_{a=\nu}^{d\,\prime} V(\mu)_a (C - c^{2k}_\mu \mathbb{1})^a_d + \sum_{a=d+1}^{\nu\,\prime} V(\mu)_a C^a_d \\ &\doteq \sum_{a=\nu}^{d\,\prime} V(\mu)_a (C - c^{2k}_\mu \mathbb{1})^a_d. \end{aligned} \tag{4.12} $$

Here the symbol $\doteq$ is used to indicate that the equation holds when both sides are applied to the state

$$ \left| \begin{matrix} m^{2\nu+1} \\ \cdot \\ \cdot \\ \cdot \\ m^{2k+1} \\ m^{2k} \end{matrix} \right\rangle $$

and we exploit the fact that C^a_d for $a > d$ annihilates that state. Next

$$ \begin{aligned} V(\mu+1)_d &\doteq V(\mu)_d (C^d_d - c^{2k}_\mu) + \sum_{a=\bar{\nu}}^{d-1\,\prime} \{C^a_d V(\mu)_a + [V(\mu)_a, C^a_d]\} \\ &\doteq V(\mu)_d \left\{ m^{2k}_d - c^{2k}_\mu + \sum_{a=\bar{\nu}}^{d-1\,\prime} (1 - \delta^{\bar{d}}_a) \right\} \\ &\quad + \sum_{a=\nu}^{d-1\,\prime} C^a_d V(\mu)_a. \end{aligned} \tag{4.13} $$

Now since by assumption

$$ V(\mu)_a \left| \begin{matrix} m^{2\nu+1} \\ \cdot \\ \cdot \\ \cdot \\ m^{2k+1} \\ m^{2k} \end{matrix} \right\rangle = 0 \quad \text{for } \mu > a, \tag{4.14} $$

the rhs of Eq. (4.13) vanishes for $\mu > d$. For $\mu = d$ we can make the rhs of Eq. (4.13) vanish too if we choose c^{2k}_μ to be

$$ c^{2k}_\mu = m^{2k}_\mu + \sum_{a=\nu}^{\mu-1\,\prime} (1 - \delta^{\bar{\mu}}_a) \tag{4.15} $$

$$ = m^{2k}_\mu + \mu + \nu - 2(\nu - k + 1)\theta_\mu, \tag{4.16} $$

where

$$ \theta_\mu = \begin{matrix} 1 \text{ if } \mu > 0, \\ 0 \text{ otherwise.} \end{matrix} \tag{4.17} $$

Thus we have shown that

$$ V(\mu+1)_d \left| \begin{matrix} m^{2\nu+1} \\ \cdot \\ \cdot \\ \cdot \\ m^{2k+1} \\ m^{2k} \end{matrix} \right\rangle = 0 \quad \text{for } \mu + 1 > d. \tag{4.18} $$

Consequently, by iteration,

$$ V(\mu)_d = \left\{ V(\bar{\nu}) \prod_{j=\bar{\nu}}^{\mu-1\,\prime} (C - c^{2k}_j \mathbb{1}) \right\}_d \tag{4.19} $$

is a solution to our problem, where the prime on Π indicates that j is in the O($2k$) range, where we define

$$ \prod_{j=\bar{\nu}}^{\bar{\nu}-1\,\prime} (C - c^{2k}_j \mathbb{1}) \equiv \mathbb{1} \tag{4.20} $$

and where $V(\bar{\nu})$ must be an O($2k$) one-tensor operator obeying

$$ V(\bar{\nu})_d \doteq 0 \quad \text{for } \bar{\nu} > d. \tag{4.21} $$

However Eq. (4.21) is *empty* since the condition $\bar{\nu} > d$ is never satisfied. It follows that the only requirement on $V(\bar{\nu})_d$ is that it be an O($2k$) one-tensor operator and the simplest nontrivial choice is

$$ V(\bar{\nu})_d = C^0_d. \tag{4.22} $$

we thus conclude that

$$ V(\mu)_d = \left\{ C \prod_{j=\bar{\nu}}^{\mu-1\,\prime} (C - c^{2k}_j \mathbb{1}) \right\}^0_d \tag{4.23} $$

and

$$ {}^{2k+1}S_\mu \left| \begin{matrix} m^{2\nu+1} \\ \cdot \\ \cdot \\ \cdot \\ m^{2k+1} \\ m^{2k} \end{matrix} \right\rangle = \left\{ C \prod_{j=\bar{\nu}}^{\mu-1\,\prime} (C - c^{2k}_j \mathbb{1}) \right\}^0_\mu \left| \begin{matrix} m^{2\nu+1} \\ \cdot \\ \cdot \\ \cdot \\ m^{2k+1} \\ m^{2k} \end{matrix} \right\rangle. \tag{4.24} $$

It is perhaps worth noting that the raising operators ($\mu < 0$) do not depend on the component of the weight being raised so that $m^{2k}_{|\mu|}$ can be raised by, say, p units by acting with $\{{}^{2k+1}S_\mu\}^p$. The lowering oprators ($\mu > 0$) depend on the entire weight, including therefore the component that is being lowered, and this must be kept

in mind. This is in contrast to the situation for the $U(n)$ groups and makes the procedure for the $O(n)$ groups somewhat more complicated.

5. O(2k) SHIFT OPERATORS

The $O(2k)$ shift operators will be denoted by ${}^{2k}S_\mu$ with the index μ in the $O(2k-1)$ range, i.e.,

$$\nu - k + 2 \leq |\mu| \leq \nu \quad \text{and} \quad \mu \neq 0. \tag{5.1}$$

Throughout this section all indices are in the $O(2k-1)$ range. The shift operators are defined by

$${}^{2k}S_\mu \begin{vmatrix} m^{2\nu+1} \\ \cdot \\ \cdot \\ \cdot \\ m^{2k} \\ m^{2k-1} \end{vmatrix} \propto \begin{vmatrix} m^{2\nu+1} \\ \cdot \\ \cdot \\ \cdot \\ m^{2k} \\ \tilde{m}^{2k-1} \end{vmatrix}, \tag{5.2}$$

where

$$m^{2k-1} \equiv (m_\nu^{2k-1}, m_{\nu-1}^{2k-1}, \ldots, m_{|\mu|}^{2k-1}, \ldots, m_{\nu-k+2}^{2k-1}) \tag{5.3}$$

and

$$\tilde{m}^{2k-1} \equiv (m_\nu^{2k-1}, m_{\nu-1}^{2k-1}, \ldots, m_{|\mu|}^{2k-1} - \epsilon_\mu, \ldots, m_{\nu-k+2}^{2k-1}), \tag{5.4}$$

with

$$\epsilon_\mu = \begin{cases} 1 & \text{for } \mu > 0 \\ 0 & \text{for } \mu = 0. \\ -1 & \text{for } \mu < 0 \end{cases} \tag{5.5}$$

It follows that ${}^{2k}S_\mu$ is a lowering, weight, or raising operator for $\mu > 0$, $\mu = 0$, or $\mu < 0$, respectively.

By following the procedure described in Sec. 4 for ${}^{2k+1}S_\mu$ modified slightly to account for the different range of the indices, we arrive at the result

$${}^{2k}S_\mu \begin{vmatrix} m^{2\nu+1} \\ \cdot \\ \cdot \\ \cdot \\ m^{2k} \\ m^{2k-1} \end{vmatrix} = \left\{ B \prod_{j=\bar{\nu}}^{\mu-1}{}' (C - c_j^{2k-1}\mathbb{1}) \right\}_\mu \begin{vmatrix} m^{2\nu+1} \\ \cdot \\ \cdot \\ \cdot \\ m^{2k} \\ m^{2k-1} \end{vmatrix}. \tag{5.6}$$

Here

$$B_d = (C_d^{\nu-k+1} - C_d^{\bar{\nu}+k-1})/\sqrt{2}, \tag{5.7}$$

$$c_j^{2k-1} = m_j^{2k-1} + \sum_{a=\bar{\nu}}^{j-1}{}' (1 - \delta_a^{\bar{j}}) \tag{5.8}$$

$$= m_j^{2k-1} + j + \nu - (2\nu - 2k + 3)\theta_j - (\nu - k + 1)\delta_j^0 \tag{5.9}$$

and the prime serves as a reminder that the indices range as is appropriate for $O(2k-1)$.

6. CONCLUSION

As the reader may have noticed our $O(n)$ shift operators, in contrast to those of Wong,[5] depend on the $O(n-1)$ weight that is being shifted. Although this may appear to be a disadvantage of the present method, the main application of the shift operators is towards the construction of an orthonormal basis for the irreps. That construction makes use of normalized lowering operators and the normalization coefficients are functions of the $O(n)$ and $O(n-1)$ weights. Consequently the additional dependence on m^{n-1} in our (unnormalized) lowering $O(n)$ operators hardly matters.

The evaluation of the normalization coefficients is quite tedious and is given in the paper[2] that follows. As mentioned earlier, our shift operators are not the same as those of Wong[5]; however, the normalization coefficients should be the same aside from different conventions on ordering and labeling of weights, and we find that to be the case.

In evaluating the normalization coefficients one obtains restrictions on the allowed values of the weight components m_j^i. These restrictions are a consequence of Eq. (1.4), which is the condition that the irrep be unitary. We may also obtain nonunitary irreps if Eq. (1.4) is replaced by certain other relations. By means of Weyl's unitary trick, one can obtain in this way orthonormal bases for certain unitary irreps (namely those in the discrete series) of noncompact orthogonal groups.

[1] A. M. Bincer, J. Math. Phys. **18**, 1870 (1977).
[2] A. M. Bincer, J. Math. Phys. **19**, 1179 (1978).
[3] J. C. Nagel and M. Moshinsky, J. Math. Phys. 6, 682 (1965); Rev. Mex. Fis. 14, 29 (1965).
[4] D. P. Zhelobenko, Usp. Mat. Nauk **17**, 27 (1962) [Russ. Math. Surveys **17**, 1 (1962)].
[5] M. K. F. Wong, J. Math. Phys. 8, 1899 (1967).
[6] I. M. Gel'fand and M. L. Zetlin, Dokl. Akad. Nauk USSR **71**, 1017 (1950).
[7] J. D. Louck and L. C. Biedenharn, J. Math. Phys. **11**, 2368 (1970).
[8] S. Okubo, J. Math. Phys. **16**, 528 (1975).
[9] C. O. Nwachuku and M. A. Rashid, J. Math. Phys. **17**, 1611 (1976).
[10] G. Racah, Springer Tracts Mod. Phys. 37, 28 (1965), Eq. (74). The relation between Racah's and our notation is $X_{ik} = C_k^i$.
[11] G. E. Baird and L. C. Biedenharn, J. Math. Phys. 4, 1449 (1963).

Normalization coefficients for the O(n) shift operators

Adam M. Bincer

Physics Department, University of Wisconsin, Madison, Wisconsin 53706
(Received 28 September 1977)

Normalization coefficients are derived for the shift operators of O(n) introduced previously. The resultant normalized lowering operator for O(n) is presented in a form analogous to the U(n) case.

1. INTRODUCTION

In this paper the O(n) shift operators of the preceding paper[1] are normalized and the normalization coefficients are evaluated. Except when otherwise stated we use the notation of the preceding paper. These normalization coefficients are proportional (with known proportionality factors) to the nonvanishing matrix elements of the O(n) generators taken between O($n-1$) states of highest weight. Thus, the calculation of this paper may also be viewed as an evaluation of all matrix elements of O(n) generators in an irrep. Another application is the construction of an orthonormal basis for an arbitrary irrep of O(n); this is accomplished by successive application of normalized lowering operators to the state of highest weight in the irrep.

It follows from the preceding paper[1] that the (not normalized) O(n) shift operators ${}^nS_\mu$ satisfy

$$
{}^nS_\mu \left| \begin{matrix} m_i^{2\nu+1} \\ \vdots \\ m_1^n \\ m_i^{n-1} \end{matrix} \right\rangle
= \left| \begin{matrix} m_i^{2\nu+1} \\ \vdots \\ m_1^n \\ m_i^{n-1} - \delta_{\mu i} + \delta_{\bar{\mu} i} \end{matrix} \right\rangle \begin{pmatrix} m_i^{n-1} \\ m_i^{n-1} - \delta_{\mu i} + \delta_{\bar{\mu} i} \end{pmatrix}, \tag{1.1}
$$

where we change the notation slightly and write m_i^p in place of m^p of the preceding paper, and where

$$
\begin{pmatrix} m_i^{n-1} \\ m_i^{n-1} - \delta_{\mu i} + \delta_{\bar{\mu} i} \end{pmatrix}
$$

are some numbers that will be called O(n) normalization coefficients provided all kets are taken to have unit norm. In the case that these coefficients are nonvanishing we may divide Eq. (1.1) by them and define *normalized* O(n) shift operators ${}^n\mathcal{S}_\mu$, which obviously satisfy

$$
{}^n\mathcal{S}_\mu \left| \begin{matrix} m_i^{2\nu+1} \\ \vdots \\ m_i^n \\ m_i^{n-1} \end{matrix} \right\rangle
= \left| \begin{matrix} m_i^{2\nu+1} \\ \vdots \\ m_i^n \\ m_i^{n-1} - \delta_{\mu i} + \delta_{\bar{\mu} i} \end{matrix} \right\rangle . \tag{1.2}
$$

We recall that μ ranges over values appropriate to O($n-1$) and that we have here a lowering, a weight, or a raising operator according to $\mu > 0$, $\mu = 0$, $\mu < 0$.

For the reader disinterested in the tedious details of the derivations we present here our main results. For the lowering ($\mu > 0$) or raising ($\mu < 0$) normalization coefficients we find

$$
\left| \begin{pmatrix} m_i^{n-1} \\ m_i^{n-1} - \delta_{\mu i} + \delta_{\bar{\mu} i} \end{pmatrix} \right|^2
= -\prod_{j=\bar{\nu}}^{\nu}{}'' (c_{\bar{\mu}}^{n-1} - c_j^n + 1) \prod_{s=\bar{\nu}}^{\mu-1}{}' (c_s^{n-1} - c_\mu^{n-1})
\times \prod_{t=\bar{\nu}}^{\bar{\mu}-1}{}' [c_{\bar{\mu}}^{n-1} - c_t^{n-1} + 1 + (-)^n \delta_{t\mu}]^{-1}, \tag{1.3}
$$

where according to the preceding paper

$$
c_\mu^n = m_\mu^n + \mu + \nu + (n - 2 - 2\nu)\theta_\mu + \left(\frac{n-1}{2} - \nu\right)\delta_{\mu 0} \tag{1.4}
$$

and where the prime and double prime on the product indicate that the dummy index is restricted to the O($n-1$) and O(n) range respectively.

For the normalized lowering operator we find

$$
{}^n\mathcal{S}_\mu = \left\{ D \prod_{j=\bar{\nu}}^{\mu-1}{}' D(j) \right\}_\mu , \quad \mu > 0, \tag{1.5}
$$

where

$$
D(j)_b^a = (C - c_j^{n-1}\,\mathbb{1})_b^a \left| (c_{\bar{\mu}}^{n-1} - c_j^{n-1} + 1) \times (c_\mu^{n-1} - c_j^{n-1}) \right|^{-1/2}, \tag{1.6}
$$

$$
D_a = V(\bar{\nu})_a \left| \prod_{j=\bar{\nu}}^{\nu}{}'' (c_{\bar{\mu}}^{n-1} - c_j^n + 1)^{-1} \prod_{s=\bar{\nu}}^{\nu}{}' \times (c_{\bar{\mu}}^{n-1} - c_s^{n-1} + 1) \right|^{1/2}, \tag{1.7}
$$

and where according to the preceding paper ($1 \le k \le \nu$)

$$
V(\bar{\nu})_a = \begin{cases} C_a^0 , & n = 2\nu + 1, \\ (C_a^{\nu-k+1} + C_a^{\bar{\nu}+k-1})/\sqrt{2} , & n = 2k - 1, \\ (C_a^{\nu-k+1} - C_a^{\bar{\nu}+k-1})/\sqrt{2} , & n = 2k. \end{cases} \tag{1.8}
$$

Lastly, we find that the labels in the Gel'fand–Zetlin patterns must satisfy the "inbetweenness" conditions:

$$
m_\nu^n \ge m_\nu^{n-1} \ge m_{\nu-1}^n \ge m_{\nu-1}^{n-1} \ge \cdots \ge m_{\nu-k+1}^n, \quad n = 2k+1, \quad 1 \le k \le \nu,
$$
$$
m_\nu^n \ge m_\nu^{n-1} \ge m_{\nu-1}^n \ge m_{\nu-1}^{n-1} \ge \cdots \ge \left| m_{\nu-k+1}^n \right|, \quad n = 2k, \quad 2 \le k \le \nu, \tag{1.9}
$$

with all m's simultaneously integer or semi-integer. This paper is organized as follows. In Sec. 2 we relate the normalization coefficients to matrix elements of generators and thus obtain a connection between raising and lowering coefficients. In Sec. 3 we use a recursion relation to express the raising coefficients in terms of so-called elementary raising coefficients. The

 0022-2488/78/1905-1179$1.00

calculation is completed in Sec. 4 where we evaluate the elementary raising coefficients by making use of the quadratic Casimir operators.

2. RELATION BETWEEN NORMALIZATION COEFFICIENTS AND MATRIX ELEMENTS OF GENERATORS

According to Eq. (4.13) of the preceding paper we have

$$V(\mu)_d \left| m_i^{n-1} \right\rangle = \left\{ (c_d - c_{\mu-1}) V(\mu-1)_d + \sum_{a=\mu-1}^{d-1}{}' C_d^a V(\mu-1)_a \right\} \left| m_i^{n-1} \right\rangle. \tag{2.1}$$

Here and in what follows we suppress the $(n-1)$ superscript on the c's and omit in the symbol for the normalized kets and bras all but the bottom row. The prime on $\sum$ serves as a reminder that all indices are in the $O(n-1)$ range. The $V(\mu-1)_d$ in the above may be eliminated in favor of $V(\mu-2)_d$ by using Eq. (2.1) with μ replaced by $\mu-1$. Repeating this procedure an appropriate number of times yields

$$V(\mu)_d \left| m_i^{n-1} \right\rangle = \prod_{j=\bar{\nu}}^{\mu-1}{}'(c_d - c_j) \left\{ V(\bar{\nu})_d + \sum_{\kappa=\bar{\nu}}^{\mu-1}{}' \prod_{s=\bar{\nu}}^{\kappa}{}' (c_d - c_s)^{-1} \sum_{a=\kappa}^{d-1}{}' C_d^a V(\kappa)_a \right\} \left| m_i^{n-1} \right\rangle. \tag{2.2}$$

We define a unitary representation by

$$C_b^{a\dagger} = C_a^b, \tag{2.3}$$

so that C_d^a in Eq. (2.2) when acting to the left is a raising generator and annihilates any $O(n-1)$ state of highest weight $\langle \hat{m}_i^{n-1} |$. Setting in Eq. (2.2) $\mu = d$ and contracting from the left with $\langle \hat{m}_i^{n-1} |$ we obtain

$$\langle \hat{m}_i^{n-1} \left| {}^nS_\mu \right| m_i^{n-1} \rangle = \prod_{j=\bar{\nu}}^{\mu-1}{}'(c_\mu - c_j) \langle \hat{m}_i^{n-1} \left| V(\bar{\nu})_\mu \right| m_i^{n-1} \rangle. \tag{2.4}$$

According to Eq. (1.1) the lhs of the above equation vanishes except if $\hat{m}_i^{n-1} = m_i^{n-1} - \delta_{\mu i} + \delta_{\bar{\mu} i}$, in which case it is equal to the normalization coefficient. It will be seen in what follows that the factor $\prod'^{\mu-1}_{j=\bar{\nu}}(c_\mu - c_j)$ never vanishes and therefore the nonvanishing matrix elements of $V(\bar{\nu})_\mu$ are proportional to the normalization coefficients. Since the generators of $O(n)$ consist of the $V(\bar{\nu})_\mu$ and the $O(n-1)$ generators, and the latter necessarily have vanishing matrix elements between $O(n-1)$ states of highest weight (except, of course, for the weight generators whose matrix elements, however, are trivial to calculate) we have here proof of the statement given in the Introduction.

Exploiting the fact that

$$V(\bar{\nu})^\dagger_\mu = (-)^n V(\nu)_{\bar{\mu}} \tag{2.5}$$

we may deduce from Eq. (2.4) the following relation between raising and lowering normalization coefficients

$$\begin{pmatrix} m_i^{n-1} \\ m_i^{n-1} - \delta_{\mu i} + \delta_{\bar{\mu} i} \end{pmatrix} \prod_{j=\bar{\nu}}^{\mu-1}{}' (c_{\bar{\mu}} - c_j + \delta_{j\mu} + 1 - \delta_{\mu 0}) = (-)^n \begin{pmatrix} m_i^{n-1} - \delta_{\mu i} + \delta_{\bar{\mu} i} \\ m_i^{n-1} \end{pmatrix}^* \prod_{j=\bar{\nu}}^{\mu-1}{}' (c_\mu - c_j). \tag{2.6}$$

We remark in passing that for $\mu = 0$, Eq. (2.6) shows that the corresponding coefficient, which is the eigenvalue of the weight operator ${}^nS_0$, is real.

3. RECURSION RELATIONS

By combining the identity (where $\alpha, \beta \geq 0$)

$$\begin{pmatrix} m_i^{n-1} \\ m_i^{n-1} + \delta_{\alpha i} \end{pmatrix} \begin{pmatrix} m_i^{n-1} + \delta_{\alpha i} \\ m_i^{n-1} \end{pmatrix} = \begin{pmatrix} m_i^{n-1} + \delta_{\beta i} \\ m_i^{n-1} + \delta_{\beta i} + \delta_{\alpha i} \end{pmatrix} \begin{pmatrix} m_i^{n-1} + \delta_{\beta i} + \delta_{\alpha i} \\ m_i^{n-1} + \delta_{\beta i} \end{pmatrix}, \quad \alpha \neq \beta, \tag{3.1}$$

which is derived in Appendix A, with the relation between raising and lowering coefficients derived in Sec. 2 we obtain the following recursion relations: for $\alpha > \beta > 0$

$$\left| \begin{pmatrix} m_i^{n-1} \\ m_i^{n-1} + \delta_{\alpha i} \end{pmatrix} \right|^2 (c_\alpha - c_\beta + 1)(c_\alpha - c_{\bar{\beta}} + 1) = \left| \begin{pmatrix} m_i^{n-1} + \delta_{\beta i} \\ m_i^{n-1} + \delta_{\beta i} + \delta_{\alpha i} \end{pmatrix} \right|^2 (c_\alpha - c_\beta)(c_\alpha - c_{\bar{\beta}} + 2), \tag{3.2}$$

and for $\beta > \alpha \geq 0$

$$\left| \begin{pmatrix} m_i^{n-1} \\ m_i^{n-1} + \delta_{\alpha i} \end{pmatrix} \right|^2 (c_\alpha - c_{\bar{\beta}} + \theta_\alpha)(c_{\bar{\alpha}} - c_{\bar{\beta}})^{-1} = \left| \begin{pmatrix} m_i^{n-1} + \delta_{\beta i} \\ m_i^{n-1} + \delta_{\beta i} + \delta_{\alpha i} \end{pmatrix} \right|^2 (c_\alpha - c_{\bar{\beta}} + \theta_\alpha + 1)(c_{\bar{\alpha}} - c_{\bar{\beta}} + 1)^{-1}. \tag{3.3}$$

Since, as shown in Appendix B, in a representation we must have

$$m_\beta^n - m_\beta^{n-1} = \text{nonnegative integer}, \quad \beta > 0, \tag{3.4}$$

we may use these recursion relations to express the coefficient involving the weight component $w_\beta = m_\beta^{n-1}$ in terms of the coefficient involving the weight component $w_\beta = m_\beta^n$. The result is

$$\left| \begin{pmatrix} m_i^{n-1} \\ m_i^{n-1} + \delta_{\alpha i} \end{pmatrix} \right|^2 = \left| {}^nN_\alpha \right|^2 \prod_{j=\gamma}^{\alpha-1} \frac{(c_j^n - c_\alpha - 2)(c_{\bar{j}}^n - c_\alpha - 1)}{(c_j - c_\alpha - 1)(c_{\bar{j}} - c_\alpha - 1)} \times \prod_{s=\bar{\nu}}^{\alpha-1}{}' \frac{(c_s - c_{\bar{\alpha}})(c_s^n - c_\alpha - \theta_\alpha)}{(c_s - c_\alpha - \theta_\alpha)(c_s^n - c_{\bar{\alpha}})} \tag{3.5}$$

where

$$\gamma = \begin{cases} \nu - k + 1 & \text{for } n = 2k+1, \\ \nu - k + 2 & \text{for } n = 2k, \end{cases} \tag{3.6}$$

and where we define the *elementary* normalization coefficient ${}^nN_\alpha$ by

$$ {}^nN_\alpha \equiv \begin{pmatrix} m_\nu^n \cdots m_{\alpha+1}^n, & m_\alpha^{n-1}, & m_{\alpha-1}^n \cdots \\ m_\nu^n \cdots m_{\alpha+1}^n, & m_\alpha^{n-1} + \theta_\alpha, & m_{\alpha-1}^n \cdots \end{pmatrix}. \tag{3.7}$$

Before proceeding to the evaluation of the elementary coefficients we note that for $\alpha = 0$, Eq. (3.5) reduces to

$$\begin{pmatrix} m_i^{n-1} \\ m_i^{n-1} \end{pmatrix} = \begin{pmatrix} m_i^n \\ m_i^n \end{pmatrix}, \tag{3.8}$$

where we have used the fact that this coefficient is real as noted in Sec. 2. We have here a direct demonstration that the weight operator ${}^nS_0$ is an O(n) invariant (n even).

4. ELEMENTARY NORMALIZATION COEFFICIENTS

By making repeated use of Eq. (2.2) and the identity

$$\sum_{\kappa=\bar{\nu}}^{a}{}' (c_b - c_\kappa)^{-1} \prod_{j=\bar{\nu}}^{\kappa-1}{}' (c_a - c_j)(c_b - c_j)^{-1} = (c_b - c_a)^{-1}, \quad b > a, \tag{4.1}$$

we can show that

$$\begin{aligned}\langle m_i^{n-1} | V(\bar{\nu})_{\bar{\beta}}^{\dagger} V(\bar{\nu})_{\bar{\beta}} - \prod_{j=\bar{\nu}}^{\bar{\beta}-1}{}' (c_{\bar{\beta}} - c_j)^{-2}\ {}^nS_{\bar{\beta}}^{\dagger}\ {}^nS_{\bar{\beta}} | m_i^{n-1}\rangle \\ = \sum_{b=\bar{\nu}}^{\bar{\beta}-1}{}' (c_{\bar{\beta}} - c_b)^{-1} \prod_{j=b+1}^{\bar{\beta}-1}{}' [1 - (c_{\bar{\beta}} - c_j)^{-1}] \\ \times \langle m_i^{n-1} | V(\bar{\nu})_b^{\dagger} V(\bar{\nu})_b | m_i^{n-1}\rangle. \end{aligned} \tag{4.2}$$

Now let in Eq. (4.2), $|m_i^{n-1}\rangle = |\alpha\rangle$, where

$$|\alpha\rangle \equiv |m_\nu^n \cdots m_{\alpha+1}^n,\ m_\alpha^{n-1},\ m_{\alpha-1}^n \cdots\rangle, \quad \alpha > 0. \tag{4.3}$$

For $\beta = \alpha$ the rhs of Eq. (4.2) vanishes because the state $V(\bar{\nu})_{\bar{b}}^{\dagger}|\alpha\rangle$, $b > \alpha$, has weight higher than the highest weight, which is impossible. Thus we obtain

$$|{}^nN_\alpha|^2 = \prod_{j=\bar{\nu}}^{\bar{\alpha}-1}{}' (c_{\bar{\alpha}} - c_j^n)^2 \langle \alpha | V(\bar{\nu})_{\bar{\alpha}}^{\dagger} V(\bar{\nu})_{\bar{\alpha}} | \alpha\rangle, \tag{4.4}$$

On the other hand, for $\alpha > \beta > 0$ we show in Appendix B that

$${}^nS_{\bar{\beta}} | \alpha > = 0, \tag{4.5}$$

which allows us to deduce from Eq. (4.2) that

$$\begin{aligned}\langle \alpha | V(\bar{\nu})_{\bar{\beta}}^{\dagger} V(\bar{\nu})_{\bar{\beta}} | \alpha\rangle = (c_{\bar{\beta}}^n - c_{\bar{\alpha}}^-)^{-1} \prod_{j=\bar{\alpha}+1}^{\bar{\beta}-1} [1 - (c_{\bar{\alpha}} - c_j^n)^{-1}] \\ \times \langle \alpha | V(\bar{\nu})_{\bar{\alpha}}^{\dagger} V(\bar{\nu})_{\bar{\alpha}} | \alpha\rangle, \quad \alpha > \beta > 0. \end{aligned} \tag{4.6}$$

To proceed we make use of the O(n) quadratic Casimir operator ${}^nQ$ defined by

$${}^nQ \equiv \tfrac{1}{2} \sum_{a,b}{}'' C_b^a C_a^b, \tag{4.7}$$

where the double prime indicates that all indices are in the O(n) [*not* O($n-1$)] range. For $n = 2k+1$ we have

$$\begin{aligned}{}^{2k+1}Q = \sum_{a=\nu-k+1}^{\nu} \{C_a^a [C_a^a + 2(a - \nu + k) - 1] + 2V(\bar{\nu})_{\bar{a}}^{\dagger} V(\bar{\nu})_{\bar{a}}\} \\ + \sum_{a>b}{}' C_b^a C_a^b. \end{aligned} \tag{4.8}$$

Since ${}^nQ$ is an O(n) invariant we may evaluate ${}^nQ|\alpha\rangle$ by setting m_α^{n-1} equal to its highest value m_α^n:

$${}^{2k+1}Q | \alpha\rangle = \sum_{a=\nu-k+1}^{\nu} m_a^{2k+1} [m_a^{2k+1} + 2(a - \nu + k) - 1] | \alpha\rangle \tag{4.9}$$

and consequently

$$\begin{aligned}2 \sum_{a=\nu-k+1}^{\alpha} V(\bar{\nu})_{\bar{a}}^{\dagger} V(\bar{\nu})_{\bar{a}} | \alpha\rangle \\ = (m_\alpha^{2k+1} - m_\alpha^{2k}) [m_\alpha^{2k+1} + m_\alpha^{2k} + 2(\alpha - \nu + k) - 1] | \alpha\rangle. \end{aligned} \tag{4.10}$$

By combining Eqs. (4.10) and (4.6) we find for $\nu - k + 1 \leq \alpha \leq \nu$, $n = 2k+1$,

$$\begin{aligned}\langle \alpha | V(\bar{\nu})_{\bar{\alpha}}^{\dagger} V(\bar{\nu})_{\bar{\alpha}} | \alpha\rangle \\ = \tfrac{1}{2}(c_\alpha^n - c_{\bar{\alpha}}) \prod_{j=\nu-k+1}^{\alpha} (c_j^n - c_\alpha - 1) \prod_{s=\nu-k+1}^{\alpha-1} (c_s^n - c_\alpha - 2)^{-1}, \end{aligned} \tag{4.11}$$

which completes the evaluation of the elementary coefficients in the case $n = 2k+1$.

The case $n = 2k$ is somewhat more complicated. In place of Eq. (4.8) we now have

$$\begin{aligned}{}^{2k}Q = \sum_{a=\nu-k+2}^{\nu} \{C_a^a [C_a^a + 2(a - \nu + k - 1)] + 2V(\bar{\nu})_{\bar{a}}^{\dagger} V(\bar{\nu})_{\bar{a}}\} \\ + V(\bar{\nu})_0 V(\bar{\nu})_0 + \sum_{a>b}{}' C_b^a C_a^b, \end{aligned} \tag{4.12}$$

where we note that

$$C_{\nu-k+1}^{\nu-k+1} = V(\bar{\nu})_0. \tag{4.13}$$

Instead of Eq. (4.9) we now have

$${}^{2k}Q | \alpha\rangle = \sum_{a=\nu-k+1}^{\nu} m_a^{2k} [m_a^{2k} + 2(a - \nu + k - 1)] | \alpha\rangle; \tag{4.14}$$

and instead of Eq. (4.10),

$$\begin{aligned}\{V(\bar{\nu})_0^{\dagger} V(\bar{\nu})_0 + 2 \sum_{a=\nu-k+2}^{\alpha} V(\bar{\nu})_{\bar{a}}^{\dagger} V(\bar{\nu})_{\bar{a}}\} | \alpha\rangle \\ = \{(m_{\nu-k+1}^{2k})^2 + (m_\alpha^{2k} - m_\alpha^{2k-1}) \\ \times [m_\alpha^{2k} + m_\alpha^{2k-1} + 2(\alpha - \nu + k - 1)]\} | \alpha\rangle. \end{aligned} \tag{4.15}$$

The added complication is due to the presence of $V(\bar{\nu})_0$ and it is resolved by using Eq. (4.2) with $\beta = 0$ to give

$$\begin{aligned}\langle \alpha | V(\bar{\nu})_0^{\dagger} V(\bar{\nu})_0 | \alpha\rangle - \left(\frac{c_0 - c_{\bar{\alpha}}^{2k}}{c_0 - c_{\bar{\alpha}}}\right)^2 (m_{\nu-k+1}^{2k})^2 \\ = (c_0 - c_{\bar{\alpha}})^{-1} \prod_{j=\nu-k+2}^{\alpha-1} [1 - (c_{\bar{\alpha}} - c_j^{2k})^{-1}] \\ \times \langle \alpha | V(\bar{\nu})_{\bar{\alpha}}^{\dagger} V(\bar{\nu})_{\bar{\alpha}} | \alpha\rangle, \end{aligned} \tag{4.16}$$

where we have also used [since ${}^{2k}S_0$ is an O($2k$) invariant]

$$\begin{aligned}\langle \alpha |{}^{2k}S_0^{\dagger}\ {}^{2k}S_0 | \alpha\rangle \\ = \langle m_i^{2k} |{}^{2k}S_0^{\dagger}\ {}^{2k}S_0 | m_i^{2k}\rangle \\ = \prod_{j=\bar{\nu}}^{\bar{\nu}+k-2}{}' (c_0 - c_j^{2k})^2 \langle m_i^{2k} | V(\bar{\nu})_0^{\dagger} V(\bar{\nu})_0 | m_i^{2k}\rangle \\ = \{m_{\nu-k+1}^{2k} \prod_{j=\bar{\nu}}^{\bar{\nu}+k-2}{}' (c_0 - c_j^{2k})\}^2. \end{aligned} \tag{4.17}$$

By combining Eqs. (4.16), (4.15), and (4.6) we now find for $\nu - k + 2 \leq \alpha \leq \nu$, $n = 2k$,

$$\begin{aligned}\langle \alpha | V(\bar{\nu})_{\bar{\alpha}}^{\dagger} V(\bar{\nu})_{\bar{\alpha}} | \alpha\rangle = \frac{(c_\alpha^n - c_{\bar{\alpha}})(c_{\nu-k+1}^n - c_{\bar{\alpha}})}{(c_0 - c_\alpha - 1)(c_\alpha - c_{\bar{\alpha}} + 2)} \\ \times \prod_{j=\nu-k+1}^{\alpha} (c_j^n - c_\alpha - 1) \\ \times \prod_{s=\nu-k+2}^{\alpha-1} (c_s^n - c_\alpha - 2)^{-1}, \end{aligned} \tag{4.18}$$

which completes the evaluation of the elementary coeff cients for the case $n = 2k$.

CONCLUSION

Collecting all these expressions yields the result for the magnitude squared of the raising and lowering normalization coefficients as stated in Eq. (1.3) of the introduction. The phases of these coefficients may be chosen arbitrarily except that the relative phase of a raising and lowering coefficient is fixed by Eq. (2.6). Since the formalism was developed in terms of states of highest weight it is the lowering operators that are of particular interest. We therefore choose all lowering normalization coefficients to be positive and thus obtain for the normalized lowering operators the expression given in the Introduction. A similar expression may be written down for the normalized raising operators. Finally the conditions that the labels m_j^n must satisfy as stated in the Introduction come about as a result of Eqs. (B9) and (B10) and the requirement that the magnitude squared of a normalization coefficient be nonnegative.

APPENDIX A

In this Appendix we prove a generalized version of Eq. (3.1) valid for all α, β subject only to the requirement $|\alpha| \neq |\beta|$. Starting from Eq. (1.1) and using Eq. (2.2) we may write

$$\begin{aligned}
&\begin{pmatrix} m_i^{n-1} + \delta_{\alpha i} + \delta_{\bar{\beta} i} \\ m_i^{n-1} + \delta_{\bar{\alpha} i} + \delta_{\bar{\beta} i} \end{pmatrix} \begin{pmatrix} m_i^{n-1} + \delta_{\alpha i} + \delta_{\beta i} \\ m_i^{n-1} + \delta_{\alpha i} + \delta_{\bar{\beta} i} \end{pmatrix} \\
&= P^{\alpha}_{\alpha\bar{\beta}} \langle m_i^{n-1} + \delta_{\bar{\alpha} i} + \delta_{\bar{\beta} i} | V(\bar{\nu})_{\alpha}\, {}^{n}S_{\beta} | m_i^{n-1} + \delta_{\alpha i} + \delta_{\beta i} \rangle \\
&= P^{\alpha}_{\alpha\bar{\beta}} P^{\beta}_{\beta\alpha} \langle m_i^{n-1} + \delta_{\bar{\alpha} i} + \delta_{\bar{\beta} i} | V(\bar{\nu})_{\alpha} \{ V(\bar{\nu})_{\beta} \\
&\quad + \sum_{\kappa=\bar{\nu}}^{\beta-1} {}' \sum_{a=\kappa}^{\beta-1} {}' C^{a}_{\beta}\, V(\kappa)_{a} (P^{\kappa+1}_{\beta\alpha})^{-1} \} | m_i^{n-1} + \delta_{\alpha i} + \delta_{\beta i} \rangle, \qquad \text{(A1)}
\end{aligned}$$

where

$$P^{\gamma}_{\alpha\beta} \equiv \prod_{j=\nu}^{\gamma-1} {}' [c_{\alpha} - c_j + \theta_{\alpha} + \delta_{j\bar{\alpha}} + \theta_{\beta}(\delta_{j\bar{\beta}} - \delta_{j\beta})]. \qquad \text{(A2)}$$

Noting that C^{a}_{β}, $a < \beta$, is a raising generator when acting to the left, that

$$[V(\bar{\nu})_{\alpha}, C^{a}_{\beta}] = \delta^{a}_{\alpha} V(\bar{\nu})_{\beta}, \quad |\alpha| \neq |\beta|, \qquad \text{(A3)}$$

and that

$$\langle m_i^{n-1} + \delta_{\bar{\alpha} i} + \delta_{\bar{\beta} i} | [V(\bar{\nu})_{\alpha}, V(\bar{\nu})_{\beta}] | m_i^{n-1} + \delta_{\alpha i} + \delta_{\beta i} \rangle = 0, \quad |\alpha| \neq |\beta|, \qquad \text{(A4)}$$

we may write for the rhs of Eq. (A1)

$$\begin{aligned}
&P^{\alpha}_{\alpha\bar{\beta}} P^{\beta}_{\beta\alpha} \langle m_i^{n-1} + \delta_{\bar{\alpha} i} + \delta_{\bar{\beta} i} | V(\bar{\nu})_{\beta} \{ V(\bar{\nu})_{\alpha} \\
&\quad + \theta_{\beta-\alpha} \sum_{\kappa=\bar{\nu}}^{\alpha} {}' V(\kappa)_{\alpha} (P^{\kappa+1}_{\beta\alpha})^{-1} \} | m_i^{n-1} + \delta_{\alpha i} + \delta_{\beta i} \rangle \\
&= P^{\alpha}_{\alpha\bar{\beta}} P^{\beta}_{\beta\alpha} [1 + \theta_{\beta-\alpha}(c_{\beta} - c_{\alpha} + \theta_{\beta} - \theta_{\alpha})^{-1}] \\
&\quad \times \langle m_i^{n-1} + \delta_{\bar{\alpha} i} + \delta_{\bar{\beta} i} | V(\bar{\nu})_{\beta} V(\bar{\nu})_{\alpha} | m_i^{n-1} + \delta_{\alpha i} + \delta_{\beta i} \rangle, \qquad \text{(A5)}
\end{aligned}$$

where we use Eq. (2.2) once more and note that

$$\theta_{\beta-\alpha} \sum_{\kappa=\bar{\nu}}^{\alpha} {}' (P^{\kappa+1}_{\beta\alpha})^{-1} P^{\kappa}_{\alpha\beta} = \theta_{\beta-\alpha} (c_{\beta} - c_{\alpha} + \theta_{\beta} - \theta_{\alpha})^{-1}. \qquad \text{(A6)}$$

Inspection of the rhs of Eq. (A5) shows that it is a symmetric function of α and β and so we arrive at the identity

$$\begin{aligned}
&\begin{pmatrix} m_i^{n-1} + \delta_{\alpha i} + \delta_{\bar{\beta} i} \\ m_i^{n-1} + \delta_{\bar{\alpha} i} + \delta_{\bar{\beta} i} \end{pmatrix} \begin{pmatrix} m_i^{n-1} + \delta_{\alpha i} + \delta_{\beta i} \\ m_i^{n-1} + \delta_{\alpha i} + \delta_{\bar{\beta} i} \end{pmatrix} \\
&= \begin{pmatrix} m_i^{n-1} + \delta_{\bar{\alpha} i} + \delta_{\beta i} \\ m_i^{n-1} + \delta_{\bar{\alpha} i} + \delta_{\bar{\beta} i} \end{pmatrix} \begin{pmatrix} m_i^{n-1} + \delta_{\alpha i} + \delta_{\beta i} \\ m_i^{n-1} + \delta_{\bar{\alpha} i} + \delta_{\bar{\beta} i} \end{pmatrix}, \quad |\alpha| \neq |\beta|. \qquad \text{(A7)}
\end{aligned}$$

Setting in Eq. (A7) $\alpha \to \bar{\alpha}$ and comparing the result with Eq. (A7) yields

$$\begin{aligned}
&\begin{pmatrix} m_i^{n-1} + \delta_{\alpha i} + \delta_{\beta i} \\ m_i^{n-1} + \delta_{\bar{\alpha} i} + \delta_{\beta i} \end{pmatrix} \begin{pmatrix} m_i^{n-1} + \delta_{\bar{\alpha} i} + \delta_{\beta i} \\ m_i^{n-1} + \delta_{\alpha i} + \delta_{\beta i} \end{pmatrix} \\
&= \begin{pmatrix} m_i^{n-1} + \delta_{\bar{\alpha} i} + \delta_{\bar{\beta} i} \\ m_i^{n-1} + \delta_{\alpha i} + \delta_{\bar{\beta} i} \end{pmatrix} \begin{pmatrix} m_i^{n-1} + \delta_{\alpha i} + \delta_{\bar{\beta} i} \\ m_i^{n-1} + \delta_{\bar{\alpha} i} + \delta_{\bar{\beta} i} \end{pmatrix}, \quad |\alpha| \neq |\beta|, \qquad \text{(A8)}
\end{aligned}$$

provided

$$\begin{pmatrix} m_i^{n-1} + \delta_{\gamma i} + \delta_{\beta i} \\ m_i^{n-1} + \delta_{\gamma i} + \delta_{\bar{\beta} i} \end{pmatrix} \neq 0 \quad \text{for} \quad \gamma = \alpha, \bar{\alpha}. \qquad \text{(A9)}$$

Taking in Eq. (A8) $\alpha, \beta \geq 0$ we obtain Eq. (3.1) of the text.

APPENDIX B

In this Appendix we show that

$${}^{n}S_{\bar{\beta}} | \alpha \rangle = 0, \quad \beta > 0, \quad \beta \neq \alpha, \qquad \text{(B1)}$$

where $|\alpha\rangle$ is the state defined by Eq. (4.3). We start with the observation that

$${}^{n}S_{\bar{\beta}} | m_i^{n-1} = m_i^{n} \rangle = 0 \qquad \text{(B2)}$$

because $|m_i^n\rangle$ is the $O(n)$ state of highest weight and must be annihilated by the raising operator ${}^{n}S_{\bar{\beta}}$, $\beta > 0$. In view of the definition, Eq. (1.1), this means

$$\begin{pmatrix} m_i^{n} \\ m_i^{n} + \delta_{\beta i} \end{pmatrix} = 0. \qquad \text{(B3)}$$

Taking Eq. (A7) with $m_i^{n-1} = m_i^{n} - \delta_{\alpha i}$ and with β relabeled $\bar{\beta}$ we get

$$\begin{pmatrix} m_i^{n} - \delta_{\alpha i} \\ m_i^{n} - \delta_{\alpha i} + \delta_{\beta i} \end{pmatrix} \begin{pmatrix} m_i^{n} \\ m_i^{n} - \delta_{\alpha i} \end{pmatrix} = 0, \quad \alpha \neq \beta, \quad \alpha, \beta > 0, \qquad \text{(B4)}$$

which means

$${}^{n}S_{\bar{\beta}} | m_i^{n} - \delta_{\alpha i} \rangle \begin{pmatrix} m_i^{n} \\ m_i^{n} - \delta_{\alpha i} \end{pmatrix} = 0, \quad \alpha \neq \beta, \quad \alpha, \beta > 0. \qquad \text{(B5)}$$

The result is only of interest if the state $|m_i^n - \delta_{\alpha i}\rangle$ is in the representation space which means

$$\begin{pmatrix} m_i^{n} \\ m_i^{n} - \delta_{\alpha i} \end{pmatrix} \neq 0 \qquad \text{(B6)}$$

and consequently

$${}^{n}S_{\bar{\beta}} | m_i^{n} - \delta_{\alpha i} \rangle = 0, \quad \alpha \neq \beta, \quad \alpha, \beta > 0. \qquad \text{(B7)}$$

Iterating this procedure we find for any positive integer p

$$ {}^{\pi}S_{\bar{\beta}} \left| m_i^{\eta} - p\delta_{\alpha i} \right\rangle = 0, \quad \alpha \neq \beta, \quad \alpha, \beta > 0 \tag{B8} $$

which is Eq. (B1).

We note the important conclusion that in a representation the weight components must obey

$$ m_{\alpha}^{\eta} - m_{\alpha}^{\eta-1} = \text{nonnegative integer}, \tag{B9} $$

$\alpha > 0$ and in $O(n-1)$ range,

and by a very similar argument one also shows

$$ m_{\alpha}^{\eta-1} - m_{\alpha-1}^{\eta} = \text{nonnegative integer}, \tag{B10} $$

$\alpha > 0$ and in $O(n-1)$ range.

[1] A. M. Bincer, J. Math. Phys. **19**, 1173 (1978).

I. Exact relations among correlation functions of inherently nondynamic interacting systems[a)]

Stephen B. Haley

Institut de Physique Théorique, Université de Lausanne, CH-1015 Lausanne-Dorigny, Switzerland
(Received 27 June 1977)

The standard basis operator (SBO) equation of motion is solved exactly for the class of interacting many-body Hamiltonians which can be constructed from a complete set of mutually orthogonal projection operators. This class includes the Ising model Hamiltonians. Two integration procedures are presented: (1) term by term integration over the poles of the SBO double-time temperature-dependent Green's functions in the spectral representation; (2) direct integration of the time-dependent equation of motion of an SBO. An ensemble average expression relating correlation functions among systems in different energy states is derived. For the Ising model this expression relates correlation functions of spin clusters of different orders, and it is analogous to a relation derived from the master equation under equilibrium conditions.

I. INTRODUCTION

Hamiltonians describing interacting quantum mechanical multilevel systems, each of which is modelled by operators that are diagonal in some representation possess no inherent dynamical properties. An important example is the Hamiltonian for Ising spin systems, on which an enormous amount of theoretical effort has been expended over the last 50 years.[1,2] Dynamical studies of Ising spin systems have been based on the phenomenological master equation.[3–5] In equilibrium, use of this equation leads to strictly algebraic relationships among spin correlation functions of different orders. These relationships were used in Ref. 5 to give explicit expressions for spin cluster correlation functions in the square lattice by making use of previous two-spin correlation function calculations.[6]

In this paper an exact expression relating correlation functions is derived for a broader class of Hamiltonians than has previously been considered. This is made possible through the utilization of Standard Basis Operators (SBO).[7] In the SBO representation the most general Hamiltonian for an ensemble of pairwise interacting systems, each of which possesses discrete energy levels, and which can be modelled by a diagonal operator is

$$\hat{H}=\sum_{\mu,l}\epsilon_{\mu}^{l}\hat{P}_{\mu}^{l}+\tfrac{1}{2}\sum_{\mu,l}\sum_{\nu,m}w_{\mu\nu}^{lm}\hat{P}_{\mu}^{l}\hat{P}_{\nu}^{m}. \tag{1.1}$$

The scalar ϵ_{μ}^{l} is the energy of a noninteracting system in the state μ, and $W_{\mu\nu}^{lm}$ is an intersystem-interlevel interaction constant, with $W_{\mu\nu}^{ll}=0$. The operator $\hat{P}_{\alpha}^{l}=\hat{L}_{\alpha\alpha}^{l}$ is a diagonal SBO, which are in general defined in terms of state vectors $|\alpha l)$ of system l in state α by

$$\hat{L}_{\alpha\beta}^{l}=|\alpha,l)(\beta,1|. \tag{1.2}$$

The operators $\hat{P}_{\mu}^{l}$ form a complete set of mutually orthogonal projection operators, and the general SBO have a simple commutator relation,[7,8]

$$[\hat{L}_{\alpha\alpha'}^{l},\hat{L}_{\beta'\beta}^{m}]=\delta^{l,m}[\delta_{\alpha'\beta'}\hat{L}_{\alpha\beta}^{l}-\delta_{\alpha\beta}\hat{L}_{\beta'\alpha'}^{l}]. \tag{1.3}$$

The equation of motion for a SBO in the Heisenberg representation follows readily from Eqs. (1.1) and (1.3). Choosing an arbitrary system labelled 0, we have that

$$i\frac{d}{dt}\hat{L}_{\alpha\alpha'}^{0}(t)=[\hat{L}_{\alpha\alpha'}(t),\hat{H}]=-(\hat{Z}_{\alpha}^{0}-\hat{Z}_{\alpha'}^{0})\hat{L}_{\alpha\alpha'}^{0}(t), \tag{1.4}$$

where

$$\hat{Z}_{\alpha}^{0}=\epsilon_{\alpha}^{0}+\tfrac{1}{2}\sum_{\mu,l}(W_{\alpha\mu}^{0l}+W_{\mu\alpha}^{l0})\hat{P}_{\mu}^{l}, \tag{1.5}$$

and

$$\hat{L}_{\alpha\alpha'}^{0}(t)=\exp(i\hat{H}t)\hat{L}_{\alpha\alpha'}^{0}(0)\exp(-i\hat{H}t). \tag{1.6}$$

Since the operator $\hat{Z}_{\alpha}^{0}$ contains only diagonal SBO, it commutes with $\hat{H}$ and is thus time independent. Hence, the solution of the equation of motion (1.4) is given immediately by

$$\hat{L}_{\alpha\alpha'}^{0}(t)=\exp[it(\hat{Z}_{\alpha}^{0}-\hat{Z}_{\alpha'}^{0})]\hat{L}_{\alpha\alpha'}(0), \tag{1.7}$$

with together with the transformation equation (1.6) yields the commutation formula

$$\exp(i\hat{H}t)\hat{L}_{\alpha\alpha'}(0)=\exp[it(\hat{Z}_{\alpha}^{0}-\hat{Z}_{\alpha'}^{0})]\hat{L}_{\alpha\alpha'}^{0}(0)\times\exp(i\hat{H}t). \tag{1.8}$$

In principle Eq. (1.8) embodies all the information about the time development of the individual system 0, but its application in calculating correlation functions is in general quite complicated. It should be noted that Eqs. (1.4)–(1.8) are valid for any complex number t; thus Eq. (1.4) may be considered as a complex operator differential equation.

II. CORRELATION FUNCTIONS

In this section exact correlation function relations of SBO's referring to different systems will be derived. The first approach taken is based on the SBO Green functions,[7] and the second on straightforward use of Eq. (1.7). The Green function method to be presented, although circuitous for this problem, has the possibility of being generalized to many-body Hamiltonians containing nondiagonal operators, i.e., interlevel transition operators $\hat{L}_{\alpha\alpha'}$ with $\alpha\neq\alpha'$.

The Green functions employed are the double-time temperature dependent functions,[9] which in the spectral

a)Work supported by the Swiss National Science Foundation, Grant No. 2.403-0.75

 0022-2488/78/1905-1184$1.00

representation satisfy the equation of motion

$$E\langle\langle \hat{A} \mid \hat{B}\rangle\rangle_E = \frac{1}{2\pi}\langle[\hat{A},\ \hat{B}]\rangle + \langle\langle[\hat{A},\hat{H}] \mid \hat{B}\rangle\rangle_E. \tag{2.1}$$

The bracket $<\cdots>$ represents a canonical ensemble average, and $\hat{A}$ and $\hat{B}$ are operators, assumed here to obey Bose statistics. The spectral Green function $\langle\langle \hat{A} \mid \hat{B}\rangle\rangle_E$ is related to the correlation function $\langle \hat{B}\hat{A}\rangle$ through the integral

$$\langle \hat{B}(t')\hat{A}(t)\rangle = i^{-1}\int_{-\infty}^{\infty} dE(\langle\langle \hat{A} \mid \hat{B}\rangle\rangle_{E-i0} - \langle\langle \hat{A} \mid \hat{B}\rangle\rangle_{E+i0}) \times f(E)\exp(iE(t-t'), \tag{2.2}$$

in which

$$f(E) = (\exp(\beta E) - 1)^{-1}, \quad \text{with } \beta = (k_\beta T)^{-1}. \tag{2.3}$$

Making use of Eqs. (1.4) and (2.2) it follows that the SBO spectral Green function equation of motion for the Hamiltonian (1.1) is

$$(E - \lambda^0_{\alpha\alpha'})\langle\langle \hat{P}^k\hat{L}^0_{\alpha\alpha'} \mid \hat{L}^m_{\beta'\beta}\rangle\rangle_E = \frac{1}{2\pi}\langle[\hat{P}^k\hat{L}^0_{\alpha\alpha'}, \hat{L}^m_{\beta'\beta}]\rangle + \langle\langle O^{0,k+1}_{\alpha\alpha'}\hat{P}^{k+1}\hat{L}^0_{\alpha\alpha'} \mid \hat{L}_{\beta'\beta}\rangle\rangle_E, \tag{2.4}$$

with

$$\lambda^0_{\alpha\alpha'} = \epsilon^0_{\alpha'} - \epsilon^0_\alpha \text{ and } O^{0,k}_{\alpha\alpha'} = \tfrac{1}{2}\sum_{\mu_k,l_k}(W^{0l}_{\alpha'\mu_k} + W^{l_k0}_{\mu_k\alpha'} - W^{0l}_{\alpha\mu_k} - W^{l_k0}_{\mu_k\alpha}). \tag{2.5}$$

The operator $\hat{P}^k$ is an arbitrary product of diagonal SBO, excluding the system operator $\hat{P}^0$, having the form

$$\hat{P}^k = \prod_{i=1}^{k}\hat{P}^{l_i}_{\mu_i}, \quad \text{for } k = 1, 2, \cdots, \tag{2.6}$$

with $\hat{P}^0 = 1$.

Since the Hamiltonian contains only diagonal operators, the only nonzero correlation functions are those of diagonal operators. This is simply the manifestation that the ensemble of systems under consideration possesses no inherent dynamics. It follows, using commutator expansion rules and Eq. (1.3), that the inhomogeneous term in Eq. (2.4) reduces to

$$\langle[\hat{P}^k\hat{L}^0_{\alpha\alpha'}, \hat{L}^m_{\beta'\beta}]\rangle = \delta^{0m}\delta_{\alpha\beta}\delta_{\alpha'\beta'}(C^k_\alpha - C^k_{\alpha'}), \tag{2.7}$$

in which C^k is a $k+1$ system SBO correlation function given by

$$C^k_\alpha = \langle P^k\hat{P}^0_\alpha\rangle. \tag{2.8}$$

The equation of motion (2.4) is a recursion formula in the subscript k, whose inhomogeneous term vanishes for all Green functions except $\langle\langle P^k\hat{L}^0_{\alpha\alpha'} \mid \hat{L}^0_{\alpha'\alpha}\rangle\rangle_E$. Iterating Eq. (2.6) for $k = 1, 2, \cdots$ yields an infinite series expansion in powers of the single system excitation energies $\lambda^0_{\alpha\alpha'}$. It is

$$2\pi\langle\langle P^k\hat{L}^0_{\alpha\alpha'} \mid \hat{L}^0_{\alpha'\alpha}\rangle\rangle_E = \sum_{l=0}^{\infty}(E - \lambda^0_{\alpha\alpha'})^{-(l+1)}\prod_{i=k+1}^{l+k}O^{0,i}_{\alpha\alpha'} \times (C^{l+k}_\alpha - C^{l+k}_{\alpha'}), \tag{2.9}$$

with $\prod_{i=k+1}^{k}O^{0,i}_{\alpha\alpha'} = 1$.

It is seen that for given values of the level subscripts, the Green function has a singularity at only one energy $\lambda^0_{\alpha\alpha'}$; thus the integrals necessary to determine the correlation functions are all of the form

$$(2\pi i)^{-1}\int_{-\infty}^{\infty}dE\{(E - \lambda - i0)^{-(l+1)} - (E - \lambda + i0)^{-(l+1)}\}g(E) = (l!)^{-1}\frac{d^l}{dx^l}g(x)\Big|_{x=\lambda} \equiv (l!)^{-1}g^{(l)}(\lambda). \tag{2.10}$$

For evaluation of this integral, see Ref. 10. Substituting Eq. (2.9) into Eq. (2.2), using (2.10), and $\hat{L}^0_{\alpha'\alpha}\hat{L}^0_{\alpha\alpha'} = \hat{P}^0_{\alpha'}$, gives the static correlation function relation

$$C^k_{\alpha'} = \sum_{l=0}^{\infty}(l!)^{-1}f^{(l)}(\lambda^0_{\alpha\alpha'})\prod_{i=k+1}^{l+k}O^{0,i}_{\alpha\alpha'}(C^{l+k}_\alpha - C^{l+k}_{\alpha'}). \tag{2.11}$$

Before proceeding it is convenient at this point to discuss the generalization of the term-by-term integration method introduced here to Hamiltonians containing nondiagonal operators. In the general case, arbitrary SBO Green functions can be expanded into sums of infinite series, each of which is similar to Eq. (2.9). These series contain products of poles at many different combinations of the single-system excitation energies, depending on the energy level structure of the individual systems. Since the argument used to obtain the reduction of the inhomogeneous term Eq. (2.7) does not hold for Hamiltonians containing nondiagonal operators, the Green function series expansions contain different types of static correlation functions constructed from nondiagonal and diagonal SBO's. Applying partial fraction exapnsions to products of unlike poles, one can integrate term-by-term over each pole to obtain correlation function relations, as done in obtaining Eq. (2.11). Since this procedure can be carried out in coordinate space, it is applicable to disordered systems.

Returning to the analysis of Eq. (2.11), it is seen that once again we have a recursion formula in the subscript k. Iterating over $k = 0, 1, 2, \cdots$ to eliminate $C^k_{\alpha'}$ on the right at each step, and carrying out the differentiation to obtain $f^{(l)}(\lambda)$ leads to the expression

$$C^k_{\alpha'} = \exp(-\beta\lambda^0_{\alpha\alpha'})\sum_{l=0}^{\infty}(l!)^{-1}(-\beta)^l\prod_{i=k+1}^{l+k}O^{0,i}_{\alpha\alpha'}C^{l+k}_\alpha. \tag{2.12}$$

Noting the definitions in Eqs. (2.5) and (2.6), the summation is readily performed giving

$$\langle\hat{P}^k\hat{P}^0_{\alpha'}\rangle \equiv \langle\prod_{i=1}^{k}\hat{P}^{n_i}_{\mu_i}\hat{P}^0_{\alpha'}\rangle = \langle\exp[\beta(\hat{Z}^0_\alpha - \hat{Z}^0_{\alpha'})]\prod_{i=1}^{k}\hat{P}^{n_i}_{\mu_i}\hat{P}^0_\alpha\rangle. \tag{2.13}$$

Equation (2.13) relates the joint probability of finding systems $n_1, \ldots, n_k$ in states $\mu_1, \ldots, \mu_k$ and system 0 in state α' to the probability that systems $n_1, \ldots, n_k$ are the same states, but that system 0 is in another state α. The operator $\hat{Z}^0_\alpha$ is defined by Eq. (1.5). Since $\hat{P}^k$ is an arbitrary product of diagonal SBO, excluding $\hat{P}^0_\alpha$, any diagonal operator which does not contain a part acting on the system 0 may be constructed from $\hat{P}^k$. It follows that Eq. (2.13) can be written in the form

$$F^0(\hat{Q}^0) = \langle\hat{Q}^0\exp(\beta\hat{Z}^0_\alpha)\hat{P}^0_\alpha\rangle = \langle\hat{Q}^0\exp(\beta\hat{Z}^0_{\alpha'})\hat{P}^0_{\alpha'}\rangle, \tag{2.14}$$

where $\hat{Q}^0$ is an arbitrary diagonal operator which can act act on any system except system 0. The correlation function F^0 depends on $\hat{Q}^0$, but is independent of the state subscripts α and α' of system 0. The operators $\hat{P}^0_\alpha$

are subject to the normalization condition

$$\sum_{\alpha} \hat{P}^0_{\alpha} = 1. \tag{2.15}$$

Equation (2.14) follows alternatively from the commutation formula (1.8). Setting $\beta = it$, and using $\hat{P}^0_{\alpha'} = \hat{L}^0_{\alpha'\alpha}(t)\hat{L}^0_{\alpha\alpha'}(t)$ and $[\hat{Z}^0_{\alpha}, \hat{H}] = 0$ in the definition of F^0 for any α yields Eq. (2.14).[11]

The normalization condition (2.15) together with Eq. (2.14) define a set of hierachy equations relating correlation functions of system clusters of different orders. Whether or not this set is in general complete in the sense that correlation functions of some given order only can be related remains unproven here. When this reduction to the same order, using the arbitrariness of $\hat{Q}^0$, is possible, the resulting equation can be solved for a given correlation function. An example of such a case is the one-dimensional random Ising chain treated in the following paper.

In the next section an explicit form of Eq. (2.14) is developed for the Ising model.

III. ISING SYSTEMS

The Hamiltonian for an ensemble of interacting Ising spin systems of spin $\hat{\sigma}^l$ is given by

$$\hat{H} = -\sum_l h^l \hat{\sigma}^l - \tfrac{1}{2}\sum_{l,m} J^{l,m}\hat{\sigma}^l\hat{\sigma}^m. \tag{3.1}$$

The local magnetic field strength at spin $\hat{\sigma}^l$ is h^l, and $J^{l,m}$ is the interaction energy between spins $\hat{\sigma}^l$ and $\hat{\sigma}^m$.

The transformation of $\hat{H}$ to SBO representation of the form (1.1) is accomplished through the relations

$$\hat{\sigma}^l = \sum_{\mu} \mu \hat{P}^l_{\mu} = \hat{P}^l_1 - \hat{P}^l_{-1}, \tag{3.2}$$

and

$$\epsilon^l_{\mu} = -\mu h^l, \quad W^{lm}_{\mu\nu} = -J^{lm}\mu\nu, \quad \text{for } \mu, \nu = \pm 1. \tag{3.3}$$

Using Eqs. (3.2) and (3.3) the operator $\hat{Z}^0_{\alpha}$ defined by Eq. (1.5) is

$$\hat{Z}^0_{\alpha} = -\alpha\hat{Z}^0 = -\alpha\{h^0 + \tfrac{1}{2}\sum_l (J^{0l} + J^{l0})\hat{\sigma}^l\}, \quad \text{for } \alpha = \pm 1. \tag{3.4}$$

Eliminating $\hat{P}_{\pm 1}$ from Eq. (2.14) with the help of Eq. (3.2) and the normalization condition, yields the following expression,

$$\langle \hat{Q}^0(\hat{\sigma}^0 \cosh\beta\hat{Z}^0 - \sinh\beta\hat{Z}^0)\rangle = 0. \tag{3.5}$$

If $\hat{Q}^0$ is chosen to be some product of spin operators, Eq. (3.5) relates spin correlation functions of different orders. Although Eq. (3.5) is in an altered form, it is analogous to the equilibrium spin correlation relations derived from the master equation.[4]

An expression useful for explicit calculation is obtained from Eq. (3.5) by introducing the expansion

$$\exp(\beta J^{0l}\hat{\sigma}^l) = \cosh\beta J^{0l}(1 + t_l\hat{\sigma}^l), \quad \text{with } t_l = \tanh\beta J^{0l}. \tag{3.6}$$

This expression for lattices having inversion symmetry is

$$\langle \hat{Q}^0(1 - V_0\hat{\sigma}^0)\{\hat{\sigma}^0\tau(\text{even}) - \tau(\text{odd})\}\rangle = 0 \tag{3.7}$$

with the operator $\hat{\tau}$ given by

$$\hat{\tau} = \prod_l (1 + t_l\hat{\sigma}^l) \quad \text{and} \quad V_0 = \tanh\beta h^0. \tag{3.8}$$

The "even" and "odd" notation means that after the product in $\hat{\tau}$ is carried out, only terms containing an even, or an odd number of operators is kept, respectively.

IV. CONCLUSIONS

The exact correlation function hierarchy equations (2.14) have not been solved for dimension $d > 1$; however they can be a useful tool in evaluating higher order correlation functions once lower order ones are known. This has been done previously for the $d = 2$ Ising model.[5] Under certain factorization assumptions, Eqs. (2.14) are solvable, e.g., the simplest is the molecular field approximation in which $\hat{Z}^0_{\alpha}$ is replaced by its thermal average value $\langle \hat{Z}^0_{\alpha}\rangle$.

In constrast to the $d > 1$ cases, the hierachy equations have an exact analytic solution for a random Ising chain in a magnetic field. This solution is given in the following paper.

The method of term-by-term integration introduced in Sec. II is quite general, and it is applicable to SBO Green function expansions for Hamiltonians containing nondiagonal operators. The resulting equations relating correlation functions are then approximated by some appropriate factorization scheme and solved by strictly algebraic manipulation.

ACKNOWLEDGMENT

The author wishes to thank Dr. Hans R. Tschudi for a number of stimulating discussions and for reading the manuscript.

[1] B.M. McCoy, C.A. Tracy, and T.T. Wu, Phys. Rev. Lett. 38, 793 (1977).

[2] B.M. McCoy and T.T. Wu, *The Two Dimensional Ising Model* (Harvard U.P. Cambridge, Mass., 1973); *Phase Transitions and Critical Phenomena*, Vol. 1, edited by C. Domb and M.S. Green (Academic, London, 1972); H.E. Stanley, *Introduction to Phase Transitions and Critical Phenomena* (Clarenden, Oxford, 1971).

[3] R.J. Glauber, J. Math. Phys. 4, 294 (1963).

[4] M. Suzuki and R. Kubo, J. Phys. Soc. Jpn. 24, 51 (1968).

[5] H. Yahata and M. Suzuki, J. Phys. Soc. Jpn. 27, 1421 (1969).

[6] B. Kaufman and L. Onsager, Phys. Rev. 76, 1224 (1949).

[7] S.B. Haley and P. Erdös, Phys. Rev. B 5, 1106 (1972).

[8] J. Hubbard, Proc. R. Soc. A285, 542 (1965).

[9] D.N. Zubarev, Usp. Fiz. Nauk 71, 71 (1960) [Sov. Phys. Usp. 3, 320 (1960)].

[10] I.M. Guelfand and G.E. Chilov, *Les Distributions* (Dunod, Paris, 1964), Tome 1.

[11] I am indebted to Dr. Hans Tschudi, at this institute, for pointing out this simple alternate derivation of Eq. (2.14).

II. Exact solutions of disordered Ising spin chains in a magnetic field[a)]

Stephen B. Haley

Institut de Physique Théorique, Université de Lausanne, CH-1015 Lausanne-Dorigny, Switzerland
(Received 27 June 1977)

The one-dimensional Ising model with arbitrary nearest neighbor interactions and a magnetic field is formulated in terms of the spin correlation function relations derived in the preceding article I. The local magnetization and the two-spin correlation functions are shown to satisfy a set of coupled algebraic summation equations. The solution of this set is given in terms of a power series in hyberbolic tangents of the magnetic field strength, for chains of arbitrary length. For cases in which the interaction constants are random variables, explicit expressions are derived for the configurational averaged magnetization and the two-spin correlation functions to the lowest order in the magnetic field.

I. INTRODUCTION

The Ising model in one dimension with nearest neighbor exchange forces does not exhibit a phase transition at a nonzero temperature,[1] and as such, it offers no insight into the understanding of critical phenomena. However, this model has often been the basis for understanding and developing theoretical techniques for use in higher dimensions. In particular, the "transfer matrix" method,[2] was the basis for the solution of the two-dimensional Ising model.[3] This method has also been used to solve the random Ising chain for narrow distributions of the interaction constants.[4] Previous methods concentrate on the direct analysis of the partition function. In this paper a completely different approach is taken.

The one-dimensional Ising model is solved exactly by manipulation of the expressions developed in the preceding article which relate correlation functions of different orders. In the next section a set of summation equations which couple the local magnetization and two spin correlation functions is derived. In the following sections this set is solved for the special cases of a homogeneous chain in a magnetic field, a random chain in the absence of an external field, and finally for the general case of chain of arbitrary exchange interactions in a magnetic field.

II. DEVELOPMENT OF THE BASIC EQUATIONS

The Hamiltonian for a chain of $N+1$ Ising spins $\hat{\sigma}^l$ coupled to a neighbor spin $\hat{\sigma}^{l+1}$ by an arbitrary exchange constant J^{l+1} is given by

$$\hat{H} = -h\sum_{l=0}^{N}\hat{\sigma}^l - \sum_{l=0}^{N} J^{l+1}\hat{\sigma}^l\hat{\sigma}^{l+1}. \tag{2.1}$$

The parameter h is proportional to a homogeneous external magnetic field.

From the preceding paper I, Eq. (I.3.7), constructed from the model Hamiltonian (2.1), for an arbitrary spin site n is

$$\langle\hat{Q}(1-V\hat{\sigma}^n)[\hat{\sigma}^n(1+t_n t_{n+1}\hat{\sigma}^{n-1}\hat{\sigma}^{n+1}) - (t_n\hat{\sigma}^{n-1}+t_{n+1}\hat{\sigma}^{n+1})]\rangle = 0, \tag{2.2}$$

where

$$V=\tanh\beta h,\quad t_n=\tanh\beta J^n \text{ with } \beta=(k_BT)^{-1}. \tag{2.3}$$

Choosing the arbitrary operator $\hat{Q}$ to be $\hat{q}$ times 1, $\hat{\sigma}^{n+1}$, $\hat{\sigma}^{n-1}$, and $\hat{\sigma}^{n-1}\times\hat{\sigma}^{n+1}$ successively, where $\hat{q}$ is another arbitrary diagonal operator, excluding $\hat{\sigma}^n$, generates four equations from Eq. (2.2). Using these four equations, which are not all independent, to eliminate the correlation functions $\langle\hat{q}\hat{\sigma}^{n-1}\hat{\sigma}^n\hat{\sigma}^{n+1}\rangle$ and $\langle\hat{q}\hat{\sigma}^{n-1}\hat{\sigma}^{n+1}\rangle$ yields the following independent, but coupled equations:

$$\xi_n q - V(\alpha_n X_{n-1} + \xi_n X_n - \alpha_{n+1}X_{n+1}) + \alpha_n Y_{n-1} - \alpha_{n+1}Y_n = 0 \quad \text{for } n=1,2,\cdots, \tag{2.4}$$

and

$$V(\gamma_n+\xi_n)q + (1-V^2)\alpha_n X_{n-1} - (\gamma_n+\xi_n V^2)X_n + (1+V^2)\alpha_{n+1}X_{n+1} - 2V\alpha_{n+1}Y_n = 0. \tag{2.5}$$

The coefficients in Eqs. (2.4) and (2.5) are

$$\gamma_n=(1-t_n^2t_{n+1}^2)(\eta_n\eta_{n+1})^{-1},\quad \alpha_n=t_n\eta_n^{-1},$$
$$\xi_n=\eta_{n+1}^{-1}-\eta_n^{-1},\quad \text{with } \eta_n=1-t_n^2, \tag{2.6}$$

and the correlation functions are

$$q=\langle\hat{q}\rangle,\quad X_n=\langle\hat{q}\hat{\sigma}^n\rangle,\quad \text{and } Y_n=\langle\hat{q}\hat{\sigma}^n\hat{\sigma}^{n+1}\rangle. \tag{2.7}$$

Equation (2.4) is a recursion relation in Y_n. Iterating Eq. (2.4) over $n=1,2,\cdots$ to determine Y_n as a function of Y_0 and X_n gives from Eq. (2.5)

$$VD(q,Y_0) + V^2\{-D(X_0,X_1) + \sum_{k=1}^{n} r_n(X_{k-1}-X_k) + g_n(X)\} - g_n(X) = 0, \quad \text{for } n=1,2,\cdots, \tag{2.8}$$

where

$$D(a,b)=(r_1-2\alpha_1)a - 2\alpha_1 b,\quad r_k=(1+t_k)(1-t_k)^{-1}, \tag{2.9}$$

and

$$g_n(X) = -\alpha_n X_{n-1} + \gamma_n X_n - \alpha_{n+1}X_{n+1}. \tag{2.10}$$

Equation (2.4) is the basic summation equation which couples correlation functions of two orders in the spins, depending on the choice of the operator $\hat{q}$. The case to be considered here is that for which $\hat{q}=1$ and $\hat{q}=\hat{\sigma}^0$. For this choice of $\hat{q}$, the correlation functions appearing in Eq. (2.8) and defined in Eq. (2.7) are given by

a)Work supported by the Swiss National Science Foundation, Grant No. 2.403-0.75.

 0022-2488/78/1905-1187$1.00

$$\begin{array}{c|c|c} q= & C_0=1 & \sigma_0 \\ \hline Y_0= & C_1 & \sigma_1 \\ \hline X_n= & \sigma_n & C_n \end{array}, \tag{2.11}$$

where $\sigma_n=\langle\hat{\sigma}^n\rangle$ and $C_n=\langle\hat{\sigma}^0\hat{\sigma}^n\rangle$.

Explicitly, using the Table (2.11), Eq. (2.8) is the set

$$VD(C_0,C_1)+V^2\{-D(\sigma_0,\sigma_1)+\sum_{k=1}^{n}r_k(\sigma_{k-1}-\sigma_k)+g_n(\sigma)\} - g_n(\sigma)=0, \tag{2.12}$$

and

$$VD(\sigma_0,\sigma_1)+V^2\{-D(C_0,C_1)+\sum_{k=1}^{n}r_k(C_{k-1}-C_k) +g_n(C)\}-g_n(C)=0.$$

It is seen explicitly in Eq. (2.12) that the presence of a nonzero magnetic field, manifested by $V\neq 0$, couples the local magnetization σ_n to the two-spin correlation function C_n. The general solution of Eq. (2.12) is quite complicated due to the presence of the summations. To obtain a feel for the difficulties involved, we consider in the next two sections the case of a homogeneous chain in a magnetic field, and that of a random chain in the absence of a magnetic field. The important simplification occurring in both these cases is that the summation does not appear in Eq. (2.12).

III. HOMOGENEOUS CHAIN IN A MAGNETIC FIELD

The homogeneous chain of spins is characterized by one exchange constant J; thus the subscript denoting a spin position in the chain may be dropped from all quantities in Eq. (2.12) except from two-spin correlation functions. Noting in particular that $\sigma_n=\sigma$ and $r_n=r$ for all n, Eq. (2.12) reduces to

$$\sigma=V[1+2\alpha(r-4\alpha)^{-1}(1-C_1)], \tag{3.1}$$

and

$$Hr+\alpha C_{n-1}-(r-2\alpha+Hr)C_n+\alpha C_{n+1}=0, \tag{3.2}$$

where

$$H=V^2(1-V^2)^{-1}=\sinh^2\beta h. \tag{3.3}$$

In order to solve Eqs. (3.1) and (3.2), one must first solve Eq. (3.2) for C_n. Since Eq. (3.2) contains the constant term Hr, we assume the solution of an infinite chain to have the form

$$C_n=A^2+(1-A^2)\phi^n, \tag{3.4}$$

which is normalized such that $C_0=1$. Substituting Eq. (3.4) into Eq. (3.2) defines A and yields ϕ as the solution of a quadratic equation. Using the result to eliminate C_1 from Eq. (2.1) gives $A=\sigma$, and the complete solution,

$$\sigma=\left\{H\left[H+\left(\frac{1-t}{1+t}\right)^2\right]^{-1}\right\}^{1/2} = \sinh\beta h[\sinh^2\beta h+\exp(-4\beta J)]^{-1/2}, \tag{3.5}$$

and

$$C_n=\sigma^2+(1-\sigma^2)\left(\frac{1-V/\sigma}{1+V/\sigma}\right)^n \quad \text{for } n=0,1,\cdots. \tag{3.6}$$

Expression (3.5) for the magnetization per spin is that originally derived by Ising.[1] The correlation functions C_n can also be derived using the transfer matrix technique[5]; however the nice form of Eq. (3.6) does not appear to have been given previously. From Eq. (3.6) it immediately follows that $C_\infty=\sigma^2$ for any value of the magnetic field and that $C_n=t^n$ in the absence of a magnetic field. Thus long range order is only present when the magnetic field is nonzero.

IV. NONHOMOGENEOUS CHAIN IN THE ABSENCE OF A MAGNETIC FIELD

When the magnetic field is absent $V=0$, and Eq. (2.12) reduces to the uncoupled equations

$$g_n^{(0)}(X)=-\alpha_nX_{n-1}^{(0)}+\gamma_nX_n^{(0)}-\alpha_{n+1}X_{n+1}^{(0)}=0, \quad \text{for } n=1,2,\cdots, \tag{4.1}$$

with $X_n=\sigma_n^{(0)}$, or $X_n^{(0)}=C_n^{(0)}$, and coefficients given in Eq. (2.6). The superscript (0) is introduced to signify that these correlation functions are the zero field functions.

Iterating Eq. (4.1) over n and using the details in Appendix A leads to the equation

$$\alpha_1\lambda_{1,n}X_0^{(0)}-\alpha_1\lambda_{0,n}X_1^{(0)}+\tau_{1,n}X_n^{(0)}=0 \quad \text{for } n=1,2,\cdots, \tag{4.2}$$

where $\tau_{l,n}$ and $\lambda_{l,n}$ are defined by

$$\tau_{l,n}=\prod_{k=l}^{n}t_k, \quad \text{with} \quad \tau_{n+1,n}=1, \tag{4.3}$$

and

$$\lambda_{l,n}=\tau_l-\tau_{l+1,n}\tau_n, \quad \text{with } \tau_n=\tau_{1,n}. \tag{4.4}$$

The complete solution of Eq. (4.2) is obtained by introducing an appropriate set of boundary conditions. The conditions usually applied are cyclic or open chain. The more easily applied cyclic boundary conditions will be assumed here. These are expressed by the equations

$$\begin{aligned} &\text{(a)}\ X_{N+1}=X_0, \quad \text{for } X_n=\sigma_n, \quad \text{and } X_n=C_n, \\ &\text{(b)}\ X_{N+2}=X_1, \quad \text{for } X_n=\sigma_n, \quad \text{only.} \end{aligned} \tag{4.5}$$

The superscript is dropped in Eq. (4.5), since these conditions hold in the presence of a magnetic field. The second boundary condition (4.5b) cannot be applied when $X_n=C_n$ because this would violate the original restriction that $n\neq 0$ when $\hat{q}=\hat{\sigma}^0$. If Eq. (4.5b) were applied for $X_n=C_n$ it would lead to a contradiction, since C_n is uniquely determined by Eq. (4.5a) and $C_0=1$. The equation for σ_n, however, is homogeneous and its solution requires two boundary conditions. Application of Eq. (4.5a) to Eq. (4.2) gives

$$X_n^{(0)}=(1+\tau_{N+1})^{-1}(\tau_n+\tau_{n+1,N+1})X_0^{(0)} \quad \text{for } n=0,1,\cdots. \tag{4.6}$$

The local magnetization $X_n^{(0)}=\sigma_n^{(0)}$ is obtained by applying the second boundary condition (4.5b) to Eq. (4.2), and solving for σ_0 using Eq. (4.6). The result, as expected from spin flip symmetry in zero magnetic field, is that $\sigma_n^{(0)}=0$ for all n.

The two-spin correlation functions $C_n^{(0)}$ are immediately given by Eq. (4.6) using $C_0=1$. They are

$$C_n^{(0)} = (1+\tau_{N+1})^{-1}(\tau_n + \tau_{n+1,N+1}) \quad \text{for } n = 0, 1, \ldots, N. \tag{4.7}$$

In the thermodynamic limit $N \to \infty$, the products τ_{N+1}, $\tau_{n+1,N+1}$ with $n \ll N$ tend to zero for nonzero temperatures; thus

$$\lim_{N\to\infty} C_n^{(0)} = \tau_n. \tag{4.8}$$

This limit is derived and discussed by Stanley.[6] The same result follows for an open chain by assuming $C_N^{(0)} \to 0$ for $N \to \infty$ in Eq. (4.2). If the exchange constants J^n are random variables, the configurational average value of the correlation functions in the thermodynamic limit is $\langle C_n^{(0)} \rangle_c = \langle t \rangle_c^n$. This is identical to the same expression as that of a homogeneous chain with exchange J whose hyperbolic tangent $t = \tanh\beta J = \langle t \rangle_c$. For homogeneous chains, the finite chain Eq. (4.7) reduces to the previously derived expression for $C_n^{(0)}$.[5] It should be noted that for values of $\beta J \gtrsim 4$, the $\tanh\beta J \approx 1$ and the values of $C_n^{(0)}$ for a finite number of spins $N \leq 10^4$ are quite different from the thermodynamic limit values. This difference is the result of the choice of cyclic boundaries, and its presence should serve a warning against the use of cyclic boundary conditions to explain physical phenomena in finite systems, such as macromolecules containing 10^3—10^4 atoms. In contrast, the application of open chain boundary conditions does not present this problem, since for $N \gtrsim 10^2$ the thermodynamic limit is essentially attained.

V. INHOMOGENEOUS CHAIN IN A MAGNETIC FIELD

The solution of Eq. (2.8) in the general case is complicated by the presence of the summations. Although direct iteration of this equation is possible (see Appendix B), evaluation of the resulting recursion relations for coefficients of the correlation functions X_0 and X_1, has not been achieved. Noting that the difficult summation term is multiplied by $V^2 = \tanh^2\beta h$ we seek solutions of Eq. (2.8) in the form of a power series in V, i.e., it is assumed that

$$X_n = \sum_{l=0}^{\infty} X_n^{(l)} V^l \quad \text{and} \quad D = \sum_{l=0}^{\infty} D^{(l)} V^l. \tag{5.1}$$

Substituting these series into Eq. (2.8) yields a set of algebraic recursion relations for the functions $X_n^{(l)}$ of order l in terms of the same functions of lower orders. This is

$$g_n^{(l)}(X) - F_n^{(l)}(X) = 0 \quad \text{for } n = 1, 2, \cdots \text{ and } l = 0, 1, \cdots, \tag{5.2}$$

where $g_n(X)$ is defined by Eq. (2.10) and $F_n^{(l)}(X)$ is

$$F_n^{(2l)}(X) = \sum_{j=1}^{l} G_n^{(2j)}(X), \quad F_n^{(2l-1)} = \sum_{j=0}^{l-1} G^{(2j+1)}(X),$$

with

$$G_n^{(l)}(X) = D^{(l-1)}(q, Y_0) - D^{(l-2)}(X_0, X_1) + \sum_{k=1}^{n} r_k [X_{k-1}^{(l-2)} - X_k^{(l-2)}]. \tag{5.3}$$

If the order superscript of any function in (5.3) is negative, that function is defined to be zero.

Iteration of Eq. (5.2) over spin site label n leads to (see Appendix A) the equation

$$\sum_{k=1}^{n-1} \lambda_{k,n} F_k^{(l)}(X) + \alpha_1 \lambda_{1,n} X_0^{(l)} - \alpha_1 \lambda_{0,n} X_1^{(l)} + \tau_n X_n^{(l)} = 0, \tag{5.4}$$

with τ_n and $\lambda_{l,n}$ defined by Eq. (4.3) and (4.4) respectively. For correlation functions $X_n = \sigma_n$ and $X_n = C_n$, Eq. (5.4) is a coupled set of equations, the coupling occurring in $D^{(l)}$ in (5.3).

Applying the cyclic boundary conditions (4.5) to Eq. (5.4) with $X_n^{(l)} = \sigma_n^{(l)}$ and $X_n^{(l)} = C_n^{(l)}$ yields the local magnetization

$$\sigma_n^{(l)} = f_n^{(l)}(\sigma) \quad \text{for } n = 0, \ldots, N, \tag{5.5}$$

and two-spin correlation functions

$$C_n^{(l)} = C_n^{(0)}[\delta^{l,0} - f_0^{(l)}(C)] + f_n^{(l)}(C) \quad \text{for } n = 0, 1, \ldots, N. \tag{5.6}$$

The function $f_n^{(l)}(X)$ is defined as

$$f_n^{(l)}(X) = (1-\tau_{N+1})^{-1}\left\{\sum_{k=1}^{n}(\tau_{k+1,n} + \tau_k\tau_{n+1,N+1})F_k^{(l)}(X) + \sum_{k=n+1}^{N+1}(\tau_{n+1,k} + \tau_n\tau_{k+1,N+1})F_k^{(l)}(X)\right\}. \tag{5.7}$$

The functions $F_k^{(l)}(X)$ are determined by substituting $\sigma_n^{(l)}$ and $C_n^{(l)}$ from Eqs. (5.5) and (5.6) into $G(X)$ in Eq. (5.3). The result is the following complicated set of coupled recursion formulas:

$$\begin{aligned}
\text{(a) } F_k^{(2l)}(C) &= \omega_-\omega_+^{-1}\left[-1 + \omega_-^{-1}\sum_{m=1}^{k}(1+t_m)(\tau_{m-1} - \tau_{m+1,N+1})\right] \\
&+ \sum_{j=1}^{l}\left\{-\omega_-^{-1}\sum_{m=1}^{N+1}(\tau_m - \tau_{m+1,N+1})F_m^{(2j-1)}(\sigma)\right. \\
&+ (\omega_-\omega_+)^{-1}\left[2\sum_{m=1}^{N}(\tau_m - \tau_{m+1,N+1}) - \sum_{l=1}^{k}(1+t_l)\right. \\
&\times (\tau_{l-1} - \tau_{l+1,N+1})\left.\sum_{m=1}^{N+1}(\tau_m + \tau_{m+1,N+1})\right]\cdot F_m^{(2j-2)}(C) \\
&+ \omega_-^{-1}\sum_{l=1}^{k}(1+t_l)\cdot\left[\sum_{m=1}^{l-1}(\tau_{m+1,l-1} - \tau_m\tau_{l+1,N+1})\right. \\
&\left.\left.- \sum_{m=l}^{N+1}(\tau_{l+1,m} - \tau_{m+1,N+1}\tau_{l-1})\right]\cdot F_m^{(2j-2)}(C)\right\},
\end{aligned} \tag{5.8}$$

$$\begin{aligned}
\text{(b) } F_k^{(2l-1)}(\sigma) &= \omega_-\omega_+^{-1} + \sum_{j=0}^{l-1}\left\{-2\omega_+^{-1}\sum_{m=1}^{N}(\tau_m - \tau_{m+1,N+1})F_m^{(2j)}(C)\right. \\
&+ \omega_-^{-1}\sum_{m=1}^{N+1}(\tau_m - \tau_{m+1,N+1})F_m^{(2j-1)}(\sigma) \\
&+ \omega_-^{-1}\sum_{l=1}^{k}(1+t_l)\cdot\left[\sum_{m=1}^{l-1}(\tau_{m+1,l-1} - \tau_m\tau_{l+1,N+1})\right. \\
&\left.\left.- \sum_{m=l}^{N+1}(\tau_{l+1,m} - \tau_{m+1,N+1}\tau_{l-1})\right]\cdot F_m^{(2j-1)}(\sigma)\right\},
\end{aligned}$$

where

$$\omega_\pm = 1 \pm \tau_{N+1}.$$

Equations (5.5)—(5.8) completely define the local magnetization σ_n and two spin correlation functions C_n for an inhomogeneous chain in a magnetic field h, expressed in the form

$$\sigma_n = \sum_{l=1}^{\infty} \sigma_n^{(2l-1)} V^{2l-1} \quad \text{and} \quad C_n = \sum_{l=0}^{\infty} C_n^{(2l)} V^{2l}, \tag{5.9}$$

with $V = \tanh\beta h$.

In view of the complexity of the expression (5.8) there appears to be little hope of obtaining closed expressions for $F_k^{(l)}(X)$, and hence for σ_n and C_n, for arbitrary distributions of interaction constants J^n. Exact expressions for first field dependent terms in the thermodynamic limit are given by

$$\sigma_n^{(1)} = \sum_{k=1}^{n} \tau_{n+1,k} + \sum_{k=n+1}^{N+1} (\tau_{n+1,k} + \tau_n \tau_{k+1,N+1}), \tag{5.10}$$

and

$$C_n^{(2)} = \left(\sum_{k=1}^{n} \tau_{k+1,n} + \sum_{k=n+1}^{N} \tau_{n+1,k} - \tau_n \sum_{k=1}^{N} \tau_k\right) \cdot \left(\sum_{m=1}^{k-1} - \sum_{m=k+1}^{N+1}\right) \times (\tau_m - \tau_{m+1,N+1}). \tag{5.11}$$

From these expressions one is in position to calculate the configurational average zero field magnetic susceptibility and energy per spin to second order in the field for any given distribution of exchange constants J^n. This can be done explicitly for random distributions, since in this case $\langle \tau_{l,n} \rangle_c = \langle t \rangle_c^{n-l+1}$ where $\langle \cdots \rangle_c$ denotes configurational average. The averages $\langle \sigma_n^{(1)} \rangle_c$ and $\langle C_n^{(2)} \rangle_c$ are arithmetico-geometric progressions which in the limit $N \to \infty$ yield the expressions

$$\langle \sigma_n^{(1)} \rangle_c = (1 + \langle t \rangle_c)(1 - \langle t \rangle_c)^{-1} \quad \text{for all } n, \tag{5.12}$$

and

$$\langle C_n^{(2)} \rangle_c = -n \langle t \rangle_c^{n+1} [(1 - \langle t \rangle_c)^{-1} + (\langle t \rangle_c - \langle t^2 \rangle_c)^{-1} + \langle t \rangle_c \times \left(\frac{1 - \langle t \rangle_c^n}{(1 - \langle t \rangle_c)^2} + \frac{(\langle t \rangle_c^n - \langle t^2 \rangle_c^n}{(\langle t \rangle_c - \langle t^2 \rangle_c)^2} \langle t^2 \rangle_c\right) + \frac{1 + \langle t \rangle_c}{1 - \langle t \rangle_c} \left(\frac{1 - \langle t \rangle_c^{n+1}}{1 - \langle t \rangle_c} - \frac{\langle t \rangle_c^{n+1} - \langle t^2 \rangle_c^{n+1}}{\langle t \rangle_c - \langle t^2 \rangle_c}\right) + \frac{(1 - \langle t^2 \rangle_c^n}{(1 - \langle t \rangle_c)^2} \langle t \rangle_c^2. \tag{5.13}$$

From Eq. (5.13) it follows that in the limit $n \to \infty$ that $\langle C_\infty^{(2)} \rangle_c = \langle \sigma_n^{(1)} \rangle_c^2$; thus it is seen from Eq. (5.12) that for ferromagnetic exchange $J^n > 0$ there is always a temperature low enough such that long range order is present in any nonzero magnetic field. It should be noted that $\tanh\beta J^n$ approaches its maximum value 1 very rapidly for $\beta J^n \gtrsim 3$. In contrast, for antiferromagnetic exchange $J^n < 0$ long-range order is absent, except at zero temperature. This same behavior is exhibited in homogeneous chains, as may be shown from Eq. (3.5).

The zero field average magnetic susceptibility $\langle \chi^{(0)} \rangle_c$ calculated from Eq. (5.12) is

$$\langle \chi^{(0)} \rangle_c = \beta \mu_B^2 (1 + \langle t \rangle_c)(1 - \langle t \rangle_c)^{-1}. \tag{5.14}$$

This expression, which also can be calculated from $\langle C_n^{(0)} \rangle_c$ in Eq. (4.8) using the fluctuation-dissipation theorem, is identical with the homogeneous chain susceptibility with t replaced by $\langle t \rangle_c$.[6]

As a final result, the configurational average energy per spin, $\langle \epsilon \rangle_c = -\langle \sigma \rangle_c h - \langle J C_1 \rangle_c$, calculated to second order in the magnetic field using Eqs. (5.11) and (5.12) is given by the expression

$$\langle \epsilon \rangle_c = -\langle Jt \rangle_c - [\beta^{-1} + \langle J \rangle_c - (\beta^{-1} \langle t \rangle_c^2 + \langle Jt^2 \rangle_c)] \times \left(\frac{\beta h}{1 - \langle t \rangle_c}\right)^2. \tag{5.15}$$

VI. CONCLUSIONS

The exact expressions give here for the local magnetization and two-spin correlation functions in a disordered Ising chain in the presence of a magnetic field appear to have only academic value. Although there exist magnetic materials in which the magnetic ions may be considered to form linear chains,[7] the Ising model is not applicable.

An important elucidation arising in the method presented here is the explicit manifestation of the close connection between the magnetization and the two-spin correlation functions when a magnetic field is present. Not only was the usual relationship between the magnetization and the infinite distance correlation function exhibited, but it was also shown that the local magnetization is intricately coupled to all the two-spin correlation functions through the nearest neighbor correlation function. Since coupling of this genre is also present in higher dimensions, it may be construed as an indication of why the problem of the Ising square lattice in a magnetic field is so complex that it has not yet been solved.

APPENDIX A

Using the definition of $g_n(X)$ in Eq. (2.10), Eq. (5.2) is explicitly

$$\alpha_n X_{n-1}^{(l)} - \gamma_n X_n^{(l)} + \alpha_{n+1} X_{n+1}^{(l)} + F_n^{(l)}(X) = 0 \quad \text{for } n = 1, 2, \cdots, \tag{A1}$$

Iteration of (A1) gives

$$\alpha_1^2 S_{2,n} X_0^{(l)} - \alpha_1 S_{1,n} X_1^{(l)} + P_n X_n^{(l)} + R_n^{(l)}(X) = 0, \tag{A2}$$

where $S_{l,n}$ and $R_n^{(l)}$ are functions satisfying the following recursion formulas:

$$S_{l,n+1} = \gamma_n S_{l,n} - \alpha_n^2 S_{l,n-1} \quad \text{with } S_{l,l} = 1, \quad S_{l+1,l} = 0, \tag{A3}$$

and

$$R_{n+1}^{(l)}(X) = \gamma_n R_n^{(l)}(X) - \alpha_n^2 R_{n-1}^{(l)}(X) + P_n F_n^{(l)}(X). \tag{A4}$$

The function P_n is defined by

$$P_n = \tau_n \Lambda_{1,n}^{-1} \quad \text{with } \Lambda_{l,n} = \prod_{k=l}^{n} (1 - t_k^2), \tag{A5}$$

and τ_n is defined in Eq. (4.3).

Iteration of (A3) and (A4) gives

$$S_{l,n} = (1 - \tau_{l,n}^2) \Lambda_{l,n}^{-1} \tag{A6}$$

and

$$R_n^{(l)}(X) = \sum_{k=1}^{n-1} S_{k+1,n} P_k F_k^{(l)}(X). \tag{A7}$$

Using (A6) and (A7) in (A2) yields Eq. (5.4).

APPENDIX B

Direct iteration of Eq. (2.8) gives

$$\alpha_1 A_n(1) B(q, Y_0) + \alpha_1^2 A_n(2) X_0 - \alpha_1 A_n(3) X_1 + P_n X_n = 0 \quad \text{for } n = 1, 2, \cdots, \tag{B1}$$

where

$$B(q, Y_0) = \sqrt{H(1+H)}[(r_1 - 2\alpha_1)q - 2\alpha_1 Y_0] + 2\alpha_1 H(X_0 + X_1),$$

$$H = \sinh^2 \beta h.$$

The functions $A_n(k)$ satisfy the four point recursion formula

$$A_{n+1}(k) = (\gamma_n + \alpha_n + Hr_n)A_n(k) - \alpha_n(\gamma_{n-1} + \alpha_n + Hr_n)A_{n-1}(k) + \alpha_{n-1}^2 \alpha_n A_{n-2}(k), \tag{B2}$$

with initial values

$$\begin{aligned} &A_1(1) = 0, && A_1(2) = 0, && A_0(3) = 0, \\ &A_2(1) = 1, && A_2(2) = 1, && A_1(3) = 1, \\ &A_3(1) = \gamma_2 + \alpha_2 + Hr_2, && A_3(2) = \gamma_2 + Hr_2, && A_2(3) = \gamma_1 + Hr_1. \end{aligned} \tag{B3}$$

When $H = 0$, (B2) reduces to the three point recursion formula (A3) for S_{1n}, but for $H \neq 0$ evaluation of $A_n(k)$ has not been achieved.

[1] E. Ising, Z. Physik **31**, 253 (1925).
[2] H.A. Kramers and G.H. Wannier, Phys. Rev. **60**, 252 (1941).
[3] L. Onsager, Phys. Rev. **65**, 117 (1944).
[4] C. Fan and B.M. McCoy, Phys. Rev. **182**, 614 (1969).
[5] B.M. McCoy and T.T. Wu, *The Two Dimensional Ising Model* (Harvard U.P., Cambridge, Mass., 1973).
[6] H.E. Stanley, *Introduction to Phase Transitions and Critical Phenomena* (Clarenden, Oxford, 1971).
[7] J. Skalyo *et al.*, Phys. Rev. B **2**, 1310 (1970).

Eigenstates of complex linear combinations of J_1, J_2, J_3 for any representation of SU(2)

H. Bacry

U.E.R. Scientifique de Luminy, Marseille, France,
Centre de Physique Théorique, CNRS Marseille, France,
and Physics Department, Technion, Haifa, Israel
(Received 27 May 1977)

The states which minimalize the uncertainty relation $\Delta J_1\,\Delta J_2 \geq (1/2)|\langle J_3\rangle|$ are eigenstates of complex linear combinations of J_1 and J_2 [S. Rushin and Y. Ben-Aryeh, Phys. Lett. A **58**, 207 (1976)]. This kind of state is shown to have a very simple geometrical interpretation in the constellation formalism. A detailed description is given in the present paper.

1. INTRODUCTION

Whenever SU(2) or the rotation group occurs in physics, two kinds of states are utilized, namely the old standard discrete states[1] $|jm\rangle$ with $m=-j,-j+1,-j+2,\ldots,+j'$ and the continuous *spin coherent states* $|j\theta\varphi\rangle$ (or equivalently $|jz\rangle$), where (θ,φ) labels points of the ordinary two-dimensional sphere S_2, and z is the image of (θ,φ) in the stereographic projection of S_2 on the complex line. The spin coherent states have been introduced by Radcliffe[2] and are known in quantum optics as Bloch states. The use of the word *coherent* is justified not only by the close analogy of their mathematical properties[3] with those of the ordinary Schrödinger—Glauber coherent states[4,5] but also by their interpretation in quantum optics.[6,7]

It has been shown in Ref. 8 that any state of spin j can be represented with the aid of S_2 as a geometrical being: a *constellation* of order $2j$. States $|jm\rangle$ and $|jz\rangle$ are only special kinds of constellations ($|jm\rangle$ is of apparent order 1 or 2, $|jz\rangle$ of apparent order 1). The advantage of the constellation concept is to visualize a state as a geometrical object on S_2 and to see very easily how the rotation group acts on it: Just rotate the sphere. Some applications have been made with the aid of this new tool, especially the classification of *kinds* of spin states[8] (that is the classification of orbits of the rotation group on spin states), the quantization of the three-dimensional harmonic oscillator,[9] the relationship between the Radcliffe—Bloch sphere and the Poincaré sphere.[5] Here we present a new application.

The application we are dealing with has been suggested by a recent work of Rushin and Ben-Arieh.[10] These authors started from the property that Radcliffe—Bloch states minimalize the uncertainty relation

$$\Delta J_1\,\Delta J_2 \geq \tfrac{1}{2}\,|\langle J_3\rangle| \tag{1}$$

and generalized this property by proving that all states which minimalize (1) are eigenstates of complex linear combinations of J_1 and J_2. Moreover, any given such linear combination, except $J_1 \pm iJ_2$ has $2j+1$ eigenstates of eigenvalues $-j,-j+1,\ldots,+j$. The operators $J_1 \pm iJ_2$ only have one eigenstate, a Radcliffe—Bloch state. All these properties are quite simple to describe in the constellation formalism. In the next section, we will recall the constellation description of a spin state. In Sec. 3, we will show that any operator of the form $J\cdot F$ (where F is a complex vector, i.e., an element of the Lie algebra of the complex rotation group) can be associated with a constellation of order 2. Finally, in Sec. 4, we will find out the constellation associated with eigenstates of a given element $J\cdot F$.

2. CONSTELLATION DESCRIPTION OF A SPIN STATE

Let us recall here some results of the work[8] where the constellation concept was introduced (the word itself has first been introduced in Ref. 5). It is well known that a spinor *up to a complex factor* can be written as

$$\psi = \begin{vmatrix} 1 \\ z \end{vmatrix} \sim \begin{vmatrix} \lambda \\ \lambda z \end{vmatrix} \quad \text{with} \quad \lambda \in \mathbb{C} - \{0\}, \tag{2}$$

provided we include the possibility of having $z=\infty$, which corresponds to the case $\lambda=0$, $\lambda z \neq 0$. The stereographic projection *from the South pole*[11] associates with ψ the point (θ,φ) on S_2 such that

$$\psi = \begin{vmatrix} \cos\frac{\theta}{2}\exp(-i\varphi/2) \\ \sin\frac{\theta}{2}\exp(i\varphi/2) \end{vmatrix} \sim \begin{vmatrix} 1 \\ \tan\frac{\theta}{2}\exp(i\varphi) \end{vmatrix}. \tag{3}$$

Since a state is defined up to a factor (or, equivalently, normalized and defined up to a phase), any $\frac{1}{2}$-spin state is uniquely characterized by one point on S_2. Such a point will be called a *constellation of order* 1. The states $|{}^1_0|$ (North pole) and $|{}^0_1|$ (South pole) are eigenstates of $J_3=\frac{1}{2}\sigma_3$ with eigenvalues $\frac{1}{2}$ and $-\frac{1}{2}$, respectively.

Any state of spin j is characterized by *a constellation of order* $2j$. By this we mean a set of $2j$ complex numbers $\{z_1,z_2,\ldots,z_{2j}\}$ *not necessarily distinct*.[12] The number of distinct values is called the *apparent order* of the constellation.[12] It is clear that through a stereographic projection, the $2j$ complex numbers are replaced by $2j$ points on S_2. The relationship between constellations of order $2j$ and states of spin j is the following one. Let

$$|\psi\rangle = S_0|j\rangle + S_1|j-1\rangle + S_2|j-2\rangle + \cdots + S_{2j}|-j\rangle \tag{4}$$

be such a state. The constellation associated with $|\psi\rangle$

 0022-2488/78/1905-1192$1.00

is given by the relations:

$$
\begin{aligned}
&S_0 = 1\\
&S_1 = \frac{z_1 + z_2 + z_3 + \cdots + z_{2j}}{\sqrt{2j}},\\
&S_2 = \frac{z_1 z_2 + z_2 z_3 + \cdots + z_{2j-1} z_{2j}}{\sqrt{j\,(2j-1)}},\\
&\text{- - - - - - - - - - - - - - - - -},\\
&S_p = \binom{2j}{p}^{-1/2} \sum (z_{i_1} z_{i_2} \cdots z_{i_p}),\\
&\text{- - - - - - - - - - - - - - - - -},\\
&S_{2j} = z_1 z_2 \cdots z_{2j},
\end{aligned}
\tag{5}
$$

where the summations are made on all combinations. If two (or more) of the z_k's are equal, the constellation is said to be degenerate. In particular, if all the z_k's are equal, the apparent order is equal to one.

3. CONSTELLATION ASSOCIATED WITH J·F

Let us now consider an element of the complex Lie algebra, say $\sigma \cdot F$. Since we are interested in its eigenstates (in the representation of spin j), we do not have to distinguish between $\sigma \cdot F$ and $\lambda \sigma \cdot F$ where λ is an arbitrary nonzero complex number. In other words we are only interested in the *class* $[F]$ of elements $\sigma \cdot F$ such that[13]

$$[F] = \{F \sim \lambda F \mid \lambda \neq 0\} \tag{6}$$

that is *a state of spin* 1 or, equivalently, *a constellation of order* 2. It is clear that constellations of order 2 are of two types, those which are of apparent order 2 (nondegenerate) and those of apparent order equal to 1 (degenerate). In order to classify these constellations, let us associate with $[F]$ the following 2×2 matrix symmetric in z_1, z_2,

$$\sigma[F] = \begin{vmatrix} z_1 + z_2 & -2 \\ 2z_1 z_2 & -(z_1 + z_2) \end{vmatrix} = (z_1 + z_2)\sigma_3 + z_1 z_2 \sigma_- - \sigma_+, \tag{7}$$

$$\det \sigma[F] = (z_1 - z_2)^2. \tag{8}$$

The associated constellation is $\{z_1, z_2\}$. We note that $\det\sigma[F]$ is zero if and only if the constellation is degenerate ($z_1 = z_2$). We also note that if F is real, $\sigma[F]$ is Hermitian (up to a factor) and $\{z_1, z_2\}$ corresponds to opposite points on the sphere: $z_1 \bar{z}_2 + 1 = 0$. The proof is as follows: Take the Hermitian conjugate of (7),

$$
\begin{aligned}
\sigma[F]^+ &= (\bar{z}_1 + \bar{z}_2)\sigma_3 + \bar{z}_1 \bar{z}_2 \sigma_+ - \sigma_-\\
&= \bar{z}_1 \bar{z}_2 \left[\left(\frac{1}{\bar{z}_1} + \frac{1}{\bar{z}_2} \right) \sigma_3 + \sigma_+ - \frac{1}{\bar{z}_1 \bar{z}_2} \sigma_- \right].
\end{aligned}
$$

By comparison with (7), one gets the conditions for $\sigma[F]$ to be Hermitian up to a factor:

$$z_1 + z_2 = -\left(\frac{1}{\bar{z}_1} + \frac{1}{\bar{z}_2} \right), \quad z_1 z_2 = -\frac{1}{\bar{z}_1 \bar{z}_2}.$$

It is a simple matter to prove that it is equivalent to $z_1 \bar{z}_2 + 1 = 0$. The associated constellation will be said to be *real*.[14]

Let us now state the following theorem.

Theorem 1: The eigenstates of $\sigma[F]$ associated with the nondegenerate constellation $\{z_1, z_2\}$ are the constellations $\{z_1\}$ and $\{z_2\}$. The eigenstates are orthogonal (then opposite on the sphere) if and only if $\{z_1, z_2\}$ is real. If $z_1 = z_2$, there is a single eigenstate, namely $\{z_1\}$.

The proof is quite easy. In fact, if $z_1 \neq z_2$

$$\sigma[F] \begin{vmatrix} 1 \\ z_1 \end{vmatrix} = (z_2 - z_1) \begin{vmatrix} 1 \\ z_1 \end{vmatrix}, \quad \sigma[F] \begin{vmatrix} 1 \\ z_2 \end{vmatrix} = -(z_2 - z_1) \begin{vmatrix} 1 \\ z_2 \end{vmatrix},$$

and the eigenstates are orthogonal: $|1\ \bar{z}_1| \begin{vmatrix} 1 \\ z_2 \end{vmatrix} = \bar{z}_1 z_2 + 1 = 0$ if and only if $\sigma[F]$ is Hermitian up to a factor. If now $z_1 = z_2 = z$, the matrix $\sigma[F]$ reads $\begin{vmatrix} z & -1 \\ z^2 & -z \end{vmatrix}$ and has determinant zero. The only eigenstate is $\begin{vmatrix} 1 \\ z \end{vmatrix}$ with eigenvalue zero.

It is of interest to consider the following particular cases:

(a) $z_1 = 0$, $z_2 = \infty$

$$\sigma[F] \approx \frac{1}{1 - z_1/z_2} \begin{vmatrix} 1 + \frac{z_1}{z_2} & -\frac{2}{z_2} \\ 2z_1 & -\left(1 + \frac{z_1}{z_2}\right) \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 0 & -1 \end{vmatrix} = \sigma_3.$$

If we remember that σ_3 generates rotations around the third axis and that *to be an eigenstate of* σ_3 *means to be invariant under a rotation around the third axis*, the fact that $\{0\}$ and $\{\infty\}$ are both invariant constellations under those rotations is quite obvious in Fig. 1. Similarly

(b) $z_1 = 1$, $z_2 = -1$

$$\sigma[F] = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = \sigma_1,$$

(c) $z_1 = i$, $z_2 = -i$

$$\sigma[F] = \begin{vmatrix} 0 & -i \\ i & 0 \end{vmatrix} = \sigma_2.$$

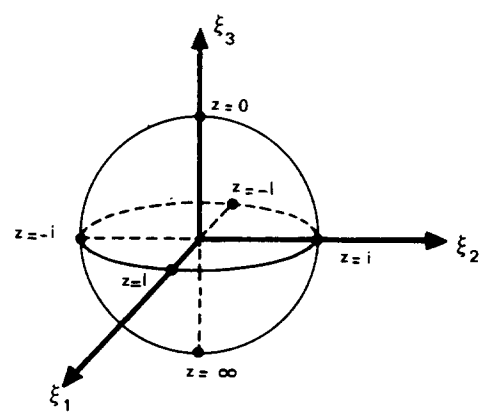

FIG. 1.

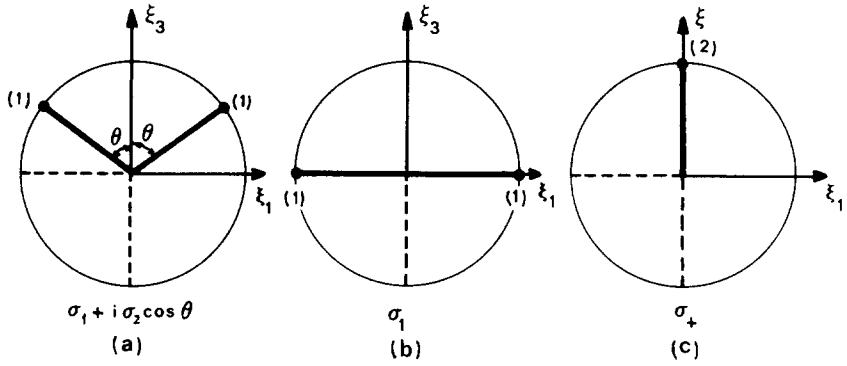

FIG. 2. Constellations $\sigma[\mathbf{F}] = \sigma_1 + \sigma_2\cos\theta$ (multiplicities are in brackets).

Again, in these two cases, the interpretation is immediately obtained from Fig. 1.

It is a simple exercise to generalize the above results: If $\{z_1\}$ and $\{z_2\}$ are opposite on the sphere, they are eigenstates of the generator of rotations around the corresponding diameter, in the spin-$\frac{1}{2}$ representation.

We are now going to generalize the above results for any spin representation.

4. EIGENCONSTELLATIONS OF J · F IN REPRESENTATION OF SPIN j

Theorem 2: Given a constellation $\{z_1, z_2\}$ associated with the matrix $\sigma[\mathbf{F}]$, the eigenconstellations of its representative in the representation of spin j are the constellations (of order $2j$) of the form $\{z_1, z_1, \ldots, z_1, z_2, z_2, \ldots, z_2\}$. If $z_1 \neq z_2$, they are $2j+1$ in number and of apparent order 2 or 1. If $z_1 = z_2$, there is a single constellation of apparent order 1 (Bloch constellation).

This simple result is obtained as follows. Let us denote by $J[\mathbf{F}]$ the representative of $\sigma[\mathbf{F}]$ in the representation of spin j,

$$J[\mathbf{F}] = (z_1 + z_2)J_3 + z_1 z_2 J_- - J_+.$$

It is a simple calculation to verify that the state

$$|\psi_\lambda\rangle = \sum_{m=-j}^{+j} \sum_{k=0}^{j-m} \left[\frac{(j+m)!\,(j-m)!}{(2j)!}\right]^{1/2} \times \frac{(j-\lambda)!\,(j+\lambda)!}{k!\,(j-\lambda-k)!\,(j-m-k)!\,(\lambda+m+k)!} \times z_1^{j-m-k} z_2^k |m\rangle$$

is an eigenstate of $J[\mathbf{F}]$,

$$J[\mathbf{F}]|\psi_\lambda\rangle = \lambda(z_2 - z_1)|\psi_\lambda\rangle.$$

Moreover with a small combinatory calculation, it is not difficult to show that the state $|\psi_\lambda\rangle$ is represented by the constellation $\{z_1, z_1, \cdots, z_1, z_2, \cdots, z_2\}$ where the multiplicity of z_1 is 2λ and the multiplicity of z_2 is $2j - 2\lambda$. Moreover if z_2 becomes equal to z_1, all $|\psi_\lambda\rangle$ collapses in a single state and the corresponding eigenvalue is zero. It is this case which corresponds to Radcliffe—Bloch states.

In order to illustrate our final theorem, let us give the constellations associated with the Rushin—Ben-Aryeh states, i.e., eigenconstellations of linear combinations of J_1 and J_2. Since these states minimalize the uncertainty relations associated with the frame $0\xi_1\xi_2\xi_3$, it is clear that they must have some geometrical relationship with the three axes. The result is represented on Fig. 2. We note that the ξ_3 axis is a symmetry axis for the constellation; the $\xi_1\xi_3$ plane is also a symmetry plane. A greater symmetry is present when

(i) F is real [maximum symmetry, Fig. 2(b)],

(ii) **F** is singular ($\det\sigma[\mathbf{F}] = 0$, Fig. 2(c)].

In the other cases, the axes ξ_1, ξ_2, ξ_3 play *distinct* privileged roles. In case $\theta = 0$ (or π) which is represented in Fig. 2(c), the uncertainty relation is minimal not only for the product $\Delta J_1 \Delta J_2$ but also for any product of the form

$$\Delta(J_1\cos\varphi - J_2\sin\varphi)\Delta(J_1\sin\varphi + J_2\cos\varphi).$$

5. CONCLUSION

Up to now, the Rushin—Ben-Aryeh states have received no interpretation. If we identify the complex rotation group with the Lorentz group, such states can be interpreted as boosted spin states of the type $|jm\rangle$. For example, the state represented in Fig. 2(b) is an eigenstate of J_1 with $j=1$, $m=0$. By boosting it in the ξ_3 direction one gets a state like the one of Fig. 2(a). Such an interpretation is merely a curiosity. There is some hope in finding more concrete applications in quantum optics.

ACKNOWLEDGMENTS

The author is grateful to Dr. Ben-Aryeh, Dr. Rushin, Dr. A. Grossmann, and Dr. J. Zak for illuminating discussions. Thanks are due to the Chairman of the Physics Department at the Technion where part of this work was done.

[1] In quantum optics they are also referred to as *Dicke states*.
[2] J. M. Radcliffe, J. Phys. A 4, 313 (1971).
[3] From the formal point of view, one must underline that the parametrization by a complex number is common to both kinds of states; the only difference is that $z = \infty$ does correspond to a spin coherent state but has no interpretation in ordinary coherent states.

[4]A. M. Perelomov, Commun. Math. Phys. **26**, 222 (1972).
[5]H. Bacry, A. Grossmann, and J. Zak, in *Lecture Notes in Physics*, *Vol.* 50, edited by A. Janner, T. Janssen, and M. Boon (Springer, New York, 1976).
[6]F. J. Arecchi, E. Courtens, R. Gilmore, and H. Thomas, Phys. Rev. A 6, 2211 (1972).
[7]I. R. Senitzky, Phys. Rev. A **15**, 284 (1977).
[8]H. Bacry, J. Math. Phys. **15**, 1686 (1974).
[9]H. Bacry, in "Proceedings of the 2nd Int. Colloquium On Group Theoretical Methods in Physics, edited by A. Janner (University of Nijmegen).
[10]S. Rushin and Y. Ben-Aryeh, Phys. Lett. A **58**, 207 (1976).
[11]The South Pole is chosen in order to recover the usual orientation of the three-dimensional space with the usual Pauli matrices and the usual spherical coordinates on the sphere.
[12]A more sophisticated definition would be as follows: A constellation of order $2j$ is a mapping f of S_2 on $\mathbb{N}_+$ (the set of nonnegative integers) such that $f^{-1}(\mathbb{N}_+ - \{0\})$ is finite and the sum of the images is $2j$. Also see Ref. 5 for another definition. The definition given here is the original one.[8]
[13]We are dealing with the projective complex Lie algebra of $SO(3, C)$.
[14]Note that the conditions for two spin-$\frac{1}{2}$ states to be orthogonal is that the associated constellations are opposite on the sphere. The concept of real constellation will be generalized later.

The projective Lie algebra of the Lorentz group and homographic transformations

H. Bacry

U.E.R. Scientifique de Luminy, Marseille, France
and Centre de Physique Théorique, CNRS, Marseille, France
(Received 21 June 1977)

The Lorentz group is used as the group of homographic transformations on the Riemann sphere. Its Lie algebra is shown to have a very simple interpretation with the aid of cross products and constellation formalism. This property is used to give a constellation description of the Clebsch–Gordan series for the product of two states of spin 1.

A homographic transformation is defined by

$$z' = \frac{az+b}{cz+d}, \tag{1}$$

where z and z' are complex numbers (including the point at infinity). It is labeled by four complex numbers a, b, c, d, satisfying $ad-bc=1$. Each such transformation is known[1] to leave two points invariant (these two points are not necessarily distinct). We will refer to this set of two fixed points as a *constellation of order two*[2]; the constellation is said to be *degenerate* whenever the two fixed points are equal. It is clear that the homographic transformations form a *three-dimensional complex group* G.[3] This group acts transitively on nondegenerate constellations of order two[4] and also transitively on degenerate constellations of order two.[4,2]

The Lie algebra of the group is the three-dimensional complex vector space V. Any two elements of the Lie algebra $\mathbf{f}_1$, $\mathbf{f}_2$ generate the same one-dimensional subgroup if and only if there exists a complex number $\lambda \neq 0$ such that $\mathbf{f}_1 = \lambda \mathbf{f}_2$. It follows that such a subgroup is characterized by a *ray* in V. In other words, there is a one-to-one mapping between the projective space $\hat{v}$ associated with V and the set of one-dimensional complex subgroups of G.

Now, given a ray in $\hat{v}$, say $\mathbf{f}$, one can associate in a unique way a constellation $\{z_1, z_2\}$ in the following way[2]:

$$\sigma(\mathbf{f}) = \begin{vmatrix} z_1+z_2 & -2 \\ 2z_1z_2 & -(z_1+z_2) \end{vmatrix}. \tag{2}$$

The converse is also true. We note that $\sigma(\mathbf{f})$ (defined up to a factor) is a traceless matrix and, therefore, is an element of the Lie algebra of $SL(2,C)$. Moreover it is symmetric with respect to z_1 and z_2, a condition which is necessary to define a constellation.

It is interesting to underline that the constellation $\{z_1, z_2\}$ is the one which is invariant under the subgroup generated by $\sigma(\mathbf{f})$. In fact[5]

$$\sigma(\mathbf{f}) \begin{vmatrix} 1 \\ z_1 \end{vmatrix} = (z_2 - z_1) \begin{vmatrix} 1 \\ z_1 \end{vmatrix}. \tag{3}$$

Let us look for the geometrical interpretation of the commutator $[\sigma(\mathbf{f}), \sigma(\mathbf{f}')]$. It is easy to get

$$[\sigma(\mathbf{f}), \sigma(\mathbf{f}')] = 2(z_1+z_2-z_1'-z_2') \begin{vmatrix} Z_1+Z_2 & -2 \\ 2Z_1Z_2 & -(Z_1+Z_2) \end{vmatrix} \tag{4}$$

with Z_1, Z_2 being the roots of a polynomial of degree two,

$$(z_1+z_2-z_1'-z_2')Z^2 - 2(z_1z_2 - z_1'z_2')Z + (z_1'+z_2')z_1z_2 - (z_1+z_2)z_1'z_2', \tag{5}$$

the roots of which are

$$Z = \frac{z_1z_2 - z_1'z_2' \pm \sqrt{(z_1-z_1')(z_2-z_1')(z_1-z_2')(z_2-z_2')}}{z_1+z_2-z_1'-z_2'}. \tag{6}$$

The constellation $\{Z_1, Z_2\}$ has a simple geometrical interpretation. In fact, a simple calculation leads to the relations

$$\frac{(Z_1-z_1)(Z_2-z_2)}{(Z_1-z_2)(Z_2-z_1)} = \frac{(Z_1-z_1')(Z_2-z_2')}{(Z_1-z_2')(Z_2-z_1')} = -1, \tag{7}$$

which show that the constellation $\{Z_1, Z_2\}$ forms with the constellations $\{z_1, z_2\}$ and $\{z_1', z_2'\}$ two harmonic quadrangles. It is a simple matter to prove that this condition uniquely defines the constellation $\{Z_1, Z_2\}$.

Let us examine the following cases:

(1) The constellations $\{z_1, z_2\}$ and $\{z_1', z_2'\}$ are nondegenerate. Since the cross ratios of Eq. (7) are equal, it is a well-known result that there exists a homographic transformation mapping $\{Z_1, Z_2, z_1, z_2\}$ on $\{Z_1, Z_2, z_1', z_2'\}$, respectively. In other words, there exists a transformation mapping the constellation $\{z_1, z_2\}$ on the constellation $\{z_1', z_2'\}$, generated by the constellation $\{Z_1, Z_2\}$ (or, equivalently, with $\{Z_1, Z_2\}$ as the fixed constellation).

As a peculiar example, we suppose that $\{z_1, z_2\}$ and $\{z_1', z_2'\}$ are real constellations,[2,6] that is, they are both symmetric with respect to the center of the Riemann sphere (diameter constellations). The commutator of these two constellations is the diameter perpendicular to the given constellations. This corresponds to the vector product in the three-dimensional real space.

(2) One of the two constellations, say $\{z_1', z_2'\}$ is degenerate; this means that $z_1' = z_2' = z'$. Then

$$\{Z_1, Z_2\} = \left\{ \frac{2z_1z_2 - (z_1+z_2)z'}{z_1+z_2-2z'}, z' \right\}.$$

(3) Both constellations are degenerate: $z_1 = z_2 = z$, $z_1' = z_2' = z'$. Then $\{Z_1, Z_2\} = \{z, z'\}$.

(4) Suppose that $z_2 = z_2'$. The commutator of $\{z_1, z_2\}$ and $\{z_1', z_2'\}$ is just $\{z_2, z_2\}$.

(5) Whenever $z_1+z_2 = z_1'+z_2'$, Eq. (5) is of degree one. Since we are working with projective spaces, it is

 0022-2488/78/1905-1196$1.00

necessary to consider the value ∞ as a solution. Therefore, the constellation associated with the commutator is $\{Z_1, Z_2\} = \{\infty, (z_1 + z_2)/2\}$.

Let us state our result in the following way.

Proposition 1: With each ray of the complex Lie algebra of SL(2, C) can be associated a constellation of order 2 denoted $\{z_1, z_2\}$. The constellation $\{Z_1, Z_2\}$ associated with the commutator $[\{z_1, z_2\}, \{z_1', z_2'\}]$ is the one given by the cross product conditions

$$[\{Z_1, Z_2\}, \{z_1, z_2\}] = [\{Z_1, Z_2\}, \{z_1', z_2'\}] = -1.$$

As a direct consequence, we get the Clebsch–Gordan decomposition of the series $D_1 \otimes D_1$ (tensor product of spin 1 states). We know that

$$D_1 \otimes D_1 = D_0 \oplus D_1 \oplus D_2. \tag{8}$$

A state of spin j is described by a constellation of order $2j$.[7] In the series, the state of spin 2 is obtained by symmetrization of the two spin 1-states and taking out the trace. If $\{z_1, z_2\}$ and $\{z_1', z_2'\}$ are the states involved in the left-hand side of (8), the state of spin 2 is[8] simply $\{z_1, z_2, z_1', z_2'\}$. Now the state of spin 1 of the right-hand side is given by the same condition as the one given in Proposition 1 (the vector product in V is the commutator).

Proposition 2: Given two elementary states of spin 1 described by the constellations $\{z_1, z_2\}$ and $\{z_1', z_2'\}$, the global state is the linear combination of a constellation of spin 2 (order 4), $\{z_1, z_2, z_1', z_2'\}$, a constellation of spin 1 (order 2), $\{Z_1, Z_2\}$, with Z_1, Z_2 given by Eq. (6) and *the* constellation of spin 0 (order 0).[9]

[1] L.R. Ford, *Automorphic Functions* (Chelsea, New York, 1929, reprinted 1951).

[2] H. Bacry, "Eigenstates of complex linear combinations of J_1, J_2, J_3 for any representation of SU(2)," J. Math. Phys. **19**, 1192 (1978).

[3] Isomorphic to SL(2, C) and locally isomorphic, *as a real group*, to the Lorentz group.

[4] In the following, all constellations are implicitly supposed to be of order two except if it is otherwise stated.

[5] The converse is true for a nondegenerate constellation: Given a nondegenerate constellation $\{z_1, z_2\}$, the transformations which leave it invariant form a one-dimensional subgroup, generated by the $\sigma(\mathbf{f})$ associated with $\{z_1, z_2\}$.

[6] This means that $z_1 z_2 + 1 = z_1' z_2' + 1 = 0$.

[7] H. Bacry, J. Math. Phys. **15**, 1686 (1974).

[8] This can be generalized as follows. If $\{z_1, z_2, \ldots, z_{2j_1}\}$ and $\{z_1', z_2', \ldots, z_{2j_2}'\}$ are states of spin j_1 and j_2, respectively, in the tensor product, the state of spin $j_1 + j_2$ is simply given by $\{z_1, z_2, \ldots, z_{2j_1}, z_1', z_2', \ldots, z_{2j_2}'\}$.

[9] It is clear that the constellation of order 0 is unique.

On the covariant equations of the relativistic electrodynamics of continua. I. General equations

G. A. Maugin

Université de Paris VI, Laboratoire de Mécanique Théorique associé au C.N.R.S., Tour 66, 75230 Paris, Cedex 05, France
(Received 6 April 1977)

The aim of this series of papers is to develop both the field and constitutive equations for general relativistic electromagnetic continua. In this paper, the formulation starts from first principles of conservation and basic thermodynamical assumptions. This axiomatic character, however, is tempered by the use of results deduced from microscopical arguments by de Groot and Suttorp insofar as relativistically invariant expressions of the ponderomotive force and couple and of the electromagnetic power are concerned. This formulation is given entirely in terms of spatial space–time geometrical objects to facilitate the comparison with the Galilean theory of electromagnetic continua. It provides the thermodynamical cornerstone, i.e., the Clausius–Duhem inequality, in various equivalent forms which will be exploited in Papers II and III for material behaviors of interest in general relativistic, cosmological, and astrophysical problems which involve both the effects of gravitational fields and electric and/or magnetic fields.

1. INTRODUCTION

Several works have been devoted in recent years to an axiomatic approach to the formulation of the covariant field equations for electromagnetic relativistic matter, for instance, the works of Grot and Eringen,[1] Boulanger and Mayne,[2] and Romano.[3] The problem then arises of choosing the source terms due to electromagnetic fields in the presence of magnetized and electrically polarized matter. The controversy about this problem is best dealt with by de Groot and Suttorp,[4,5] Penfield and Haus,[6] de Groot,[7] Brevik,[8] Sedov,[9] and Pao.[10] Certain authors favor the *a priori* choice of a certain electromagnetic energy—momentum tensor and justify their choice by the successes obtained in deducing some well-known simple (i.e., linear) constitutive equations and force expressions and in accounting for the results of some experiments in electromagnetic optics. This is the case of Brevik. Boulanger and Mayne base their choice on pure Lorentz invariance regulations. Another possibility, however, is to search for some microscopic justification for the choice of the terms looked for. The most extensive microscopic treatment that has appeared so far is that of de Groot and Suttorp.[11] Our conviction is that sensible macroscopic expressions for the total matter-plus-field quantities must emerge from a compromise between the *a priori* choice and the relevance of microscopic considerations. Thus our first aim in this series of papers is to combine a somewhat axiomatic approach (needed to obtain nonlinear constitutive equations) insofar as the contributions labeled "material" fields are concerned, with the acceptance of the expressions deduced by de Groot and Suttorp in their formulation for the electromagnetic contributions, adjustments being made by means of the (in general nonlinear) constitutive equations that remain to be constructed with the help of thermodynamic arguments nowadays accepted in continuum thermodynamics. The field equations thus obtained in this first part, devoted to media without intrinsic spin, are Lorentz invariant and are equally applicable in the general relativistic framework. This approach avoids the use of variational principles,[12,13] and is based on the postulate of local balance laws for mechanics and thermodynamics. It is the relativistic analog of the Galilean invariant approach recently given by Maugin and Eringen.[14]

In the first part of this work, after introducing the basic concepts and notation needed in the four parts, we construct the local covariant conservation laws and the local entropy inequality, which will be exploited in Parts II and III. We acknowledge the ambiguity of splitting in matter and field contributions, but show in an example how the underdetermination is removed by the fact that source terms which differ slightly from those proposed by de Groot and Suttorp in fact yield the same mechanical and thermodynamical equations thanks to the automatic adjustment of the "material" contributions.

Part II will be devoted to deducing sets of nonlinear or linear constitutive equations for electromagnetic fluids. Part III will be devoted to constructing nonlinear or linearized constitutive equations for electromagnetic elastic solids with an emphasis on the cases which prove useful in practical applications where the relativistic framework is needed, e.g., nonlinear elastic magnetized bodies in general relativity with the special case of magnetoelasticity under very high pressure (in neutron stars), and piezoelectric bodies (which can be used in piezoelectric detectors of gravitational waves). Part IV is devoted to showing that the present formulation accomodates the relativistic theory of magnetized continua with intrinsic spin previously developed by the author,[15] as well as the theory of electromagnetic continua endowed with a rigid microstructure.

2. PRELIMINARIES, NOTATION

The arena of relativistic events is a space—time. A space—time $M=(\mathcal{V}^4, g_{\alpha\beta})$ is a differentiable manifold of dimension four, of continuity class C^p, $p \geq 2$, equipped with a normal hyperbolic Riemannian metric $g_{\alpha\beta}=g_{\beta\alpha}$, hence with Lorentzian signature +2. That is, the squared element of distance in $\mathcal{V}^4$ in general is given by $ds^2=g_{\alpha\beta}(x^\lambda)\,dx^\alpha\,dx^\beta$, where $\{x^\alpha;\alpha,\beta,\lambda=1,2,3,4\}$ is a local chart of $\mathcal{V}^4$. In special relativity, we have the

 0022-2488/78/1905-1198$1.00

global reduction $ds^2=\eta_{\alpha\beta}dx^\alpha dx^\beta$, where $\eta_{\alpha\beta}$ $=\mathrm{diag}(+1,+1,+1,-1)$. The space–time then is the *flat* Minkowskian space–time M^4. The same reduction holds good, but only *locally* at each event point, in the standard theory of general relativity (Einstein, 1916), where the space–time is a *curved* manifold V^4 in the presence of matter and energy. Expressions valid in an inertial frame are indicated by the sign $\stackrel{*}{=}$. Such expressions are written in "rectangular" coordinate systems $x^\alpha \stackrel{*}{=} \{x^i=(x,y,z),\ i=1,2,3;x^4=ct\}$.

All rectifiable curves in V^4 being parametrized proportionally to the length of their arc in the metric associated with V^4, the proper time τ, defined by

$$(d\tau)^2=-c^{-2}g_{\alpha\beta}dx^\alpha dx^\beta, \tag{2.1}$$

is the relevant timelike parameter for the parametrization of timelike world lines of particles. c is the velocity of light in a vacuum. Timelike 4-vectors V^α are such that $g_{\alpha\beta}V^\alpha V^\beta<0$. The world line of a pointlike particle X (of nonzero mass) in V^4,

$$C(X,\tau): x^\alpha=\chi^\alpha(X,\tau), \tag{2.2}$$

is such that its tangent, the world velocity of X, defined by $u^\alpha\equiv\partial\chi^\alpha/\partial\tau$, is everywhere timelike, for $g_{\alpha\beta}u^\alpha u^\beta=-c^2$ by virtue of Eq. (2.1). We note $\partial_\alpha=\partial/\partial x^\alpha$, and ∇_α is the covariant derivative. $D\equiv u^\alpha\nabla_\alpha$ is the invariant derivative in the direction of u^α. $a^\alpha\equiv Du^\alpha$, with $a^\alpha u_\alpha=0$, is the 4-acceleration of X. A relativistic motion for which $a^\alpha=0$ at all proper times is said to be *inertial*. 4-vectors such as the 4-acceleration which admit the world velocity as a zero vector will be called *spatial* 4-vectors. The 4-vector $c^{-2}a^\alpha$ is none other than the curvature vector of $C(X,\tau)$. If the motion of X is always inertial in V^4, then $C(X,\tau)$ is a straight line in V^4. A frame attached to a moving particle X is called its *proper frame* or *comoving* frame $R_C(X)$. The *rest* frame of a particle is the frame in which the particle (three-dimensional) velocity vanishes.

3. MATERIAL CONTINUUM[16,17]

Let $\mathcal{M}^3$ be the three-dimensional manifold which serves to describe a material continuum, and $\{X^K;K=1,2,3\}$ a local chart of $\mathcal{M}^3$. A material continuous body B is an open region of $\mathcal{M}^3$; its constituents, the material particles, are the points $\mathbf{X}$ of $B\subset\mathcal{M}^3$. In Newtonian mechanics $\mathcal{M}^3$ is related to the Galilean space–time by a one-parameter (the absolute Newtonian time t) family of differential embeddings. However, as V^4 in general does not possess a canonical time in general relativity, it is preferable to describe the relationship existing between $\mathcal{M}^3$ and V^4 in relativistic continuum mechanics by means of a *canonical differentiable projection* P such that $P:\mathcal{T}\to\mathcal{M}^3$ of an open $\mathcal{T}$ of V^4 onto $\mathcal{M}^3$. $\mathcal{T}[B;\mathbf{X}\in B]$ may be thought of as the region (a tube) of V^4 swept out by B at time goes on. We have $\mathbf{X}=P(\mathbf{x})$, where $\mathbf{x}$ is the variable event of V^4 described by the "particle" $\mathbf{X}\in\mathcal{M}^3$. We can write

$$X^K=X^K(x^\alpha),\quad \mathbf{x}\in C(\mathbf{X},\tau), \tag{3.1}$$

and

$$\tau=\tau(x^\alpha), \tag{3.2}$$

where τ is the proper time of $\mathbf{X}$.

P has the following property: The inverse image $P^{-1}(\mathbf{X})\subset V^4$ of any $\mathbf{X}\in B\subset\mathcal{M}^3$ is the timelike world line of $\mathbf{X}$, i.e., the space–time curve

$$x^\alpha=\chi^\alpha(X^K,\tau). \tag{3.3}$$

X^K and τ are independent variables, so that

$$DX^K=0. \tag{3.4}$$

The X^K are referred to either as Lagrangian coordinates (in the case of fluids) or as material coordinates (in the case of deformable solids).

The projection P determines the fundamental flow of vectors u^α. The latter in turn defines the fundamental field of projection operators $P_{\alpha\beta}$, which acts as a positive metric tensor on the tangent subspace orthogonal to u^α. The (covariant) projection operator $P_{\alpha\beta}(\mathbf{x})$ at $\mathbf{x}\in C(\mathbf{X},\tau)$, is such that

$$P_{\alpha\beta}(\mathbf{x})\equiv g_{\alpha\beta}(\mathbf{x})+c^{-2}u_\alpha(\mathbf{x})\mu_\beta(\mathbf{x})=P_{\beta\alpha}(x), \tag{3.5}$$

$$P_{\alpha\beta}P^{\beta\gamma}=P^\gamma_\alpha,\quad P_{\alpha\beta}u^\beta=0,\quad P^\alpha_{\ \alpha}=3. \tag{3.6}$$

The introduction of this operator leads to the notion of space and time decomposition.

Space and time decomposition[17,18]: Consider the general space–time $M=(V^4,g_{\alpha\beta})$. Let x, T_x, U_x, and $M_\perp(\mathbf{x})$ be respectively an event point on $C(\mathbf{X},\tau)$, the tangent vector space to M at $\mathbf{x}$, the vector subspace of 4-vectors collinear with u^α at $\mathbf{x}$, and the three-dimensional hyperplane orthogonal to $u^\alpha(\mathbf{x})$. Let T^*_x, U^*_x, and $M^*_\perp(\mathbf{x})$ be the dual spaces of these spaces. Then $T_x=U_x\oplus M_\perp(\mathbf{x})$, where $\oplus$ indicates the direct sum. An analogous relationship holds good for the dual spaces. A tensor field defined at $\mathbf{x}\in C(\mathbf{X},\tau)$, which is m times contravariant and n times covariant, belongs to

$$T_x^{\otimes m}\otimes T_x^{*\otimes n}=(U_x\oplus M_\perp(\mathbf{x}))^{\otimes m}\otimes(U_x^*\oplus M_\perp^*(\mathbf{x}))^{\otimes n}, \tag{3.7}$$

where $\otimes$ indicates the tensor product. Expanding the right-hand side of this expression shows that such a tensor can be *canonically decomposed*, or projected in various manners. A tensor field $\mathbf{A}$, m times contravariant and n times covariant, is said to be *spatial* at $\mathbf{x}\in C(\mathbf{X},\tau)$ if and only if

$$\mathbf{A}\in M_\perp^{m,n}(\mathbf{x})\equiv(M_\perp(\mathbf{x}))^{\otimes m}\otimes(M_\perp^*(\mathbf{x}))^{\otimes n}. \tag{3.8}$$

In these conditions the world velocity is a zero vector of all spatial tensor fields. The symbolic notation $(\cdots)_\perp$ is used to indicate the projection onto $M_\perp^{m,n}$. For example,

$$(A^\alpha_{\cdot\beta})_\perp\equiv P^\alpha_{\cdot\sigma}P^{\cdot\mu}_\beta A^\sigma_{\cdot\mu},\quad (A^\alpha_{\cdot\beta})_\perp u_\alpha=0,\quad (A^\alpha_{\cdot\beta})_\perp u^\beta=0. \tag{3.9}$$

Three examples of canonical decomposition performed in accordance with Eq. (3.7) are particularly useful, those of a contravariant vector field $\mathbf{A}$, of a general second-order contravariant tensor field $\mathbf{T}$, and of a two-form $\mathbf{G}$ of components $G^{\alpha\beta}=-G^{\beta\alpha}$. We have

$$A^\alpha=\mathcal{A}^\alpha+Au^\alpha,\quad A\equiv-c^{-2}A^\alpha u_\alpha,\quad \mathcal{A}^\alpha\equiv(A^\alpha)_\perp, \tag{3.10}$$

$$T^{\alpha\beta}=c^{-2}\omega u^\alpha u^\beta+c^{-2}u^\alpha Q^\beta+p^\alpha u^\beta-t^{\beta\alpha}, \tag{3.11}$$

where

$$\omega\equiv c^{-2}u_\alpha T^{\alpha\beta}u_\beta,\quad t^{\beta\alpha}\equiv-(T^{\alpha\beta})_\perp,$$
$$Q^\beta\equiv-(u_\alpha T^{\alpha\beta})_\perp,\quad p^\alpha\equiv-c^{-2}(T^{\alpha\beta}u_\beta)_\perp, \tag{3.12}$$

and

$$G^{\alpha\beta}=\frac{2}{c}u^{[\alpha}\mathcal{D}^{\beta]}+(G^{\alpha\beta})_{\perp} \tag{3.13}$$

with

$$\mathcal{D}^{\beta}\equiv\frac{1}{c}G^{\beta\alpha}u_{\alpha},\quad (G^{\alpha\beta})_{\perp}=\frac{1}{c}\eta^{\alpha\beta\gamma\delta}\mathcal{H}_{\gamma}u_{\delta},$$
$$\mathcal{H}_{\alpha}\equiv\frac{1}{2c}\eta_{\alpha\beta\gamma\delta}(G^{\beta\gamma})_{\perp}u^{\delta}=\tfrac{1}{2}\overset{*}{\eta}{}_{\alpha\beta\gamma}(G^{\beta\gamma})_{\perp}. \tag{3.14}$$

Here, $\mathcal{H}_{\alpha}$ is the spatial axial 4-vector which is associated in a unique way with $(G^{\beta\gamma})_{\perp}$, $\eta^{\alpha\beta\gamma\delta}$ and $\eta_{\alpha\beta\gamma\delta}$ are the tensorial densities ($g=\det g_{\alpha\beta}$)

$$\eta_{\alpha\beta\gamma\delta}=\sqrt{-g}\epsilon_{\alpha\beta\gamma\delta},\quad \eta^{\alpha\beta\gamma\delta}=-\frac{1}{\sqrt{-g}}\epsilon^{\alpha\beta\gamma\delta}, \tag{3.15}$$

where $\epsilon_{\alpha\beta\gamma\delta}$ and $\epsilon^{\alpha\beta\gamma\delta}$ are four-dimensional alternation symbols such that $\epsilon_{1234}=\epsilon^{1234}=+1$, and

$$\overset{*}{\eta}_{\alpha\beta\gamma}\equiv c^{-1}\eta_{\alpha\beta\gamma\delta}u^{\delta},\quad \overset{*}{\eta}{}^{\gamma\alpha\beta}\equiv c^{-1}\eta^{\gamma\alpha\beta\delta}u_{\delta}. \tag{3.16}$$

Brackets around a set of indices indicate alternation while parentheses indicate symmetrization. For further use we also define the "$*$" product of two spatial covariant 4-vectors **A** and **B** by

$$(\mathbf{A}*\mathbf{B})^{\alpha}\equiv\overset{*}{\eta}{}^{\alpha\beta\gamma}A_{\beta}B_{\gamma}, \tag{3.17}$$

so that

$$u^{\alpha}\overset{*}{\eta}_{\alpha\beta\gamma}=0,\quad u_{\alpha}(\mathbf{A}*\mathbf{B})^{\alpha}=0. \tag{3.18}$$

The "$*$" product of two spatial contravariant 4-vectors is defined in a similar fashion.

Although it is expressed in a fully invariant form, a spatial tensor field has essentially spatial and three-dimensional values on M. It reduces to the equivalent three-dimensional object of classical physics in a local instantaneous rest frame. The decomposition procedure of which the essence is summarized in Eq. (3.7), is a technique of intrinsic character in curved space–time, which provides definitions for *relative*, but nontheless *covariant*, quantities and operations. In particular, the covariant derivative ∇_{α} can be decomposed according to

$$\nabla_{\alpha}=\overset{\perp}{\nabla}_{\alpha}-c^{-2}u_{\alpha}D, \tag{3.19}$$

where $\overset{\perp}{\nabla}_{\alpha}\equiv P_{\alpha}^{\circ\beta}\nabla_{\beta}$ is the *transverse* or *spatial* covariant derivative. Then the relativistic velocity gradient $e_{\alpha\beta}$, the relativistic rate of strain $d_{\alpha\beta}$, the relativistic rate of rotation $\omega_{\alpha\beta}$, and the spatial vorticity 4-vector ω^{α} are defined by

$$e_{\alpha\beta}\equiv\overset{\perp}{\nabla}_{\beta}u_{\alpha}=(\nabla_{\beta}u_{\alpha})_{\perp}, \tag{3.20}$$

$$d_{\alpha\beta}\equiv e_{(\alpha\beta)},\quad \omega_{\alpha\beta}\equiv e_{[\alpha\beta]}, \tag{3.21}$$

and

$$\omega^{\alpha}\equiv\tfrac{1}{2}\overset{*}{\eta}{}^{\alpha\beta\gamma}e_{\gamma\beta}, \tag{3.22}$$

respectively. Hence[19]

$$\nabla_{\beta}u_{\alpha}=d_{\alpha\beta}-\overset{*}{\eta}_{\alpha\beta\gamma}\omega^{\gamma}-c^{-2}a_{\alpha}u_{\beta}. \tag{3.23}$$

A relativistic motion for which $\omega^{\alpha}=0$ for all τ is said to be irrotational. A relativistic motion for which $d_{\alpha\beta}=0$ for all τ is called a Herglotz–Born rigid body motion.[20]

4. RELATIVISTIC CONSERVATION LAWS IN ABSENCE OF SPIN

The basic *local* conservation laws expressed in an *inertial* frame can be written in the following form[21]:

conservation of mass:

$$\partial_{\alpha}\rho^{\alpha}\overset{*}{=}0, \tag{4.1}$$

balance of energy–momentum:

$$\partial_{\beta}\,{}_{m}T^{\alpha\beta}\overset{*}{=}F^{\alpha}, \tag{4.2}$$

balance of moment of energy-momentum

$$\partial_{\mu}\mathcal{S}^{\alpha\beta\mu}\overset{*}{=}x^{[\alpha}F^{\beta]}+L^{\alpha\beta}, \tag{4.3}$$

second principle of thermodynamics[22]

$$\partial_{\alpha}\eta^{\alpha}\geq 0. \tag{4.4}$$

These are supplemented with Maxwell's equations (cf. Appendix A). Here, F^{α} is a 4-force and $L^{\alpha\beta}=-L^{\beta\alpha}$ are the components of the couple 2-form. Both include contributions due to the interactions exerted between matter and electromagnetic fields. The mass flux ρ^{α}, the "matter" energy–momentum tensor ${}_{m}T^{\alpha\beta}$ (a general second-order tensor), the total (orbital plus intrinsic) spin tensor $\mathcal{S}^{\alpha\beta\mu}$ and the entropy-flux 4-vector η^{α} have the following decompositions:

$$\rho^{\alpha}=\rho u^{\alpha}, \tag{4.5}$$

$${}_{m}T^{\alpha\beta}=\rho(1+c^{-2}\epsilon)u^{\alpha}u^{\beta}+c^{-2}u^{\alpha}Q^{\beta}+p^{\alpha}u^{\beta}-t^{\beta\alpha}, \tag{4.6}$$

$$\mathcal{S}^{\alpha\beta\mu}\overset{*}{=}x^{[\alpha}\,{}_{m}T^{\beta]\mu}+s^{\alpha\beta\mu}=-\mathcal{S}^{\beta\alpha\mu}, \tag{4.7}$$

$$\eta^{\alpha}=\rho\eta u^{\alpha}+N^{\alpha}. \tag{4.8}$$

ρ is the proper mass density which is related to the number density n of particles by the equation $\rho=nm_{0}$, where m_{0} is the mean rest mass per particle. The decomposition (4.6)—compare (3.11)—is that considered primarily by Eckart[23] and Taub.[24] ϵ is a density of internal energy, the factor one representing the contribution of the rest energy. $t^{\beta\alpha}$ is the spatial relativistic stress tensor. p^{α} is the nonmechanical momentum spatial 4-vector. The total 4-momentum is

$$P^{\alpha}=\rho(1+c^{-2}\epsilon)u^{\alpha}+p^{\alpha}, \tag{4.9}$$

and obviously is not collinear with the world velocity except when $p^{\alpha}\equiv 0$. The spatial 4-vector Q^{β} is the energy 4-current. It necessarily contains the spatial heat-flux 4-vector q^{β} in order that the mixed component ${}_{m}T^{4j}$ at least contains, in an inertial frame, the classical expression $q^{j}-t^{ji}v_{i}+\rho\epsilon v^{j}$.[25] Thus, we shall write

$$Q^{\beta}=q^{\beta}+\hat{Q}^{\beta}, \tag{4.10}$$

where the spatial 4-vector $\hat{Q}^{\beta}$ remains to be chosen. The spin tensor $s^{\alpha\beta\mu}=-s^{\beta\alpha\mu}$ can be decomposed as

$$s^{\alpha\beta\mu}=\tfrac{1}{2}\rho S^{\alpha\beta}u^{\mu}-M^{\alpha\beta\mu}, \tag{4.11}$$

where

$$S^{\alpha\beta}=-S^{\beta\alpha}=-2\rho^{-1}c^{-2}(s^{\alpha\beta\mu}u_{\mu})_{\perp} \tag{4.12}$$

and

$$M^{\alpha\beta\mu}=-M^{\beta\alpha\mu}=(s^{\alpha\beta\mu})_{\perp} \tag{4.13}$$

are the spatial intrinsic spin 2-form and the spatial relativistic couple-stress tensor, respectively. In ab-

sence of these effects, we shall set

$$S^{\alpha\beta} \equiv 0, \quad M^{\alpha\beta\mu} \equiv 0. \tag{4.14}$$

Finally, η is the entropy per unit of proper mass and N^α is the spatial entropy-flux 4-vector. According to classical thermodynamics, N^α and q^α are related by

$$N^\alpha = q^\alpha/\theta, \tag{4.15}$$

where θ is the proper thermodynamical temperature such that $\theta > 0$, $\inf\theta = 0$, and[26]

$$\theta \overset{*}{=} \theta_{\text{lab}}\left(1 - \frac{\mathbf{v}^2}{c^2}\right)^{1/2}, \tag{4.16}$$

where θ_{lab} is the temperature in a laboratory frame with respect to which $\mathbf{v}$ is the matter three-dimensional velocity.

Taking account of Eq. (4.2) to reduce Eq. (4.3) and passing to an arbitrary frame by replacing partial derivatives with covariant derivatives, we obtain the following set of local conservation laws:

$$\nabla_\beta(\rho u^\beta) = 0, \tag{4.17}$$

$$\nabla_\beta\, {}_m T^{\alpha\beta} = F^\alpha, \tag{4.18}$$

$$\nabla_\mu s^{\alpha\beta\mu} - {}_m T^{[\alpha\beta]} = L^{\alpha\beta}, \tag{4.19}$$

$$\nabla_\alpha \eta^\alpha \geq 0. \tag{4.20}$$

If Eqs. (4.14) hold true, then Eq. (4.19) reduces to

$${}_m T^{[\alpha\beta]} + L^{\alpha\beta} = 0. \tag{4.21}$$

We consider that the hypotheses (4.14) are valid in the subsequent analysis of this paper (no spin). Thus Eqs. (4.17), (4.18), (4.21), and (4.20) form the set of basic local covariant conservation laws. We assume in the sequel that F^α and $L^{\alpha\beta}$ result solely from the presence of electromagnetic fields, and are thus denoted by ${}_M F^\alpha$ and ${}_M L^{\alpha\beta}$ in conformity with the notation adopted in a previous paper.[27] These geometrical objects admit general canonical decompositions of the type (3.10) and (3.13). Explicitly,

$${}_M F^\alpha = {}_M f^\alpha + c^{-2}\, {}_M\bar{w} u^\alpha, \tag{4.22}$$

$${}_M L^{\alpha\beta} = {}_M \angle^{[\alpha} u^{\beta]} + {}_M C^{\alpha\beta}, \tag{4.23}$$

with

$${}_M f^\alpha \equiv ({}_M F^\alpha)_\perp, \quad {}_M C^{\alpha\beta} \equiv ({}_M L^{\alpha\beta})_\perp = -\, {}_M C^{\beta\alpha}, \tag{4.24}$$

$${}_M\bar{w} \equiv -\, {}_M F^\alpha u_\alpha, \quad {}_M \angle^\alpha \equiv -\, 2c^{-2}\, {}_M L^{\alpha\beta} u_\beta. \tag{4.25}$$

We shall call *ponderomotive 4-force* and *ponderomotive couple 2-form, per se*, the purely spatial fields ${}_M f^\alpha$ and ${}_M C^{\alpha\beta}$, respectively. ${}_M\bar{w}$ is the electromagnetic power developed by electromagnetic fields in presence of matter. The spatial vector ${}_M \angle^\alpha$ has the dimension of a 4-momentum. It may be either zero or nonzero depending on the theory considered.

Projecting now Eqs. (4.18) and (4.21) onto $M_\perp$ and along u_α, using Eq. (4.7) and the various canonical decompositions, we obtain the 4-vector form of the conservation laws:

conservation of mass:

$$D\rho + \rho e^\alpha_{\cdot\alpha} = 0, \tag{4.26}$$

local balance of 4-momentum (Euler—Cauchy equations of motion):

$$\rho\left(1 + \frac{\epsilon}{c^2}\right)(Du^\alpha)_\perp + \rho\left[D\left(\frac{p^\alpha}{\rho}\right)\right]_\perp = (\nabla_\beta t^{\beta\alpha})_\perp + {}_M f^\alpha - c^{-2} Q^\beta e^\alpha_{\cdot\beta}, \tag{4.27}$$

local balance of energy:

$$\rho D\epsilon + \nabla_\beta Q^\beta + p^\alpha Du_\alpha = t^{\beta\alpha} e_{\alpha\beta} + {}_M\bar{w}, \tag{4.28}$$

local balance of moment of momentum:

$$t^{[\beta\alpha]} = {}_M C^{\alpha\beta}, \tag{4.29}$$

and

$$p^\alpha = c^{-2} Q^\alpha - {}_M \angle^\alpha, \tag{4.30}$$

local balance of entropy:

$$\rho D\eta + \nabla_\beta N^\beta \geq 0. \tag{4.31}$$

Introducing the free energy per unit of proper mass, ψ, by

$$\psi \equiv \epsilon - \eta\theta, \tag{4.32}$$

and eliminating $D\eta$ between Eqs. (4.28) and (4.31), we obtain the *Clausius—Duhem inequality* in the following form:

$$-\rho(D\psi + \eta D\theta) + \theta\nabla_\beta\left(N^\beta - \frac{Q^\beta}{\theta}\right) + \theta Q^\beta \overset{\perp}{\nabla}_\beta\left(\frac{1}{\theta}\right) - p^\alpha Du_\alpha + t^{\beta\alpha} e_{\alpha\beta} + {}_M\bar{w} \geq 0. \tag{4.33}$$

Now, taking account of Eqs. (4.29) and (4.30), of the decomposition $e_{\alpha\beta} = d_{\alpha\beta} + \omega_{\alpha\beta}$, and assuming that $\hat{Q}^\beta \equiv 0$ (or $Q^\beta \equiv q^\beta$) and that Eq. (4.15) holds true, Eq. (4.33) yields

$$-\rho(D\psi + \eta D\theta) - \theta^{-1} q^\alpha \overset{*}{\theta}_\alpha + t^{(\beta\alpha)} d_{\alpha\beta} + \{{}_M C^{\alpha\beta}\omega_{\alpha\beta} + {}_M \angle^\alpha Du_\alpha + {}_M\bar{w}\} \geq 0, \tag{4.34}$$

where we have introduced the field

$$\overset{*}{\theta} \equiv \overset{\perp}{\nabla}_\alpha \theta + c^{-2}\theta a_\alpha. \tag{4.35}$$

Using the latter instead of $\overset{\perp}{\nabla}_\alpha\theta$ means that we consider that thermal equilibrium corresponds to constant redshifted temperature in a stationary gravitational field instead of constant temperature.[28]

By the same token we can rewrite Eqs. (4.27) and (4.28) as

$$\rho\left(1 + \frac{\epsilon}{c^2}\right)(Du^\alpha)_\perp + c^{-2}(D_c q^\alpha + 2q^\beta e^\alpha_{\cdot\beta}) = (\nabla_\beta t^{\beta\alpha})_\perp + \{{}_M f^\alpha + \rho[D({}_M \angle^\alpha/\rho)]_\perp\}, \tag{4.36}$$

and

$$\rho D\epsilon + (\overset{\perp}{\nabla}_\beta q^\beta + 2c^{-2} q^\beta a_\beta) = t^{\beta\alpha} d_{\alpha\beta} + \{{}_M C^{\alpha\beta}\omega_{\alpha\beta} + {}_M \angle^\alpha a_\alpha + {}_M\bar{w}\}, \tag{4.37}$$

respectively. The latter equation will provide the equation that governs heat propagation. Herein above we have used the fact that, on account of Eq. (4.26),

$$\rho D(q^\alpha/\rho) = Dq^\alpha + q^\alpha(\nabla_\beta u^\beta), \tag{4.38}$$

and we have introduced the contravariant convective time derivative—noted by the symbol D_C—by

$$D_C q^\alpha \equiv (\underset{u}{\pounds} q^\alpha)_\perp + q^\alpha(\nabla_\beta u^\beta), \tag{4.39}$$

where the Lie derivative noted $\underset{u}{\pounds}$ is given by (after projection)

$$(\underset{u}{\pounds} q^\alpha)_\perp(\mathbf{x}) \equiv p^{-1}\left[\frac{d}{d\tau} p(q^\alpha)(\mathbf{X},\tau)\right](\mathbf{x})$$
$$= (Dq^\alpha)_\perp - q^\beta \overset{\perp}{\nabla}_\beta u^\alpha, \tag{4.40}$$

in terms of the canonical projection p.[29]

To complete the present general scheme of the equations of relativistic continua in absence of spin, it remains to specify the expressions of ${}_M f^\alpha$, ${}_M C^{\alpha\beta}$, ${}_M \mathcal{L}^\alpha$—or, equivalently, of ${}_M F^\alpha$ and ${}_M L^{\alpha\beta}$—which contribute to the expressions placed within braces in Eqs. (4.36), (4.37), and (4.34). In these equations, the factor c^{-2} places in evidence the purely relativistic contributions. For further use we note

$${}_M \hat{f}^\alpha \equiv {}_M f^\alpha + \rho[D({}_M \mathcal{L}^\alpha/\rho)]_\perp. \tag{4.41}$$

5. ELECTROMAGNETIC INTERACTIONS WITH MATTER

In order to avoid a fully arbitrary choice of a 4-force ${}_M F^\alpha$ and of a couple ${}_M L^{\alpha\beta}$, a microscopic model of interactions must be considered as a starting point. The relativistically moving electromagnetic matter may be considered as an aggregate of stable groups of pointlike particles labeled $k = 1, 2, \cdots$, of electric charge $\delta q^{(k)}$ in their own rest frame and of world velocity $u^{(k)\alpha}$, and situated at the position $R^{(k)\beta}$ in flat space–time, upon which a relativistically invariant Lorentz force $\delta f^{(k)\alpha}$ is acting. This force reads

$$\delta f^{(k)\alpha} \overset{*}{=} \frac{1}{c}\delta q^{(k)} f^{\alpha\beta}(\mathbf{R}^{(k)}) D^{(k)} R^{(k)}_\beta, \tag{5.1}$$

where $D^{(k)} \equiv u^{(k)\alpha}\partial_\alpha$, and $f_{\alpha\beta} = -f_{\beta\alpha}$ is the "microscopic" magnetic flux tensor evaluated at $\mathbf{R}^{(k)}$. Performing a Lorentz invariant phase-space averaging of the relativistic equations of motion that govern the individual particles $k = 1, 2, \cdots$, de Groot and Suttorp[30] were able to smooth out the force expressions (5.1) in order to obtain a sensible expression for both ${}_M F^\alpha$ and ${}_M L^{\alpha\beta}$ which act per unit of proper volume on a piece of electromagnetic matter considered as a continuum. These expressions read[30] (in our notation and in an arbitrary frame)

$${}_M F^\alpha = \frac{1}{c} F^{\alpha\beta} J_\beta + \tfrac{1}{2}\pi_{\beta\gamma} g^{\alpha\mu}\nabla_\mu F^{\beta\gamma}$$
$$- c^{-2}\rho D[\rho^{-1}(F^{\alpha\beta}\pi_{\beta\gamma} - \pi^{\alpha\beta}F_{\beta\gamma})u^\gamma]$$
$$+ c^{-4}\rho D[\rho^{-1}u^\alpha u^\beta F_{\beta\gamma}\pi^{\gamma\epsilon}u_\epsilon], \tag{5.2}$$

and

$${}_M L^{\alpha\beta} = \pi^{[\alpha}_{\cdot\mu} F^{\beta]\mu} + c^{-2}u^{[\beta}(F^{\alpha]\gamma}\pi_{\gamma\epsilon} - \pi^{\alpha]\gamma}F_{\gamma\epsilon})u^\epsilon. \tag{5.3}$$

In these equations, J_β is the electric current 4-vector, $F_{\alpha\beta} = -F_{\beta\alpha}$ is the macroscopic magnetic flux tensor, and $\pi_{\alpha\beta} = -\pi_{\beta\alpha}$ is the polarization-magnetization tensor. We refer to the original work of de Groot and Suttorp for the definition of these macroscopic fields in terms of microscopic fields. $F_{\alpha\beta}$ and $\pi^{\alpha\beta}$ admit decompositions of the type (3.13). That is,

$$F_{\alpha\beta} = \frac{1}{c}(u_\alpha \mathcal{E}_\beta - u_\beta \mathcal{E}_\alpha + \eta_{\alpha\beta\mu\nu}\mathcal{B}^\mu u^\nu), \tag{5.4}$$

$$\pi^{\alpha\beta} = \frac{1}{c}(p^\alpha u^\beta - p^\beta u^\alpha + \eta^{\alpha\beta\mu\nu}\mathcal{M}_\mu u_\nu), \tag{5.5}$$

where the spatial 4-vectors $\mathcal{E}_\beta$, $\mathcal{B}^\mu$, p^α, and $\mathcal{M}_\mu$ are the spatial electric-field, magnetic-induction, polarization, and magnetization 4-vectors, respectively. These are fields evaluated in a comoving frame. Furthermore, J_β admits a decomposition of the type (3.10), which reads

$$J_\beta = \mathcal{J}_\beta + q u_\beta, \tag{5.6}$$

where $\mathcal{J}_\beta$ is the spatial electric current, or *conduction current*, 4-vector, and q is the density of free charges per unit of proper volume.

Substituting for (5.4), (5.5), and (5.6) in Eqs. (5.2) and (5.3), it is found after a somewhat lengthy calculation that

$${}_M f^\alpha = q\mathcal{E}^\alpha + \frac{1}{c}(\mathcal{J} * \mathcal{B})^\alpha + P^{\alpha\beta}[\, p^\mu \overset{\perp}{\nabla}_\beta \mathcal{E}_\mu$$
$$+ \mathcal{M}^\mu \overset{\perp}{\nabla}_\beta \mathcal{B}_\mu + {}_M p^\mu \overset{\perp}{\nabla}_\beta u_\mu$$
$$+ \rho D(\rho^{-1}\, {}_M p_\beta)] + c^{-2}(\mathcal{E}\cdot p)(Du^\alpha)_\perp, \tag{5.7}$$

$${}_M \bar{w} = \rho \mathcal{E}_\alpha D\pi^\alpha - \mathcal{M}^\alpha D\mathcal{B}_\alpha + \mathcal{J}\cdot\mathcal{E}, \tag{5.8}$$

$${}_M \mathcal{L}^\alpha \equiv 0, \tag{5.9}$$

and

$${}_M C^{\alpha\beta} = p^{[\alpha}\mathcal{E}^{\beta]} + \mathcal{M}^{[\alpha}\mathcal{B}^{\beta]}, \tag{5.10}$$

where

$$\pi^\alpha \equiv p^\alpha/\rho, \tag{5.11}$$

and

$${}_M p^\alpha \equiv \frac{1}{c}(p * \mathcal{B} + \mathcal{E} * \mathcal{M})^\alpha, \quad {}_M p^\alpha u_\alpha = 0. \tag{5.12}$$

The definitions (3.15)–(3.19) have been used to write down Eqs. (5.7)–(5.13).

Define now the *covariant* convective time derivative—noted D_C—by

$$D_C A_\alpha \equiv (\underset{u}{\pounds} A_\alpha)_\perp + A_\alpha \nabla_\beta u^\beta$$
$$= (DA_\alpha)_\perp + A_\beta \overset{\perp}{\nabla}_\alpha u^\beta + A_\alpha \nabla_\beta u^\beta, \tag{5.13}$$

a new electromagnetic power ${}_M \tilde{w}$ by

$${}_M \tilde{w} \equiv \mathcal{J}\cdot\mathcal{E} - \mathcal{M}^\alpha D\mathcal{B}_\alpha - p^\alpha D\mathcal{E}_\alpha = {}_M \bar{w} - \rho D(\mathcal{E}_\alpha \pi^\alpha), \tag{5.14}$$

a new internal energy $\tilde{\epsilon}$ and a new free energy $\tilde{\psi}$ by

$$\tilde{\epsilon} = \epsilon - \mathcal{E}_\alpha \pi^\alpha, \quad \tilde{\psi} = \tilde{\epsilon} - \eta\theta, \tag{5.15}$$

and the ponderomotive force ${}_M \tilde{f}^\alpha$ by

$${}_M \tilde{f}^\alpha = q\mathcal{E}^\alpha + \frac{1}{c}(\mathcal{J} * \mathcal{B})^\alpha + P^{\alpha\beta}[\, p^\mu \overset{\perp}{\nabla}_\beta \mathcal{E}_\mu$$
$$+ \mathcal{M}^\mu \overset{\perp}{\nabla}_\beta \mathcal{B}_\mu + D_C({}_M p_\beta)], \tag{5.16}$$

on account of Eq. (5.13) and of an equation of the type (4.38) for ${}_M p_\beta$. Substituting from Eqs. (5.7)–(5.10) in Eqs. (4.36), (4.37), and (4.34), and taking account of the definitions (5.14)–(5.16), we find that the local balance laws of momentum and energy read

$$\rho\left(1 + \frac{\tilde{\epsilon}}{c^2}\right)(Du^\alpha)_\perp + c^{-2}(D_C q^\alpha + 2q^\beta e^\alpha_{\cdot\beta})$$
$$= (\nabla_\beta t^{\beta\alpha})_\perp + {}_M \tilde{f}^\alpha, \tag{5.17}$$

and

$$\rho D\tilde{\epsilon} + (\overset{\perp}{\nabla}_\beta q^\beta + 2c^{-2} q^\beta a_\beta)$$
$$= t^{(\beta\alpha)} d_{\alpha\beta} + (\mathcal{P}^{[\alpha}\mathcal{E}^{\beta]} + \mathcal{M}^{[\alpha}\mathcal{B}^{\beta]})\,\omega_{\alpha\beta} + {}_M\tilde{w}, \tag{5.18}$$

while the Clausius—Duhem inequality now reads

$$-\rho(D\tilde{\psi} + \eta D\theta) - \theta^{-1} q^\alpha \overset{*}{\theta}_\alpha + t^{(\beta\alpha)} d_{\alpha\beta}$$
$$+ (\mathcal{P}^{[\alpha}\mathcal{E}^{\beta]} + \mathcal{M}^{[\alpha}\mathcal{B}^{\beta]})\,\omega_{\alpha\beta} + {}_M\tilde{w} \geq 0. \tag{5.19}$$

Equations (5.15) represent Legendre transformations of the internal and free energies. ψ now depends on $\mathcal{E}_\alpha$ and $\mathcal{B}_\alpha$ as electromagnetic constitutive arguments. Other possibilities exist, which will be fully exploited in Part II of the present study. For instance, defining

$$\hat{\psi} \equiv \psi + \mu^\alpha \mathcal{B}_\alpha = \tilde{\psi} + \mathcal{E}_\alpha \pi^\alpha + \mu^\alpha \mathcal{B}_\alpha, \tag{5.20}$$

where

$$\mu^\alpha \equiv \mathcal{M}^\alpha/\rho, \tag{5.21}$$

introducing a new (symmetrical) spatial stress tensor, ${}_E t^{\beta\alpha}$, by

$${}_E t^{\beta\alpha} \equiv t^{(\beta\alpha)} + \mathcal{E}^{(\alpha} p^{\beta)} + \mathcal{B}^{(\alpha}\mathcal{M}^{\beta)} = {}_E t^{\alpha\beta}, \tag{5.22}$$

in such a way that

$$t^{\beta\alpha} = {}_E t^{\beta\alpha} - p^\beta \mathcal{E}^\alpha - \mathcal{M}^\beta \mathcal{B}^\alpha, \tag{5.23}$$

on account of Eqs. (4.29) and (5.10), and using definition (4.39), we can rewrite Eq. (5.19) in the following form:

$$-\rho(D\hat{\psi} + \eta D\theta) + {}_E t^{\beta\alpha} d_{\alpha\beta} + \mathcal{J}\cdot\mathcal{E} + \mathcal{E}_\alpha (D_C p^\alpha)$$
$$+ \mathcal{B}_\alpha (D_C \mathcal{M}^\alpha) - \theta^{-1} q^\alpha \overset{*}{\theta}_\alpha \geq 0. \tag{5.24}$$

Similarly, defining another symmetrical spatial stress tensor, $\tilde{t}^{\beta\alpha}$ by

$$\tilde{t}^{\beta\alpha} \equiv t^{(\beta\alpha)} - \mathcal{E}^{(\beta} p^{\alpha)} - \mathcal{B}^{(\beta}\mathcal{M}^{\alpha)} + (\mathcal{E}\cdot p + \mathcal{B}\cdot\mathcal{M})\,P^{\alpha\beta} = \tilde{t}^{\alpha\beta}, \tag{5.25}$$

in such a way that

$$t^{\beta\alpha} = \tilde{t}^{\beta\alpha} + \mathcal{E}^\beta p^\alpha + \mathcal{B}^\beta \mathcal{M}^\alpha - (\mathcal{E}\cdot p + \mathcal{B}\cdot\mathcal{M})\,P^{\alpha\beta}, \tag{5.26}$$

inequality (5.19) transforms to

$$-\rho(D\tilde{\psi} + \eta D\theta) + \tilde{t}^{\beta\alpha} d_{\alpha\beta} - p_\alpha (D_C \mathcal{E}^\alpha) - \mathcal{M}_\alpha (D_C \mathcal{B}^\alpha)$$
$$+ \mathcal{J}\cdot\mathcal{E} - \theta^{-1} q^\alpha \overset{*}{\theta}_\alpha \geq 0. \tag{5.27}$$

If we adopt the contemporary viewpoint on continuum thermodynamics,[31] inequality (5.24) or (5.27) constitutes a constraint which must be satisfied for any thermodynamical process. This constraint being placed upon the constitutive equations which are needed to close the system of field equations, it is a tenet for the constitutive theory. Since constitutive equations must be *objective*,[32] the cofactors of the dependent constitutive functions in the Clausius—Duhem inequality must also be *objective*. This is the case of

$$d_{\alpha\beta},\ \mathcal{E},\ \overset{*}{\theta}_\alpha,$$

and of the contravariant convective time derivative of objective vector fields according to a previous study of the author.[33] Hence, the inequalities (5.24) and (5.27) are in a ready-for-use form for the constitutive theory of media without spin developed in Part II. The choice of either one of these inequalities depends on the behavior studied. The form (5.24) is more convenient for the study of elastic solids, whereas form (5.27) allows one to study the case of electromagnetic fluids in a simpler fashion.

6. REMARKS ON THE PONDEROMOTIVE FORCE AND COUPLE

A. Comparison with the Galilean invariant formulation

Because of the systematic use of spatial 4-vectors, the expressions obtained above are readily compared to the Galilean expressions arrived at in other works. For instance, using the conventional three-dimensional notation, the spatial 4-force (5.16) yields the spatial component

$${}_M\mathbf{f} = q\mathbf{E}' + \frac{1}{c}\mathcal{J}\times\mathbf{B}' + (\nabla\mathbf{E}')\cdot\mathbf{P}' + (\nabla\mathbf{B}')\cdot\mathbf{M}'$$
$$+ \frac{1}{c}(\nabla\mathbf{v})\cdot(\mathbf{P}'\times\mathbf{B}' - \mathbf{M}'\times\mathbf{E}')$$
$$+ \frac{\rho}{c}\frac{d}{dt}(\pi'\times\mathbf{B}') - \frac{\rho}{c}\frac{d}{dt}(\mu'\times\mathbf{E}'), \tag{6.1}$$

where d/dt indicates the usual material derivative, and the primes indicate that the fields are measured in a comoving frame (i.e., by an observer moving with the velocity v of the medium), at the nonrelativistic limit. The expression (6.1) is none other than the Galilean invariant ponderomotive force obtained by de Groot and Mazur.[34] As to the Lorentz invariant electromagnetic power (5.8), it has exactly the same structure as the Galilean invariant expression obtained by Maugin and Eringen.[35] In fact, in terms of the expression ${}_M w$ obtained by these authors, we have up to relativistic terms of the order of v^2/c^2,

$${}_M\bar{w} = {}_M w - {}_M\mathbf{f}\cdot\mathbf{v}. \tag{6.2}$$

That is, ${}_M\bar{w}$ is the total electromagnetic power minus the power developed by the ponderomotive force. The comparison can be carried on. To that purpose, it is convenient to introduce the notion of electromagnetic energy—momentum tensor.

B. Electromagnetic energy-momentum tensor

de Groot and Suttorp[36] have shown that the expressions (5.2) and (5.3) are expressible in terms of a (nonsymmetric) electromagnetic energy—momentum tensor, ${}_M T^{\alpha\beta}$, in such a way that

$${}_M F^\alpha = -\nabla_\beta\, {}_M T^{\alpha\beta}, \quad {}_M L^{\alpha\beta} = {}_M T^{[\alpha\beta]}. \tag{6.3}$$

Introducing the electric displacement-magnetic intensity tensor $G^{\alpha\beta} = -G^{\beta\alpha}$, so that

$$G^{\alpha\beta} = F^{\alpha\beta} - \pi^{\alpha\beta}, \tag{6.4}$$

${}_M T^{\alpha\beta}$ has the following expression:

$${}_M T^{\alpha\beta} = F^{\alpha\gamma} G^\beta_{\cdot\gamma} + \tfrac{1}{4} F_{\mu\nu} F^{\nu\mu} g^{\alpha\beta}$$
$$+ c^{-2} u^\beta (F^{\alpha\gamma}\pi_{\gamma\epsilon}$$
$$- \pi^{\alpha\gamma} F_{\gamma\epsilon}) u^\epsilon - c^{-4} u^\alpha u^\beta u^\gamma F_{\gamma\epsilon}\pi^{\epsilon\xi} u_\xi. \tag{6.5}$$

The equations (6.3), (4.18), and (4.21) show that the present formulation is compatible with the standard theory of general relativity (in absence of spin), for we can define a total energy—momentum tensor, $T^{\alpha\beta}$, in such a way that

$$T^{\alpha\beta} \equiv {}_m T^{\alpha\beta} + {}_M T^{\alpha\beta}, \tag{6.6}$$

and

$$\nabla_\beta T^{\alpha\beta}=0, \quad T^{[\alpha\beta]}=0. \tag{6.7}$$

Decomposing $G^{\alpha\beta}$ as

$$G^{\alpha\beta}=\frac{1}{c}(u^\alpha \mathcal{D}^\beta - u^\beta \mathcal{D}^\alpha + \eta^{\alpha\beta\lambda\mu}\mathcal{H}_\lambda u_\mu), \tag{6.8}$$

where $\mathcal{D}^\beta=\mathcal{E}^\beta+\mathcal{P}^\beta$ and $\mathcal{H}_\lambda=\mathcal{B}_\lambda-\mathcal{M}_\lambda$ are the spatial electric-displacement and magnetic-intensity 4-vectors, respectively, we deduce from Eq. (6.5) the following canonical space and time decomposition of ${}_M T^{\alpha\beta}$,

$${}_M T^{\alpha\beta}=\tfrac{1}{2}(\mathcal{E}^2+\mathcal{B}^2)\frac{u^\alpha u^\beta}{c^2}+c^{-2}(u^\alpha \mathcal{S}^\beta + u^\beta \mathcal{S}^\alpha) - {}_M t^{\beta\alpha}, \tag{6.9}$$

where

$$\mathcal{S}^\beta \equiv c(\mathcal{E}*\mathcal{H})^\alpha, \tag{6.10}$$

and

$${}_M t^{\beta\alpha}=\mathcal{D}^\beta \mathcal{E}^\alpha + \mathcal{B}^\beta \mathcal{H}^\alpha - \tfrac{1}{2}(\mathcal{E}^2+\mathcal{B}^2-2\mathcal{M}\cdot\mathcal{B})P^{\alpha\beta}, \tag{6.11}$$

are the spatial Poynting 4-vector and the spatial electromagnetic stress tensor, respectively. The latter yields the spatial components

$${}_M t_{ji} \overset{*}{=} D'_j E'_i + B'_j H'_i - \tfrac{1}{2}(\mathbf{E}'^2+\mathbf{B}'^2-2\mathbf{M}'\cdot\mathbf{B}')\delta_{ji}, \tag{6.12}$$

in a local inertial frame. The electromagnetic stress tensor thus defined is the relativistic analog of the Galilean invariant one considered by various authors.[37]

We note that the skew symmetry of ${}_M T^{\alpha\beta}$ results solely from the skew symmetry of ${}_M t^{\beta\alpha}$. This is in agreement with the fact that in the present formulation ${}_M L^{\alpha\beta}$ is purely spatial [compare Eqs. (5.9) and (5.10)]. Obviously, the tensor (6.5) is none of the well-known energy–momentum tensors considered by Minkowski, Abraham, Einstein, and Laub and Dällenbach[38] at the beginning of this century. Nonetheless, it is intimately related to the electromagnetic energy–momentum tensor considered by Grot and Eringen[39] in special relativity, Maugin[40] in general relativity, and Israel[41] in the kinetic theory of spinning continua. This tensor reads

$${}_M \bar{T}^{\alpha\beta}=F^{\alpha\gamma}G^\beta_{\cdot\gamma}+\tfrac{1}{4}F_{\mu\nu}F^{\nu\mu}g^{\alpha\beta}. \tag{6.13}$$

Its much simpler form than that of the tensor ${}_M T^{\alpha\beta}$ allows a straightforward deduction of ${}_M \bar{T}^{\alpha\beta}$ from a variational principle.[42,43] Its canonical space and time decomposition is easily shown to be

$${}_M \bar{T}^{\alpha\beta}=\tfrac{1}{2}(\mathcal{E}^2+\mathcal{B}^2+2\mathcal{E}\cdot\mathcal{P})\frac{u^\alpha u^\beta}{c^2} + \frac{1}{c}u^\alpha(\mathcal{E}*\mathcal{H})^\beta + \frac{1}{c}(\mathcal{D}*\mathcal{B})^\alpha u^\beta - {}_M t^{\beta\alpha}, \tag{6.14}$$

where ${}_M t^{\beta\alpha}$ is the same as that defined by Eq. (6.11). Grot and Eringen considered that ${}_M F^\alpha$ and ${}_M L^{\alpha\beta}$ were given by

$${}_M F^\alpha = -\nabla_\beta {}_M \bar{T}^{\alpha\beta}, \quad {}_M L^{\alpha\beta} = {}_M \bar{T}^{[\alpha\beta]}. \tag{6.15}$$

A long calculation allows one to show that, in this case,

$${}_M F^\alpha = {}_M \check{f}^\alpha + \frac{1}{c^2}\,{}_M \check{w}\, u^\alpha, \tag{6.16}$$

where

$${}_M \check{f}^\alpha = q\mathcal{E}^\alpha + \frac{1}{c}(\mathcal{J}*\mathcal{B})^\alpha + P^{\alpha\beta}(\mathcal{P}^\mu \overset{\perp}{\nabla}_\beta \mathcal{E}_\mu + \mathcal{M}^\mu \overset{\perp}{\nabla}_\beta \mathcal{B}_\mu + {}_M p^\mu \overset{\perp}{\nabla}_\beta u_\mu), \tag{6.17}$$

and

$${}_M \check{w} = \mathcal{J}\cdot\mathcal{E} - \mathcal{P}^\mu D\mathcal{E}_\mu - \mathcal{M}^\mu D\mathcal{B}_\mu - {}_M p^\mu Du_\mu, \tag{6.18}$$

and

$${}_M L^{\alpha\beta}=\mathcal{P}^{[\alpha}\mathcal{E}^{\beta]}+\mathcal{M}^{[\alpha}\mathcal{B}^{\beta]}+{}_M p^{[\alpha}u^{\beta]}, \tag{6.19}$$

where ${}_M p^\alpha$ is defined as in Eq. (5.12). Then Eqs. (5.9) and (4.30) are replaced by

$${}_M \check{L}^\alpha \equiv {}_M p^\alpha, \tag{6.20}$$

$$p^\alpha = c^{-2}q^\alpha - {}_M p^\alpha, \tag{6.21}$$

whereas Eqs. (5.10) and (4.29) remain unaltered. It is then a simple matter to show that if one uses the above expressions one obtains the same final form for the field equations (5.17) and (5.18) and the Clausius–Duhem inequality (5.24) or (5.27). Indeed, in this new formulation, the spatial 4-force defined by Eq. (4.41) is easily shown to be equal to the 4-force defined by (5.16) on account of Eqs. (6.18) and (6.20). The energy density present in the left-hand side of Eq. (4.36) must depend on $\mathcal{E}_\alpha$ and $\mathcal{B}_\alpha$ in agreement with Eq. (6.19). Thus the local conservation of momentum is the same in both formulations. The same holds true as far as the energy equation is concerned, since

$${}_M \check{w} \equiv {}_M \tilde{w} - {}_M p^\alpha Du_\alpha, \tag{6.22}$$

and on account of the general expression (4.28) and the peculiar Eqs. (4.30), (5.9), and (6.21). Finally, the same Clausius–Duhem inequality (5.18) obtains on account of the general expression (4.34) and of Eqs. (4.30), (5.9), (6.21), and (6.22). The choice of the expressions (6.13) and (6.15) to start with seems to be much simpler but rather formal, while the choice (6.5) allows one to exhibit a strong relationship with an accepted microscopic model. As far as an axiomatic theory is concerned, this choice clearly is irrelevant.

APPENDIX A: COVARIANT FORMULATION OF MAXWELL'S EQUATIONS IN MATTER

The covariant formulation of Maxwell's equations at the event points $\mathbf{x}$ of the tube $\mathcal{T}$ considered in Sec. 3 is usually given in terms of the field $F_{\alpha\beta}$, $G^{\alpha\beta}$, and J^α. Indeed, the conservation of magnetic flux and, together, Ampère's and Gauss' laws are represented by the space–time equations (in Lorentz–Heaviside units)

$$\nabla_\beta \hat{F}^{\alpha\beta}=0, \tag{A1}$$

and

$$\nabla_\beta G^{\alpha\beta}=\frac{1}{c}J^\alpha, \tag{A2}$$

where $\hat{F}^{\alpha\beta}$ is the dual form of $F_{\alpha\beta}$, defined by

$$\hat{F}^{\alpha\beta}=-\tfrac{1}{2}\eta^{\alpha\beta\mu\nu}F_{\mu\nu}. \tag{A3}$$

It follows from Eq. (A3) that Eq. (A1) can also be written in the more usual form

$$\nabla_\gamma F_{\alpha\beta}+\nabla_\alpha F_{\beta\gamma}+\nabla_\beta F_{\gamma\alpha}=0. \tag{A4}$$

On account of Eqs. (3.15) and of the algebra of the alternation symbols, it is a simple matter to show that

$$\hat{F}^{\alpha\beta}=\frac{1}{c}(\mathcal{B}^\alpha u^\beta - \mathcal{B}^\beta u^\alpha - \eta^{\alpha\beta\gamma\delta}\mathcal{E}_\gamma u_\delta). \tag{A5}$$

Then using the decompositions (A5), (6.8), and (5.6), projecting Eqs. (A1) and (A2) onto $M_\perp$ and along u_α,

and using the definitions (3.17), (3.19), (3.22), and (4.39), it is possible to obtain a covariant formulation of Maxwell's equations of which the expressions are very close to those of the classical three-dimensional formulation. We obtain after some algebra[44]:

conservation of magnetic flux:

$$\overset{\perp}{\nabla}_\beta \mathcal{B}^\beta - \frac{2}{c}\omega^\alpha \mathcal{E}_\alpha = 0, \tag{A6}$$

Faraday's equations:

$$[(\overset{\perp}{\nabla} + c^{-2}\mathbf{a}) * \mathcal{E}]^\alpha + \frac{1}{c} D_C \mathcal{B}^\alpha = 0, \tag{A7}$$

Gauss' equation:

$$\overset{\perp}{\nabla}_\alpha \mathcal{D}^\alpha + \frac{2}{c}\omega^\alpha \mathcal{H}_\alpha = 0, \tag{A8}$$

Ampère's equations:

$$[(\overset{\perp}{\nabla} + c^{-2}\mathbf{a}) * \mathcal{H}]^\alpha - \frac{1}{c} D_C \mathcal{D}^\alpha = \frac{1}{c}\mathcal{J}^\alpha. \tag{A9}$$

Taking the covariant divergence of Eq. (A2), we obtain the equation of conservation of charge in the form

$$\nabla_\alpha J^\alpha = 0. \tag{A10}$$

Upon use of Eq. (5.6), this can be rewritten as

$$(D_C q + c^{-2}\mathcal{J}^\alpha Du_\alpha) + \overset{\perp}{\nabla}_\alpha \mathcal{J}^\alpha = 0. \tag{A11}$$

For an irrotational ($\omega^\alpha = 0$) and inertial ($a^\alpha = 0$) relativistic motion, Eqs. (A6)—(A11) take on the same form as the three-dimensional equations obtained in the Galilean invariant formulation of Maxwell's equations for matter.[45]

The system of differential equations (A6)—(A11) can be closed if and only if constitutive equations are given, for instance, for the fields $\mathcal{P}_\alpha$, $\mathcal{M}_\alpha$, and $\mathcal{J}_\alpha$. The Clausius—Duhem inequality (5.27) shows that these constitutive equations can be studied by the method of the thermodynamical admissibility (cf. Part II).

[1]R.A. Grot and A.C. Eringen, Int. J. Eng. Sci. 4, 639 (1966).
[2]Ph. Boulanger and G. Mayne, Bull. Acad. R. Sci. Belg. Cl. Sci. **57**, 872 (1971); C.R. Acad. Sci. Ser. A **274**, 591 (1972); Arch. Rat. Mech. Anal. **53**, 295 (1974).
[3]A. Romano, Meccanica **9**, 244 (1974).
[4]S.R. de Groot and L.G. Suttorp, Physica **39**, 84 (1968).
[5]S.R. de Groot and L.G. Suttorp, *Foundations of Electrodynamics* (North-Holland, Amsterdam, 1972).
[6]P. Penfield and H.A. Haus, *Electrodynamics of Moving Media* (M.I.T. Press, Cambridge, Mass., 1967).
[7]S.R. de Groot, *The Maxwell Equations—Nonrelativistic and Relativistic Derivations from Electron Theory* (North-Holland, Amsterdam, 1969).
[8]I. Brevik, Mat. Fys. Medd. Dan. Vidensk. Selsk. **37**, Nos. 11 and 13 (1970).
[9]L.I. Sedov, Appl. Math. Mech. (English translation of P.M.M.) **29**, 4 (1965); *Mécanique des Milieux Continus* (Editions MIR, Moscow, 1974), Vol. I, Chap. VI.
[10]Y.H. Pao, "Material Science Center Report" No. 2508, Cornell University, Ithaca, N.Y. (1975).
[11]S.R. de Groot and L.G. Suttorp, Physica **37**, 284, 297 (1967); **39**, 28, 41, 61, 84 (1968).
[12]Compare R.A. Grot, J. Math. Phys. **11**, 109 (1970); G.A. Maugin and A.C. Eringen, J. Math. Phys. **13**, 1777 (1972); G.A. Maugin, in *Continuum Physics*, edited by A.C. Eringen (Academic, New York, 1976), Vol. III, pp. 221—312.
[13]G.A. Maugin, Ann. Inst. H. Poincaré **15**, 275 (1971).
[14]G.A. Maugin and A.C. Eringen, J. Mécanique (Paris), **16**, 101 (1977).
[15]Cf. G.A. Maugin, J. Phys. A: Math. Nucl. Gen. **5**, 786 (1972); **6**, 306, 1647 (1973); 7, 818 (1974); Ann. Inst. H. Poincaré **20**, 41 (1974).
[16]B. Carter and H. Quintana, Proc. R. Soc. London A **331**, 57 (1972).
[17]G.A. Maugin, *Thèse de Doctorat d'Etat ès Sci. Math.* (Université de Paris, France, 1975).
[18]C. Cattaneo, *Mimeographed Lecture Notes* (Collège de France, Paris, 1962).
[19]Compare J. Ehlers, Abh. Math. Akad. W. Mainz **11**, 793 (1961).
[20]G.A. Maugin, C.R. Acad. Sci. Ser. A **273**, 65 (1971).
[21]The formulation of integral balance laws in special relativistic continuum physics is discussed in R.A. Grot and A.C. Eringen, Int. J. Eng. Sci. **4**, 611 (1966); and in G. Lianis and R.S. Rivlin, Arch. Rat. Mech. Anal. **48**, 64 (1972).
[22]This form of the second principle of thermodynamics is stated in agreement with the relativistic kinetic theory. Cf. J. Ehlers in *General Relativity and Cosmology*, edited by K. Sachs (Academic, New York, 1971).
[23]C. Eckart, Phys. Rev. **58**, 919 (1940).
[24]A.H. Taub, in *Proceedings of the First Symposium of Applied Mathematics* (American Math. Soc., Providence, R.I., 1949), pp. 148—57.
[25]Cf. Ref. 1.
[26]The controversy about this transformation is dealt with in the review of D. Ter Haar and H. Wergeland, Physics Rep. **1**, 31 (1971).
[27]See Ref. 14.
[28]Cf. G.A. Maugin, C.R. Acad. Sci. Ser. A **278**, 185 (1974).
[29]See Ref. 17.
[30]See Refs. 11.
[31]Cf. C. Truesdell, *Rational Thermodynamics* (McGraw-Hill, New York, 1969).
[32]Although a unique definition of objectivity in classical continuum mechanics (time-dependent rotational invariance in three-dimensional Euclidean space) is agreed upon, there is, at present, no such unique definition in relativistic continuum mechanics. Cf. in special relativity, L. Soderholm, Arch. Rat. Mech. Anal. **39**, 89 (1970); and in general relativity, G.A. Maugin, C.R. Acad. Sci. Ser. A **275**, 319 (1972) and G. Lianis, Nuovo Cimento B **14**, 57 (1973). The general relativistic definition which is closest to the classical one is that given by G.A. Maugin, in *Ondes et Radiations gravitationnelles* (Editions C.N.R.S., Paris, 1974), pp. 331—338.
[33]See Ref. 17.
[34]Compare Eq. (23) in Chap. XIV of S.R. de Groot and P. Mazur, *Non-Equilibrium Thermodynamics* (North-Holland, Amsterdam, 1962).
[35]Cf. Eq. (3.45) in Ref. 14.
[36]Cf. Refs. 11.
[37]Cf. G.A. Maugin and B. Collet, C.R. Acad. Sci. Ser. B **279**, 379, 439 (1974) and also Ref. 14 and H.F. Tiersten and C.F. Tsai, J. Math. Phys. **13**, 361 (1972). The same expression is given in a fixed Galilean inertial frame by G.H. Livens, *The Theory of Electricity* (Cambridge U.P., London, 1962), 2nd ed.
[38]See Ref. 8 for an analysis of these different tensors.
[39]See Ref. 1.
[40]See Ref. 13.
[41]W. Israel, in *Ondes et Radiations Gravitationnelles* (Editions C.N.R.S., Paris, 1974).
[42]See Refs. 12 and 13.
[43]A brief comparison of ${}_M\overline{T}^{\alpha\beta}$ and ${}_M T^{\alpha\beta}$ has already been given in G.A. Maguin's thesis, Princeton University, 1971.
[44]Cf. G.A. Maugin, J. Franklin Inst., "On Maxwell's Covariant Equations in Matter," in press.
[45]Cf. Ref. 14.

On the covariant equations of the relativistic electrodynamics of continua. II. Fluids

G. A. Maugin

Université de Paris VI, Laboratoire de Mécanique Théorique associé au C.N.R.S., Tour 66, 75230 Paris, Cedex 05, France
(Received 6 April 1977)

Based on the thermodynamical equations and conservation laws of a previous work which synthesized the axiomatic approach to relativistic continuum physics with the acceptance of the relativistically invariant ponderomotive force and couple and electromagnetic power derived by de Groot and Suttorp from a microscopical treatment, this paper develops a constitutive theory for general relativistic electromagnetic fluids. These include nonlinear and linear fluids with electrostriction and magnetostriction effects, dissipative fluids, and the "perfect magnetohydrodynamical" scheme as the extreme simplified case.

1. INTRODUCTION

Having established a useful form of the Clausius–Duhem inequality in Part I[1]—on the basis of axiomatic relativistic continuum thermodynamics and of the expressions derived by de Groot and Suttorp[2] from a microscopical approach for the electromagnetic "source terms" in the relativistic conservation laws—we propose in this part a brief thermodynamical constitutive theory for relativistic electromagnetic fluids. The equations deduced provide, also, a basis for various Galilean approximations in which small material velocities and weak gravitational fields are involved. First, exact nonlinear constitutive equations are derived on account of the Lorentz invariance of the free energy density for general nondissipative electromagnetic fluids (Sec. 2). Then the attention is focussed on the special case of constitutive equations linear in the electromagnetic fields (Sec. 3). This model exhibits both electrostrictive and magnetostrictive isotropic effects. Recombination with the "material" contributions which were separated in a somewhat arbitrary way from the "ponderomotive" contributions singled out by the deliberate choice made in Part I in favor of de Groot and Suttorp's expressions, allows us to answer some of the critical comments made by Brevik,[3] who favored Minkowski's electromagnetic energy–momentum tensor on the basis of various arguments. Here, however, it is shown, on the one hand, that the various static or quasistatic approximations yield the expected force densities (which are all in agreement), and, on the other hand, that the expected momentum densities of interest for the description of optical properties emerge quite naturally. Elementary dissipative phenomena are easily described on the same basis (Sec. 4). Finally, the extremely simplified scheme of perfect mangetohydrodynamics is shown to follow in a straightforward manner (Sec. 5). The main tool used to arrive at these various approximations is that provided by Legendre—i.e., contact—transformations of the energy densities. This is illustrated in the Appendix by showing that similar results can be obtained if one selects a different set of independent constitutive variables.

2. NONDISSIPATIVE ELECTROMAGNETIC FLUIDS

The general thermodynamics of electromagnetic fluids is governed by the local statement of the second principle of thermodynamics known as the *Clausius–Duhem inequality*. This inequality which has been obtained as Eq. (5.27) in Part I reads

$$-\rho(D\tilde{\psi}+\eta D\theta)+\tilde{t}^{\beta\alpha}d_{\alpha\beta}-\mathcal{P}_\alpha(D_C\mathcal{E}^\alpha)-\mathcal{M}_\alpha(D_C\mathcal{B}^\alpha)+\mathcal{J}_\alpha\mathcal{E}^\alpha-\theta^{-1}q^\alpha\overset{*}{\theta}_\alpha\geq 0. \tag{2.1}$$

In this equation ρ is the invariant proper mass density, $\tilde{\psi}$ is the proper density of free energy, η is the proper density of entropy, θ is the thermodynamical temperature, and $\overset{*}{\theta}_\alpha$ is the relativistic gradient of temperature defined by Eq. (I.4.35). $\mathcal{E}^\alpha$, $\mathcal{J}_\alpha$, $\mathcal{P}_\alpha$, $\mathcal{M}_\alpha$, and q^α are the spatial electric-field, conduction-current, polarization, magnetization, and heat-flux 4-vectors, respectively. $d_{\alpha\beta}$ is the spatial relativistic rate-of-strain tensor. $\tilde{t}^{\beta\alpha}$ is a symmetrical stress tensor which is related to the (in general, nonsymmetrical) Cauchy stress tensor $t^{\beta\alpha}$ by

$$t^{\beta\alpha}=\tilde{t}^{\beta\alpha}+\mathcal{E}^\beta\mathcal{P}^\alpha+\mathcal{B}^\beta\mathcal{M}^\alpha-(\mathcal{E}_\gamma\mathcal{P}^\gamma+\mathcal{B}_\gamma\mathcal{M}^\gamma)P^{\alpha\beta}. \tag{2.2}$$

Finally, $D\equiv u^\alpha\nabla_\alpha$ is the invariant time derivative if u^α is the world velocity, and the symbolism D_C indicates the contravariant convective time derivative such that

$$D_C\mathcal{E}^\alpha\equiv(D\mathcal{E}^\alpha)_\perp-\mathcal{E}^\beta\overset{\perp}{\nabla}_\beta u^\alpha+\mathcal{E}^\alpha\nabla_\beta u^\beta. \tag{2.3}$$

Here $\overset{\perp}{\nabla}_\beta$ is the spatial convariant derivative and the symbolism $(\cdots)_\perp$ indicates the spatial projection.

For a nondissipative electromagnetic fluid, $\mathcal{J}_\alpha=0$ and $q^\alpha=0$, so that inequality (2.1) reduces to Gibbs' equation

$$\rho D\tilde{\psi}=-\rho\eta D\theta+\tilde{t}^{\beta\alpha}d_{\alpha\beta}-\mathcal{P}_\alpha(D_C\mathcal{E}^\alpha)-\mathcal{M}_\alpha(D_C\mathcal{B}^\alpha). \tag{2.4}$$

Perfect electromagnetic fluids have constitutive equations which are derivable from the potential

$$F\equiv\rho\tilde{\psi}=\tilde{F}(\rho,\theta,\mathcal{E},\mathcal{B}). \tag{2.5}$$

The Lorentz invariance of $\tilde{F}$ requires that $\tilde{F}$ be form invariant under the connected Lie group $\mathcal{L}^\uparrow_+$ (orthochronous Lorentz transformations), hence in particular under the infinitesimal transformations (in rectangular coordinates)

$$x'_\alpha\overset{*}{=}(\delta_{\alpha\beta}+\epsilon L_{\alpha\beta})x_\beta\quad(L_{\alpha\beta}=-L_{\beta\alpha}), \tag{2.6}$$

of M^4 onto M^4, where ϵ is infinitesimally small and $L_{\alpha\beta}$ has arbitrary components. Applying this form invariance to $\tilde{F}$, we obtain the following constraint

 0022-2488/78/1905-1206$1.00

which must be satisfied by $\tilde{F}$ (and here rewritten in an arbitrary frame):

$$\frac{\partial \tilde{F}}{\partial \mathcal{E}_{[\alpha}}\mathcal{E}^{\beta]} + \frac{\partial \tilde{F}}{\partial \mathcal{B}_{[\alpha}}\mathcal{B}^{\beta]} = 0. \tag{2.7}$$

In virtue of the spatial character of $\mathcal{E}^\beta$ and $\mathcal{B}^\beta$, this constitutes a set of three linear differential equations of the first order, which, for the time being, does not need to be integrated. Then, computing $D\tilde{F}$ on account of the continuity equation (I.4.26),

$$D\rho + \rho \nabla_\beta u^\beta = 0, \tag{2.8}$$

and of Eqs. (2.3) and (2.7), and substituting the result in Eq. (2.4) posited to be valid for arbitrary non-vanishing objective[5] time rates $D\theta$, $d_{\alpha\beta}$, $D_C\mathcal{E}^\alpha$, and $D_C\mathcal{B}^\alpha$, we are led to the following constitutive equations:

$$\tilde{t}^{\beta\alpha} = \left[\left(\tilde{F} - \rho\frac{\partial \tilde{F}}{\partial \rho}\right) - \frac{\partial \tilde{F}}{\partial \mathcal{E}^\gamma}\mathcal{E}^\gamma - \frac{\partial \tilde{F}}{\partial \mathcal{B}^\gamma}\mathcal{B}^\gamma\right] P^{\alpha\beta} + \frac{\partial \tilde{F}}{\partial \mathcal{E}_{(\alpha}}\mathcal{E}^{\beta)} + \frac{\partial \tilde{F}}{\partial \mathcal{B}_{(\alpha}}\mathcal{B}^{\beta)}, \tag{2.9}$$

$$\mathcal{P}_\alpha = -\frac{\partial \tilde{F}}{\partial \mathcal{E}^\alpha}, \tag{2.10}$$

$$\mathcal{M}_\alpha = -\frac{\partial \tilde{F}}{\partial \mathcal{B}^\alpha}, \tag{2.11}$$

and

$$\eta = -\rho^{-1}\frac{\partial \tilde{F}}{\partial \theta}. \tag{2.12}$$

Substituting for Eqs. (2.9)—(2.11) into Eq. (2.2), we obtain

$$t^{\beta\alpha} = -pP^{\alpha\beta} = t^{\alpha\beta}, \tag{2.13}$$

where the thermodynamical pressure p is defined by

$$p = \tilde{p}(\rho, \theta, \mathcal{E}, \mathcal{B}) = \rho\frac{\partial \tilde{F}}{\partial \rho} - \tilde{F}. \tag{2.14}$$

We remark that the Cauchy stress tensor $t^{\beta\alpha}$ is now symmetrical. Indeed, on account of Eqs. (2.10) and (2.12) and of Eq. (I.5.10), we see that the pondero-motive-couple 2-form vanishes identically in perfect electromagnetic fluids, hence the symmetry of $t^{\beta\alpha}$ in virtue of Eq. (I.4.29). Furthermore, integrating the system (2.7) of linear partial differential equations by the classical method of characteristics and accounting for the fact that $\mathcal{B}^\alpha$ is an axial 4-vector while the invariance required for $\tilde{F}$ is that under the *proper* Lorentz group, we deduce that $\tilde{F}$ necessarily has the following functional dependence:

$$F = \tilde{F}(\rho, \theta, I_{(k)}), \tag{2.15}$$

where the invariants $I_{(k)}$, $k = 1, 2, 3$, are such that

$$I_{(1)} = \tfrac{1}{2}\mathcal{E}^2, \quad I_{(2)} = \tfrac{1}{2}\mathcal{B}^2, \quad I_{(3)} = \tfrac{1}{2}(\mathcal{E}_\gamma\mathcal{B}^\gamma)^2. \tag{2.16}$$

It follows that Eqs. (2.10) and (2.11) yield

$$\mathcal{P}^\alpha = \tilde{\alpha}_{(1)}\mathcal{E}^\alpha + \tilde{\alpha}_{(3)}\mathcal{B}^\alpha, \quad \mathcal{M}^\alpha = \tilde{\alpha}_{(2)}\mathcal{B}^\alpha + \tilde{\alpha}_{(3)}\mathcal{E}^\alpha, \tag{2.17}$$

where

$$\tilde{\alpha}_{(1)} = -\frac{\partial \tilde{F}}{\partial I_{(1)}}, \quad \tilde{\alpha}_{(2)} = -\frac{\partial \tilde{F}}{\partial I_{(2)}}, \quad \tilde{\alpha}_{(3)} = -\frac{\partial \tilde{F}}{\partial I_{(3)}}\sqrt{2I_{(3)}}. \tag{2.18}$$

Of course, $\tilde{\alpha}_{(i)} = \tilde{\alpha}_{(i)}(\rho, \theta, I_{(k)})$, $i = 1, 2, 3$. These scalar coefficients are positive provided $\tilde{F}$ has the appropriate variational behavior with respect to the variables $I_{(k)}$. Equations (2.17) are *exact nonlinear* electromagnetic constitutive equations. By the same token Eq. (2.14) takes on the form

$$p = \tilde{p}(\rho, \theta, I_{(k)}) \equiv \rho\frac{\partial \tilde{F}}{\partial \rho} - \tilde{F}, \tag{2.19}$$

where $\tilde{F}$ is given by the general invariant function (2.15).

3. LINEAR ELECTROMAGNETIC CONSTITUTIVE EQUATIONS

Linear approximations of Eqs. (2.17) for weak electromagnetic fields are obtained by discarding terms containing $\tilde{\alpha}_{(3)}$—which are at least quadratic in the fields—and setting

$$\chi^e(\rho, \theta) \equiv \tilde{\alpha}_{(1)}\big|_{I_{(k)}=0}, \quad \alpha_2(\rho, \theta) \equiv \tilde{\alpha}_{(2)}\big|_{I_{(k)}=0}, \tag{3.1}$$

so that

$$\mathcal{P}^\alpha = \chi^e(\rho, \theta)\mathcal{E}^\alpha, \quad \mathcal{M}^\alpha = \alpha_2(\rho, \theta)\mathcal{B}^\alpha.$$

Now consider the following special case of Eq. (2.15):

$$F = F_0(\rho, \theta) - [\chi^e I_{(1)} + \alpha_2 I_{(2)}]. \tag{3.2}$$

Then Eq. (2.13) yields

$$t^{\beta\alpha} = -\left\{p_0 + \tfrac{1}{2}\mathcal{E}^2\left[\chi^e - \rho\left(\frac{\partial \chi^e}{\partial \rho}\right)_\theta\right] + \tfrac{1}{2}\mathcal{B}^2\left[\alpha_2 - \rho\left(\frac{\partial \alpha_2}{\partial \rho}\right)_\theta\right]\right\} P^{\alpha\beta}, \tag{3.3}$$

with

$$p_0 \equiv \rho\frac{\partial F_0}{\partial \rho} - F_0.$$

Define the dielectric "constant" ϵ and the magnetic permeability μ by

$$\epsilon \equiv 1 + \chi^e, \quad \mu^{-1} \equiv 1 - \alpha_2, \tag{3.4}$$

so that the spatial electric displacement 4-vector, $\mathcal{D}^\alpha$ and the spatial magnetic field 4-vector $\mathcal{H}^\alpha$ are given by

$$\mathcal{D}^\alpha = \mathcal{E}^\alpha + \mathcal{P}^\alpha = \epsilon(\rho, \theta)\mathcal{E}^\alpha, \tag{3.5a}$$

$$\mathcal{H}^\alpha = \mathcal{B}^\alpha - \mathcal{M}^\alpha = \mu^{-1}(\rho, \theta)\mathcal{B}^\alpha. \tag{3.5b}$$

The general expression (I.6.9) of de Groot and Suttorp's electromagnetic energy—momentum tensor then reads

$${}_M T^{\alpha\beta} = \tfrac{1}{2}(\mathcal{E}^2 + \mathcal{B}^2)\frac{u^\alpha u^\beta}{c^2} + \mu^{-1}(u^\alpha \mathcal{G}^\beta + u^\beta \mathcal{G}^\alpha) - [\epsilon\mathcal{E}^\alpha\mathcal{E}^\beta + \mu^{-1}\mathcal{B}^\alpha\mathcal{B}^\beta - \tfrac{1}{2}(\mathcal{E}^2 + \mathcal{B}^2 - 2\alpha_2\mathcal{B}^2)P^{\alpha\beta}], \tag{3.6}$$

where

$$\mathcal{G}^\alpha \equiv \frac{1}{c}(\mathcal{E} * \mathcal{B})^\alpha, \quad \mathcal{G}^\alpha u_\alpha = 0. \tag{3.7}$$

By the same token the "matter" energy—momentum tensor (I.4.6) can be written as

$${}_m T^{\alpha\beta} = \rho[c^2 + \epsilon_0(v, \eta) + \frac{v\chi^e}{2}\mathcal{E}^2 - \frac{v\alpha_2}{2}\mathcal{B}^2]\frac{u^\alpha u^\beta}{c^2} - t^{\beta\alpha} \tag{3.8}$$

on account of Eqs. (I.4.30) and (I.5.9) and account of the fact that $q^\alpha = 0$ and the excess-energy-current $\hat{Q}^\beta$ introduced in Part I has been set equal to zero. In Eq. (3.8), $v \equiv \rho^{-1}$ is the proper specific volume, and the internal energy $\epsilon_0(v, \eta)$ has been introduced on account of the definition (I.4.32) and of Eqs. (3.2) and (I.5.15). I.e.,

$$\epsilon = \tilde{\epsilon} + \mathcal{E}_\alpha \pi^\alpha = \tilde{\psi} + \eta\theta + v\mathcal{E}_\alpha P^\alpha = \epsilon_0 - \tfrac{1}{2}(v\alpha_2 \mathcal{B}^2 - v\chi^e \mathcal{E}^2), \tag{3.9}$$

where

$$\epsilon_0(v, \eta) = vF_0(\rho^{-1}, \theta) + \eta\theta. \tag{3.10}$$

On account of this Legendre transformation p_0 and θ are now given by the constitutive equations

$$p_0(v, \eta) = -\left(\frac{\partial \epsilon_0}{\partial v}\right)_\eta, \quad \theta(v, \eta) = \left(\frac{\partial \epsilon_0}{\partial \eta}\right)_v \tag{3.11}$$

in terms of the equation of state $\epsilon_0 = \epsilon_0(v, \eta)$.

Using now $\mathcal{H}^\alpha$ instead of $\mathcal{B}^\alpha$ on account of Eq. (3.5a), the total "matter plus field" energy–momentum tensor (I.6.6) has the following space and time decomposition

$$T^{\alpha\beta} = {}_m T^{\alpha\beta} + {}_M T^{\alpha\beta} = c^{-2}\omega_{(\mathrm{tot})} u^\alpha u^\beta + \mu^{-1}(u^\alpha \mathcal{G}^\beta + u^\beta \mathcal{G}^\alpha) - t^{\beta\alpha}_{(\mathrm{tot})}, \tag{3.12}$$

where the total energy density $\omega_{(\mathrm{tot})}$ and the total spatial (symmetrical) relativistic stress tensor $t^{\beta\alpha}_{(\mathrm{tot})}$ are given by

$$\omega_{(\mathrm{tot})} = \rho\left(c^2 + \epsilon_0 + \frac{v\epsilon}{2}\mathcal{E}^2 + \frac{v\mu}{2}\mathcal{H}^2\right) \tag{3.13}$$

and

$$t^{\beta\alpha}_{(\mathrm{tot})} = -\left\{p_0 + \tfrac{1}{2}\mathcal{E}^2\left[\epsilon - \rho\left(\frac{\partial \epsilon}{\partial \rho}\right)_\theta\right] + \tfrac{1}{2}\mathcal{H}^2 \times \left[\mu - \rho\left(\frac{\partial \mu}{\partial \rho}\right)_\theta\right]\right\} P^{\alpha\beta} + \epsilon \mathcal{E}^\alpha \mathcal{E}^\beta + \mu \mathcal{H}^\alpha \mathcal{H}^\beta, \tag{3.14}$$

respectively. The last expression contains isotropic effects of electrostriction and magnetostriction via the dependence of ϵ and μ upon ρ. These effects are also expressible in terms of derivatives with respect to v taken at constant entropy if, in agreement with Eqs. (3.11), we consider the functional dependence $\epsilon(v, \eta)$ and $\mu(v, \eta)$. The expresssion (3.14) is important for, in agreement with Brevik,[6] its electrostatic approximation, or, rather, the electrostatic approximation of the total spatial 4-force

$$f^\alpha_{(\mathrm{tot})} = [\nabla_\beta t^{\beta\alpha}_{(\mathrm{tot})}]_\perp, \tag{3.15}$$

allows one to test the validity of the expression chosen for the electromagnetic energy–momentum tensor. Indeed, in the *electrostatic* approximation where the magnetic fields are discarded and $\mathcal{E}$ can be replaced by the three-dimensional electric field $\mathbf{E}$ expressed in a fixed Galilean inertial frame, while the remaining Maxwell equations read

$$\nabla \cdot \mathbf{D} = q, \tag{3.16a}$$

$$\nabla \times \mathbf{E} = 0, \tag{3.16b}$$

Eq. (3.15) yields the total three-dimensional force acting on the fluid in the form

$$\mathbf{f}_{(\mathrm{tot})} = -\nabla p_0 + \left\{q\mathbf{E} + \nabla\left[\rho\frac{\mathbf{E}^2}{2}\left(\frac{\partial \epsilon}{\partial \rho}\right)_\theta\right] - \frac{\mathbf{E}^2}{2}\nabla\epsilon\right\}. \tag{3.17}$$

The expression within braces is the so-called Helmholtz force.[7] For a dielectric fluid $q = 0$. Taking account of the fact that $\epsilon = \epsilon(\rho, \theta)$, Eq. (3.17) then takes the form

$$\mathbf{f}_{(\mathrm{tot})} = -\nabla p_0 + \frac{\rho}{2}\nabla\left[\frac{\mathbf{E}^2}{2}\left(\frac{\partial \epsilon}{\partial \rho}\right)_\theta\right] - \frac{\mathbf{E}^2}{2}\left(\frac{\partial \epsilon}{\partial \theta}\right)_\rho \nabla\theta, \tag{3.18}$$

which is the expression given by Landau and Lifshitz.[8] In the case of a dielectric liquid where the free charge density q is due to injected charges (through electrodes injecting ions or with an electron beam), and where the electrostriction effect can be neglected so that $\epsilon = \epsilon(\theta)$ only, Eq. (3.17) reduces to

$$\mathbf{f}_{(\mathrm{tot})} = -\nabla p_0 + {}_e\mathbf{f}, \tag{3.19}$$

where ${}_e\mathbf{f}$ is the so-called Korteweg–Helmholtz electrostatic force[9] defined as

$${}_e\mathbf{f} = q\mathbf{E} - \frac{\mathbf{E}^2}{2}\nabla\epsilon. \tag{3.20}$$

This expression is of central importance in electrohydrodynamics. On account of Eqs. (3.16), we have

$${}_e f_i = {}_e t_{ji,j}, \tag{3.21}$$

where

$${}_e t_{ji} \equiv \epsilon(E_j E_i - \tfrac{1}{2}\mathbf{E}^2\delta_{ji}) = {}_e t_{ij} \tag{3.22}$$

may be referred to as the electric stress tensor.

Another way of arriving at the expression of the total force is to consider the *electrostatic* approximation of the right-hand side of Eq. (I.5.7). On account of Eqs. (2.13) and (I.5.16), and (3.16b), we obtain

$$\mathbf{f}_{(\mathrm{tot})} = -\nabla p + {}_K\mathbf{f}, \tag{3.23}$$

where ${}_K\mathbf{f}$ is Kelvin's force density,[10] which is given by

$${}_K\mathbf{f} = q\mathbf{E} + (\mathbf{P} \cdot \nabla)\mathbf{E}. \tag{3.24}$$

It is easily shown that Eqs. (3.23) and (3.17) are in agreement by noting that, in electrostatics, p and p_0 are related by

$$p = p_0 + \tfrac{1}{2}\mathbf{E}^2\left[\chi^e - \rho\left(\frac{\partial \chi^e}{\partial \rho}\right)_\theta\right] \tag{3.25}$$

on account of Eqs. (2.19) and (3.2).

Other consequences of the linear constitutive equations (3.5) can be noted by way of conclusion of this section. First, it is a simple matter to show that, on account of the decompositions (I.5.4) and (I.6.8), the 4-vectorial constitutive equations (3.5) are equivalent to the single tensorial constitutive equation

$$G^{\alpha\beta} = C^{\alpha\beta\gamma\delta} F_{\gamma\delta}, \tag{3.26}$$

where

$$C^{\alpha\beta\gamma\delta} = -C^{\beta\alpha\gamma\delta} = -C^{\alpha\beta\delta\gamma}$$

and

$$C^{\alpha\beta\gamma\delta} = \mu^{-1}\left[P^{\alpha\gamma} - \left(\frac{n}{2}\right)^2 u^\alpha u^\gamma\right]\left[P^{\beta\delta} - \left(\frac{n}{c}\right)^2 u^\beta u^\delta\right]$$
$$= \mu^{-1}\left[g^{\alpha\gamma} - \left(\frac{n^2-1}{c}\right)\frac{u^\alpha u^\gamma}{c}\right]\left[g^{\beta\delta} - \left(\frac{n^2-1}{c}\right)\frac{u^\beta u^\delta}{c}\right], \tag{3.27}$$

$n \equiv \sqrt{\epsilon\mu}$ being the refractive index of the medium. Finally, substituting from Eqs. (3.5) into the definition (I.5.12) of the induced momentum density, here a spatial 4-vector, we find

$$_M p^\alpha = \frac{(n^2-1)}{c}(\mathcal{E} * \mathcal{H})^\alpha. \tag{3.28}$$

This is nothing but the induced momentum density invoked in the case of an optical wave in the experiment of Jones and Richard,[11] but expressed in covariant form in a comoving frame. This momentum runs away with the electromagnetic field whose energy 4-current is given by the spatial Poynting vector $\mathcal{S}^\alpha \equiv c(\mathcal{E} * \mathcal{H})^\alpha$ and whose spatial 4-momentum $_M\bar{p}^\alpha$ is given by

$$_M\bar{p}^\alpha = \frac{1}{c}(\mathcal{E} * \mathcal{H})^\alpha = c^{-2}\mathcal{S}^\alpha \tag{3.29}$$

on account of the space and time decomposition (I.6.9). The difference between $_M p^\alpha$ and $_M\bar{p}^\alpha$ then is

$$_M p^\alpha - {}_M\bar{p}^\alpha = \frac{n^2}{c}(\mathcal{E} * \mathcal{H})^\alpha \equiv \frac{1}{c}(\mathcal{D} * \mathcal{B})^\alpha. \tag{3.30}$$

This is Minkowski's momentum density expressed in covariant form in a comoving frame.[12]

4. ELEMENTARY DISSIPATIVE PROCESSES

Discarding any electromagnetic dissipative processes such as hysteresis effects (the media described are either dielectrics or paramagnetic or diamagnetic media) without light absorption, we find on account of Eqs. (2.1) and (2.4) that the dissipative contribution (if it exists), $_D t^{\beta\alpha} = {}_D t^{\alpha\beta}$, to the relativistic stress tensor, the conduction current, and the heat flux satisfy the following dissipation inequality:

$$\Phi \equiv {}_D t^{\beta\alpha} d_{\alpha\beta} + \mathcal{J}_\alpha \mathcal{E}^\alpha - \theta^{-1} q^\alpha \overset{*}{\theta}_\alpha \geq 0. \tag{4.1}$$

Linear *isotropic* constitutive equations which satisfy both Eq. (4.1) and the Onsager–Casimir pinciple, are

$$_D t^{\beta\alpha} = \eta^{\beta\alpha\lambda\mu} d_{\lambda\mu}, \tag{4.2}$$

$$\mathcal{J}_\alpha = \sigma(\rho, \theta)\mathcal{E}_\alpha + A(\rho, \theta)\overset{*}{\theta}_\alpha, \tag{4.3}$$

and

$$q_\alpha = -[\kappa(\rho, \theta)\overset{*}{\theta}_\alpha + \theta A(\rho, \theta)\mathcal{E}_\alpha] \tag{4.4}$$

with

$$\eta^{\beta\alpha\lambda\mu} = \eta_1(\rho, \theta)P^{\beta\alpha}P^{\lambda\mu} + \eta_2(\rho, \theta)(P^{\beta\lambda}P^{\alpha\mu} + P^{\beta\mu}P^{\alpha\lambda}) \tag{4.5}$$

and

$$3\eta_1 + 2\eta_2 \geq 0, \quad \eta_2 \geq 0,$$
$$\sigma \geq 0, \quad \kappa \geq 0, \quad \sigma\kappa - \theta^{-1}A^2 \geq 0. \tag{4.6}$$

η_1 and η_2 are viscosities, κ is the thermal conductivity, σ is the electrical conductivity, and A is the material coefficient which accounts for the Thomson and Peltier effects. If there is no such couplings, $A \equiv 0$, then

$$\mathcal{J}^\alpha = \sigma(\rho, \theta)\mathcal{E}^\alpha, \tag{4.7a}$$

$$q_\alpha = -\kappa(\rho, \theta)\overset{*}{\theta}_\alpha. \tag{4.7b}$$

The last equation is a relativistic version of Fourier's law.[13] For a perfect electricity conductor, σ goes to infinity. The magnitude of the total electric current (I.5.6) then remains finite if and only if

$$\mathcal{E}^\alpha \equiv 0. \tag{4.8}$$

5. PERFECT RELATIVISTIC MAGNETOHYDRODYNAMICS

The "perfect magnetohydrodynamic" scheme of relativistically moving fluids corresponds to the following hypotheses: The material continuum is an isotropic fluid with

$$\eta_1 = \eta_2 = \kappa = A = 0, \quad \sigma \to \infty, \quad \mu = \text{const.} \tag{5.1}$$

The magnetostriction effect then is discarded while condition (4.8) must be fulfilled. It then follows that $\mathcal{S}^\alpha = \mathcal{J}^\alpha = 0$, and Eq. (3.12) reduces to

$$T^{\alpha\beta} = T^{\alpha\beta}_{(\rho_0 f)} + T^{\alpha\beta}_{(\mathcal{H})}, \tag{5.2}$$

where

$$T^{\alpha\beta}_{(\rho_0 f)} \equiv \rho(c^2 + \epsilon_0)\frac{u^\alpha u^\beta}{c^2} + p_0 P^{\alpha\beta} \tag{5.3}$$

and

$$T^{\alpha\beta}_{(\mathcal{H})} \equiv \mu\mathcal{H}^2(\tfrac{1}{2}g^{\alpha\beta} + c^{-2}u^\alpha u^\beta) - \mu\mathcal{H}^\alpha\mathcal{H}^\beta \tag{5.4}$$

represents the energy–momentum tensor of a relativistic perfect fluid (the magnetic field being turned off) and the energy–momentum tensor due to the magnetic field alone, respectively. The latter is the same as that which is deduced from Minkowski's tensor under the same hypotheses.[14] The splitting achieved in Eq. (5.2) is not as arbitrary as it would appear to be, since $T^{\alpha\beta}_{(\rho_0 f)}$ contains no magnetic field at all. Introducing the thermodynamical function known as the *index* f of the fluid and defined by

$$f \equiv 1 + c^{-2}(\epsilon_0 + vp_0), \tag{5.5}$$

we can rewrite Eq. (5.2) in the equivalent form[15]

$$T^{\alpha\beta} = (\rho f + \mu c^{-2}\mathcal{H}^2)u^\alpha u^\beta + (p_0 + \tfrac{1}{2}\mu\mathcal{H}^2)g^{\alpha\beta} - \mu\mathcal{H}^\alpha\mathcal{H}^\beta. \tag{5.6}$$

Remarks: (i) The equations derived in Secs. 2–5 have been obtained by considering the formulation (2.1) of the Clausius–Duhem inequality. There, from the magnetic viewpoint, it is the magnetic induction 4-vector $\mathcal{B}^\alpha$ which is taken as a constitutive independent variable. Other choices, however, are possible as, for instance, those of $\mathcal{M}^\alpha$ or $\mathcal{H}^\alpha$. In the Appendix we show that identical results can be obtained if one considers $\mathcal{H}^\alpha$ (under the perfect magnetohydrodynamic hypothesis).

(ii) The fluids described by Eq. (2.15) or (3.2) are isotropic. That is, they exhibit no preferred direction in their mechanical and electromagnetic behaviors. Recent studies,[16] however, have been devoted to what is referred to as relativistic *anisotropic* magnetohydrodynamics. The main peculiarity of such fluids

is that, at rest, they do not obey Euler's hydrodynamics. This behavior which closely resembles that of liquid crystals,[17] can be reproduced by considering that, in addition to the already present independent variables, Eqs. (2.15) and (3.2) depend also on a unit vector field d^α, ($d^\alpha u_\alpha = 0$, $P_{\alpha\beta} d^\alpha d^\beta = 1$), which points instantaneously in the preferred direction around which the properties of the fluid are invariant by rotation. In relativistic magnetohydrodynamics, this vector field may be none other than the magnetic field itself, normalized to unity. Developing such a scheme presents no major difficulties and will not be done here.

APPENDIX

On account of the relationship $H^\alpha = \mathcal{B}^\alpha - \mathcal{M}^\alpha$, we note that

$$\mathcal{M}_\alpha D_c \mathcal{B}^\alpha = \mathcal{M}_\alpha D_c H^\alpha + \rho D \frac{v}{2} \mathcal{M}^2 - (\mathcal{M}^\alpha \mathcal{M}^\beta - \mathcal{M}^2 P^{\alpha\beta}) d_{\alpha\beta}. \tag{A1}$$

Defining a new free enegy density $\check{\psi}$ and a new symmetrical relativistic stress tensor $\check{t}^{\beta\alpha}$ by

$$\check{\psi} = \tilde{\psi} + \frac{v}{2} \mathcal{M}^2, \tag{A2}$$

and

$$\check{t}^{\beta\alpha} = \tilde{t}^{\beta\alpha} + \mathcal{M}^\alpha \mathcal{M}^\beta - \mathcal{M}^2 P^{\alpha\beta} = \check{t}^{\alpha\beta}, \tag{A3}$$

and using the hypotheses of the perfect magnetohydrodynamic scheme ($\mathcal{P}^\alpha = \mathcal{G}^\alpha = \mathcal{J}^\alpha = q^\alpha = 0$), we can rewrite Gibbs' equation (2.4) in the following manner:

$$\rho D\check{\psi} = -\rho\eta D\theta + \check{t}^{\beta\alpha} d_{\alpha\beta} - \mathcal{M}_\alpha D_c H^\alpha. \tag{A4}$$

By the same token the Cauchy stress tensor (2.2) is given by

$$t^{\beta\alpha} = \check{t}^{\beta\alpha} + \mathcal{M}^\alpha H^\beta - \mathcal{M}^\gamma H_\gamma P^{\alpha\beta}. \tag{A5}$$

Constitutive equation for the present scheme are derivable from the potential per unit volume,

$$\check{F} = \rho\check{\psi} = \check{F}(\rho, \theta, H). \tag{A6}$$

The Lorentz invariance of $\check{F}$ leads to the linear first-order partial differential equation [compare Eq. (2.7)]

$$\frac{\partial \check{F}}{\partial H_{(\alpha}} H^{\beta)} = 0, \tag{A7}$$

which has the integral

$$F = \check{F}(\rho, \theta, \tfrac{1}{2}H^2). \tag{A8}$$

Computing $\rho D\check{\psi}$ from Eq. (A6), taking account of Eq. (A7) and of the definition (2.3) applied to H^α, we find that Eq. (A4) is identically satisfied for arbitrary (nonvanishing) independent objective time rates $D\theta$, $d_{\alpha\beta}$, and $D_c H^\alpha$, if and only if the constitutive equations are

$$\check{t}^{\beta\alpha} = -\left(p + \frac{\partial \check{F}}{\partial H^\gamma} H^\gamma\right) P^{\alpha\beta} + \frac{\partial \check{F}}{\partial H_{(\alpha}} H^{\beta)}, \tag{A9}$$

$$\mathcal{M}_\alpha = -\partial \check{F}/\partial H^\alpha, \tag{A10}$$

$$\eta = -v\partial \check{F}/\partial\theta, \tag{A11}$$

with

$$p \equiv \rho \frac{\partial \check{F}}{\partial \rho} - \check{F}. \tag{A12}$$

If follows from Eqs. (A7), (A9), and (A10) that Eq. (A5) yields

$$t^{\beta\alpha} = -pP^{\alpha\beta} = t^{\alpha\beta}. \tag{A13}$$

Consider now the special case,

$$\check{F} = F_0(\rho, \theta) - \tfrac{1}{2}\chi^m(\rho, \theta) H^2, \tag{A14}$$

of Eq. (A8). Then Eq. (A10) gives

$$\mathcal{M}^\alpha = \chi^m(\rho, \theta) H^\alpha, \tag{A15}$$

while Eqs. (A13) and (I.6.11) yield

$$t^{\beta\alpha} = -\left\{p_0 + \tfrac{1}{2}H^2\left[\chi^m - \rho\left(\frac{\partial \chi^m}{\partial \rho}\right)_\theta\right]\right\} P^{\alpha\beta}, \tag{A16}$$

and

$${}_M t^{\beta\alpha} = \mu H^\alpha H^\beta - \tfrac{1}{2}\mu H^2 P^{\alpha\beta}, \tag{A17}$$

where

$$p_0 \equiv \rho \frac{\partial F_0}{\partial \rho} - F_0, \quad \mu \equiv 1 + \chi^m. \tag{A18}$$

Moreover, defining the internal energy ϵ_0 as in Eq. (3.10) and noting that the internal energy ϵ present in the general equation (I.4.6) is related to ϵ_0 by

$$\epsilon = \epsilon_0 - \tfrac{1}{2}v\chi^m \mu H^2 \tag{A19}$$

on account of Eqs. (3.10), (5.15), (A2), (A6), and (A14), neglecting the dependence of χ^m upon ρ and gathering the results (A16), (A17), and (A19) in the expression (I.6.6) of the total energy—momentum tensor, we are led to the expression (5.2). Q.E.D

[1] G. A. Maugin, Part I, J. Math. Phys. **19**, 1198 (1978).
[2] S. R. de Groot and L. G. Suttorp, Physica **37**, 284, 297 (1967); **39**, 28, 41, 61, 84 (1968); *Foundations of Electrodynamics*, (North-Holland, Amsterdam, 1972).
[3] I. Brevik, Mat. Fys. Medd. Dans. Vidensk. Selsk. **37**, No. 13 (1970).
[4] The notation is that of Part I to which we refer. The equations of Part I are referred to by I followed by their number, e.g., Eq. (I.5.27).
[5] These fields are "objective" in the following sense. They obey the so-called principle of objectivity or principle of material frame indifference in general relativity as formulated by the author [G. A. Maugin, in *Ondes et Radiations Gravitationnelles* (Editions C.N.R.S., Paris, 1974), pp. 331—8]. Their tetrad components are form invariant under proper time dependent rotation of the tetrad—this invariance is related to spinorial algebra. They are also "rheologically invariant" according to the invariance principle set forth by J. G. Oldroyd, Proc. R. Soc. London Ser. A **316**, 1 (1970).
[6] Cf. Ref. 3.
[7] Compare Eq. (2.20) in Ref. 3. For the experimental validity of Helmholtz' force we refer to the comments of Brevik in Ref. 3, p. 13.
[8] Cf. L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media* (Pergamon, Oxford, 1960), Sec. 15.
[9] This expression was deduced from a variational principle by Korteweg and Helmholtz; cf. J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill, New York, 1941), pp. 137—40. It is also referred to as Minkowski's force density by Brevik [cf. Ref. 3, Eq. (2.19)].
[10] Cf. Eq. (2.15) in Ref. 3, and also Lord Kelvin (Sir W. Thomson), *Reprints of Papers on Electrostatics and Magnetism* (MacMillan, London, 1884), 2nd ed.
[11] R. V. Jones and J. C. S. Richards, Proc. R. Soc. London Ser. A **221**, 480 (1954).

[12]H. Minkowski, Nach. Ges. Wiss. Göttingen 53 (1908); Math. Ann. **68**, 472 (1910).

[13]We know that the oversimplified constitutive equation (4.7b) violates the principle of relativistic causality in that it yields a parabolic equation of heat. The solution to this paradox requires considerably more involved relativistic heat conduction laws. A review of such laws (there is no unique solution) is given in G.A. Maugin, J. Phys. A Math. Nucl. Gen. **7**, 465 (1974).

[14]Cf A. Lichnerowicz, *Rdativistic Hydrodynamics and Magnetohydrodynamics* (Benjamin, New York, 1967).

[15]Up to the signature chosen for the space—time metric, Eq. (5.8) is the same as that on which several authors have based their mathematical studies of relativistic magnetohydrodynamics; e.g., in Ref. 14 and in K.O. Friedrichs, Commun. Pure Appl. Math. **27**, 749 (1974). Equation (5.8) and the corresponding Einstein field equations have been deduced simultaneously from various variational formulations; see A.H. Taub, in *Relativistic Fluid Dynamics*, edited by C. Cattaneo (Cremonese, Rome, 1971), and G.A. Maugin, C.R. Acad. Sci. Paris **274A**, 602 (1972); Ann. Inst. H. Poincaré **16**, 133 (1972).

[16]Cf. M. Cissoko, C.R. Acad. Sci. Ser. A **278**, 463 (1974); *Thèse de Doctorat ès. Sci. Math.* (Université de Paris VI, 1975), and other references quoted by this author.

[17]P.G. De Gennes, *Liquid Crystals* (Oxford U.P., London, 1974).

On the covariant equations of the relativistic electrodynamics of continua. III. Elastic solids

G. A. Maugin

Université de Paris VI, Laboratoire de Mécanique Théorique associé au C.N.R.S., Tour 66, 75230 Paris, Cedex 05, France
(Received 6 April 1977)

Based on the thermodynamical equations of a previous work, which synthesized the axiomatic approach to relativistic continuum physics with the acceptance of the relativistically invariant ponderomotive force and couple and electromagnetic power derived by de Groot and Suttorp from a microscopical treatment, this paper develops a constitutive theory for general relativistic deformable solids. After a review of the essentials of deformation theory of general relativistic continua, particular attention is paid to nonlinear elastic electromagnetic solids in normal conditions of pressure and temperature, then to the linear piezoelectric scheme which can be used for the theoretical analysis of gravitational-wave detectors that use a piezoelectric device, and, finally, to magnetoelasticity under high pressure which describes the magnetomechanical behavior of matter in certain astrophysical objects (neutron stars).

1. INTRODUCTION

This paper is a continuation of two other papers herein after referred to as Parts I and II.[1,2] Upon accepting the general thermodynamical inequality—the Clausius–Duhem inequality—derived in Part I and which resulted from a compromise between a purely axiomatic approach to relativistic continuum physics and the reliance upon specific electromagnetic source terms as obtained by de Groot and Suttorp[3] from a microscopical approach, we develop in the present part a thermodynamical constitutive theory for general relativistic electromagnetic deformable solids. Instead of constructing a complete logico-deductive theory of the behavior of such materials, we content ourselves with the study of specific schemes of matter, which can be of practical use in problems involving both the effects of gravitational fields and of electric and/or magnetic fields.

Having exposed the essentials of deformation theory of general relativistic continua in sec. 2, we construct in sec. 3 a quite general case, that of nonlinear thermoelastic electromagnetic solids. This is done for normal conditions of pressure, density, and temperature, when the deformation mapping from an ideally relaxed, unstressed state to the present state is sufficiently smooth and involves no structural change. Although rigorous and generalizing to the general relavistic framework the Galilean theory of thermo-electro-magneto-elasticity, the description thus obtained is of weak practical interest. Next, in sec. 4, we concentrate upon the case of solid dielectrics in general relativity. There the whole description is made without recourse to a three-dimensional background. The constitutive equations obtained at first are nonlinear. However, by studying the superimposition of infinitesimal strains, electric fields and space–time metric variation on a finite initial state of strain and fields, we are able to construct constitutive equations valid for initially strained ferroelectrics and pyroelectrics. Such equations must enable one to study photoelastic and electro-optical effects in the material if the latter is transparent. Furthermore, if this linearization procedure is performed about an initially nonpolarized and negligibly strained state, then there follows the constitutive equations of piezoelectric bodies, which can be used in gravitational-wave detectors in which the influence of possibly incident waves can be detected via the induced piezoelectric effect. These equations are a general relativistic generalization of Voigt's classical equations of linear piezoelectricity in that they account for infinitesimal perturbations in the space–time metric. In the last section, considering the anomalous conditions of pressure and magnetic field existing in astrophysical objects such as neutron stars, we develop a theory of general relativistic magnetoelasticity valid under conditions of extremely high pressure. For this purpose the hypotheses set forth by Carter and Quintana[4] are used. In particular, all fields depend upon the matter density as can be expected.

2. ELEMENTS OF DEFORMATION THEORY

A. Nonlinear deformations[5,6]

The relativistic motion of a continuum is described either by means of a canonical differentiable projection $\mathcal{P}$ such that $\mathcal{P}: \mathcal{T}[B] \to \mathcal{H}^3$, or with the aid of the space–time parametrized congruence of world lines $\mathcal{C}: \mathbf{x} = \chi(\mathbf{X}, \tau)$, $\mathbf{X} \in B$, $\tau \in \mathbb{R}$. Here $\mathcal{T}[B]$ is the open tube of the space–time $\mathcal{V}^4$ which is swept out by the material body B (whose constituents are the material "particles" $\mathbf{X}$) and $\mathcal{H}^3$ is the three-dimensional manifold which serves to describe the material continuum. B is an open region of $\mathcal{H}^3$. τ is the proper time of $\mathbf{X}$. In terms of local charts x^α, $(\alpha = 1, 2, 3, 4)$ and X^K $(K = 1, 2, 3)$, these two manifolds are equipped with the background metrics $g_{\alpha\beta}$ and G_{KL}, respectively, and we have

$$\mathcal{P}: X^K = X^K(x^\alpha), \tag{2.1a}$$

$$\tau = \tau(x^\alpha), \tag{2.1b}$$

and

$$\mathcal{C}: x^\alpha = \chi^\alpha(X^K, \tau). \tag{2.2}$$

Both relations (2.1a) and (2.2) are assumed to be sufficiently differentiable, so that the direct and inverse relativistic deformation gradients,

$$x^\alpha_K \equiv \left(\frac{\partial \chi^\alpha}{\partial X^K}\right)_\perp, \quad (x^\alpha_K u_\alpha = 0), \tag{2.3}$$

 0022-2488/78/1905-1212$1.00

and

$$\partial_\alpha X^K \equiv \frac{\partial X^K}{\partial x^\alpha} \qquad (u^\alpha \partial_\alpha X^K = 0), \tag{2.4}$$

are well defined. The last equation within parentheses is none other than $DX^K = 0$, which signifies that X^K and τ are independent variables. Here $D \equiv u^\alpha \nabla_\alpha$, where u^α is the world velocity of **X**, and the symbolism $(\circ\circ\circ)_\perp$ indicates the spatial projection obtained with the help of the projection operator $P^\alpha_{\circ\beta} \equiv \delta^\alpha_\beta + c^{-2} u^\alpha u_\beta$.

$g^{\alpha\beta}$ being the reciprocal of $g_{\alpha\beta}$, a space–time invariant (but tensor field on $\mathcal{M}^3$) is formed from the spatial 4-vector $\partial_\alpha X^K$ by

$$\overset{1}{C}{}^{KL} \equiv g^{\alpha\beta} \partial_\alpha X^K \partial_\beta X^L = P^{\alpha\beta} \partial_\alpha X^K \partial_\beta X^L = \overset{1}{C}{}^{LK}. \tag{2.5}$$

This defines the relativistic analog of the Piola strain tensor of classical continuum mechanics.[7] The geometrical significance of definition (2.5) is clear: The relativistic Piola strain tensor is the image of the space–time metric $g^{\alpha\beta}$ by the projection of the space–time on its quotient by the congruence (2.2). Noting from Eqs. (2.3) and (2.4) that we have

$$(\partial_\beta X^K) x^\alpha_K = P^\alpha_{\circ\beta}, \quad x^\alpha_L (\partial_\alpha X^K) = \delta^K_L, \tag{2.6}$$

it is also possible to construct the relativistic version of the Cauchy, Lagrange, Green, and Euler strain tensors by

$$C_{KL} \equiv g_{\alpha\beta} x^\alpha_K x^\beta_L = P_{\alpha\beta} (\partial_K x^\alpha)(\partial_L x^\beta) = C_{LK}, \tag{2.7}$$

$$E_{KL} \equiv \tfrac{1}{2}(C_{KL} - G_{KL}) = E_{LK}, \tag{2.8}$$

$$c_{\alpha\beta} \equiv G_{KL} \partial_\alpha X^K \partial_\beta X^L = c_{\beta\alpha}, \tag{2.9}$$

and

$$\mathcal{E}_{\alpha\beta} \equiv \tfrac{1}{2}(P_{\alpha\beta} - c_{\alpha\beta}) = \mathcal{E}_{\beta\alpha}, \tag{2.10}$$

respectively. $c_{\alpha\beta}$ and $\mathcal{E}_{\alpha\beta}$ are spatial symmetrical tensors. It is readily checked that

$$\mathcal{E}_{\alpha\beta} = E_{KL} \partial_\alpha X^K \partial_\beta X^L, \quad E_{KL} = \mathcal{E}_{\alpha\beta} x^\alpha_K x^\beta_L. \tag{2.11}$$

Let $G = \det \|G_{KL}\|$ and $\eta^{KLM} \equiv G^{-1/2} \mathcal{E}^{KLM}$, where $\mathcal{E}^{KLM}$ is the three-dimensional alternation symbol. Then, with the notation of Part I, the Jacobian determinant of the mapping (2.2) is given by

$$J = \frac{1}{6c} \eta_{\alpha\beta\gamma\delta} \eta^{KLM} x^\alpha_K x^\beta_L x^\gamma_M u^\delta,$$

i.e.,

$$J = \frac{1}{3!} \overset{*}{\eta}_{\alpha\beta\gamma} \eta^{KLM} x^\alpha_K x^\beta_L x^\gamma_M = (-g/G)^{1/2} \det \|x^\alpha_K\|. \tag{2.12}$$

Let $\rho_0(\mathbf{X})$ be the matter density at $\mathbf{X} \in B \subset \mathcal{M}^3$ at a certain proper time $\tau = \tau_0$. We have $D\rho_0(\mathbf{X}) = 0$. Then the invariant relativistic density of proper mass, ρ, can be formally defined as being the image of ρ_0 by the projection $\mathcal{P}$. Hence

$$\rho(\mathbf{X}, \tau) = \rho_0(\mathbf{X}) J^{-1}(\mathbf{X}, \tau). \tag{2.13}$$

B. Rates of deformation

We recall that

$$e_{\alpha\beta} \equiv (\nabla_\beta u_\alpha)_\perp \tag{2.14}$$

and

$$d_{\alpha\beta} \equiv e_{(\alpha\beta)} \tag{2.15}$$

defines the spatial relativistic velocity gradient and the rate-of-strain tensor, respectively. From Eqs. (2.3), (2.4), and (2.12) it then is a simple matter to show that

$$[D(\partial_\alpha X^K)]_\perp = - e^\beta_{\circ\alpha} \partial_\beta X^K, \tag{2.16a}$$

$$(D x^\alpha_K)_\perp = e^\alpha_{\circ\lambda} x^\lambda_K, \tag{2.16b}$$

$$DJ = J d^\alpha_{\circ\alpha}. \tag{2.16c}$$

Noting that $DG_{KL} = 0$ and $D\rho_0 = 0$, it follows from Eqs. (2.7), (2.8), (2.5), and (2.13) that

$$DE_{KL} = x^\alpha_K d_{\alpha\beta} x^\beta_L, \tag{2.17}$$

$$D\overset{1}{C}{}^{KL} = -2 P^{\mu\alpha} d_{\alpha\beta} P^{\beta\nu} (\partial_\mu X^K)(\partial_\nu X^L), \tag{2.18}$$

and

$$D\rho = -\rho d^\alpha_{\circ\alpha} = -\rho P^{\alpha\beta} d_{\alpha\beta}. \tag{2.19}$$

Whereas on account of the definitions of $P_{\alpha\beta}$, $c_{\alpha\beta}$, and $\mathcal{E}_{\alpha\beta}$, and that of the Lie derivative with respect to the world velocity u^α, we have

$$d_{\alpha\beta} \equiv \tfrac{1}{2} (\underset{u}{\pounds} P_{\alpha\beta})_\perp, \tag{2.20}$$

$$(\underset{u}{\pounds} c_{\alpha\beta})_\perp \equiv 0, \tag{2.21}$$

and

$$(\underset{u}{\pounds} \mathcal{E}_{\alpha\beta})_\perp \equiv d_{\alpha\beta}. \tag{2.22}$$

The result (2.21) and Eq. (2.9) indicate that $c_{\alpha\beta}$ serves as a local background metric to measure strains in space–time (in the same way as G_{KL} serves as a local background metric to measure strains on $\mathcal{M}^3$).

Finally, let us note the following result. Let A^α be a contravariant spatial 4-vector field. If we construct the space–time scalars

$$A^K \equiv J A^\alpha (\partial_\alpha X^K) = J D_A X^K, \tag{2.23}$$

where D_A indicates the invariant derivative in the direction of A^α, we deduce from Eq. (2.16) and the definition (I.4.39) of the contravariant convective time derivative, noted D_C, that

$$DA^K = J (D_c A^\alpha)(\partial_\alpha X^K). \tag{2.24}$$

C. Infinitesimal deformations in space-time

The spatial tensor field which measures infinitesimal deformations in space–time can be obtained by considering the infinitesimal variation $(\delta \mathcal{E}_{\alpha\beta})_\perp$ of $\mathcal{E}_{\alpha\beta}$, which results from both a variation of the space–time metric and a variation of the space–time event. It follows from the definition of $P_{\alpha\beta}$ that

$$(\delta P_{\alpha\beta})_\perp = (\delta g_{\alpha\beta})_\perp \equiv h_{\alpha\beta} = h_{\beta\alpha}. \tag{2.25}$$

$h_{\alpha\beta}$ is the spatial perturbation in the space–time metric. Noting that [compare Eq. (2.16a)]

$$[\delta(\partial_\alpha X^K)]_\perp = - P^{\mu\lambda} (\partial_\mu X^K)(\nabla_\alpha \xi_\lambda)_\perp, \tag{2.26}$$

where $\xi^\lambda \equiv \delta x^\lambda$, we deduce from Eq. (2.9) and the fact that G_{KL} remains invariant in the variation procedure that

$$(\delta c_{\alpha\beta})_\perp = - P^{\lambda\mu} [c_{\mu\beta} (\nabla_\alpha \xi_\lambda)_\perp + c_{\alpha\mu} (\nabla_\beta \xi_\lambda)_\perp]. \tag{2.27}$$

Upon again using Eq. (2.10) and carrying the resulting expression and Eq. (2.25) in the variation $(\delta\mathcal{E}_{\alpha\beta})_\perp$ computed from Eq. (2.10), we arrive at the variation

$$(\delta\mathcal{E}_{\alpha\beta})_\perp = \tfrac{1}{2}\{h_{\alpha\beta} + P^{\lambda\mu}[(P_{\mu\beta} - 2\mathcal{E}_{\mu\beta}) \times (\nabla_\alpha \xi_\lambda)_\perp + (P_{\alpha\mu} - 2\mathcal{E}_{\alpha\mu})(\nabla_\beta \xi_\lambda)_\perp]\}. \tag{2.28}$$

Linearizing the right-hand side of Eq. (2.28) by discarding the contributions of $\mathcal{E}_{\mu\beta}$, we obtain the spatial infinitesimal-strain tensor

$$E_{\alpha\beta} = \tfrac{1}{2}[h_{\alpha\beta} + (\nabla_\alpha \xi_\beta + \nabla_\beta \xi_\alpha)_\perp] = E_{\beta\alpha}. \tag{2.29}$$

Clearly, the contribution of $h_{\alpha\beta}$ is a purely general relativistic effect.[8] In special relativity we thus have $E_{\alpha\beta} = (\nabla_{(\alpha}\xi_{\beta)})_\perp = \tilde{E}_{\alpha\beta}$, an expression which is entirely similar to that of classical continuum mechanics, ξ_α playing the role of the infinitesimal displacement. Define the spatial rotation tensor by

$$\tilde{\Omega}_{\lambda\gamma} = -(\nabla_{[\gamma}\xi_{\lambda]})_\perp = -\tilde{\Omega}_{\gamma\lambda}. \tag{2.30}$$

Then using the decomposition of $(\nabla_\gamma \xi_\lambda)_\perp$ in symmetrical and antisymmetrical parts, we deduce from Eq. (2.29) that

$$(\nabla_\gamma \xi_\lambda)_\perp = E_{\gamma\lambda} - (\tfrac{1}{2}h_{\lambda\gamma} + \tilde{\Omega}_{\lambda\gamma}). \tag{2.31}$$

3. THERMOELASTIC ELECTROMAGNETIC INSULATORS

The general thermodynamics of relativistically moving electromagnetic deformable solids is governed by the local statement of the second principle of thermodynamics, or Clausius–Duhem inequality, Eq. (I.5.24),

$$-\rho(D\hat{\psi} + \eta D\theta) + {}_E t^{\beta\alpha} d_{\alpha\beta} + \mathcal{J}^\alpha \mathcal{E}_\alpha + \mathcal{E}_\alpha (D_c \mathcal{P}^\alpha) + \mathcal{B}_\alpha (D_c \mathcal{M}^\alpha) - \theta^{-1} q^\alpha \overset{*}{\theta}_\alpha \geq 0. \tag{3.1}$$

Here $\hat{\psi}$ is the free energy per unit of proper mass, η is the entropy density per unit of proper mass, θ is the (proper) thermodynamical temperature. $\mathcal{J}^\alpha$, $\mathcal{E}_\alpha$, $\mathcal{B}_\alpha$, $\mathcal{P}^\alpha$, and $\mathcal{M}^\alpha$ are, respectively, the conduction-current, electric-field, magnetic-induction, polarization, and magnetization spatial 4-vectors. q^α is the spatial heat-flux 4-vector and $\overset{*}{\theta}_\alpha$ is the relativistic temperature gradient defined by Eq. (I.4.35). Finally, ${}_E t^{\beta\alpha}$ is a spatial symmetrical stress tensor which is related to the relativistic Cauchy stress tensor by

$$t^{\beta\alpha} = {}_E t^{\beta\alpha} - \mathcal{P}^\beta \mathcal{E}^\alpha - \mathcal{M}^\beta \mathcal{B}^\alpha \neq t^{\alpha\beta}. \tag{3.2}$$

It is convenient at this point to define various fields on $\mathcal{M}^3$ in terms of the spatial tensor fields which appear in Eq. (3.1) with the help of projection (2.1a). Indeed, define ${}_E T^{KL}$, $\mathcal{P}^K$, $\mathcal{M}^K$, $\mathcal{J}^K$, Q^K, $\mathcal{E}_K$, $\mathcal{B}_K$, and Θ_K by:

$${}_E T^{KL} \equiv J_E t^{\beta\alpha} \partial_\beta X^K \partial_\alpha X^L = {}_E T^{LK}, \tag{3.3}$$

$$\mathcal{P}^K \equiv J\mathcal{P}^\alpha \partial_\alpha X^K, \quad \mathcal{M}^K \equiv J\mathcal{M}^\alpha \partial_\alpha X^K, \tag{3.4}$$

$$\mathcal{J}^K \equiv J\mathcal{J}^\alpha \partial_\alpha X^K, \quad Q^K \equiv Jq^\alpha \partial_\alpha X^K, \tag{3.5}$$

and

$$\mathcal{E}_K \equiv \mathcal{E}_\alpha X^\alpha_K, \tag{3.6a}$$

$$\mathcal{B}_K \equiv \mathcal{B}_\alpha X^\alpha_K, \tag{3.6b}$$

$$\Theta_K \equiv \overset{*}{\theta}_\alpha X^\alpha_K. \tag{3.6c}$$

Then, on account of Eqs. (3.3)–(3.6), of Eqs. (2.13) and (2.17), and of identities of the type (2.24), we can rewrite Eq. (3.1) in the following form:

$$-\rho_0(D\hat{\psi} + \eta D\theta) + {}_E T^{KL} DE_{KL} + \mathcal{J}^K \mathcal{E}_K + \mathcal{E}_K D\mathcal{P}^K + \mathcal{B}_K D\mathcal{M}^K - \theta^{-1} Q^K \Theta_K \geq 0. \tag{3.7}$$

A homogeneous thermoelastic electromagnetic insulator is such that $\mathcal{J}^K = 0$ and its constitutive equations have the following *a priori* functional dependence,

$$\psi = \hat{\psi}(\partial_\alpha X^K, \mathcal{P}^\alpha, \mathcal{M}^\alpha, \theta, \overset{*}{\theta}_\alpha). \tag{3.8}$$

An analogous dependence holds good to start with for ${}_E T^{KL}$, $\mathcal{E}_K$, $\mathcal{B}_K$, and Q^K. The Lorentz invariance of the free energy function (3.8) is identically satisfied if ψ has the following reduced functional dependence:

$$\psi = \hat{\psi}(\bar{C}^{-1}_{KL}, \mathcal{P}^K, \mathcal{M}^K, \theta, \Theta_K). \tag{3.9}$$

We note from Eqs. (2.8), (2.17), and (2.18) and the fact that $\bar{C}^{-1}_{KL} C_{LM} = \delta^K_M$ that

$$D\bar{C}^{-1}_{MN} = -2\bar{C}^{-1}_{MK} \bar{C}^{-1}_{NL} DE_{KL}. \tag{3.10}$$

Now we compute $D\hat{\psi}$ from Eq. (3.9) on account of Eq. (3.10), carrying the resulting expression in Eq. (3.7), and assuming that the latter is to be satisfied for arbitrary independent time rates $D\theta$, DE_{KL}, $D\mathcal{P}^K$, and $D\mathcal{M}^K$ while noting that the factors of these rates do not depend on the rates themselves—by virtue of the very reduction (3.9)—whereas Q^K depends on Θ_K. This is realized if and only if we have the following invariant constitutive equations[9]:

$${}_E T^{KL} = -2\rho_0 \frac{\partial \bar{\psi}}{\partial \bar{C}^{-1}_{MN}} \bar{C}^{-1}_{MK} \bar{C}^{-1}_{NL}, \tag{3.11}$$

$$\mathcal{E}_K = \rho_0 \frac{\partial \bar{\psi}}{\partial \mathcal{P}^K}, \quad \mathcal{B}_K = \rho_0 \frac{\partial \bar{\psi}}{\partial \mathcal{M}^K}, \quad \eta = -\frac{\partial \bar{\psi}}{\partial \theta}, \tag{3.12}$$

where the dependence of ψ has been reduced to

$$\psi = \bar{\psi}(\bar{C}^{-1}_{KL}, \mathcal{P}^K, \mathcal{M}^K, \theta) \tag{3.13}$$

and Q^M, which still is of the form

$$Q^M = \bar{Q}^M(\bar{C}^{-1}_{KL}, \mathcal{P}^K, \mathcal{M}^K, \theta, \Theta_K), \tag{3.14}$$

satisfies the remaining dissipation inequality

$$\Phi = -\theta^{-1} Q^M \Theta_M \geq 0. \tag{3.15}$$

Furthermore, if $\bar{Q}^M$ is assumed to be of class C^1 in its arguments Θ_K, $K = 1, 2, 3$, then Eq. (3.15) implies by continuity that

$$\bar{Q}^M(\bar{C}^{-1}_{KL}, \mathcal{P}^K, \mathcal{M}^K, \theta, 0) = 0. \tag{3.16}$$

The spatial constitutive equations corresponding to the results thus derived are obtained by inverting Eqs. (3.3), (3.4), and (3.5b) and using Eq. (3.2). We thus have

$$t^{\beta\alpha} = -\left(2\frac{\partial \bar{F}}{\partial \bar{C}^{-1}_{KL}} P^{\beta\mu} \partial_\mu X^L + \frac{\partial \bar{F}}{\partial \mathcal{P}^K} \mathcal{P}^\beta + \mathcal{M}^\beta \frac{\partial \bar{F}}{\partial \mathcal{M}^K}\right) P^{\alpha\nu}(\partial_\nu X^K), \tag{3.17}$$

$$\mathcal{E}^\alpha = P^{\alpha\mu}(\partial_\mu X^K) \frac{\partial \bar{F}}{\partial \mathcal{P}^K}, \quad \mathcal{B}^\alpha = P^{\alpha\mu}(\partial_\mu X^K) \frac{\partial \bar{F}}{\partial \mathcal{M}^K}, \tag{3.18}$$

$$\eta = -\rho_0^{-1} \frac{\partial \bar{F}}{\partial \theta}, \tag{3.19}$$

and

$$q^\alpha = J^{-1} x^\alpha_M \bar{Q}^M(\overset{1}{C}{}^{KL},\ \rho^K,\ \mathcal{M}^K,\ \theta,\ \Theta_M), \tag{3.20}$$

where

$$\bar{F} = \rho_0 \bar{\psi} = \bar{F}(\overset{1}{C}{}^{KL},\ \rho^K,\ \mathcal{M}^K,\ \theta,\ \rho_0). \tag{3.21}$$

A linear approximation of q^α in the neighborhood of $\Theta_M = 0$ then yields a heat flux constitutive equation of the form[10]

$$q^\alpha = -K^{\alpha\beta} \overset{*}{\theta}_\beta \tag{3.22}$$

on account of Eq. (3.6c) and of a definition of the type

$$K^{\alpha\beta} = -J^{-1} x^\alpha_M x^\beta_L \bar{Q}^{ML}(\overset{1}{C}{}^{PQ},\ \rho^Q,\ \mathcal{M}^Q,\ \theta) \tag{3.23}$$

for the spatial heat conductivity tensor.

Equations (3.17)—(3.19) and Eq. (3.22) are constitutive equations which are *nonlinear* in the deformation field and in the electromagnetic fields. They are valid for large deformation fields and strong electromagnetic fields in normal conditions of pressure and temperature. In particular, it has been implicitly assumed that the projection (2.1a) is well behaved and that the Jacobian determinant J keeps the same sign during the deformation process, no peculiar change of material structure occuring in these normal conditions. Hence exotic situations such as those encountered in certain astrophysical objects (white dwarfs, neutron stars) are not accounted for by the equations above. Such situations require particular attention and the development of special models which account for the influence of the density in an appropriate manner. (See Sec. 5 below.) Given the general functional dependence (3.22)—note that the material symmetry has not been specified—Eqs. (3.17)—(3.18) account in an intricate way for coupled nonlinear effects of elasticity, piezoelectricity, and piezomagnetism, pyroelectricity, pyromagnetism, electrostriction, and magnetostriction. They clearly are too general to be of direct practical use. However, they show that such effects can be reproduced in the general relativistic frame. They generalize to electromagnetic bodies the equations recently developed, for instance, by Barrabes.[11] In the absence of heat conduction, they coincide with the equations deduced previously (along with Maxwell's equations and Einstein field equations) by the author from a variational principle.[12] They contain as special cases both the case of thermoelastic nonpolarizable ($\mathcal{P}^\alpha \equiv 0$) nonlinear paramagnetic and diamagnetic insulators and the case of nonmagnetizable ($\mathcal{M}^\alpha \equiv 0$) dielectrics. In virtue of their nonlinearity, they also contain the case of so-called soft ferromagnetic elastic bodies (in which spin effects are discarded). However, they are insufficient to describe hard ferromagnetic elastic bodies (in which spin and exchange effects must be taken into account.[13] The formulation given above places in evidence the role of the strain state described on the material manifole $\mathcal{M}^3$. To arrive at sensible spatial linearized expressions in the case of dielectrics we shall avoid any reference to such a state in the next section.

4. PIEZOELECTRIC SOLIDS

A. Exact nonlinear theory

For the sake of simplicity we consider the case of nonmagnetizable ($\mathcal{M}^\alpha = 0$), nondissipative ($\mathcal{J}^\alpha = q^\alpha = 0$, no hysteresis) dielectrics. Then Eq. (3.1) yields Gibbs' equation,

$$\rho D\hat{\psi} = -\rho\eta D\theta + {}_E t^{\beta\alpha} d_{\alpha\beta} + \mathcal{E}_\alpha (D_C \mathcal{P}^\alpha). \tag{4.1}$$

In order to be closer to the conventional formulation of piezoelectricity[14] it is preferable to express the right-hand side of Eq. (4.1) in terms of $\mathcal{E}_\alpha$ as an independent constitutive variable while keeping ${}_E t^{\alpha\beta}$ as the thermodynamical dual of $d_{\alpha\beta}$. This is achieved by noting that

$$\mathcal{E}_\alpha (D_C \mathcal{P}^\alpha) = \rho D(\mathcal{E}_\alpha \pi^\alpha) - \mathcal{P}^\beta (\underset{u}{\pounds} \mathcal{E}_\beta)_\perp, \tag{4.2}$$

where $\pi^\alpha \equiv \mathcal{P}^\alpha / \rho$, and $\underset{u}{\pounds} \mathcal{E}_\beta$ is the covariant Lie derivative defined by

$$\underset{u}{\pounds} \mathcal{E}_\beta = D\mathcal{E}_\beta + \mathcal{E}^\lambda \nabla_\beta u_\lambda. \tag{4.3}$$

Then, on account of Eq. (2.22) and of the Legendre transformation

$$\bar{\psi} = \hat{\psi} - \mathcal{E}_\alpha \pi^\alpha, \tag{4.4}$$

Eq. (4.1) takes the form

$$\rho D\bar{\psi} = -\rho\eta D\theta + {}_E t^{\beta\alpha} (\underset{u}{\pounds} \mathcal{E}_{\alpha\beta}) - \mathcal{P}^\beta (\underset{u}{\pounds} \mathcal{E}_\beta)_\perp. \tag{4.5}$$

The constitutive equations of nonmagnetizable nondissipative piezoelectric solids are derivable from the potential

$$\psi = \bar{\psi}(\mathcal{E}_{\alpha\beta},\ \mathcal{E}_\alpha,\ \theta). \tag{4.6}$$

To be valid in special relativity this scalar expression must be Lorentz invariant. Let us define $\psi^{\alpha\beta}$ and ψ^α by

$$\psi^{\alpha\beta} \equiv \frac{\partial \bar{\psi}}{\partial \mathcal{E}_{\alpha\beta}}, \quad \psi^\alpha \equiv \frac{\partial \bar{\psi}}{\partial \mathcal{E}_\alpha}. \tag{4.7}$$

The form invariance of $\bar{\psi}$ under infinitesimal transformations of M^4 onto M^4 of the type (in rectangular coordinates)

$$x'_\alpha \overset{*}{=} (\delta_{\alpha\beta} + \mathcal{E} L_{\alpha\beta}) x_\beta \quad (L_{\alpha\beta} = -L_{\beta\alpha}), \tag{4.8}$$

where $\mathcal{E}$ is infinitesimally small and $L_{\alpha\beta}$ has arbitrary components, is satisfied by discarding terms of the order of $\mathcal{E}^2$ if and only if $\bar{\psi}$ satisfies the following set of first-order linear partial differential equations [compare Eqs. (II. 2.7)],

$$2\psi^{\gamma[\alpha} \mathcal{E}^{\beta]}_{\cdot\gamma} + \psi^{[\alpha} \mathcal{E}^{\beta]} = 0. \tag{4.9}$$

These equations are rewritten in an arbitrary frame. Projecting them orthogonally to u^α and along the direction of u_α then yields the constraints

$$\psi^{\gamma\alpha} \equiv (\psi^{\gamma\alpha})_\perp, \quad \psi^\alpha \equiv (\psi^\alpha)_\perp \tag{4.10}$$

after a little algebra. We thus have

$$D\bar{\psi} = \psi^{\alpha\beta} (D\mathcal{E}_{\alpha\beta})_\perp + \psi^\alpha (D\mathcal{E}_\alpha)_\perp + \left(\frac{\partial \bar{\psi}}{\partial \theta}\right) D\theta. \tag{4.11}$$

On account of Eq. (4.3) of the fact that $\mathcal{E}_{\alpha\beta}$ is symmetrical,

$$(D\mathcal{E}_{\alpha\beta})_\perp = (\underset{u}{\pounds} \mathcal{E}_{\alpha\beta})_\perp - 2\mathcal{E}_{\lambda(\alpha} e^\lambda_{\cdot\beta)}, \tag{4.12}$$

and using the decomposition of $e_{\lambda\beta}$ into symmetrical and skew symmetrical parts and Eq. (4.9), we transform Eq. (4.11) to

$$D\bar{\psi} = [\psi^{\lambda(\beta} (P^{\alpha)}_{\cdot\lambda} - 2\mathcal{E}^{\alpha)}_{\cdot\lambda}) - \psi^{(\beta} \mathcal{E}^{\alpha)})] (\underset{u}{\pounds} \mathcal{E}_{\alpha\beta})_\perp + \psi^\alpha (\underset{u}{\pounds} \mathcal{E}_\alpha)_\perp + \left(\frac{\partial \bar{\psi}}{\partial \theta}\right) D\theta. \tag{4.13}$$

Gibbs' equation (4.5) being posited to be valid for arbitrary, independent, nonvanishing, objective time rates $D\theta$, $(\overset{u}{\mathcal{L}}\mathcal{E}_{\alpha\beta})_\perp$, and $(\overset{u}{\mathcal{L}}\mathcal{E}_{\alpha})_\perp$, we are led to the constitutive equations

$$ {}_E t^{\beta\alpha} = {}_E\hat{t}^{\beta\alpha} + \rho^{(\beta}\mathcal{E}^{\alpha)}, \tag{4.14} $$

$$ \rho^\beta = -\rho\psi^\beta, \quad \eta = -\frac{\partial\bar{\psi}}{\partial\theta} \tag{4.15} $$

with[15]

$$ {}_E\hat{t}^{\beta\alpha} \equiv \rho\psi^{\lambda(\beta}(P^{\alpha)}_{\cdot\lambda} - 2\mathcal{E}^{\alpha)}_{\cdot\lambda}) = \rho(\psi^{\alpha\beta} - 2\psi^{\lambda(\alpha}\mathcal{E}^{\beta)}_{\cdot\lambda}) = {}_E\hat{t}^{\alpha\beta}. \tag{4.16} $$

Then Eq. (3.2) yields

$$ t^{\beta\alpha} = {}_E\hat{t}^{\beta\alpha} - \rho^{[\beta}\mathcal{E}^{\alpha]}. \tag{4.17} $$

Taking the skew symmetric part of both sides of this equation yields $t^{[\beta\alpha]} = \rho^{[\alpha}\mathcal{E}^{\beta]}$ in accordance with the general equation (I.5.10) and the fact that $\mathcal{M}^\alpha \equiv 0$. Equations (4.14)—(4.17) are *exact nonlinear constitutive equations* for the theory of piezoelectricity in *relativistic* continua. In most experimental cases of interest (e.g., in gravitational-wave detectors), however, infinitesimally small strains, space—time metric variations, and electric field must be considered. Thus, there is need for a linearized version of the equations just obtained. For this purpose we shall consider an infinitesimal variation of Eqs. (4.14)—(4.17) about an initially well-defined state of temperature, strains, and electric and gravitational fields. That is, we superimpose infinitesimal variations on finite fields.

B. Linearized theory

We consider infinitesimal variations $(\delta t^{\beta\alpha})_\perp$, $(\delta\rho^\beta)_\perp$, and $\delta\eta$ which result from infinitesimal variations $(\delta\mathcal{E}_{\alpha\beta})_\perp$, $(\delta\mathcal{E}_{\alpha})_\perp$, $\delta\theta$, and $(\delta g_{\alpha\beta})_\perp$. Noting that

$$ \delta g^{\mu\nu} = -g^{\mu\alpha}g^{\nu\beta}\delta g_{\alpha\beta} $$

so that

$$ (\delta g^{\mu\nu})_\perp = -P^{\mu\alpha}P^{\nu\beta}h_{\alpha\beta} \tag{4.18} $$

on account of Eq. (2.25), defining the instantaneous values of the components of the elastic stiffness tensor, $c_E^{\beta\alpha\gamma\delta}$, of the piezoelectric tensor, $e^{\gamma\alpha\beta}$, of the thermoelastic tensor, $\Theta^{\beta\alpha}$, of the electric susceptibility tensor, $\chi^{\alpha\beta}$, of the pyroelectric vector, A^β, and of the specific heat C by

$$ c_E^{\beta\alpha\gamma\delta} \equiv \rho\left(\frac{\partial\psi^{\alpha\beta}}{\partial\mathcal{E}_{\gamma\delta}}\right)_\perp = \rho\left(\frac{\partial\psi^{\gamma\delta}}{\partial\mathcal{E}_{\alpha\beta}}\right)_\perp, \tag{4.19} $$

$$ e^{\gamma\alpha\beta} \equiv -\rho\left(\frac{\partial\psi^{\alpha\beta}}{\partial\mathcal{E}_{\gamma}}\right)_\perp = -\rho\left(\frac{\partial\psi^{\gamma}}{\partial\mathcal{E}_{\alpha\beta}}\right)_\perp, \tag{4.20} $$

$$ \Theta^{\beta\alpha} \equiv \rho\left(\frac{\partial\psi^{\alpha\beta}}{\partial\theta}\right)_\perp = -\rho\left(\frac{\partial\eta}{\partial\mathcal{E}_{\alpha\beta}}\right)_\perp, \tag{4.21} $$

$$ \chi^{\beta\alpha} \equiv -\rho\left(\frac{\partial\psi^{\alpha}}{\partial\mathcal{E}_{\beta}}\right)_\perp = -\rho\left(\frac{\partial\psi_{\beta}}{\partial\mathcal{E}_{\alpha}}\right)_\perp, \tag{4.22} $$

$$ A^\beta \equiv -\rho\left(\frac{\partial\psi^{\beta}}{\partial\theta}\right)_\perp = -\rho\left(\frac{\partial\eta}{\partial\mathcal{E}_{\beta}}\right)_\perp, \tag{4.23} $$

and

$$ C \equiv -\theta\frac{\partial^2\bar{\psi}}{\partial\theta^2} = \theta\frac{\partial\eta}{\partial\theta}, \tag{4.24} $$

respectively, and noting that

$$ \delta\rho = -\rho P^{\alpha\beta}(\delta\mathcal{E}_{\alpha\beta})_\perp, \tag{4.25} $$

we obtain the following variations after a somewhat lengthy, but simple, calculation[16]:

$$ \begin{aligned} (\delta t^{\beta\alpha})_\perp = {} & (c_E^{\beta\alpha\gamma\delta} - e^{[\beta|\gamma\delta|}\mathcal{E}^{\alpha]} - 2c_E^{\lambda(\alpha|\gamma\delta|}\mathcal{E}^{\beta)}_{\cdot\lambda} \\ & - 2\psi^{\delta(\alpha}P^{\beta)\gamma} - t^{\beta\alpha}P^{\gamma\delta})(\delta\mathcal{E}_{\gamma\delta})_\perp \\ & - (e^{\gamma\alpha\beta} + \chi^{\gamma[\beta}\mathcal{E}^{\alpha]} + \rho^{[\beta}P^{\alpha]}{}_\gamma + 2e^{\gamma\lambda(\alpha}\mathcal{E}^{\beta)}_{\cdot\lambda})(\delta\mathcal{E}_{\gamma})_\perp \\ & + (\Theta^{\beta\alpha} - 2\Theta^{\lambda(\alpha}\mathcal{E}^{\beta)}_{\cdot\lambda} - A^{[\beta}\mathcal{E}^{\alpha]})\delta\theta \\ & + (2\mathcal{E}^{\mu}_{\cdot\lambda}\psi^{\lambda(\beta}P^{\alpha)\nu} + \mathcal{E}^{\mu}\rho^{[\beta}P^{\alpha]\nu})h_{\mu\nu}, \end{aligned} \tag{4.26} $$

$$ (\delta\rho^\beta)_\perp = (e^{\beta\gamma\delta} - \rho^\beta P^{\gamma\delta})(\delta\mathcal{E}_{\gamma\delta})_\perp + \chi^{\beta\gamma}(\delta\mathcal{E}_{\gamma})_\perp + A^\beta\delta\theta, \tag{4.27} $$

and

$$ \delta\eta = \theta^{-1}C\delta\theta + \rho^{-1}[A^\gamma(\delta\mathcal{E}_{\gamma})_\perp - \Theta^{\gamma\delta}(\delta\mathcal{E}_{\gamma\delta})_\perp]. \tag{4.28} $$

On account of obvious symmetries the spatial tensorial fields defined by Eqs. (4.19)—(4.23) have, respectively, 21, 18, 6, 6, and 3 independent components. The expression for $(\delta\mathcal{E}_{\alpha\beta})_\perp$ is provided by Eq. (2.28). As for the variation $(\delta\mathcal{E}_{\gamma})_\perp$, it results from a self variation of the electric field, $E_\gamma = (E_\gamma)_\perp$, and a variation due to the infinitesimal variation in the space—time event, $\xi^\lambda = \delta x^\lambda$. That is, similarly to Eq. (4.3), we have

$$ (\delta\mathcal{E}_{\gamma})_\perp = E_\gamma + \mathcal{E}^\lambda(\nabla_\gamma\xi_\lambda)_\perp. \tag{4.29} $$

Let us denote by the subscript (i) the fields evaluated at an initial state of strains, temperature, electric field and gravitational field. Then the equations

$$ t^{\beta\alpha} = t^{\beta\alpha}_{(i)} + (\delta t^{\beta\alpha})_\perp, \tag{4.30a} $$

$$ \rho^\beta = \rho^\beta_{(i)} + (\delta\rho^\beta)_\perp, \tag{4.30b} $$

$$ \eta = \eta_{(i)} + \delta\eta, \tag{4.30c} $$

describe the infinitesimally varying state with $(\delta t^{\beta\alpha})_\perp$, $(\delta\rho^\beta)_\perp$, and $\delta\eta$ given by Eqs. (4.26)—(4.28) in which all factors of $h_{\alpha\beta}$, $(\nabla_\alpha\xi_\beta)_\perp$, E_γ, and $\delta\theta$ obtained after substitution of the expressions (2.28) and (4.29) must be evaluated at the state (i).

For nonvanishing initial strains and electric fields the equations obtained can be used to study photoelastic and electro-optical effects (in the absence of light absorption and dichroism since dissipative effects have been discarded) since we have infinitesimal strains and infinitesimal electric fields (the latter, for instance, due to light propagating in the transparent medium) superimposed on initial finite states of strain and electrical polarization. The existence of such an initial polarization also allows us to account for ferroelectric and pyroelectric effects. In addition, Eqs. (4.26)—(4.28) contain the effect of a variation in the space—time metric (the general relativistic effect). These equations thus are appropriate for studying the influence of an incident gravitational wave of infinitesimal amplitude on a piezoelectric sample. This will result in infinitesimally small displacement gradients $(\nabla_\alpha\xi_\beta)_\perp$ coupled to an infinitesimally small electric field E_γ (which needs amplification to be practically detected[17]).

Consider now, for instance, an initially unstrained body, $(\mathcal{E}_{\alpha\beta})_{(i)}=0$, i.e., $P_{\alpha\beta(i)}=c_{\alpha\beta(i)}$ according to Eq. (2.10). Then Eqs. (4.30b) and (4.27) yield

$$\rho^{\beta}=\rho^{\beta}_{(i)}+\chi^{\beta\gamma}_{(i)}E_{\gamma}+(e^{\beta\gamma\delta}_{(i)}-\rho^{\beta}_{(i)}P^{\gamma\delta}_{(i)}+\chi^{\beta\gamma}_{(i)}\mathcal{E}^{\delta}_{(i)})E_{\gamma\delta}$$
$$-\chi^{\beta\gamma}_{(i)}\mathcal{E}^{\delta}_{(i)}(\tfrac{1}{2}h_{\delta\gamma}+\tilde{\Omega}_{\delta\gamma})+A^{\beta}_{(i)}\delta\theta \qquad (4.31)$$

on account of Eqs. (2.29), (4.29), and (2.31). For an isothermal variation in special relativity, Eq. (4.31) reduces to

$$\rho^{\beta}=\rho^{\beta}_{(i)}+\chi^{\beta\gamma}_{(i)}E_{\gamma}+(e^{\beta\gamma\delta}_{(i)}-\rho^{\beta}_{(i)}P^{\gamma\delta}_{(i)}+\chi^{\beta\gamma}_{(i)}\mathcal{E}^{\delta}_{(i)})\tilde{E}_{\gamma\delta}$$
$$-\chi^{\beta[\gamma}_{(i)}\mathcal{E}^{\delta]}_{(i)}\tilde{\Omega}_{\delta\gamma}. \qquad (4.32)$$

In agreement with the noncovariant equations recently obtained by Lax and Nelson,[18] this equation shows that both infinitesimal strain $\tilde{E}_{\gamma\delta}$ and infinitesimal rotation $\tilde{\Omega}_{\gamma\delta}$ are involved, and not only $\tilde{E}_{\gamma\delta}$, in the constitutive equation of ρ^{β} in the presence of initial electric and polarization fields (i.e., in ferroelectric and pyroelectric crystals). The same comment holds true if we write down the complete expression (4.30a) in the presence of initial strains and initial electric and polarization fields in special relativity. This means that, in these conditions, the photoelastic effect involves rotation as well as shear.[19]

In the study of gravitational-wave detectors that use a piezoelectric device, it is sufficient to consider the approximation of Eqs. (4.30) for vanishing initial fields and isothermal processes. Then, on account of Eqs. (4.26) and (4.27) and the fact that ρ can be considered as being constant without loss of generality, we obtain the linearized constitutive equations

$$t^{\beta\alpha}=c_{E}^{\beta\alpha\gamma\delta}(0)E_{\gamma\delta}-e^{\gamma\alpha\beta}(0)E_{\gamma}=t^{\alpha\beta} \qquad (4.33)$$

and

$$\mathcal{D}^{\beta}=e^{\beta\gamma\delta}(0)E_{\gamma\delta}+\epsilon^{\beta\gamma}(0)E_{\gamma} \qquad (4.34)$$

on account of $\mathcal{D}^{\beta}=\mathcal{E}^{\beta}+\rho^{\beta}$. We have defined

$$c_{E}^{\beta\alpha\gamma\delta}(0)\equiv\left(\frac{\partial^2 F}{\partial E_{\alpha\beta}\partial E_{\gamma\delta}}\right)_{\perp}^{(0)},\quad e^{\gamma\alpha\beta}(0)\equiv-\left(\frac{\partial^2 F}{\partial E_{\gamma}\partial E_{\alpha\beta}}\right)_{\perp}^{(0)},$$
$$\epsilon^{\beta\gamma}(0)\equiv P^{\beta\gamma}(0)-\left(\frac{\partial^2 F}{\partial E_{\gamma}\partial E_{\beta}}\right)_{\perp}^{(0)}, \qquad (4.35)$$

where $F\equiv\rho\psi$ and (0) indicates that the tensorial coefficients are evaluated at the zero value of the fields $E_{\gamma\delta}$ and E_{γ}. The free energy F per unit of proper volume then has the quadratic expression

$$F=\tfrac{1}{2}c_{E}^{\beta\alpha\gamma\delta}(0)E_{\alpha\beta}E_{\gamma\delta}-\tfrac{1}{2}\chi^{\beta\alpha}(0)E_{\beta}E_{\alpha}$$
$$-e^{\beta\gamma\delta}(0)E_{\beta}E_{\gamma\delta}. \qquad (4.36)$$

On account of the definition (2.29), Eqs. (4.33) and (4.34) are the equations that we have considered in a previous more naive approach.[20] They are the general relativistic generalization of the equations of Voigt's linear theory of piezoelectricity.[21] In the application of the linearized equations (4.33) and (4.34) to wave propagation the relevant material tensor to be considered is the spatial *piezoelectricity stiffened stiffness tensor* $\bar{c}^{\beta\alpha\gamma\delta}$. If λ_{α}, $\lambda_{\alpha}u^{\alpha}=0$, is a unit spatial 4-vector in the direction of travel of the piezoelectric vibrations, this tensor can be defined by[22]

$$\bar{c}^{\beta\alpha\gamma\delta}\equiv c_{E}^{\beta\alpha\gamma\delta}(0)+[\lambda_{\mu}\epsilon^{\mu\nu}(0)\lambda_{\nu}]^{-1}$$
$$\times\lambda_{\kappa}e^{\kappa\beta\alpha}(0)\lambda_{\sigma}e^{\sigma\gamma\delta}(0). \qquad (4.37)$$

Clearly, this tensor possesses the same symmetries as $c_{E}^{\beta\alpha\gamma\delta}(0)$.

Finally, let us comment upon the use of Eq. (4.33) in the relativistic Euler–Cauchy equation of motion (I.5.17). Equation (4.33) is linearized in the infinitesimal electric field E_{γ}. Furthermore, the Cauchy stress tensor thus approximated is symmetrical. This means that, in accordance with these approximations, the ponderomotive force and the ponderomotive couple are neglected since they are of second order in the electromagnetic fields. Electrostriction effects, which yield contributions of the second order in the Cauchy stress, have been accordingly neglected. These are the hypotheses upon which Voigt's classical linear theory of piezoelectricity rely. It remains to specify the material symmetry (which, of course, allows for the piezoelectric effect) for a given material (e.g., tetragonal barium titanate or trigonal lithium niobate).

5. MAGNETOELASTICITY UNDER HIGH PRESSURE

The developments given in Secs. 3 and 4 are based on the clear cut definition of the spatial space–time background metric $c_{\alpha\beta}$ or, equivalently, on that of the metric G_{KL} on $\mathcal{M}^3$. In exceptional conditions such as those thought to occur in neutron stars, extremely high densities and pressures are developed. These are responsible for the crystal-like structure of the neutron star crust. One cannot relax this crystalline structure to reach a zero state of stress without breaking down the structure. Thus $c_{\alpha\beta}$ and G_{KL} are not even defined in these conditions, while they defined a relaxed, ideal nonstressed state in the foregoing developments. To cope with this type of situation where strong magnetic fields may be expected to be present,[23] we need to devise a general relativistic magnetoelasticity under high pressure. Following Carter and Quintana,[24] we may assume, however, that for any deformation at constant volume, there exists a minimum of free energy denoted by $\bar{\psi}(\rho)$. Let $\hat{c}_{\alpha\beta}(\rho)$ be the spatial space–time metric corresponding to this state. Then, similarly to Eq. (2.10), we define the relativistic Euler strain tensor by

$$\hat{\mathcal{E}}_{\alpha\beta}(\rho)\equiv\tfrac{1}{2}[P_{\alpha\beta}-\hat{c}_{\alpha\beta}(\rho)]. \qquad (5.1)$$

The deformation which yields $\hat{c}_{\alpha\beta}(\rho)$ being isochoric, the tensor $\hat{\mathcal{E}}_{\alpha\beta}$ represents shearing effects and, consequently, must be tracefree with respect to $\hat{c}_{\alpha\beta}$. That is,

$$(\hat{\mathbf{c}}^{-1})^{\alpha\beta}\hat{\mathcal{E}}_{\alpha\beta}=0 \qquad (5.2)$$

as can be checked if $\hat{c}^{-1}$ is such that $(\hat{c}^{-1})^{\alpha\beta}\hat{c}_{\beta\gamma}=P^{\alpha}_{\cdot\gamma}$. Equation (2.20) still holds true as well as Eq. (I.4.26), i.e.,

$$D\rho=-\rho P^{\alpha\beta}d_{\alpha\beta}. \qquad (5.3)$$

We consider nonpolarized, nondissipative media in isothermal evolution ($\mathcal{P}^{\alpha}=\mathcal{J}^{\alpha}=q^{\alpha}=0$, $D\theta=0$). Then Eq. (3.1) yields Gibbs' equation

$$\rho D\psi={}_{E}t^{\beta\alpha}d_{\alpha\beta}+\mathcal{B}_{\alpha}(D_c\mathcal{M}^{\alpha}). \qquad (5.4)$$

Noting, on account of the definition of D_c given in Part I,

that

$$\mathcal{B}_\alpha(D_c\mathcal{H}^\alpha)=\mathcal{B}^\alpha(\underset{u}{\pounds}\mathcal{H}_\alpha)_\perp+\mathcal{B}^\gamma\mathcal{H}_\gamma(\nabla_\beta v^\beta)-2\mathcal{B}^{(\alpha}\mathcal{H}^{\beta)}d_{\alpha\beta}, \tag{5.5}$$

we can rewrite Eq. (5.4) in the form

$$\rho D\hat{\psi}={}_E\hat{t}^{\beta\alpha}d_{\alpha\beta}+\mathcal{B}^\alpha(\underset{u}{\pounds}\mathcal{H}_\alpha)_\perp. \tag{5.6}$$

Here,

$${}_E\hat{t}^{\beta\alpha}\equiv{}_E t^{\beta\alpha}+\mathcal{B}^\gamma\mathcal{H}_\gamma P^{\alpha\beta}-2\mathcal{B}^{(\alpha}\mathcal{H}^{\beta)}={}_E\hat{t}^{\alpha\beta}. \tag{5.7}$$

It follows from Eqs. (5.7) and (3.2) that the relativistic Cauchy stress tensor is given by

$$t^{\beta\alpha}={}_E\hat{t}^{\beta\alpha}-\mathcal{B}^\gamma\mathcal{H}_\gamma P^{\alpha\beta}+\mathcal{B}^{(\alpha}\mathcal{H}^{\beta)}-\mathcal{H}^{[\beta}\mathcal{B}^{\alpha]}. \tag{5.8}$$

We naturally assume in general that

$$\hat{\psi}=\hat{\psi}(\rho,\hat{\mathcal{E}}_{\alpha\beta},\hat{\mathcal{H}}_\alpha), \tag{5.9}$$

and define the fields

$$\hat{\psi}_\rho\equiv\frac{\partial\hat{\psi}}{\partial\rho},\quad \hat{\psi}^{\alpha\beta}\equiv\frac{\partial\hat{\psi}}{\partial\hat{\mathcal{E}}_{\alpha\beta}},\quad \hat{\underset{u}{\psi}}^\alpha\equiv\frac{\partial\hat{\psi}}{\partial\hat{\mathcal{H}}_\alpha}. \tag{5.10}$$

The Lorentz invariance of $\hat{\psi}$ requires [compare Eqs. (4.9) and (4.10)] that $\hat{\psi}^{\alpha\beta}$ and $\underset{u}{\psi}$ be spatial objects, while they satisfy a constraint of the type (4.9). We have assumed that $\mathcal{H}_\alpha=\hat{\mathcal{H}}_\alpha(\rho)$, i.e., the magnetic dipole depends on the proper density in the same way as $\hat{\mathcal{E}}_{\alpha\beta}$. We thus define

$$\hat{\rho}_{\alpha\beta}\equiv\left(\frac{\partial\hat{\mathcal{E}}_{\alpha\beta}}{\partial\rho}\right)_\perp=-\tfrac{1}{2}\left(\frac{d\hat{c}_{\alpha\beta}}{d\rho}\right)_\perp,\quad \hat{\rho}_\alpha\equiv(\partial\hat{\mathcal{H}}^\alpha/\partial\rho)_\perp. \tag{5.11}$$

Hence,

$$(D\hat{\mathcal{E}}_{\alpha\beta})_\perp=(D\hat{\mathcal{E}}_{\alpha\beta})_\perp^{(\rho)}+\hat{\rho}_{\alpha\beta}D\rho,$$
$$(D\hat{\mathcal{H}}_\alpha)_\perp=(D\hat{\mathcal{H}}_\alpha)_\perp^{(\rho)}+\hat{\rho}_\alpha D\theta, \tag{5.12}$$

where the superscript (ρ) indicates that the quantities thus labelled are evaluated at constant density. According to Eqs. (5.8), (5.1), (2.20), and an equation of the type of Eq. (4.12) written for $\hat{\mathcal{E}}_{\alpha\beta}$, we have

$$(D\hat{\mathcal{H}}_\alpha)_\perp^{(\rho)}=(\underset{u}{\pounds}\hat{\mathcal{H}}_\alpha)_\perp-\hat{\mathcal{H}}_\beta e^\beta{}_{\circ\alpha} \tag{5.13}$$

and

$$(D\hat{\mathcal{E}}_{\alpha\beta})_\perp^{(\rho)}=d_{\alpha\beta}-2\hat{\mathcal{E}}_{\lambda(\alpha}e^\lambda{}_{\circ\beta)}. \tag{5.14}$$

We compute $D\hat{\psi}$ on account of Eqs. (5.3), (5.10)—(5.14), and of the Lorentz invariance condition (4.9)—written for $\hat{\psi}^{\alpha\beta}$ and $\hat{\underset{u}{\psi}}^\alpha$. Substituting from the resulting expression in Eq. (5.6), posited to be valid for arbitrary independent time rates $d_{\alpha\beta}$ and $(\underset{u}{\pounds}\mathcal{H}_\alpha)_\perp$, we are led to the following constitutive equations:

$${}_E\hat{t}^{\beta\alpha}=-\hat{p}P^{\alpha\beta}+\rho(\hat{\psi}^{\beta\alpha}-2\hat{\psi}^{\lambda(\beta}\hat{\mathcal{E}}^{\alpha)}_{\circ\lambda})-\rho\hat{\underset{u}{\psi}}^{(\alpha}\mathcal{H}^{\beta)}, \tag{5.15}$$

$$\mathcal{B}^\alpha=\rho\hat{\underset{u}{\psi}}^\alpha, \tag{5.16}$$

where the effective pressure $\hat{p}$ is defined by

$$\hat{p}\equiv-\rho^2(\hat{\psi}_\rho+\hat{\psi}^{\alpha\beta}\hat{\rho}_{\alpha\beta}+\hat{\underset{u}{\psi}}^{\alpha\beta}\hat{\rho}_\alpha)+\mathcal{B}^\gamma\mathcal{H}_\gamma. \tag{5.17}$$

Then Eq. (5.8) provides the Cauchy stress

$$t^{\beta\alpha}=-\hat{p}P^{\alpha\beta}+\rho(\hat{\psi}^{\beta\alpha}-2\hat{\psi}^{\lambda(\beta}\hat{\mathcal{E}}^{\alpha)}_{\circ\lambda})-\hat{\mathcal{H}}^{[\beta}\mathcal{B}^{\alpha]}. \tag{5.18}$$

Equations (5.16) and (5.18) are *exact nonlinear* constitutive equations for relativistic magnetoelasticity under high pressure. An approximation closely related to the classical Hookean approximation is obtained by considering an idealized equation of state of the form

$$\hat{\psi}=\bar{\psi}(\rho)+\tfrac{1}{2}\hat{C}^{\beta\alpha\gamma\delta}(\rho)\hat{\mathcal{E}}_{\alpha\beta}(\rho)\hat{\mathcal{E}}_{\gamma\delta}(\rho)+\tfrac{1}{2}\hat{K}^{\beta\alpha}(\rho)\hat{\mathcal{H}}_\beta(\rho)\hat{\mathcal{H}}_\alpha(\rho)+\hat{F}^{\gamma\beta\alpha}(\rho)\hat{\mathcal{H}}_\gamma(\rho)\hat{\mathcal{E}}_{\beta\alpha}(\rho). \tag{5.19}$$

Then Eqs. (5.18) and (5.16) give

$$t^{\beta\alpha}=-\hat{p}P^{\alpha\beta}+[\hat{c}_E^{\beta\alpha\mu\nu}(\rho)-2\hat{c}^{\lambda(\beta\mu\nu}(\rho)\hat{\mathcal{E}}^{\alpha)}_{\circ\lambda}(\rho)]\hat{\mathcal{E}}_{\mu\nu}(\rho)+[\hat{f}^{\gamma\beta\alpha}(\rho)-2\hat{f}^{\gamma\lambda(\beta}(\rho)\hat{\mathcal{E}}^{\alpha)}_{\circ\lambda}(\rho)-\mathcal{H}^{[\alpha}P^{\beta]\gamma}]\hat{\mathcal{H}}_\gamma(\rho) \tag{5.20}$$

and

$$\mathcal{H}^\alpha=\hat{\chi}^{-1\alpha\beta}(\rho)\hat{\mathcal{H}}_\beta(\rho)+\hat{f}^{\alpha\gamma\delta}(\rho)\hat{\mathcal{E}}_{\gamma\delta}(\rho) \tag{5.21}$$

on account of the relationship $\mathcal{H}^\alpha=\mathcal{B}^\alpha-\mathcal{H}^\alpha$. We have not formulated the pressure formula (5.17). In spite of the approximation (5.19) we have kept quadratic terms in $\hat{\mathcal{E}}_{\alpha\beta}$ in $t^{\beta\alpha}$ because of the possibly large deformation. The spatial tensors

$$\hat{c}_E^{\beta\alpha\mu\nu}(\rho)\equiv\rho\hat{C}^{\beta\alpha\mu\nu}(\rho),\quad \hat{f}^{\gamma\beta\alpha}(\rho)\equiv\rho\hat{F}^{\gamma\beta\alpha}(\rho),$$
$$\hat{\chi}^{-1\alpha\beta}(\rho)\equiv\rho\hat{K}^{\alpha\beta}(\rho)-P^{\alpha\beta}, \tag{5.22}$$

may be referred to as the elastic stiffness tensor, the piezomagneto tensor and the reciprocal magnetic suceptibility tensor. In virtue of their spatial character, of obvious symmetries, and of the fact that $\hat{\mathcal{E}}_{\alpha\beta}$ has only five independent components as a consequence of constraint (5.2), they have, respectively, 15, 15, and 6 independent components for a general material symmetry. Thus, with $\bar{\psi}(\rho)$, $\hat{c}_{\alpha\beta}(\rho)$, and $\hat{\mathcal{H}}_\beta(\rho)$ in addition, there are in all and at most, 45 independent scalar functions of ρ which must be specified to construct the present model. In addition, to study dynamical problems, time evolution equations must be constructed for $\hat{c}_E^{\beta\alpha\mu\nu}$, $\hat{f}^{\gamma\beta\alpha}$, and $\hat{\chi}^{-1\alpha\beta}$ in order to be able to compute the time evolutions $(\underset{u}{\pounds}t^{\beta\alpha})_\perp$ and $(\underset{u}{\pounds}\mathcal{H}^\alpha)_\perp$. Such computations can be carried out by generalizing the works of Carter, Quintana, and Barrabes.

[1] G.A. Maugin, Part I, J. Math. Phys. **19**, 1198 (1978).
[2] G.A. Maugin, Part II, J. Math. Phys. **19**, 1206 (1978).
[3] S.R. de Groot and L.G. Suttorp, Physica **37**, 284, 297 (1967); **39**, 28, 41, 61, 84 (1968).
[4] B. Carter and H. Quintana, Proc. R. Soc. London Ser. A **331**, 57 (1972).
[5] The notation is that of Parts I and II. The definitions not recalled are to be found in Part I whose equations are referred to by I followed by their number, e.g., Eq. (I.3.1).
[6] Elements of the relativistic theory of nonlinear deformations are to be found, for instance, in A. Bressan, Ann. Mat. Pura Appl. **62**, 99 (1963); R.A. Grot and A.C. Eringen, Int. J. Eng. Sci. **4**, 611 (1966); G.A. Raminez and G. Lianis, Acta Mechanica **6**, 326 (1968); **7**, 58 (1969); G.A. Maugin, C.R. Acad. Sci. Ser. A **272**, 1482 (1971); **273**, 65 (1971); J. Phys. A: Gen. Phys. **5**, 786 (1972); Gen. Rel. Grav. J. **4**, 251 (1973); G. Lianis, Nuovo Cimento B **14**, 57 (1973) and Ref. 4. The most extensive treatment is that given in G.A. Maugin, *Thèse de Doctorat ès. Sci. Math.* (Université de Paris VI, Paris, 1975).
[7] Cf. A.C. Eringen, *Mechanics of Continua* (Wiley, New York, 1967), Chap. I. Indeed, in an inertial frame $DX^K=0$ yields
$$\partial_4 X^K\overset{*}{=}-(v^k/c)\,\partial_k X^K$$

where v^k is the 3-velocity. Then Eq. (2.5) can be written as

$$C^{-1KL} \overset{*}{=} (\delta_{ij} - c^{-2} v_i v_j)\, \partial_i X^K \partial_j X^L.$$

This reduces to the classical definition in a local instantaneous rest frame.

[8]An expression of the type (2.29) to describe infinitesimal deformations in general relativistic continuum mechanics has been proposed independently and almost simultaneously by several authors. The derivation given above, however, is original. See F. J. Dyson, Astrophys. J. **156**, 529 (1969); E. N. Glass and J. Winicour, J. Math. Phys. **13**, 1934 (1972); G. A. Maugin, Gen. Rel. Grav. J. **4**, 251 (1973); A. Papapetrou, Ann. Inst. H. Poincaré **16**, 63 (1972). See also D. E. Soper, *Classical Theory of Fields* (Wiley, New York, 1976), p. 205.

[9]This is the argument known as the thermodynamical admissibility. Initiated by B. Coleman in the early sixties, it is the accepted standard argument in "rational" thermodynamics; cf. C. Truesdell, *Rational Thermodynamics* (McGraw-Hill, New York, 1969).

[10]An equation of this sort, however, is not entirely adequate, for it is known to yield a propagation of thermal disturbances at infinite speed. To avoid this contradiction with the principle of relativistic causality, a time-functional heat-flux constitutive equation must be considered. In this respect see G. A. Maugin, J. Phys. A: Math. Nucl. Gen. **7**, 465 (1974).

[11]C. Barrabes, Nuovo Cimento B **28**, 377 (1975).

[12]G. A. Maugin, Ann. Inst. H. Poincaré **15**, 275 (1971). In this paper 2-forms are used instead of spatial 4-vectors as far as electromagnetic fields are concerned.

[13]Cf. Part IV of the present series.

[14]Cf. W. Cady, *Piezoelectricity* (MacGraw-Hill, New York, 1946).

[15]In the absence of electrical effects, Eq. (4.16) is the relativistic version of the classical constitutive equation of nonlinear elasticity; cf. A. C. Eringen, *Nonlinear Theory of Continuous Media* (McGraw-Hill, New York, 1962). This is known as Murnaghan's spatial formulation of nonlinear elasticity.

[16]The variation symbol used in Eqs. (4.26)—(4.27) can be replaced by an operator which satisfies both the linearity property and the Leibnitz rule, i.e., any operator known under the generic name of *derivative*. If one considers the Lie derivative, then equations such as (4.26) may be referred to as time evolution equations.

[17]This provides the theoretical basis for recording signals on "elastic" gravitational-wave detectors; cf. J. Weber, *General Relativity and Gravitational Waves* (Wiley Interscience, New York, 1961), Chap. 8.

[18]M. Lax and D. F. Nelson, Phys. Rev. B **13**, 1759, 1785 (1976).

[19]M. Lax and D. F. Nelson, in *The Photoelastic Effect and Its Applications* (I. U. T. A. M Coll., Ottignies, Belgium, 1973), edited by J. Kestens (Springer-Verlag, Berlin, 1975); R. D. Mindlin, *ibid.*

[20]G. A. Maugin, Gen. Rel. Grav. J. **4**, 251 (1973).

[21]See Ref. 14.

[22]See Ref. 14 and W. P. Mason, *Piezoelectric Crystals and their Applications to Ultrasonics* (Van Nostrand, Princeton, N. J., 1950).

[23]See M. A. Ruderman, Phys. Rev. Lett. **27**, 1306 (1971); lectures at the "Ettore Majorana" Second School of Cosmic Physics, Erice, Sicily (1971).

On the covariant equations of the relativistic electrodynamics of continua. IV. Media with spin

G. A. Maugin

Université de Paris VI, Laboratoire de Mécanique Théorique associé au C.N.R.S., Tour 66, 75230 Paris, Cedex 05, France
(Received 6 April 1977)

The general scheme given in a previous work (Paper I) for the conservation laws and the thermodynamics of general relativistic electromagnetic continua is reexamined in the presence of intrinsic spin and couple stresses (i.e., when the *total* energy–momentum tensor is not symmetrical). Following two different sets of hypotheses concerning the nature of this spin and of the interactions responsible for the couple stresses, two descriptions of continuous matter emerge from the kinematical and thermodynamical study: first, that of paramagnetic and/or dielectric continua endowed with a rigid microstructure (hence a granular structure of the media) and, second, that of continua endowed with a continuous distribution of electronic spins. The latter scheme provides a phenomenological description of quantum-mechanical effects, which has been previously studied in detail by the author.

1. INTRODUCTION

This paper is a continuation of three other papers herein after referred to as Parts I, II and III.[1] We recall that the purpose of this series is to provide a firm basis for the covariant formulation of the electrodynamics of continua, from the results of a microscopical treatment serving as a background for the electromagnetic sources in the mechanical and energetical conservation laws. That is, the final phenomenological formulation emerges from a compromise between an entirely formal logico-deductive approach (as often favored nowadays in continuum mechanics) and the reliance upon microscopical concepts, the compromise with the first type of approach being necessary for the construction of quite general, but nonetheless exploitable, constitutive equations. The microscopical treatment of which the results have been adopted is that of de Groot and Suttorp,[2] of which the main features and results have been recalled in Part I.

In this part we consider the *a priori* existence of a density of intrinsic spin and of couple stresses. Depending on the nature of this spin and of the interactions for which the couple stresses account, it is shown that we can deduce, on the one hand, the basic equations of the theory of relativistic continua with rigid microstructure, as initiated in the works of Kafadar, Eringen, and Maugin,[3,4] but modify them so as to account for electromagnetic effects (e.g., in dielectrics and in linear or nonlinear paramagnets), and on the other hand, the basic equations of the relativistic theory of continua with electronic spin. In the first case we settle a covariant formulation, hence a basis and justification for the Galilean equations of the thermo-electro-magneto-elasticity of micropolar continua,[5] where the new internal degrees of freedom (rotation of infinitesimally small rigid particles) account for the granular structure of the medium. Couple stresses then arise naturally as the macroscopic continuum representation of the interactions exerted between neighboring rotating particles. In the second case both intrinsic-spin density and couple stresses find their origin in quantum mechanical effects (electron spin and exchange forces), which are smoothed out in a continuous manner and may contribute at the macroscopical scale to the mechanical and electromagnetic behavior of the material, as is the case in the phenomenological theory of deformable ferromagnets. We thus provide a basis for the relativistic theories of magnetoelastic interactions and spinning fluids developed previously by the author.[6]

In Sec. 2, we recall the expression of the relativistic conservation laws in the presence of spin. These are transformed in the form of the conservation laws of 4-momentum, moment of 4-momentum, and the local statement of the second principle of thermodynamics in its form known as the Clausius–Duhem inequality, and this for arbitrary electromagnetic sources of ponderomotive force and couple and electromagnetic power. These sources are specified in Sec. 3 on the basis of the results of de Groot and Suttorp. The basic equations of the theory of relativistic continua with rigid microstructure are developed in Sec. 4, while those of the theory of relativistic continua with electronic spin are deduced in Sec. 5.

2. CONSERVATION LAWS IN PRESENCE OF SPIN

The general local conservation laws recalled in Part I are[7]:

conservation of mass:

$$\nabla_\beta(\rho u^\beta)=0, \tag{2.1}$$

conservation of energy–momentum:

$$\nabla_\beta\, {}_m T^{\alpha\beta} = {}_M F^\alpha, \tag{2.2}$$

conservation of moment of energy–momentum:

$$\nabla_\mu S^{\alpha\beta\mu} - {}_m T^{[\alpha\beta]} = {}_M L^{\alpha\beta}, \tag{2.3}$$

second principle of thermodynamics:

$$\nabla_\alpha \eta^\alpha \geq 0. \tag{2.4}$$

In these equations the symbols used bear the following significance: ρ is the proper mass density, u^β is the world velocity ($g_{\alpha\beta}u^\alpha u^\beta + c^2 = 0$), ${}_m T^{\alpha\beta}$ is the "matter" energy–momentum tensor (${}_m T^{[\alpha\beta]} \neq 0$), ${}_M F^\alpha$ is the electromagnetic 4-force, ${}_M L^{\alpha\beta}$ is the electromagnetic

 0022-2488/78/1905-1220$1.00

couple 2-form, $s^{\alpha\beta\mu} = -s^{\beta\alpha\mu}$ is the total intrinsic spin tensor, and η^α is the entropy-flux 4-vector. The fields ${}_mT^{\alpha\beta}$, $s^{\alpha\beta\mu}$, η^α, ${}_MF^\alpha$, and ${}_ML^{\alpha\beta}$ have the following decompositions:

$$ {}_mT^{\alpha\beta} = \rho(1 + c^{-2}\epsilon)u^\alpha u^\beta + c^{-2}u^\alpha Q^\beta + p^\alpha u^\beta - t^{\beta\alpha}, \tag{2.5} $$

$$ s^{\alpha\beta\mu} = \tfrac{1}{2}\rho S^{\alpha\beta}u^\mu - M^{\alpha\beta\mu}, \tag{2.6} $$

$$ \eta^\alpha = \rho\eta u^\alpha + N^\alpha, \tag{2.7} $$

$$ {}_MF^\alpha = {}_Mf^\alpha + c^{-2}\,{}_M\overline{w}u^\alpha, \tag{2.8} $$

$$ {}_ML^{\alpha\beta} = {}_M\mathcal{L}^{[\alpha}u^{\beta]} + {}_MC^{\alpha\beta}. \tag{2.9} $$

Here ϵ is the proper internal energy density, Q^β is the spatial energy-current 4-vector, p^α is the spatial "nonmechanical" momentum 4-vector, $t^{\beta\alpha} \neq t^{\alpha\beta}$ is the spatial relativistic stress tensor, $S^{\alpha\beta} = -S^{\beta\alpha}$ is the spatial intrinsic-spin 2-form, $M^{\alpha\beta\mu} = -M^{\beta\alpha\mu}$ is the spatial relativistic couple-stress tensor, η is the proper entropy density, N^α is the spatial entropy-flux 4-vector, ${}_Mf^\alpha$ is the (spatial) ponderomotive 4-force, ${}_M\overline{w}$ is the electromagnetic power density per unit of proper volume, ${}_MC^{\alpha\beta}$ is the (spatial) ponderomotive couple 2-form, and ${}_M\mathcal{L}^\alpha$ is a spatial electromagnetic "momentum."

On account of the definitions (2.6), (2.7), and (2.9), Eqs. (2.3) and (2.4) can be rewritten in the following form:

$$ \tfrac{1}{2}\rho DS^{\alpha\beta} - \nabla_\mu M^{\alpha\beta\mu} - {}_mT^{[\alpha\beta]} = {}_ML^{\alpha\beta} \tag{2.10} $$

and

$$ \rho D\eta + \nabla_\alpha N^\alpha \geq 0. \tag{2.11} $$

$D \equiv u^\alpha\nabla_\alpha$ is the invariant derivative in the direction of the world velocity.

Projecting Eqs. (2.2) and (2.10) in the direction of u_α and onto the three-dimensional hyperplane $M_\perp$ orthogonal to u^α on account of Eq. (2.1) and account of the defintions (2.5)—(2.9), we obtain the following local balance equations:

local balance of 4-*momentum* (Euler—Cauchy equations of motion):

$$ \rho\left(1 + \frac{\epsilon}{c^2}\right)(Du^\alpha)_\perp + \rho\left[D\left(\frac{p^\alpha}{\rho}\right)\right]_\perp = (\nabla_\beta t^{\beta\alpha})_\perp + {}_Mf^\alpha - c^{-2}Q^\beta e^\alpha_{\cdot\beta} \tag{2.12} $$

local balance of energy:

$$ \rho D\epsilon + \nabla_\beta Q^\beta + p^\alpha Du_\alpha = t^{\beta\alpha}e_{\alpha\beta} + {}_M\overline{w}, \tag{2.13} $$

local balance of moment of 4-*momentum*:

$$ \tfrac{1}{2}\rho(DS^{\alpha\beta})_\perp - (\nabla_\mu M^{\alpha\beta\mu})_\perp + t^{[\beta\alpha]} = {}_MC^{\alpha\beta} \tag{2.14} $$

and

$$ p^\beta = c^{-2}[Q^\beta + \rho u_\alpha(DS^{\alpha\beta}) - 2u_\alpha(\nabla_\mu M^{\alpha\beta\mu})] - {}_M\mathcal{L}^\beta. \tag{2.15} $$

In these equations the symbolism $(\cdots)_\perp$ indicates projection onto $M_\perp$, and

$$ e_{\alpha\beta} \equiv \overset{\downarrow}{\nabla}_\beta u_\alpha = d_{\alpha\beta} + \omega_{\alpha\beta}, \tag{2.16} $$

$$ d_{\alpha\beta} \equiv e_{(\alpha\beta)}, \quad \omega_{\alpha\beta} = e_{[\alpha\beta]}, \tag{2.17} $$

are respectively the spatial gradient of the world velocity, the rate of strain tensor, and the rate of rotation tensor.

Equation (2.14) provides the value of $t^{[\beta\alpha]}$ in terms of the intrinsic spin, the couple-stress and the ponderomotive couple. That is,

$$ t^{[\beta\alpha]} = {}_MC^{\alpha\beta} + (\nabla_\mu M^{\alpha\beta\mu})_\perp - \tfrac{1}{2}\rho(DS^{\alpha\beta})_\perp. \tag{2.18} $$

The quantity $(DS^{\alpha\beta})_\perp$ is none other than the Fermi derivative of the intrinsic spin.[8] As for Eq. (2.15), on account of the fact that $S^{\alpha\beta}$ and $M^{\alpha\beta\mu}$ are purely spatial (i.e., they admit u_α as zero vector), it can be written in the form

$$ p^\beta = c^{-2}(Q^\beta + \rho S^{\beta\alpha}Du_\alpha + 2M^{\alpha\beta\mu}e_{\alpha\mu}) - {}_M\mathcal{L}^\beta. \tag{2.19} $$

The energy 4-current can be decomposed as

$$ Q^\beta = q^\beta + \hat{Q}^\beta, \tag{2.20} $$

where q^β is the spatial heat-flux 4-vector and the spatial 4-vector $\hat{Q}^\beta$ can be called the "excess-energy-flux" spatial 4-vector. We have shown in Part I that this vector can be set equal to zero without loss of generality when spin effects are discarded. Substituting for p^β and Q^β in Eqs. (2.12) and (2.13) yields the field equations

$$ \begin{aligned} &\rho J^{\alpha\beta}(Du_\beta)_\perp + c^{-2}\rho S^{\alpha\beta}(D^2u_\beta)_\perp + c^{-2}(D_cq^\alpha + 2q^\beta e^\alpha_{\cdot\beta}) \\ &+ c^{-2}\rho[D\{\rho^{-1}(\hat{Q}^\alpha + 2M^{\beta\alpha\mu}e_{\beta\mu})\}]_\perp + c^{-2}\hat{Q}^\beta e^\alpha_{\cdot\beta} \\ &= (\nabla_\beta t^{\beta\alpha})_\perp + \{{}_Mf^\alpha + \rho[D({}_M\mathcal{L}^\alpha/\rho)]_\perp\} \end{aligned} \tag{2.21} $$

and

$$ \begin{aligned} &\rho D\epsilon + (\overset{\downarrow}{\nabla}_\beta q^\beta + 2c^{-2}q^\beta Du_\beta) + c^{-2}[\nabla_\beta\hat{Q}^\beta + (\hat{Q}^\beta + 2M^{\alpha\beta\mu}e_{\alpha\mu})Du_\beta] \\ &\quad = t^{\beta\alpha}e_{\alpha\beta} + \{{}_M\mathcal{L}^\alpha Du_\alpha + {}_M\overline{w}\}. \end{aligned} \tag{2.22} $$

In writing down these equations we have taken account of definition (I. 3.19) of the spatial covariant derivative noted $\overset{\downarrow}{\nabla}_\alpha$; of the definition (I. 4.39) of the contravariant convective time derivative noted D_C, and of the obvious result

$$ S^{\alpha\beta}Du_\alpha Du_\beta \equiv 0 \tag{2.23} $$

which follows from the skew symmetry of $S^{\alpha\beta}$. Finally, we have defined a kind of spatial inertia tensor, $J^{\alpha\beta}$, per unit of proper mass by

$$ J^{\alpha\beta} \equiv \left(1 + \frac{\epsilon}{c^2}\right)P^{\alpha\beta} + \frac{1}{c^2}(DS^{\alpha\beta})_\perp, \tag{2.24} $$

so that, obviously,

$$ J^{(\alpha\beta)} = \left(1 + \frac{\epsilon}{c^2}\right)P^{\alpha\beta}, \quad J^{[\alpha\beta]} = \frac{1}{c^2}(DS^{\alpha\beta})_\perp, \quad J^\alpha_{\cdot\alpha} = 3\left(1 + \frac{\epsilon}{c^2}\right). \tag{2.25} $$

The right-hand side of Eq. (2.22) can be transformed further if we introduce the spatial precession-velocity 2-form of the intrinsic spin, $\Omega_{\alpha\beta}$, such that

$$ \Omega_{\alpha\beta} = -\Omega_{\beta\alpha}, \quad \Omega_{\alpha\beta}u^\alpha = 0. \tag{2.26} $$

Then we can write

$$ t^{\beta\alpha}e_{\alpha\beta} = t^{(\beta\alpha)}d_{\alpha\beta} + t^{[\beta\alpha]}(\omega_{\alpha\beta} - \Omega_{\alpha\beta}) + t^{[\beta\alpha]}\Omega_{\alpha\beta} \tag{2.27} $$

on account of the decomposition of $e_{\alpha\beta}$. The last contribution in this expression can be evaluated by saturating the indices of Eq. (2.18) by applying $\Omega_{\alpha\beta}$ to both sides of this equation. We thus have

$$ \begin{aligned} t^{\beta\alpha}e_{\alpha\beta} &= t^{(\beta\alpha)}d_{\alpha\beta} + t^{[\beta\alpha]}(\omega_{\alpha\beta} - \Omega_{\alpha\beta}) + {}_MC^{\alpha\beta}\Omega_{\alpha\beta} \\ &\quad + [(\nabla_\mu M^{\alpha\beta\mu})_\perp - \tfrac{1}{2}\rho(DS^{\alpha\beta})_\perp]\Omega_{\alpha\beta}. \end{aligned} \tag{2.28} $$

Substituting this result into Eq. (2.22), we obtain the local energy balance in the following form:

$$\rho D\epsilon + \tfrac{1}{2}\rho(DS^{\alpha\beta})_{\perp}\Omega_{\alpha\beta} + (\overset{\perp}{\nabla}_{\beta}q^{\beta} + 2c^{-2}q^{\beta}Du_{\beta}) = l^{(\beta\alpha)}d_{\alpha\beta} + l^{[\beta\alpha]}\nu_{\alpha\beta} + \nabla_{\mu}(M^{\alpha\beta\mu}\Omega_{\alpha\beta} - \hat{Q}^{\mu}) - \{M^{\alpha\beta\mu}[(\nabla_{\mu}\Omega_{\alpha\beta})_{\perp} + c^{-2}e_{\alpha\mu}Du_{\beta}] + c^{-2}\hat{Q}^{\beta}Du_{\beta}\} + \{{}_{M}C^{\alpha\beta}\Omega_{\alpha\beta} + {}_{M}\angle^{\alpha}Du_{\alpha} + {}_{M}\overline{w}\}. \quad (2.29)$$

The quantities

$$\nu_{\alpha\beta} \equiv \omega_{\alpha\beta} - \Omega_{\alpha\beta}, \quad (\nu_{\alpha\beta}u^{\beta} = 0), \quad (2.30)$$

are the components of a 2-form which represents the (spatial) precession velocity of the intrinsic spin with respect to the deformation matter in rotation. In general $\nu_{\alpha\beta}$ differs from zero, for there is no reason that in a local frame the intrinsic spin precesses at the local rotational velocity of the matter (i.e., the vorticity). The condition $\nu_{\alpha\beta}=0$ would be a constraint imposed on the spin precession (i.e., the spin would be "frozen in" the deformable matter), somewhat analogous to the condition $d_{\alpha\beta}=0$ of local rigid-body motion. Indeed, in agreement with the duality inherent in thermodynamics, the time rates $d_{\alpha\beta}$ and $\nu_{\alpha\beta}$ are the duals of $t^{(\beta\alpha)}$ and $t^{[\beta\alpha]}$, respectively.

Introducing the proper free energy density ψ by

$$\psi \equiv \epsilon - \eta\theta, \quad (2.31)$$

then eliminating $D\epsilon$ between Eqs. (2.28) and (2.11), and introducing the proper thermodynamical temperature θ, we obtain the local statement of the second principle of thermodynamics, known as the *Clausius–Duhem inequality*, in the form

$$-\rho(D\psi + \eta D\theta) + \theta\nabla_{\beta}\left(N^{\beta} - \frac{q^{\beta}}{\theta}\right) - \theta^{-1}q^{\alpha}\overset{*}{\theta}_{\alpha} + t^{(\beta\alpha)}d_{\alpha\beta} + t^{[\beta\alpha]}\nu_{\alpha\beta} + \tfrac{1}{2}\rho(DS^{\alpha\beta})_{\perp}\Omega_{\alpha\beta} + \nabla_{\mu}(M^{\alpha\beta\mu}\Omega_{\alpha\beta} - \hat{Q}^{\mu}) + \{M^{\beta\alpha\mu}[\nabla_{\mu}\Omega_{\alpha\beta})_{\perp} + c^{-2}e_{\alpha\mu}Du_{\beta}] - c^{-2}\hat{Q}^{\mu}Du_{\mu}\} + \{{}_{M}C^{\alpha\beta}\Omega_{\alpha\beta} + {}_{M}\angle^{\alpha}Du_{\alpha} + {}_{M}\overline{w}\} \geq 0, \quad (2.32)$$

where $\overset{*}{\theta}_{\alpha}$ is the spatial relativistic gradient of temperature defined by Eq. (I.4.35).

Equations (2.21), (2.22), and (2.32) are the local statements of the balance laws of momentum, of energy and of entropy for a general electromagnetic continuum of which the total intrinsic spin tensor has the form (2.6). We still must select the expression of the spatial 4-vectors N^{β} and Q^{μ}, provide the expressions of the electromagnetic "sources" ${}_{M}C^{\alpha\beta}$, ${}_{M}f^{\alpha}$, and ${}_{M}\overline{w}$, and construct constitutive equations for the fields ψ, η, q^{α}, $\mathcal{J}^{\alpha}$, $t^{(\beta\alpha)}$, $t^{[\beta\alpha]}$, and $M^{\alpha\beta\mu}$ to complete the theory.

3. ELECTROMAGNETIC INTERACTIONS WITH MATTER

The expressions of ${}_{M}F^{\alpha}$, ${}_{M}f^{\alpha}$, ${}_{M}\overline{w}$, ${}_{M}\angle^{\alpha}$, and ${}_{M}C^{\alpha\beta}$ deduced from the microscopic treatment of de Groot and Suttorp[9] and written down in terms of spatial 4-vectors have been stated in Part I. They read

$${}_{M}F^{\alpha} = -\nabla_{\beta}\,{}_{M}T^{\alpha\beta}, \quad (3.1)$$

$${}_{M}f^{\alpha} = {}_{M}\tilde{f}^{\alpha} + \rho c^{-2}(\mathcal{E}_{\gamma}\pi^{\gamma})(Du^{\alpha})_{\perp}, \quad (3.2)$$

$${}_{M}\overline{w} = {}_{M}\tilde{w} + \rho D(\mathcal{E}_{\gamma}\pi^{\gamma}), \quad (3.3)$$

$${}_{M}\angle^{\alpha} = 0, \quad {}_{M}C^{\alpha\beta} = {}_{M}T^{[\alpha\beta]} = P^{[\alpha}\mathcal{E}^{\beta]} + \mathcal{M}^{[\alpha}\mathcal{B}^{\beta]}, \quad (3.4)$$

where

$${}_{M}\tilde{f}^{\alpha} = q\mathcal{E}^{\alpha} + \frac{1}{c}(\mathcal{J}*\mathcal{B})^{\alpha} + P^{\alpha\beta}[P^{\mu}\overset{\perp}{\nabla}_{\beta}\mathcal{E}_{\mu} + \mathcal{M}^{\mu}\overset{\perp}{\nabla}_{\beta}\mathcal{B}_{\mu} + \mathcal{D}_{C}({}_{M}p_{\beta})], \quad (3.5)$$

$${}_{M}\tilde{w} = \mathcal{J}_{\alpha}\mathcal{E}^{\alpha} - \mathcal{M}^{\alpha}D\mathcal{B}_{\alpha} - P^{\alpha}D\mathcal{E}_{\alpha}, \quad (3.6)$$

$${}_{M}p^{\alpha} = \frac{1}{c}(P*\mathcal{B} + \mathcal{E}*\mathcal{M})^{\alpha}, \quad (3.7)$$

$${}_{M}T^{\alpha\beta} = \tfrac{1}{2}(\mathcal{E}^{2} + \mathcal{B}^{2})\frac{u^{\alpha}u^{\beta}}{c^{2}} + c^{-2}(u^{\alpha}\mathcal{S}^{\beta} + u^{\beta}\mathcal{S}^{\alpha}) - {}_{M}t^{\beta\alpha}, \quad (3.8)$$

$$\mathcal{S}^{\alpha} = c(\mathcal{E}*\mathcal{H})^{\alpha}, \quad (3.9)$$

and

$${}_{M}t^{\beta\alpha} = \mathcal{D}^{\beta}\mathcal{E}^{\alpha} + \mathcal{B}^{\beta}\mathcal{H}^{\alpha} - \tfrac{1}{2}(\mathcal{E}^{2} + \mathcal{B}^{2} - 2\mathcal{M}*\mathcal{B})P^{\alpha\beta}. \quad (3.10)$$

In these equations q is the proper free-charge density, $\mathcal{J}_{\alpha}$ is the conduction-current 4-vector and $\mathcal{E}, \mathcal{B}, \mathcal{M}, P, \mathcal{D} = \mathcal{E} + P$, and $\mathcal{H} = \mathcal{B} - \mathcal{M}$ are the spatial electric field, magnetic induction, magnetization, polarization, electric displacement, and magnetic field 4-vectors, respectively. ${}_{M}T^{\alpha\beta}$ is the electromagnetic energy–momentum tensor, ${}_{M}t^{\beta\alpha}$ is the spatial electromagnetic stress tensor, and $\mathcal{S}^{\alpha}$ is the spatial Poynting 4-vector. The symbolism $(\circ * \bullet)^{\alpha}$ has been defined by Eq. (I. 3.17), and $\mathcal{D}_{C}$ indicates the covariant convective time derivative defined by Eq. (I.5.13). We have set

$$\pi^{\alpha} \equiv P^{\alpha}/\rho, \quad \mu^{\alpha} \equiv \mathcal{M}^{\alpha}/\rho. \quad (3.11)$$

Equations (3.1) and (3.4) show that Eqs. (2.2) and (2.3) can be rewritten in the condensed form

$$\nabla_{\beta}T^{\alpha\beta} = 0 \quad (3.12)$$

and

$$\nabla_{\mu}s^{\alpha\beta\mu} = T^{[\alpha\beta]}, \quad (3.13)$$

where $T^{\alpha\beta}$ is the *nonsymmetric* total energy–momentum tensor,

$$T^{\alpha\beta} \equiv {}_{m}T^{\alpha\beta} + {}_{M}T^{\alpha\beta}. \quad (3.14)$$

In order to keep the standard Riemannian geometrical description in the general relativistic framework a symmetrization procedure must be used as regards $T^{\alpha\beta}$. The Belinfante–Rosenfeld[10] procedure fulfills this purpose. Hence, define a new total energy–momentum tensor, $\overline{T}^{\alpha\beta}$, by

$$\overline{T}^{\alpha\beta} \equiv T^{\alpha\beta} + \nabla_{\gamma}(s^{\alpha\gamma\beta} + s^{\beta\gamma\alpha} - s^{\alpha\beta\gamma}). \quad (3.15)$$

Upon using Eq. (3.13), it is immediately checked that $\overline{T}^{\alpha\beta}$ is symmetrical and divergence free, i.e.,

$$\overline{T}^{[\alpha\beta]} = 0, \quad \nabla_{\beta}\overline{T}^{\alpha\beta} = 0, \quad (3.16)$$

Thus $\overline{T}^{\alpha\beta}$ can be used as a source of nongravitational energy–momentum in the standard Einstein field equations. However, Eq. (3.12) now is modified and must read

$$\nabla_{\beta}T^{\alpha\beta} + R^{\alpha}_{\cdot\beta\gamma\delta}s^{\gamma\delta\beta} = 0, \quad (3.17)$$

where $R^{\alpha}_{\cdot\beta\gamma\delta}$ is the Riemannian curvature tensor of space–time.

4. RELATIVISTIC CONTINUA WITH RIGID MICROSTRUCTURE

Each particle of the material continuum is assumed to be an infinitesimally small rigid body endowed with rotational inertia. There is thus defined at each event point a density of inertia (per unit of proper mass), $j^{\alpha\beta}$, which is a *symmetric* spatial tensor and acts as a linear operator between the precessional velocity of the spin and the spin itself. We thus define the spatial intrinsic spin 4-vector s_α and the precessional velocity 4-vector Ω^α by

$$s_\alpha \equiv \overset{*}{\eta}_{\alpha\beta\gamma} S^{\beta\gamma} \quad (S^{\alpha\beta} = \tfrac{1}{2}\overset{*}{\eta}^{\alpha\beta\gamma} s_\gamma), \tag{4.1}$$

and

$$\Omega^\gamma \equiv -\tfrac{1}{2}\overset{*}{\eta}^{\gamma\mu\nu}\,\Omega_{\mu\nu} \quad (\Omega_{\alpha\beta} = -\overset{*}{\eta}_{\alpha\beta\gamma}\Omega^\gamma). \tag{4.2}$$

The alternation symbols (in fact, tensorial densities) $\overset{*}{\eta}_{\alpha\beta\gamma}$ and $\overset{*}{\eta}^{\alpha\beta\gamma}$ have been defined in Part I. Then s^α and Ω_β are related by

$$s^\alpha = j^{\alpha\beta}\Omega_\beta. \tag{4.3}$$

It follows that

$$\tfrac{1}{2}S^{\alpha\beta}\Omega_{\beta\alpha} = \tfrac{1}{2}s^\alpha\Omega_\alpha = \tfrac{1}{2}j^{\alpha\beta}\,\Omega_\alpha\Omega_\beta. \tag{4.4}$$

If an orthonormal triad of spatial 4-vectors is attached to each particle, then, in virtue of the rigidity of the "particle," the spatial tensor $j^{\alpha\beta}$ has constant components in this triad. This can be restated as: The corotational time derivative (with respect to the precessional motion) of $j^{\alpha\beta}$ vanishes. That is,[11]

$$D_\Omega j^{\alpha\beta} \equiv (Dj^{\alpha\beta})_\perp - \Omega^\alpha_{\cdot\gamma}\, j^{\gamma\beta} - \Omega^\beta_{\cdot\gamma}\, j^{\alpha\gamma} = 0. \tag{4.5}$$

As a result of this identity and of Eqs. (4.1)—(4.4), it is easily shown that

$$\rho D(\tfrac{1}{2}S^{\alpha\beta}\Omega_{\beta\alpha}) = \tfrac{1}{2}\rho(DS^{\alpha\beta})_\perp\,\Omega_{\beta\alpha}. \tag{4.6}$$

Then the theory of electromagnetic relativistic continua with rigid microstructure is obtained under the following hypotheses:

H_1: Eqs. (4.3) and (4.5) hold good;

H_2: Like in classical thermodynamics, the spatial entropy-flux 4-vector N^α and the heat-flux 4-vector q^α are related by

$$N^\alpha = q^\alpha/\theta; \tag{4.7}$$

H_3: The excess-energy-flux $\hat{Q}^\mu$ is chosen to be

$$\hat{Q}^\mu = M^{\alpha\beta\mu}\Omega_{\alpha\beta}. \tag{4.8}$$

Indeed, defining new internal and free energy densities $\bar{\epsilon}$ and $\bar{\psi}$ by

$$\bar{\epsilon} = \epsilon - \mathcal{E}_\alpha\pi^\alpha - \tfrac{1}{2}S^{\alpha\beta}\Omega_{\beta\alpha}, \quad \bar{\psi} = \bar{\epsilon} - \eta\theta, \tag{4.9}$$

defining the spatial third-order tensor $\mathcal{A}_{\alpha\beta\gamma}$ by

$$\begin{aligned}\mathcal{A}_{\alpha\beta\mu} &\equiv \{\nabla_\mu\Omega_{\alpha\beta} + c^{-2}(\Omega_{\alpha\beta}Du_\mu + 2\nabla_\mu u_{[\alpha}Du_{\beta]}\}_\perp \\ &= (\overset{1}{\nabla}_\mu\Omega_{\alpha\beta})_\perp + c^{-2}(\Omega_{\alpha\beta}Du_\mu + 2e_{[\alpha|\mu|}Du_{\beta]}) \\ &= -\mathcal{A}_{\beta\alpha\mu},\end{aligned} \tag{4.10}$$

and the corotational time derivatives (with respect to the precessional motion[11]) of $\mathcal{E}_\alpha$ and $\mathcal{B}_\alpha$ by

$$D_\Omega\mathcal{E}_\alpha \equiv (D\mathcal{E}_\alpha)_\perp - \Omega_{\alpha\beta}\mathcal{E}^\beta, \quad D_\Omega\mathcal{B}_\alpha \equiv (D\mathcal{B}_\alpha)_\perp - \Omega_{\alpha\beta}\mathcal{B}^\beta, \tag{4.11}$$

and taking account of H_1, H_2, and H_3 and of Eqs. (3.2), (3.3), (3.4), and (3.6), we can rewrite Eqs. (2.21), (2.29), and (2.32) in the form

$$\begin{aligned}&\rho\bar{J}^{\alpha\beta}(Du_\beta)_\perp + c^{-2}\{\rho S^{\alpha\beta}(D^2u_\beta)_\perp + (D_C q^\alpha + 2q^\beta e^\alpha_{\cdot\beta} \\ &\quad + M^{\mu\nu\beta}\Omega_{\mu\nu}e^\alpha_{\cdot\beta} + \rho(D[\rho^{-1}(M^{\mu\nu\alpha}\Omega_{\mu\nu} + 2M^{\beta\alpha\mu}e_{\alpha\mu})])_\perp\} \\ &\quad = (\nabla_\beta t^{\beta\alpha})_\perp + {}_M\tilde{f}^\alpha\end{aligned} \tag{4.12}$$

with

$$\bar{J}^{\alpha\beta} \equiv [1 + c^{-2}(\bar{\epsilon} + \tfrac{1}{2}S^{\mu\nu}\Omega_{\nu\mu})]P^{\alpha\beta} + c^{-2}(DS^{\alpha\beta})_\perp, \tag{4.13}$$

and

$$\begin{aligned}&\rho D\bar{\epsilon} + (\overset{1}{\nabla}_\beta q^\beta + 2c^{-2}q^\beta Du_\beta) \\ &\quad = t^{(\beta\alpha)}d_{\alpha\beta} + t^{[\beta\alpha]}\nu_{\alpha\beta} + M^{\beta\alpha\mu}\mathcal{A}_{\alpha\beta\mu} \\ &\quad + \mathcal{J}_\alpha\mathcal{E}^\alpha - \mathcal{P}^\alpha(D_\Omega\mathcal{E}_\alpha) - \mathcal{M}^\alpha(D_\Omega\mathcal{B}_\alpha),\end{aligned} \tag{4.14}$$

$$\begin{aligned}&-\rho(D\bar{\psi} + \eta D\theta) - \theta^{-1}q^\alpha\overset{*}{\theta}_\alpha + t^{(\beta\alpha)}d_{\alpha\beta} + t^{[\beta\alpha]}\nu_{\alpha\beta} \\ &\quad + M^{\beta\alpha\mu}\mathcal{A}_{\alpha\beta\mu} + \mathcal{J}_\alpha\mathcal{E}^\alpha - \mathcal{P}^\alpha(D_\Omega\mathcal{E}_\alpha) - \mathcal{M}^\alpha(D_\Omega\mathcal{B}_\alpha) \geq 0.\end{aligned} \tag{4.15}$$

By the same token the space and time decomposition of the total energy—momentum tensor (3.14) is shown to read, on account of Eqs. (2.5), (2.19), (2.20), and (3.8),

$$T^{\alpha\beta} = \rho\omega_{(\mathrm{tot})}u^\alpha u^\beta + c^{-2}u^\alpha Q^\beta_{(\mathrm{tot})} + u^\beta p^\alpha_{(\mathrm{tot})} - t^{\beta\alpha}_{(\mathrm{tot})}, \tag{4.16}$$

where

$$\omega_{(\mathrm{tot})} = 1 + c^{-2}[\bar{\epsilon} + \tfrac{1}{2}S^{\alpha\beta}\Omega_{\beta\alpha} + \frac{1}{2\rho}(\mathcal{E}^2 + \mathcal{B}^2 + 2\mathcal{E}\cdot\mathcal{P})], \tag{4.17}$$

$$Q^\beta_{(\mathrm{tot})} = q^\beta + \mathcal{J}^\beta + M^{\mu\nu\beta}\Omega_{\mu\nu}, \tag{4.18}$$

$$p^\alpha_{(\mathrm{tot})} = c^{-2}(q^\alpha + \mathcal{J}^\alpha + \rho S^{\alpha\gamma}Du_\gamma + 2M^{\gamma\alpha\mu}e_{\gamma\mu} + M^{\mu\nu\alpha}\Omega_{\mu\nu}), \tag{4.19}$$

and

$$t^{\beta\alpha}_{(\mathrm{tot})} = t^{\beta\alpha} + \mathcal{D}^\beta\mathcal{E}^\alpha + \mathcal{B}^\beta\mathcal{H}^\alpha - \tfrac{1}{2}(\mathcal{E}^2 + \mathcal{B}^2 - 2\mathcal{M}\cdot\mathcal{B})P^{\alpha\beta}. \tag{4.20}$$

Comments: Equations (4.2) and (4.14) are the local balance laws of 4-momentum and energy for an electromagnetic continuum with rigid microstructure. In the absence of electromagnetic phenomena, they reduce the to the equations obtained by Kafadar and Eringen[12] from a formal approach and by Maugin and Eringen[13] from a variational principle. The Clausius—Duhem inequality (4.15) is the constraint to be placed upon the constitutive equations for the fields $\bar{\psi}$, η, q^α, $t^{(\beta\alpha)}$, $t^{[\beta\alpha]}$, $M^{\beta\alpha\mu}$, $\mathcal{J}_\alpha$, $\mathcal{P}^\alpha$, and $\mathcal{M}^\alpha$. The thermodynamical dual of the couple stress tensor $M^{\beta\alpha\mu}$ is the spatial tensor $\mathcal{A}_{\alpha\beta\mu}$ defined by Eq. (4.10), which clearly is a relativistic generalization of the notion of gradient of precessional velocity. Indeed, if we take the nonrelativistic limit of the spatial component of Eq. (4.10) in a local inertial frame, we obtain $(i, j, k = 1, 2, 3)$

$$\mathcal{A}_{ijk} \overset{*}{=} \Omega_{ij,k}. \tag{4.21}$$

The fact that corotational time derivatives replace convective time derivatives (compare Part I) as far as time rates of electromagnetic field are concerned, places in evidence the role of rotational effects in this model, which represents some kind of granular medium whose particles possess internal degrees of freedom (of rotation). As in classical continuum mechanics,[14] the thermodynamical duality between $t^{[\beta\alpha]}$ and the relative precessional velocity $\nu_{\alpha\beta}$ indicates the possible existence of rotational viscosity. The presence of couple stresses greatly complicates the equations and,

in particular, the coupling between the Euler–Cauchy equations of motion and the moment of 4-momentum equations. The skew symmetry of $t^{\beta\alpha}$ being left aside, these couplings are purely relativistic as shown by the presence of the factor c^{-2}. However, even in the absence of heat conduction, electromagnetic phenomena, and couple stresses, it must be noted that the 4-momentum defined by

$$P^\alpha \equiv \rho\omega_{(\mathrm{tot})}u^\alpha + p^\alpha_{(\mathrm{tot})} = \rho\omega_{(\mathrm{tot})}u^\alpha + \rho c^{-2}S^{\alpha\gamma}Du_\gamma, \tag{4.22}$$

is not colinear with the world velocity. It possesses a spatial part which involves the spin density. This fact was already acknowledged in simpler theories of spinning continua.[15] The development of a constitutive theory is straightforward and will not be reported here. Of course, the new rotational degrees of freedom are accounted for by considering a rigid tetrad attached to each "particle." We refer the reader to the above quoted papers for the exposé of the corresponding kinematics and theory of deformations[16] and for the constitutive equations of elastic bodies in the absence of electromagnetic phenomena. Further comments on this relativistic theory will be made in Sec. 5. In the nonrelativistic limit ($c \to \infty$), Eqs. (4.12), (2.14), (4.14), and (4.15) yield, in Cartesian coordinates, the equations[17]

$$\rho\frac{dv_i}{dt} = t_{ji,j} + {}_M\tilde{f}_i, \tag{4.23}$$

$$\rho\frac{dS_{ij}}{dt} = M_{ijk,k} + t_{[ij]} + \mathcal{P}_{[i}\mathcal{E}_{j]} + \mathcal{M}_{[i}\mathcal{B}_{j]}, \tag{4.24}$$

$$\rho\frac{d\bar{\epsilon}}{dt} + \nabla\cdot\mathbf{q} = t_{(ji)}d_{ij} + t_{[ji]}(\omega_{ij} - \Omega_{ij}) + M_{jik}\mathcal{A}_{ijk} + \mathcal{J}\cdot\mathcal{E} - \mathcal{P}_i(D_\Omega\mathcal{E}_i) - \mathcal{M}_i(D_\Omega\mathcal{B}_i), \tag{4.25}$$

and

$$-\rho\left(\frac{d\bar{\psi}}{dt} + \eta\frac{d\theta}{dt}\right) - \theta^{-1}\mathbf{q}\cdot\nabla\theta + t_{(ji)}d_{ij} + t_{[ji]}(\omega_{ij} - \Omega_{ij}) + M_{jik}\mathcal{A}_{ijk} + \mathcal{J}\cdot\mathcal{E} - \mathcal{P}_i(D_\Omega\mathcal{E}_i) - \mathcal{M}_i(D_\Omega\mathcal{B}_i) \geq 0, \tag{4.26}$$

which are none other than the equations of the theory of *micropolar* continua[18] modified so as to account for the fact that the continua considered can be electrically polarized and magnetized, and can conduct electricity. That is, they are, for instance, the equations of the thermo-electro-magneto-elasticity of micropolar continua.[19] Thus Eqs. (4.12), (2.14), (4.14), and (4.15) are the corresponding covariant equations for general nonlinear constitutive equations.

5. RELATIVISTIC CONTINUA WITH ELECTRON SPIN

In Sec. 4 we have assigned a real microstructure to each "particle" of the material continuum. The latter then is some kind of aggregate of rigid particles of which a smooth representation is obtained by allowing for the existence of a density of inertia (i.e., a geometrical distribution of matter). This results in the fact that a precessional velocity of these rigid "particles" different from the mean vorticity causes the existence of a skew symmetric part for the stress tensor, whereas the spatial variation of this precessional velocity from one particle to its nearest neighbors "causes" the existence of couple stresses. Here, however, we consider a somewhat different model which accounts, not for a granular structure of the medium but rather for quantum mechanical effects, namely, the existence of an intrinsic spin with no reference to any structure. In agreement with Uhlenbeck, Goudsmit, and Frenkel, this intrinsic spin is purely spatial and has its origin in the magnetic dipole moments. We are thus led to put forward the following hypotheses:

H_1': The intrinsic-spin 4-vector defined by Eq. (4.1) is linearly related to the magnetization per unit of proper mass;

H_2': The intrinsic spin has a magnitude which does not depend on proper time, but depends eventually on which particle it is attached to;

H_3': The hypotheses H_2 and H_3 of Sec. 4 hold good.

The hypothesis H_1' is materialized by the (isotropic) gyromagnetic relationship

$$s^\alpha = \gamma^{-1}\mu^\alpha, \tag{5.1}$$

where γ is the gyromagnetic ratio (assumed to be a constant for the sake of simplicity). Using the notation of Part I, H_2' can be restated as

$$P_{\alpha\beta}\mu^\alpha\mu^\beta = \mu^2(\mathbf{X}). \tag{5.2}$$

Obviously, s^α has constant components in a spatial triad rigidly attached to it. Hence the corotational time derivative of μ^α with respect to this triad (which necessarily precesses at the spatial time rate $\Omega^{\alpha\beta}$) vanishes. I.e., on account of Eq. (5.1)

$$D_\Omega\mu^\alpha \equiv (D\mu^\alpha)_\perp - \Omega^{\alpha\beta}\mu_\beta = 0, \tag{5.3}$$

whereas the invariant time derivative of the spatial 4-vector μ^α of constant magnitude in general reads[20]

$$D\mu^\alpha = g^{\alpha\beta}[\Omega_{\beta\gamma} + c^{-2}u_\beta(Du_\gamma)]\,\mu^\gamma. \tag{5.4}$$

That is, the latter is made of two parts, a precession measured, for instance, in a Fermi frame, and a Fermi–Walker dragging along the world line of the particle to which μ^α is attached.[20] Taking the spatial part of Eq. (5.4) yields anew Eq. (5.3), which can also be rewritten as [using the notation (I. 3.17)]

$$(D\mu^\alpha)_\perp = (\Omega * \mu)^\alpha. \tag{5.5}$$

It then follows from Eqs. (4.1), (4.2) and (5.3) or (5.5) that

$$\tfrac{1}{2}\rho(DS^{\alpha\beta})_\perp\Omega_{\beta\alpha} = \rho\gamma^{-1}(D\mu^\alpha)_\perp\Omega_\alpha \equiv 0. \tag{5.6}$$

This means that in contrast to the case considered in Sec. 4, the intrinsic spin is here *gyroscopical* in that in it does not produce any power in a real precessional velocity field [compare Eq. (4.6)].

Defining new internal and free energy densities $\hat{\epsilon}$ and $\hat{\psi}$ by

$$\hat{\epsilon} = \epsilon - \mathcal{E}_\alpha\pi^\alpha + \mathcal{B}_\alpha\mu^\alpha, \quad \hat{\psi} = \hat{\epsilon} - \eta\theta, \tag{5.7}$$

taking account of the hypotheses H_1', H_2', H_3', of definition (4.10), of Eqs. (3.2)–(3.6), of Eqs. (4.11), and of the peculiar result (5.3)—which results from H_2', we can rewrite the general equation (2.21) in the form (4.12), but now with

$$\bar{J}^{\alpha\beta} \equiv [1 + c^{-2}(\hat{\epsilon} - \beta_\gamma \mu^\gamma)] P^{\alpha\beta} + c^{-2}(DS^{\alpha\beta})_\perp , \tag{5.8}$$

while Eqs. (2.29) and (2.32) become

$$\rho D\hat{\epsilon} + (\overset{1}{\nabla}_\beta q^\beta + 2c^{-2} q^\beta D u_\beta) = t^{(\beta\alpha)} d_{\alpha\beta} + t^{[\beta\alpha]} \nu_{\alpha\beta} + M^{\beta\alpha\mu} \mathcal{A}_{\alpha\beta\mu} + \mathcal{J}^\alpha \mathcal{E}_\alpha - \mathcal{P}^\alpha (D_\Omega \mathcal{E}_\alpha), \tag{5.9}$$

and

$$-\rho(D\hat{\psi} + \eta D\theta) + t^{(\beta\alpha)} d_{\alpha\beta} + t^{[\beta\alpha]} \nu_{\alpha\beta} - \theta^{-1} q^\alpha \overset{*}{\theta}_\alpha + M^{\beta\alpha\mu} \mathcal{A}_{\alpha\beta\mu} + \mathcal{J}^\alpha \mathcal{E}_\alpha - \mathcal{P}^\alpha (D_\Omega \mathcal{E}_\alpha) \geq 0. \tag{5.10}$$

By the same token the space and time decomposition of the total energy–momentum tensor (3.14) can be written in the form (4.16), with Eqs. (4.18), (4.19), and (4.20) left unchanged, but with Eq. (4.17) replaced by

$$\omega_{(\text{tot})} = 1 + c^{-2}\, \hat{\epsilon} + \frac{1}{2\rho}(\mathcal{E}^2 + \mathcal{B}^2 + 2\mathcal{E}\cdot\mathcal{P} - 2\mathcal{M}\cdot\mathcal{B}) . \tag{5.11}$$

Comments: Clearly, the presence of electrical polarization can be discarded without too much loss of generality since, as is shown by Eqs. (5.9) and (5.10), this polarization is thermodynamically determined in the same way as in the theory outlined in Sec. 4 (i.e., $\mathcal{P}^\alpha$ has for its thermodynamical dual the corotational time derivative $D_\Omega \mathcal{E}_\alpha$ in both cases). However, in the present case, the magnetization density is not provided by a constitutive equation constrained to satisfy the Clausius–Duhem inequality. Rather, the magnetization is subjected to satisfy an *evolution equation*, Eq. (5.5), the magnitude of the 4-vector μ^α, or s^α, being only possibly dependent on the local thermodynamical temperature. That is, we have a situation analogous to that met in the three-dimensional theory of deformable ferromagnets at low temperature,[21] where the magnetization is considered to be saturated. On account of this comment we may consider that the equations developed in Sec. 4 correspond to the behavior of (linear or nonlinear) diamagnetic and paramagnetic bodies. There are other points which differ from one theory to the other one. In the theory of Sec. 4 the energy densities may depend on the magnetization (compare Part II), but not on the magnetization gradients, for the couple stresses are caused by the spatial disuniformities in the fields which describe the internal degrees of freedom of the microstructured bodies. This microstructure is not related to the magnetization. In the theory of this section, it can be shown that the couple stresses result from the fact that the energy densities do depend, along with the magnetization itself, on the gradient of the magnetization. Comparing this fact to the classical theory of deformable ferromagnets,[22] we see that the couple stresses then account in a phenomenological manner for the Heisenberg exchange forces of quantum electrostatic origin. It can be shown that Eqs. (5.5) and (2.14) are entirely equivalent. This requires that Ω_α be related to some dependent constitutive arguments from which $M^{\beta\alpha\mu}$ can be deduced.[23]

When the medium is not electrically polarized, the equations obtained above are those which have been proposed in previous papers and on which we have based a detailed constitutive theory.[24] Furthermore, if the medium is a fluid of infinite electrical conductivity (which imposes that $\mathcal{E}_\alpha \equiv 0$) subject to no other type of dissipation, then Eqs. (4.12), (5.5), and (5.9) are those that we have derived with the corresponding constitutive equations from two different variational principles in the special relativistic framework.[25] In the case of elastic bodies with vanishing electrical conduction, these equations had already been deduced from a variational principle by Maugin and Eringen.[26] In these two cases the energy equation (5.9) reduces to the corresponding Gibbs' equation[27]

$$\rho D\hat{\epsilon} = t^{(\beta\alpha)} d_{\alpha\beta} + t^{[\beta\alpha]} \nu_{\alpha\beta} + M^{\beta\alpha\mu} \mathcal{A}_{\alpha\beta\mu}. \tag{5.12}$$

Finally, it can be shown that, like in the classical theory of deformable ferromagnets,[28] the rotational viscosity which may exist according to the Clausius–Duhem inequality (5.10), is equivalent to a relaxation of the intrinsic spin towards its equilibrium direction.[29] In the nonrelativistic limit ($c \to \infty$) and in the absence of polarization in the quasimagnetostatic approximation, Eqs. (4.12), (2.14), (5.5), (5.9), and (5.10) take on the Galilean "couple-stress-theory" form of the equations that govern deformable ferromagnets as given, for instance, by the author.[30]

[1] G.A. Maugin, Parts I, II, and III, J. Math. Phys. **19**, 1198, 1206, 1212 (1978).

[2] S.R. de Groot and L.G. Suttorp, Physica 37, 284, 297 (1967); 39, 28, 41, 61, 84 (1968); *Foundations of Electrodynamics* (North-Holland, Amsterdam, 1972).

[3] C.B. Kafadar and A.C. Eringen, Int. J. Eng. Sci. 9, 307 (1971).

[4] G.A. Maugin and A.C. Eringen, J. Math. Phys. 13, 1788 (1972)

[5] Cf. W. Nowacki, "Coupled Fields in Mechanics of Solids," in *Proceedings of the 14th International Congress of Theoretical and Applied Mechanics* (North-Holland, Amsterdam, 1976), Vol. I.

[6] G.A. Maugin, J. Phys. A: Math. Nucl. Gen. 5, 786 (1972); 6, 306, 1647 (1973); 7, 818 (1974).

[7] The notation is that of Parts I and II, Refs. 1, to which we refer. Equations of these parts are accordingly referred to by I or II followed by their number, e.g., Eq. (I.3.25).

[8] Cf. A. Schild, "Lectures in General Relativity," in *Relativity Theory and Astrophysics*, Lectures in Appl. Math., Vol. 8, (American Math. Soc., Providence, R.I., 1967).

[9] Cf. Refs. 2.

[10] F.J. Belinfante, Physica 7, 449 (1940); L. Rosenfeld, Mém. Acad. R. Belg. Cl. Sci. 18, No. 6 (1940); see also D.E. Soper, *Classical Theory of Fields* (Wiley, New York, 1976).

[11] The justification for the definition of the space–time corotational derivative (with respect to any tetrad) is given in Part I, Chap. 2 of G.A. Maugin, *Thèse de Doctorat d'Etat ès Sci. Math.* (Université de Paris, France, 1975).

[12] See Ref. 3.

[13] See Ref. 4.

[14] Compare Eqs. (18) and (23) of Chap. XII in S.R. de Groot and P. Mazur, *Non-Equilibrium Thermodynamics* (North-Holland, Amsterdam, 1962).

[15] Cf. F. Halbwachs, *Théorie relativiste des fluides à spin* (Gauthier-Villars, Paris, 1960).

[16] See Refs. 3 and 4 and Part I, Chap. 4 in the reference quoted in 11.

[17] Here $(d/dt) \equiv (\partial/\partial t) + \mathbf{v}\cdot\nabla$ is the usual material time derivative. The corotational derivative D_Ω of a vector field can be defined symbolically by $D_\Omega \equiv (d/dt) - \Omega\times$, if Ω is the vector dual of Ω_{ij}.

[18] Cf. A.C. Eringen, in *Fracture*, Vo. 2, pp. 621–729, edited by H. Liebowitz (Academic, New York, 1968); C.B. Kafadar and A.C. Eringen, Int. J. Eng. Sci. 9, 271 (1971).

[19]In the absence of polarization and magnetization and for linear constitutive equations, compare the equations given in Ref. 5.
[20]See Ref. 8 and the second of the papers quoted in Ref. 5.
[21]Cf. G.A. Maugin, in *Continuum Physics*, edited by A.C. Eringen (Academic, New York, 1976), Vol. 3, pp. 221—312; G.A. Maugin, J. Mécanique (Paris) **13**, 75 (1974); G.A. Maugin and A.C. Eringen, J. Math. Phys. **13**, 143, 1334 (1972).
[22]See the second reference quoted in Ref. 21.
[23]See Eq. (2.21) in the fourth paper quoted in Ref. 6.
[24]See the second, third, and fourth papers quoted in Ref. 6.
[25]G.A. Maugin, Ann. Inst. H. Poincaré **20**, 41 (1974); Ann. Math. Pura Appl. **110**, 247 (1976).
[26]G.A. Maugin and A.C. Eringen, J. Math. Phys. **13**, 1777 (1972).
[27]Compare Eq. (4.77) in the first paper quoted in Ref. 25.
[28]G.A. Maugin, J. Phys. A: Gen. Phys. **5**, 1550 (1972); Physica (Holland) A **81**, 454 (1975).
[29]See the third of the papers quoted in Ref. 6.
[30]See the second of the references quoted in Ref. 21.

Entropy production for quantum dynamical semigroups

Herbert Spohn[a)]

Department of Physics, Princeton University, Princeton, New Jersey 08540
(Received 15 August 1977)

In analogy to the phenomenological theory of irreversible thermodynamics we define the entropy production for an arbitrary quantum dynamical semigroup with a stationary state. We prove that the entropy production is convex and positive and that the entropy production is a measure of dissipativity of the semigroup. The entropy production is used to prove the approach to equilibrium and to classify the stationary states of semigroups arising in the weak coupling limit.

1. DEFINITION OF THE ENTROPY PRODUCTION

Under certain limiting conditions,[1,2] the dynamics of an *open* quantum mechanical system $\mathcal{S}$ may be described by a quantum dynamical semigroup.[3]

We associate with $\mathcal{S}$ a separable Hilbert space $\mathcal{H}$. We denote by $T(\mathcal{H})$ the Banach space of trace class operators on $\mathcal{H}$ and by $B(\mathcal{H})$ the C^*-algebra of bounded operators on $\mathcal{H}$. A quantum dynamical semigroup is then a family of linear mappings $\Lambda_t: T(\mathcal{H}) \to T(\mathcal{H})$, $t \geq 0$, with the properties: (i) The dual map $\Lambda_t^*: B(\mathcal{H}) \to B(\mathcal{H})$ is completely positive, (ii) Λ_t is trace-preserving, (iii) $\{\Lambda_t \mid t \geq 0\}$ form a semigroup, $\Lambda_{t+s} = \Lambda_t \Lambda_s$, (iv) $\lim_{t \to 0+} \|\Lambda_t \rho - \rho\|_1 = 0$ for all $\rho \in T(\mathcal{H})$. Λ_t is a contraction, $\|\Lambda_t\| = 1$, and the strong continuity of $\{\Lambda_t \mid t \geq 0\}$ implies by the Hille–Yosida theorem the existence of a densely defined generator L: $T(\mathcal{H}) \to T(\mathcal{H})$ such that for ρ in the domain of L $\lim_{t \to 0+} \|L\rho - (1/t)(\Lambda_t \rho - \rho)\|_1 = 0$.

To motivate the definition of the entropy production, we start from the local form of the balance equation for the entropy, well known from nonequilibrium thermodynamics,[4]

$$\frac{\partial S}{\partial t} = -\operatorname{div} \mathbf{J}_s + \sigma, \tag{1.1}$$

where S is the local entropy and $\mathbf{J}_s$ is the vector field of the entropy flow per unit area per unit time. The source term σ is the entropy production. In the phenomenological theory it is asserted that $\sigma \geq 0$, since entropy can only be created, never destroyed. For reversible processes $\sigma = 0$. This is a local formulation of the second law of thermodynamics.

In the microscopic picture, the total entropy of $\mathcal{S}$ in the state ρ is given by

$$S(\rho) = -\operatorname{tr}(\rho \log \rho). \tag{1.2}$$

If the surroundings of $\mathcal{S}$ are at temperature β^{-1}, then the total entropy flow is $\beta đ Q$, where $đ Q$ is the heat transfer from $\mathcal{S}$ to its surroundings. In other words—$đ Q$ is the change of energy of $\mathcal{S}$ due to the coupling to its surroundings and therefore—$đ Q = \operatorname{tr}[\rho L^*(H)]$ with H the Hamiltonian of $\mathcal{S}$. For certain semigroups (e.g., those obtained in the weak coupling limit[1]) the canonical ensemble $\rho_\beta = Z^{-1} e^{-\beta H}$ is invariant and for them the total entropy flow is given by

$$J_s(\rho) = \frac{d}{dt} \operatorname{tr}((\Lambda_t \rho) \log \rho_\beta)\big|_{t=0} . \tag{1.3}$$

Inserting (1.2) and (1.3) in (1.1), we obtain for the entropy production

$$\begin{aligned} \sigma(\rho) &= \frac{d}{dt} [\operatorname{tr}((\Lambda_t \rho) \log \rho_\beta) \\ &\quad - \operatorname{tr}((\Lambda_t \rho) \log(\Lambda_t \rho))]\big|_{t=0} \\ &= \frac{d}{dt} S(\Lambda_t \rho \,|\, \rho_\beta)\big|_{t=0}. \end{aligned} \tag{1.4}$$

$S(\rho \,|\, \rho_\beta)$ is the relative entropy.[5] We generalize to arbitrary quantum dynamical semigroups with a stationary state.

Definition 1: Let $\{\Lambda_t \mid t \geq 0\}$ be a quantum dynamical semigroup, and let $\rho^\circ \in T(\mathcal{H})$ be an Λ_t-invariant state, $\Lambda_t \rho^\circ = \rho^\circ$ for all $t \geq 0$. Then the *entropy production* σ (relative to ρ°) is defined by

$$\sigma(\rho) = -\frac{d}{dt} S(\Lambda_t \rho \,|\, \rho^\circ)\big|_{t=0}, \quad \rho \in T_+(\mathcal{H}), \tag{1.5}$$

whenever the derivative exists.

In the present paper we want to investigate a few properties of the entropy production and indicate its usefulness for the study of quantum dynamical semigroups. We will prove that σ is nonnegative and that σ is a convex functional. The second result turns out to be an important stability property of the entropy production which allows one to show the so-called principle of minimal entropy production. This together with other applications to irreversible thermodynamics of quantum mechanical systems weakly coupled to thermal reservoirs can be found in Ref. 6. The entropy production is shown to be a measure for the irreversibility (dissipation) of the time evolution of the system. For semigroups arising in the weak coupling limit the entropy production can be used to study their approach to stationarity (limit as $t \to \infty$) and to classify all their stationary states (cf. Sec. 3).

2. SOME PROPERTIES OF THE ENTROPY PRODUCTION

Lemma 1: $t \mapsto S(\Lambda_t \rho \,|\, \rho^\circ)$ is decreasing and continuous from the right.

Proof: Since Λ_t is completely positive and since ρ° is invariant, by Ref. 7, $S(\Lambda_t \rho \,|\, \rho^\circ) \leq S(\rho \,|\, \rho^\circ)$ for all $t \geq 0$. By the semigroup property

[a)]On leave of absence of the Fachbereich Physik der Universität München. Work supported by a DFG-research scholarship.

 0022-2488/78/1905-1227$1.00

$$S(\Lambda_{t+s}\rho\,|\,\rho^\circ) \le S(\Lambda_t\rho\,|\,\rho^\circ) \tag{2.1}$$

for all $t, s \ge 0$. $\Lambda_{t+s}\rho \to \Lambda_t\rho$ uniformly for $s \to 0_+$. Therefore, by Ref. 7, Lemma 1,

$$S(\Lambda_t\rho\,|\,\rho^\circ) \le \lim_{s\to 0} \inf S(\Lambda_{t+s}\rho\,|\,\rho^\circ), \tag{2.2}$$

which implies $\lim_{s\to 0_+} S(\Lambda_{t+s}\rho\,|\,\rho^\circ) = S(\Lambda_t\rho\,|\,\rho^\circ)$ by (2.1).

Corollary 2: Let $P\mathcal{H} = \overline{\mathrm{Ran}\rho^\circ}$. Then $T(P\mathcal{H})$ is an invariant subspace for $\{\Lambda_t \,|\, t \ge 0\}$.

Proof: Let $\mathcal{D} = \{\rho \in T_+(\mathcal{H}) \,|\, S(\rho\,|\,\rho^\circ) < \infty\}$. Since the finite rank operators in $T_+(P\mathcal{H})$ are in $\mathcal{D}$, $\bar{\mathcal{D}} = T_+(P\mathcal{H})$. By Lemma 1, $\Lambda_t\mathcal{D} \subset \mathcal{D}$ and by continuity $\Lambda_t T_+(P\mathcal{H}) \subset T_+(P\mathcal{H})$, which extends by linearity to $T(P\mathcal{H})$.

By Corollary 2, we may without loss of generality assume that $\mathrm{Ran}\rho^\circ = \mathcal{H}$.

Theorem 3: Let $\dim\mathcal{H} < \infty$ and let $\mathrm{Ran}\rho^\circ = \mathcal{H}$. Then the entropy production σ is defined on $T_+(\mathcal{H})$ and is given by

$$\sigma(\rho) = \mathrm{tr}(L(\rho)(\log\rho^\circ - \log\rho)), \tag{2.3}$$

$\rho \in T_+(\mathcal{H})$, where $\Lambda_t = e^{Lt}$, $t \ge 0$. We choose the convention that in the eigenbasis of ρ

$$\mathrm{tr}(L(\rho)\log\rho) = \sum_j (\log\rho_j)\ \langle\psi_j\,|\,L(\rho)\psi_j\rangle$$

with

$$(\log\rho_j)\langle\psi_j\,|\,L(\rho)\psi_j\rangle = \begin{cases} -\infty, & \text{if } \langle\psi_j\,|\,L(\rho)\psi_j\rangle \neq 0 \text{ and } \rho_j = 0, \\ 0 & \text{if } \langle\psi_j\,|\,L(\rho)\psi_j\rangle = 0. \end{cases}$$

We have the properties

(i) σ is convex,

(ii) σ takes value in $[0, \infty]$.

Proof: Since $\log\rho^\circ$ is bounded, $(d/dt)\,\mathrm{tr}((\Lambda_t\rho)\log\rho^\circ) = \mathrm{tr}(L(\rho)\log\rho^\circ)$. Using the spectral decomposition of $\Lambda_t\rho$, $\Lambda_t\rho = \sum_j \rho_j(t)\,|\psi_j(t)\rangle\langle\psi_j(t)|$, we obtain in the basis of ρ

$$\frac{d}{dt}\mathrm{tr}((\Lambda_t\rho)\log(\Lambda_t\rho))\Big|_{t=0}$$

$$= \sum_k \{\rho_k \log\rho_k(\langle\dot\psi_k\,|\,\psi_k\rangle + \langle\psi_k\,|\,\dot\psi_k\rangle) + \dot\rho_k\log\rho_k\}, \tag{2.4}$$

where we denoted the time derivative at $t=0$ by a dot. If $\rho_k = 0$ and $\dot\rho_k > 0$, then $\dot\rho_k\log\rho_k = -\infty$. If $\rho_k = 0$ and $\dot\rho_k = 0$, then using a Taylor expansion of $\rho_k(t)$ and performing the limit as $t \to 0_+$, one obtains $\dot\rho_k\log\rho_k = 0$. Since

$$\langle\psi_k\,|\,L(\rho)\psi_k\rangle = \rho_k(\langle\dot\psi_k\,|\,\psi_k\rangle + \langle\psi_k\,|\,\dot\psi_k\rangle) + \dot\rho_k, \tag{2.5}$$

we have $(d/dt)\mathrm{tr}((\Lambda_t\rho)\log(\Lambda_t\rho))\,|_{t=0} = \mathrm{tr}(L(\rho)\log\rho)$ together with our convention.

(i) By a theorem of Lindblad,[3] the general form of L is

$$L(\rho) = -i[H,\rho] + \tfrac{1}{2}\sum_{j\in I}[V_j, \rho V_j^*] + [V_j\rho, V_j^*], \tag{2.6}$$

where $H = H^* \in B(\mathcal{H})$ and $\sum_{j\in I} V_j^* V_j \in B(\mathcal{H})$, $V_j \in B(\mathcal{H})$. The first term in (2.3) is linear. The second term reads

$$-\mathrm{tr}(L(\rho)\log\rho)$$

$$= \sum_{j\in I} \mathrm{tr}(V_j^* V_j\rho\log\rho) - \mathrm{tr}(V_j\rho V_j^*\log\rho). \tag{2.7}$$

By a theorem of Lieb,[8] the function

$$\rho \mapsto -\mathrm{tr}(\rho^q V_j \rho^{1-q} V_j^*), \tag{2.8}$$

$0 \le q \le 1$, is convex. For $q = 0$, (2.8) is affine. Therefore, its derivative at $q = 0$

$$\rho \mapsto \mathrm{tr}(V_j\rho\log\rho V_j^*) - \mathrm{tr}(\log\rho\, V_j\rho V_j^*) \tag{2.9}$$

is convex. One checks that our convention is satisfied. This implies then the convexity of $-\mathrm{tr}(L(\rho)\log\rho)$.

(ii) $\sigma \ge 0$ follows from Lemma 1. We give here an alternative proof. By the convexity of $-\mathrm{tr}(L(\rho)\log\rho)$

$$-\mathrm{tr}(L(\lambda\rho + (1-\lambda)\rho^\circ)\log(\lambda\rho + (1-\lambda)\rho^\circ))$$

$$\le -\lambda\,\mathrm{tr}(L(\rho)\log\rho) - (1-\lambda)\mathrm{tr}(L(\rho^\circ)\log\rho^\circ)\ . \tag{2.10}$$

By the invariance of ρ°

$$0 \le \mathrm{tr}(L(\rho)\log(\lambda\rho + (1-\lambda)\rho^\circ)) - \mathrm{tr}(L(\rho)\log\rho), \tag{2.11}$$

from which the derived result follows by taking the limit as $\lambda \to 0$.

Remark: The extension of Theorem 3 to $\dim\mathcal{H} = \infty$ is most desirable, where, by Lemma 1, the natural domain of σ should be $\mathcal{D} = \{\rho \,|\, S(\rho\,|\,\rho^\circ) < \infty\}$. Of course, one can define σ by (2.3). If the quantum dynamical semigroup is uniformly continuous, $\lim_{t\to 0_+}\|\Lambda_t^* - 1\| = 0$, then L is bounded and Lindblad's theorem[3] still holds. Therefore, the so-defined σ is positive and convex on a suitable domain. However, the so-defined σ would not serve as a tool to study the approach to stationarity for which it is essential to know that σ is the time derivative of $-S(\Lambda_t\rho\,|\,\rho^\circ)$.

Corollary 4: (i) $\{\rho \in T_+(\mathcal{H}) \,|\, L(\rho) = 0\} \subset \{\rho \in T_+(\mathcal{H}) \,|\, \sigma(\rho) = 0\}$. (In general, we have a proper inclusion.) If $\sigma(\rho) = 0$ only for $\rho = \rho^\circ$, then

$$\lim_{t\to\infty} \Lambda_t\rho = \rho^\circ \mathrm{tr}\rho \tag{2.12}$$

for all $\rho \in T_+(\mathcal{H})$.

(ii) $\sigma = 0$ if and only if $L = -i[H, \cdot\,]$.

Proof: If $L = -i[H, \cdot\,]$, then $[H, \log\rho^\circ] = 0$ and therefore $\sigma = 0$. If $\sigma = 0$, then $\mathrm{tr}(L(\rho)\log\rho) = \mathrm{tr}(L(\rho)\log\rho^\circ)$. For $\rho = |\psi\rangle\langle\psi|$, $\psi \in \mathcal{H}$, according to our convention, the left side vanishes. Therefore, $\langle\psi\,|\,L^*(\log\rho^\circ)\psi\rangle = 0$ which implies $\mathrm{tr}(L(\rho)\log\rho) = 0$ for all $\rho \in T_+(\mathcal{H})$ and by (2.7)

$$\sum_{j\in I}\mathrm{tr}(V_j^* V_j\rho\log\rho) = \sum_{j\in I}\mathrm{tr}(V_j\rho V_j^*\log\rho). \tag{2.13}$$

For $\rho = |\psi\rangle\langle\psi|$, $\psi \in \mathcal{H}$, by our convention, $\langle\psi\,|\,V_j\phi\rangle = 0$ for all $\phi \in \mathcal{H}$ orthogonal to ψ and all $j \in I$ which implies $V_j = 0$. Therefore, $V_j\rho V_j^* = 0$, $V_j^* V_j\rho = 0$, $\rho V_j^* V_j = 0$ for all $\rho \in T(\mathcal{H})$. Therefore, the non-Hamiltonian part of L in (2.6) vanishes.

Remark: Lindblad introduced the dissipation function $D(L^*; X, Y) = L^*(X^*Y) - L^*(X^*)Y - X^*L^*(Y)$, which has some properties similar to the entropy production σ. $D(L^*; X, X) = 0$ is equivalent to $L = -i[H, \cdot\,]$. If $D(L^*; X, X) = 0$ implies $X = c1$ (i.e., if the dissipation function is as positive as possible), then there exists

a unique stationary state ρ° such that $\lim_{t\to\infty}\Lambda_t\rho=\rho$ for every state $\rho\in T(\mathcal{H})$.[9] (This result has recently been extended by Frigerio[10] to $\dim\mathcal{H}=\infty$ under certain technical restrictions on the generator L.)

It is surprising that the properties of the entropy production anticipated by the phenomenological theory can be proved in a rather general context. The entropy production is nonnegative. If $\sigma=0$, we have reversible motion; if $\sigma>0$ except at ρ°, we have complete dissipation. The convexity of σ is a stability property (comparable to the concavity of the entropy).

3. AN APPLICATION

The situation when the quantum mechanical system $\mathcal{S}$ is weakly coupled to a thermal reservoir $\mathcal{R}$ is of particular physical interest. In the weak coupling limit, $\lambda\to 0$, $t\to\infty$, but $\lambda^2 t=\tau$ fixed, the reduced dynamics of $\mathcal{S}$ in the interaction picture is described by a quantum dynamical semigroup $\{\Lambda_t\,|\,t\geq 0\}$. Let $H=\sum_j\epsilon_j P_j$ be the self-adjoint Hamiltonian of $\mathcal{S}$ such that $e^{-\beta H}\in T(\mathcal{H})$ for $\beta>0$. If the interaction between $\mathcal{S}$ and $\mathcal{R}$ is choosen to be $\sum_{\alpha\in I}V_\alpha\otimes V_\alpha^R$, where $V_\alpha=V_\alpha^*\in B(\mathcal{H})$, $\sum_{\alpha\in I}\|V_\alpha\|^2<\infty$, V_α^R bounded, self-adjoint operators pertaining to the reservoir with thermal average $\langle V_\alpha^R\rangle=0$, then the bounded generator $L:T(\mathcal{H})\to T(\mathcal{H})$ of Λ_t, $t\geq 0$, is of the form[3,11]

$$L(\rho)=\sum_{\omega\in Sp([H,\circ])}\sum_{\alpha,\beta\in I}\{-is_{\alpha\beta}(\omega)[V_\beta(\omega)^*V_\alpha(\omega),\rho]$$
$$+h_{\alpha\beta}(\omega)([V_\alpha(\omega)\rho,V_\beta(\omega)^*]+[V_\alpha(\omega),\rho V_\beta(\omega)^*])\}$$
$$=L_a(\rho)+L_s(\rho). \tag{3.1}$$

The convergence on the right-hand side is in the weak topology of $T(\mathcal{H})$. $\mathrm{Sp}([H,\cdot])$ denotes the spectrum of the Liouville–von Neumann operator $i[H,\cdot]$. It is the set $\{\epsilon_m-\epsilon_n\,|\,\epsilon_m,\epsilon_n\in\mathrm{Sp}(H)\}$. The operators $V_\alpha(\omega)\in B(\mathcal{H})$ are defined by

$$V_\alpha(\omega)=\sum_{\epsilon_m-\epsilon_n=\omega}P_nV_\alpha P_m=V_\alpha(-\omega)^*. \tag{3.2}$$

$h_{\alpha\beta}$ is the Fourier transform of the time correlation function $\langle V_\beta^R V_\alpha^R(t)\rangle$ of the reservoir and therefore satisfies the KMS boundary condition

$$h_{\beta\alpha}(-\omega)=e^{-\beta\omega}h_{\alpha\beta}(\omega). \tag{3.3}$$

$\{h_{\alpha\beta}(\omega)\}$ is a positive matrix for all $\omega\in\mathbb{R}$. $s_{\alpha\beta}$ is the Hilbert transform of $h_{\alpha\beta}$. Using (3.3), one checks that the canonical ensemble $\rho_\beta=[\mathrm{tr}(e^{-\beta H})]^{-1}e^{-\beta H}$ is Λ_t-invariant. Furthermore, the detailed balance property[12] $L_a(\rho_\beta)=0$, $L_s^*(A\rho_\beta)=L_s(A)\rho_\beta$ holds, $A\in B(\mathcal{H})$. [Equivalently, L_a^* is antisymmetric and L_s^* is symmetric in the pre-Hilbert space $B(\mathcal{H})$ equipped with the scalar product $\langle A\,|\,B\rangle_\beta=\mathrm{tr}(A^*B\rho_\beta)$.]

Now let $\dim\mathcal{H}=N<\infty$. Then one obtains for the entropy production σ (relative to ρ_β)

$$\sigma(\rho)=\mathrm{tr}[L(\rho)(\log\rho_\beta-\log\rho)]$$
$$=\sum_{\omega\in\mathrm{Sp}(H,\circ])}\sum_{j,k=1}^{N}\left\{\sum_{\alpha,\beta\in I}h_{\alpha\beta}(\omega)\langle\psi_j\,|\,V_\alpha(\omega)\psi_k\rangle\langle\psi_k\,|\,V_\beta(\omega)^*\psi_j\rangle\right\}$$
$$\times\{(\rho_k-\rho_je^{-\beta\omega})(\log\rho_k-\log\rho_j+\beta\omega\} \tag{3.4}$$

in the eigenbasis of $\rho=\sum_{j=1}^N\rho_j|\psi_j\rangle\langle\psi_j|$.

Lemma 5: Insert $\sigma(\rho)=0$ if and only if $L(\rho)=0$.

Proof: Let $\sigma(\rho)=0$. Since $\{h_{\alpha\beta}(\omega)\}$ is a positive matrix, we can diagonalize it. Let $h_\gamma(\omega)>0$ be its eigenvalues, and let $\{\tilde V_\gamma(\omega)\}$ be $\{V_\alpha(\omega)\}$ expressed in the eigenbasis of $\{h_{\alpha\beta}(\omega)\}$. Then

$$\sigma(\rho)=\sum_{\omega\in\mathrm{Sp}([H,\circ])}\sum_{j,k=1}^{N}\sum_{\gamma\in I}h_\gamma(\omega)\,|\,\langle\psi_j\,|\,\tilde V_\gamma(\omega)\psi_k\rangle\,|^2$$

$$(\rho_k-\rho_je^{-\beta\omega})(\log\rho_k-\log\rho_j+\beta\omega)=0$$

implies

$$h_\gamma(\omega)\langle\psi_j\,|\,\tilde V_\gamma(\omega)\psi_k\rangle(\rho_k-\rho_je^{-\beta\omega})=0 \tag{3.5}$$

for all $\gamma\in I$, $\omega\in\mathrm{Sp}([H,\cdot])$ and $j,k=1,\ldots,N$. By (3.2)

$$[\tilde V_\gamma(\omega),\rho e^{\beta H}]=0 \tag{3.6}$$

for all $\gamma\in I$ and $\omega\in\mathrm{Sp}([H,\cdot])$. Let

$$\mathcal{N}=\{\tilde V_\gamma(\omega)\,|\,\gamma\in I,\ \omega\in\mathrm{Sp}([H,\circ])\}', \tag{3.7}$$

where the prime denotes the commutant in $B(\mathcal{H})$. Then $\rho=A\rho_\beta$ with $A\in\mathcal{N}$. We write the generator (3.1) in terms of the $\tilde V_\gamma(\omega)$. Then $[\tilde V_\gamma(\omega)^*V_\gamma(\omega),A\rho_\beta]=0$ since $A\in\mathcal{N}$, which implies $L_a(\rho)=0$. By the detailed balance property $L_s(A\rho_\beta)=L_s^*(A)\rho_\beta=0$ since $A\in\mathcal{N}$. Therefore, $L(\rho)=0$.

Lemma 5 gives a classification of all stationary states of L. They are of the form

$$\rho=A\rho_\beta \tag{3.8}$$

with $A\in\mathcal{N}$. We have $L_a(A\rho_\beta)=0$ and $L_s(BA\rho_\beta)=L_s^*(BA)\rho_\beta=L_s^*(B)A\rho_\beta$. Therefore, L satisfies the detailed balance property with respect to any stationary state. (We note that by defining the entropy production through (3.4) all statements made so far remain valid for $\dim\mathcal{H}=\infty$.)

Since $\sigma(\rho)=-(d/dt)S(\Lambda_t\rho\,|\,\rho_\beta)$ and by Lemma 5,

$$\lim_{t\to\infty}\Lambda_t\rho=A(\rho)\rho_\beta \tag{3.9}$$

for all $\rho\in T_+(\mathcal{H})$, where $A(\rho)\in\mathcal{N}$ depends on ρ. If $\mathcal{N}=\{\mathbb{C}1\}$, the unique stationary state is ρ_β and every initial state relaxes to it. We obtain a condition more directly related to the Hamiltonian and the interaction, if we choose $\{h_{\alpha\beta}(\omega)\}>0$ for all $\omega\in\mathrm{Sp}([H,\cdot])$. In that case $\{H,V_\alpha\,|\,\alpha\in I\}'=\{\mathbb{C}1\}$ implies (cf. Ref. 9) $\{V_\alpha(\omega)\,|\,\alpha\in I,\ \omega\in\mathrm{Sp}([H,\cdot])\}'=\{\mathbb{C}1\}$, which is equivalent to $\mathcal{N}=\{\mathbb{C}1\}$. The latter conditions have a direct physical interpretation: If all relevant frequencies of $\mathcal{S}$ are coupled to the reservoir and if the Hamiltonian H and the interaction operators V_α, $\alpha\in I$, together are sufficiently incompatible, then the reservoir drives the system to thermal equilibrium.

[1] E.B. Davies, Comm. Math. Phys. 39, 91 (1974).
[2] V. Gorini and A. Kossakowski, J. Math. Phys. 17, 1298 (1976).
[3] G. Lindblad, Comm. Math. Phys. 48, 119 (1976).
[4] S.R. de Groot and P. Mazur, *Non-Equilibrium Thermodynamics* (North-Holland, Amsterdam, 1962).
[5] G. Lindblad, Comm. Math. Phys. 33, 305 (1973).

[6]H. Spohn and J. L. Lebowitz, Irreversible Thermodynamics for Quantum Systems Weakly Coupled to Thermal Reservoirs, to appear in Advan. Phys. Chem.
[7]G. Lindblad, Comm. Math. Phys. 40, 147 (1975).
[8]E. Lieb, Advan. Math. 11, 267 (1973).
[9]H. Spohn, "An algebraic condition for the approach to equilibrium of an open N-level system," Lett. Math. Phys. 2, 33 (1977).
[10]A. Frigerio, "Quantum dynamical semigroups and approach to equilibrium," to appear in Lett. Math. Phys.
[11]V. Gorini, A. Frigerio, M. Verri, A. Kossakowski, and E. C. G. Sudarshan, "Properties of quantum markovian master equations," preprint, University of Milan, 1976.
[12]A. Kossakowski, A. Frigerio, V. Gorini, and M. Verri, "Detailed balance and KMS condition," preprint, University of Milan, 1976.

ERRATA

Erratum: Continuum calculus and Feynman's path integrals [J. Math. Phys. 17, 1988 (1976)]

L. L. Lee

School of Chemical Engineering and Materials Science, University of Oklahoma, Norman, Oklahoma 73019
(Received 31 October 1977)

The definition (2.1.) giving the rational derivative of a function f on a Banach space was not precise. A proper construction in a complex Banach space is given in the following:

Let C^B be the set of functions from B, the Banach space, to C, the complex number field. The function, f, maps B into C, i.e., $f \in C^B$. Let S be the support of f, $(S \neq \phi)$, and $S_i = \bar{S} - \partial S$ be the interior of S (the overbar denotes the closure and ∂ denotes the boundary). Given a point $t \in S_i$, there exists a neighborhood, V of t and $V \subset S_i$. We first define a linear form LRf_t: $B \to C$, for f at the point t. In fact, LRf_t will be recognized as the Fréchet differential[1,2] for $\ln f$.

Definition 1: The Fréchet differential LRf_t for $\ln f$ at t:

(i) LRf_t is a linear form on B.

(ii) For any $\epsilon > 0$, there exist $\delta > 0$, $\ni$

$$|\ln f(t+b) - \ln f(t) - LRf_t(b)| < \epsilon \|b\|$$

whenever

the norm of $b \in V$, $\|b\| < \delta$.

We note that $\ln f(t)$ is defined since $t \in S$. If such LRf_t exists, we can define the r derivative of f at t as follows:

Definition 2: The r derivative of f in a Banach space:

The r derivative of f, $Rf(t)/Rt$, at t is a function on $B \to C$ defined in terms of the Fréchet differential LRf_t, $\forall\, s \in B$, where LRf_t is defined, by

$$Rf(t)/Rt(s) = \exp[LRf_t(s)].$$

[1] N. Dunford and J. T. Schwartz, *Linear Operators, Vol. I: General Theory* (Interscience, New York, 1967), pp. 92ff.
[2] M. A. Krasnosel'skii, *Topological Methods in the Theory of Nonlinear Integral Equations* (Pergamon, New York, 1964), pp. 68ff.

Erratum: Post-Newtonian two-body and *n*-body problems with electric charge in general relativity [J. Math. Phys. 18, 1818 (1977)]

B. M. Barker

Department of Physics and Astronomy, The University of Alabama, University, Alabama 35486

R. F. O'Connell

Department of Physics and Astronomy, Louisiana State University, Baton Rouge, Louisiana 70803
(Received 3 November 1977)

In this paper the three terms

$$-(\alpha_g + \alpha_p)\left(\frac{1}{r_{ij}^2} + \frac{1}{r_{ik}^2}\right)\frac{\mathbf{r}_{ij}\circ\mathbf{r}_{ik}}{r_{ij}\,r_{ik}},$$

$$-(\alpha_g + \alpha_p)\left(\frac{1}{r_{ji}^2} + \frac{1}{r_{jk}^2}\right)\frac{\mathbf{r}_{ji}\cdot\mathbf{r}_{jk}}{r_{ji}\,r_{jk}},$$

$$-(\alpha_g + \alpha_p)\left(\frac{1}{r_{ki}^2} + \frac{1}{r_{kj}^2}\right)\frac{\mathbf{r}_{ki}\cdot\mathbf{r}_{kj}}{r_{ki}\,r_{kj}},$$

in the last three lines of Eq. (60) are in error and should be replaced, respectively, by

$$-\left(\frac{\alpha_p}{r_{ji}^2} + \frac{\alpha_g}{r_{jk}^2}\right)\frac{\mathbf{r}_{ji}\cdot\mathbf{r}_{jk}}{r_{ji}\,r_{jk}} - \left(\frac{\alpha_p}{r_{ki}^2} + \frac{\alpha_g}{r_{kj}^2}\right)\frac{\mathbf{r}_{ki}\cdot\mathbf{r}_{kj}}{r_{ki}\,r_{kj}},$$

$$-\left(\frac{\alpha_p}{r_{ij}^2} + \frac{\alpha_g}{r_{ik}^2}\right)\frac{\mathbf{r}_{ij}\circ\mathbf{r}_{ik}}{r_{ij}\,r_{ik}} - \left(\frac{\alpha_g}{r_{ki}^2} + \frac{\alpha_p}{r_{kj}^2}\right)\frac{\mathbf{r}_{ki}\cdot\mathbf{r}_{kj}}{r_{ki}\,r_{kj}},$$

$$-\left(\frac{\alpha_g}{r_{ij}^2} + \frac{\alpha_p}{r_{ik}^2}\right)\frac{\mathbf{r}_{ij}\cdot\mathbf{r}_{ik}}{r_{ij}\,r_{ik}} - \left(\frac{\alpha_g}{r_{ji}^2} + \frac{\alpha_p}{r_{jk}^2}\right)\frac{\mathbf{r}_{ji}\circ\mathbf{r}_{jk}}{r_{ji}\,r_{jk}}.$$

These changes do not affect any of the other results or conclusions of this paper.

0022-2488/78/1905-1231$0.00