Symmetry and generating relations for Clebsch–Gordan coefficients arising from automorphisms

R. Dirl
Institut für Theoretische Physik, TU Wien, A-1040 Wien, Karlsplatz 13, Austria

(Received 17 December 1981; accepted for publication 1 April 1983)

A new class of symmetry and generating relations for Clebsch–Gordan coefficients is derived. It is associated with certain elements of the full automorphism group. Our approach turns out to be useful for an efficient computation of Clebsch–Gordan coefficients.

PACS numbers: 02.20. + b

INTRODUCTION

The importance of CG coefficients for physical applications is uncontested. For instance, they are employed to define model Hamiltonians, to introduce tensor order parameters for structural phase transitions, to determine selection rules of transitions, or to calculate transition probabilities. To obtain an efficient computation and tabulation of CG coefficients, we utilize reasonably as far as possible each admissible "symmetry" and "generating" relation for these coefficients. We define symmetry relations for CG vectors $\mathbf{C}_j^{\alpha\alpha';\alpha''w}$ as correlations between CG vectors that belong to the *same* triplet $(\alpha,\alpha',\alpha'')$ of equivalence classes. Generating relations correlate CG vectors that belong to *different* triplets of equivalence classes. The purpose of the present paper is to introduce a new class of symmetry and generating relations for CG vectors. These relations are engendered by certain elements of the full automorphism group[1] of the group in question. In some cases these relations involve other symmetries of CG vectors, such as their behavior under complex conjugation or their permutational symmetries. We make general estimates and give some examples of the utility of the "automorphism approach." We argue under what circumstances generating relations are preferred to a direct calculation of CG vectors. We are primarily interested in an application of this approach to the systematization of the computation of space group CG coefficients.

The material of the present article is organized as follows: In Sec. 1 we recall briefly the basic properties of the automorphism group $A(G)$ and its representation theory for which G should be finite or countable. For countable groups G it is assumed that each unirrep is finite-dimensional, as for space groups. Compact continuous groups could be considered, but we wish to avoid topological questions. Since CG vectors of G shall be inspected with respect to its automorphism group $A(G)$, we embed G and $A(G)$ in a supergroup that contains G as normal subgroup. If the center $Z(G)$ of G is trivial, we choose $A(G)$ as this supergroup. If the center $Z(G)$ is nontrivial, we choose the so-called "holomorph"[1] $H(G) = G \circledS A(G)$ as the supergroup. We also point out that projective matrix representations of the "little cogroups"[2] $A(G)^\alpha/G$ of $A(G)$ [or $A(G)^\alpha \simeq H(G)^\alpha/G$ of $H(G)$] occur in general. We denote "little groups"[1,2] of $A(G)$ by $A(G)^\alpha$ and those of $H(G)$ by $H(G)^\alpha$. Since we also take complex conjugation into account, we are able to make predictions concerning the reality of factor systems. In Sec. 2 we recall some properties of CG vectors. We derive in Sec. 3 simple but nontrivial relations between "multiplicities" $m(\alpha,\alpha';\alpha'')$. These relations are associated with certain "outer" automorphisms. Multiplicities indicate how often a given unirrep of G is contained as constituent in the considered Kronecker product. Schur's lemma with respect to G is used to Sec. 4 to derive symmetry relations for CG vectors of G. However, they are only nontrivial when the subgroup $A(G)_i$ of "inner" automorphisms is a proper subgroup of the so-called "triple intersection group" $A(G)^{\alpha\alpha'\alpha''} = A(G)^\alpha \cap A(G)^{\alpha'} \cap A(G)^{\alpha''}$. Complex conjugation gives further features of the factor systems and special representations of the little cogroups. In Sec. 5 three different classes of new generating relations are defined. They are associated with specific outer automorphisms. We derive in Sec. 6 correlations between CG vectors, which correspond to "special" symmetries of $3jm$ factors. These relations are associated with the multiplication of a Kronecker product by one-dimensional nontrivial unirreps. Finally in Sec. 7 we apply the automorphism approach to space group CG vectors in order to demonstrate its utility and to give some striking examples. In addition these examples show that automorphisms, complex conjugation, and multiplication by one-dimensional nontrival unirreps lead to *different* symmetry and generating relations for CG vectors.

1. AUTOMORPHISM GROUP AND HOLOMORPH

We recall briefly some properties of automorphism groups. Since CG coefficients of a given group G are to be investigated with regard to their behavior concerning its automorphism group $A(G)$, we have to embed G and $A(G)$ in a supergroup that contains G as a normal subgroup. $A(G)$ can be chosen as the supergroup if an only if the group of inner automorphisms $A(G)_i$ of G is isomorphic to G, because this requires that the center $Z(G)$ be trivial. For each group with a nontrivial center $Z(G)$ cannot be embedded in its automorphism group, and we extend G and its full automorphism group $A(G)$ to a semidirect product group $H(G)$, which is usually called the "holomorph"[1] of G.

Since in general G cannot be regarded as subgroup of automorphism group $A(G)$, let us start with the definition of the holomorph $H(G)$ of G:

$$H(G) = G \circledS A(G). \tag{1.1}$$

To be more specific, the elements of $A(G)$ are denoted by

 0022-2488/83/081935-10$02.50

$a \in A(G)$ and satisfy

$$a(g)a(g') = a(gg') \quad \text{for all } g,g' \in G,$$
$$a(e) = e \quad \text{for all } a \in A(G), \tag{1.2}$$
$$a_j(a_k(g)) = (a_j a_k)(g) \quad \text{for all } g \in G,$$

where $a_j a_k \in A(G)$. Moreover, adopting the same notation for group elements of $H(G)$ as introduced in Ref. 1, i.e., $(g;a) \in H(G)$, the multiplication law reads

$$(g_m;a_j)(g_n;a_k) = (g_m a_j(g_n);a_j a_k). \tag{1.3}$$

We assume in addition that a complete set $D = \{D^\alpha : \alpha \in A_G\}$ of ordinary matrix unirreps of G is given, where α indicates an equivalence class of G. The set of equivalence classes is denoted by A_G. For each $a \in A(G)$ the following equivalence relation holds:

$$D^\alpha(a(g)) \sim D^{a(\alpha)}(g) \quad \text{for all } g \in G. \tag{1.4}$$

Note that $a(\alpha) \in A_G$ indicates a well-defined element of A_G for each $\alpha \in A_G$ and $a \in A(G)$, where $n_\alpha \equiv \dim D^\alpha = \dim D^{a(\alpha)} \equiv n_{a(\alpha)}$ must be satisfied. The elements of $A(G)$ not only map G onto G, but also A_G onto A_G. The mapping $a:\alpha \to a(\alpha)$, $\alpha \in A_G$, must not depend on the chosen form of the matrix unirreps D^α of G, even though it is defined by (1.4). We may write Eq. (1.4) as an equality by inserting an n_α-dimensional unitary matrix $Z^\alpha(a)$:

$$D^\alpha(a(g)) = Z^\alpha(a) D^{a(\alpha)}(g) Z^\alpha(a)^+ \quad \text{for all } g \in G. \tag{1.5}$$

For obvious reasons we define

$$A(G)^\alpha = \{a \in A(G) \mid a(\alpha) = \alpha\}, \tag{1.6}$$

which is just the corresponding "little cogroup"[2] of D^α with respect to $H(G)$. Therefore,

$$H(G)^\alpha = G \circledS A(G)^\alpha \tag{1.7}$$

presents the corresponding "little group" of D^α with respect to $H(G)$, and the factor group $H(G)^\alpha/G$ is isomorphic to $A(G)^\alpha$. The set of matrices $Z^\alpha = \{Z^\alpha(a) : a \in A(G)^\alpha\}$ defines a n_α-dimensional representation of $A(G)^\alpha$, which is in general a projective one.[3,4] This means

$$Z^\alpha(a) Z^\alpha(a') = Q^\alpha(a,a') Z^\alpha(aa') \quad \text{for all } a,a' \in A(G)^\alpha, \tag{1.8}$$

where $Q^\alpha : A(G)^\alpha \times A(G)^\alpha \to \mathbb{C}$ is a standard factor system of $A(G)^\alpha$, which (up to equivalence[4]) does not depend on the chosen form of the unirreps D^α.

A factor system for inner automorphisms occurs if the center $Z(G)$ of G is nontrivial, as can be seen from the following. Denoting the subgroup of inner automorphisms by $A(G)_i$, we know that

$$A(G)_i \simeq G/Z(G) \tag{1.9}$$

holds. We adopt the following notation for coset representatives:

$$G_j \in G:Z(G) \quad \text{and} \quad z_k \in Z(G), \tag{1.10}$$

where $G:Z(G)$ denotes the set of (left) coset representatives of G with respect to $Z(G)$; each element $g \in G$ can be uniquely written as $g = G_j z_k; G_j \in G:Z(G)$ and $z_k \in Z(G)$. Hence each $a_j \in A(G)_i$ can be identified with a particular $G_j \in G:Z(G)$; i.e.,

$$a_j = G_j, \quad a_{G_j} = G_j g G_j^{-1} \quad \text{for all } g \in G. \tag{1.11}$$

Defining

$$Z^\alpha(a_j) = D^\alpha(G_j) \quad \text{for all } = a_{G_j} \in A(G)_i \tag{1.12}$$

and using $G_j G_k = G_{jk} z_{jk}$, $G_{jk} \in G:Z(G)$ and $z_{jk} \in Z(G)$, it follows that

$$Z^\alpha(a_j) Z^\alpha(a_k) = D^\alpha(z_{jk}) Z^\alpha(a_{G_{jk}}) \quad \text{for all } a_j,a_k \in A(G)_i. \tag{1.13}$$

Introducing $D^\alpha(z) = e^{i\gamma^\alpha(z)} 1_\alpha$, $z \in Z(G)$, we have

$$Q^\alpha(a_{G_j},a_{G_k}) = e^{i\gamma^\alpha(z_{jk})} \quad \text{for all } a_j,a_k \in A(G)_i, \tag{1.14}$$

which forms a standard factor system of $A(G)_i$.

As has already been pointed out, each group G can be embedded in its automorphism group $A(G)$, if $Z(G)$ is trivial. This means that G forms a normal subgroup of $A(G)$. For such cases $A(G)^\alpha$ is now the "little group" of D^α with respect to $A(G)$ and the factor group $A(G)^\alpha/G$ is the corresponding "little cogroup" of D^α. Observe that in comparison to $G \lhd H(G)^\alpha \subseteq H(G)$ now $G \lhd A(G)^\alpha \subseteq A(G)$ holds, irrespective of which α has been considered. But note also that $A(G)^\alpha$ need not be a normal subgroup of $A(G)$.

Irrespective of whether G can be embedded in $A(G)$ or whether the holomorph must be considered for the following, a left coset decomposition of $A(G)$ with respect to $A(G)^\alpha$ is chosen,

$$a_j^\alpha \in A(G):A(G)^\alpha, \quad j = 1,2,\ldots,|A(G)|/|A(G)^\alpha| \tag{1.15}$$

so that each element $a \in A(G)$ can be uniquely expressed in terms of $a_j^\alpha \in A(G):A(G)^\alpha$ and $a^\alpha \in A(G)^\alpha$. The subsets

$$\Delta(\alpha) = \{\alpha' \in A_G \mid \alpha' = a_j^\alpha(\alpha);\ a_j^\alpha \in A(G):A(G)^\alpha\} \subset A_G \tag{1.16}$$

of A_G define orbits, where each orbit is associated with a certain $\alpha \in A_G$. Accordingly, A_G is decomposable into disjoint subsets $\Delta(\alpha)$, $\alpha \in A_G$, where ΔA_G is a subset of A_G and is usually called the "representation domain" of G with respect to $A(G)$ and hence also with respect to $H(G)$. Therefore, outer automorphisms correlate inequivalent unirreps of G in a generic way whenever $A(G)^\alpha$ is a proper subgroup of $A(G)$.

Since Eq. (1.5) holds for each $a \in A(G)$, let us take $a = a_j a_k$ for some $a_j,a_k \in A(G)$. Equation (1.5) yields

$$D^\alpha((a_j a_k)(g)) = Z^\alpha(a_j a_k) D^{(a_j a_k)(\alpha)}(g) Z^\alpha(a_j a_k)^+ = D^\alpha(a_j(a_k(g))) \quad \text{for all } g \in G, \tag{1.17}$$

which implies

$$(a_j a_k)(\alpha) = a_k(a_j(\alpha)) \quad \text{for all } \alpha \in A_G, \tag{1.18}$$
$$Z^\alpha(a_j) Z^{a_j(\alpha)}(a_k) = Q^\alpha(a_j,a_k) Z^\alpha(a_j a_k). \tag{1.19}$$

When the matrices $Z^\alpha(a)$, $a \in A(G)$ and $\alpha \in A_G$, have been determined independently, it is necessary to consider the unimodular quantities $Q^\alpha : A(G) \times A(G) \to \mathbb{C}$, which satisfy

$$Q^\alpha(a_i,a_j) Q^\alpha(a_i a_j,a_k) = Q^\alpha(a_i,a_j a_k) Q^{a_i(\alpha)}(a_j,a_k). \tag{1.20}$$

These quantities generalize the concept of factor systems, since Eqs. (1.19) cannot be compared with the usual definition of projective representations. Certainly when restricting the elements of $A(G)$ to those of $A(G)^\alpha$ the corresponding quantities Q^α form a standard factor system of $A(G)^\alpha$.

Since the projective matrix representations Z^α of

$A(G)^\alpha$ will play an essential role for the subsequent discussion, it is useful to gain more insight into the structure of their factor systems. For this purpose let us take also into account complex conjugation in order to give predictions about the factor systems Q^α. It is well known[1,2] that complex conjugation classifies unirreps into three classes:

$$D^\alpha(g)^* = U^\alpha + D^{\alpha^*}(g) U^\alpha \quad \text{for all } g \in G; \tag{1.21}$$

type I(*)—orthogonal: $U^\alpha U^{\alpha *} = +1_\alpha, \quad \alpha = \alpha^*;$ (1.22)

type II(*)—symplectic: $U^\alpha U^{\alpha *} = -1_\alpha, \quad \alpha = \alpha^*;$ (1.23)

type III(*)—complex: $U^\alpha U^{\alpha +} = 1_\alpha, \quad \alpha \neq \alpha^*;$ (1.24)

When (1.22) or (1.23) is realized, Schur's lemma with respect to G requires

$$Z^\alpha(a)^* = e^{i\omega^\alpha(a)} U^{\alpha +} Z^\alpha(a) U^\alpha \quad \text{for all } a \in A(G)^\alpha \tag{1.25}$$

for some well-defined unimodular factors $e^{i\omega^\alpha(a)}$, $a \in A(G)^\alpha$, since U^α and $Z^\alpha(a)$, $a \in A(G)^\alpha$, are assumed to be known. Now from (1.8) and (1.25) the factor system satisfies

$$Q^\alpha(a,a')^* = e^{i[\omega^\alpha(a) + \omega^\alpha(a') - \omega^\alpha(aa')]} Q^\alpha(a,a') \quad \text{for all } a,a' \in A(G)^\alpha, \tag{1.26}$$

that is, the factor system is *real* up to equivalence. However, since, on the other hand, the matrices $Z^\alpha(a)$, $a \in A(G)^\alpha$, can be chosen so that $e^{i\omega^\alpha(a)} = 1$ for each $a \in A(G)^\alpha$, the relevant set Z^α of matrices $Z^\alpha(a)$, $a \in A(G)^\alpha$, belongs to a real factor system. Their behavior under complex conjugation is described by the same unitary matrix U^α as occurs in (1.25). But note the difference that D^α is a unirrep of G and Z^α is a projective representation of $A(G)^\alpha$. Moreover, Z^α forms a projective unirrep of $A(G)^\alpha$, since it follows from a character test that $Z^\alpha(a)$, $a \in A(G)_i \sim G/Z(G)$ forms a projective unirrep of $A(G)_i$.

Taking finally the case of (1.24), where α and α^* distinguish two inequivalent unirreps of G related by

$$A(G)^{\alpha^*} = A(G)^\alpha, \quad a \in A_{\mathrm{III}(*)} \tag{1.27}$$

we may derive

$$Z^{\alpha^*}(a) = e^{-i\upsilon^\alpha(a)} U^\alpha Z^\alpha(a)^* U^{\alpha +} \quad \text{for all } a \in A(G)^\alpha. \tag{1.28}$$

As above, the phase factors $e^{i\upsilon^\alpha(a)}$, $a \in A(G)^\alpha$, cannot be omitted, if the matrices $Z^\alpha(a)$ and $Z^{\alpha^*}(a)$, $a \in A(G)^\alpha$, have been determined independently. On the other hand, Eqs. (1.28) can be utilized to define $Z^{\alpha^*}(a)$ from $Z^\alpha(a)$ by choosing the phase factors $e^{-i\upsilon^\alpha(a)}$ in some way. If this is done Z^α and Z^{α^*} are type III(*) projective unirreps of $A(G)^\alpha$, whose factor systems are correlated by

$$Q^{\alpha^*}(a,a') = e^{-i[\upsilon^\alpha(a) + \upsilon^\alpha(a') - \upsilon^\alpha(aa')]} Q^\alpha(a,a')^* \quad \text{for all } a,a' \in A(G)^\alpha. \tag{1.29}$$

In contrast to the previous cases, Eqs. (1.29) do not predict whether the factor systems for type III(*) projective unirreps of $A(G)^\alpha$ are real or complex. But note once again that the behavior of $Z^\alpha(a)$, $a \in A(G)^\alpha$, under complex conjugation is described by the same unitary matrix U^α as occurs in (1.24).

2. CG VECTORS FOR G

In a general method of systematically computing CG coefficients,[5] the columns ("CG vectors") of the CG matrices are interpreted as symmetry adapted vectors under the action of the Kronecker product, i.e.,

$$D^{\alpha\alpha'}(g) \mathbf{C}_j^{\alpha\alpha';\alpha'' w} = \sum_k D^{\alpha''}_{kj}(g) \mathbf{C}_k^{\alpha\alpha';\alpha'' w},$$

$$\alpha'' \in A_G, \quad w = 1,2,..\, m(\alpha,\alpha';\alpha''), \quad j = 1,2,..\, n_{\alpha''}, \tag{2.1}$$

where $D^{\alpha\alpha'}(g) = D^\alpha(g) \otimes D^{\alpha'}(g)$, $g \in G$. $m(\alpha,\alpha';\alpha'')$ denotes the so-called "multiplicity" and indicates how often the unirrep $D^{\alpha''}$ occurs in $D^{\alpha\alpha'}$. $m(\alpha,\alpha';\alpha'')$ is given by the character formula[1,2,6]

$$m(\alpha,\alpha';\alpha'') = |G|^{-1} \sum_{g \in G} \chi^\alpha(g) \chi^{\alpha'}(g) \chi^{\alpha''}(g)^*. \tag{2.2}$$

Note the asymmetry in this formula, the third simple character is complex conjugated. By changing the definition, the more symmetric $3jm$ symbols may be used.[6,7] The CG coefficients and $3jm$ symbols are related by unitary transformations, which always can be chosen real, if certain conditions are satisfied.[8,9]

3. AUTOMORPHISM SYMMETRIES OF MULTIPLICITIES

It is readily verified from (1.5) and (2.2) that for each triplet $(\alpha,\alpha',\alpha'')$ the relation

$$m(\alpha,\alpha';\alpha'') = m(a(\alpha),a(\alpha');a(\alpha'')) \quad \text{for all } a \in A(G) \tag{3.1}$$

holds. Restricting the elements $a \in A(G)$ to elements of the "triple intersection group"

$$A(G)^{\alpha\alpha'\alpha''} = A(G)^\alpha \cap A(G)^{\alpha'} \cap A(G)^{\alpha''}, \tag{3.2}$$

Eqs. (3.1) are trivial. However, if $A(G)^{\alpha\alpha'\alpha''}$ is a proper subgroup of $A(G)$, it follows that for each nontrivial left coset representative

$$a_t \in A(G) : A(G)^{\alpha\alpha'\alpha''} \tag{3.3}$$

(3.1) is nontrivial, since at least one of the quantities $a_t(\alpha), a_t(\alpha'), a_t(\alpha'')$ is different from α, α', α'', respectively. Taking into account the notations (1.15), Eqs. (3.1) are readily extended to

$$\begin{aligned} &m(a_i^\alpha(\alpha),a_j^{\alpha'}(\alpha');a_k^{\alpha''}(\alpha'')) \\ &\quad = m(a(a_i^\alpha(\alpha)),a(a_j^{\alpha'}(\alpha'));a(a_k^{\alpha''}(\alpha''))) \\ &\quad = m((a_i^\alpha a)(\alpha),(a_j^{\alpha'} a)(\alpha');(a_k^{\alpha''} a)(\alpha'')) \quad \text{for all } a \in A(G), \end{aligned} \tag{3.4}$$

where $a_1^\gamma \in A(G) : A(G)^\gamma$ denotes corresponding left coset representatives. Summarizing, each element of the triple intersection group (3.2) maps the corresponding triplet $(\alpha, \alpha', \alpha'')$ onto itself, whereas each nontrivial left coset representative $a_t \in A(G) : A(G)^{\alpha\alpha'\alpha''}$ maps $(\alpha,\alpha',\alpha'')$ on a different triplet $(a_t(\alpha),a_t(\alpha'),a_t(\alpha''))$. Thus the latter give rise to nontrivial symmetry relations of the corresponding multiplicities and ensure that there exists a one-to-one correspondence

between the decomposition
$D^{\alpha\alpha'}(g) \sim \Sigma_{\alpha''} \oplus m(\alpha,\alpha';\alpha'')D^{\alpha''}(g)$ and

$$D^{a_t(\alpha),a_t(\alpha')}(g) \sim \sum_{\alpha''} \oplus m(\alpha,\alpha';\alpha'')D^{a_t(\alpha'')}(g). \tag{3.5}$$

We call the set
$\Delta(\alpha\alpha'\alpha'') = \{(a_t(\alpha),a_t(\alpha'),a_t(\alpha''):a_t \in A(G):A(G)^{\alpha\alpha'\alpha''}\}$ a "triplet orbit." This suggests that the product set $A_G \times A_G \times A_G$ decomposes into mutually disjoint subsets $\Delta(\alpha\alpha'\alpha'')$. We have no proof of this conjecture.

We introduce the "pair intersection group"

$$A(G)^{\alpha\alpha'} = A(G)^{\alpha} \cap A(G)^{\alpha'}. \tag{3.6}$$

Clearly,

$$A(G)^{\alpha\alpha'\alpha''} \subseteq A(G)^{\alpha\alpha'}. \tag{3.7}$$

When $A(G)^{\alpha\alpha'\alpha''}$ is a proper subgroup of $A(G)^{\alpha\alpha'}$, each left coset representative

$$b_t \in A(G)^{\alpha\alpha'}:A(G)^{\alpha\alpha'\alpha''} \tag{3.8}$$

maps the triplets $(\alpha,\alpha',\alpha'')$ onto $(\alpha,\alpha',b_t(\alpha''))$, i.e., only the third quantity α'' is affected. Therefore,

$$m(\alpha,\alpha';\alpha'') = m(\alpha,\alpha';b_t(\alpha'')) \quad \text{for all } b_t \in A(G)^{\alpha\alpha'}:A(G)^{\alpha\alpha'\alpha''}. \tag{3.9}$$

Since the subgroup relation

$$A(G)^{\alpha\alpha'\alpha''} \subseteq A(G)^{\alpha''} \tag{3.10}$$

is always valid, it is useful to consider the case where $A(G)^{\alpha\alpha'\alpha''}$ is a proper subgroup of $A(G)^{\alpha''}$. Each left coset representative

$$c_t \in A(G)^{\alpha''}:A(G)^{\alpha\alpha'\alpha''} \tag{3.11}$$

maps $(\alpha,\alpha',\alpha'')$ on a different triplet $(c_t(\alpha),c_t(\alpha'),\alpha'')$, and therefore we have

$$m(\alpha,\alpha';\alpha'') = m(c_t(\alpha),c_t(\alpha');\alpha'') \quad \text{for all } c_t \in A(G)^{\alpha''}:A(G)^{\alpha\alpha'\alpha''}. \tag{3.12}$$

For convenience, let us give now a more precise definition of "symmetry" and "generating" relations for CG vectors of G. For this purpose we recall that CG vectors $\mathbf{C}_j^{\alpha\alpha';\alpha''w}$ are labeled by a triplet $(\alpha,\alpha',\alpha'')$, the multiplicity index w $[w = 1,2,\ldots,m(\alpha,\alpha';\alpha'')]$ and the row index j $(j = 1,2,\ldots,n_{\alpha''})$ of the unirrep $D^{\alpha''}$. By definition we call each correlation between CG vectors of G, which belong to a given triplet $(\alpha,\alpha',\alpha'')$ a "symmetry" relation, if only their multiplicity index and/or their row index are affected. Accordingly, each element of the triple intersection group $A(G)^{\alpha\alpha'\alpha''}$ gives rise to a symmetry relation for corresponding CG vectors of G, since each $a \in A(G)^{\alpha\alpha'\alpha''}$ leaves $(\alpha,\alpha',\alpha'')$ unaltered. In contrast to symmetry relations, we call a correlation between CG vectors of G a "generating" relation, if the correlated CG vectors belong to two *different* triplets $(\alpha,\alpha',\alpha'')$ and $(\gamma,\gamma',\gamma'')$, respectively. Accordingly, each *nontrivial* left coset representative $a_t \in A(G):A(G)^{\alpha\alpha'\alpha''}$ or $b_t \in A(G)^{\alpha\alpha'}:A(G)^{\alpha\alpha'\alpha''}$ or $c_t \in A(G)^{\alpha''}:A(G)^{\alpha\alpha'\alpha''}$ must generate a generating relation for the corresponding CG vectors of G. Hence in contrast to the preceding case at least one alpha in $(\alpha,\alpha',\alpha'')$ must be affected by a nontrivial left coset representative a_t, b_t, or c_t. In other words, nontrivial left coset representatives correlate CG vectors which belong to *different* triplets, namely such ones which belong to $(\alpha,\alpha',\alpha'')$ with such ones which belong to $(a_t(\alpha),a_t(\alpha'),a_t(\alpha''))$, or $(\alpha,\alpha',b_t(\alpha''))$ or $(c_t(\alpha),c_t(\alpha'),\alpha'')$.

4. SYMMETRY RELATIONS FOR CG VECTORS OF G

Starting from the basic equations (2.1) for the definition of CG vectors of G, the crucial point is that

$$D^{\alpha\alpha'}(a(g))\mathbf{C}_j^{\alpha\alpha';\alpha''w} = \sum_k D_{kj}^{\alpha''}(a(g))\mathbf{C}_k^{\alpha\alpha';\alpha''w} \quad \text{for all } a \in A(G) \tag{4.1}$$

must remain valid in any case. Inserting (1.5) for the respective unirreps, Eqs. (4.1) become

$$Z^{\alpha\alpha'}(a)D^{a(\alpha),a(\alpha')}(g)Z^{\alpha\alpha'}(a)^{+}\mathbf{C}_j^{\alpha\alpha';\alpha''w} = \sum_k (Z^{\alpha''}(a)D^{a(\alpha'')}(g)Z^{\alpha''}(a)^{+})_{kj}\mathbf{C}_k^{\alpha\alpha';\alpha''w}, \tag{4.2}$$

where the notation $Z^{\alpha\alpha'}(a) = Z^{\alpha}(a) \otimes Z^{\alpha'}(a), a \in A(G)$, is used. A simple manipulation yields

$$D^{a(\alpha),a(\alpha')}(a)\sum_i Z_{ij}^{\alpha''}(a)Z^{\alpha\alpha'}(a)^{+}\mathbf{C}_i^{\alpha\alpha';\alpha''w} = \sum_k D_{kj}^{a(\alpha'')}(g)\sum_i Z_{ik}^{\alpha''}(a)Z^{\alpha\alpha'}(a)^{+}\mathbf{C}_i^{\alpha\alpha';\alpha''w}. \tag{4.3}$$

Finally Schur's lemma with respect to G requires

$$\sum_i Z_{ij}^{\alpha''}(a)Z^{\alpha\alpha'}(a)^{+}\mathbf{C}_i^{\alpha\alpha';\alpha''w} = \sum_v B_{vw}(a)^{+}\mathbf{C}_j^{a(\alpha),a(\alpha');a(\alpha''),v} \text{ for all } a \in A(G), \tag{4.4}$$

where CG vectors of G, which belong to the triplet $(a(\alpha), a(\alpha'),a(\alpha''))$ are denoted correspondingly and where, for the moment, $B(a)$, $a \in A(G)$, represent arbitrary $m(\alpha,\alpha';\alpha'')$-dimensional unitary matrices.

Since we are interested in "symmetry" relations for CG vectors of G, we restrict for a given triplet $(\alpha,\alpha',\alpha'')$ the group elements of $A(G)$ to those of the corresponding triple intersection group $A(G)^{\alpha\alpha'\alpha''}$. Provided that this is done, Eqs. (4.4) become by definition symmetry relations

$$Z^{\alpha\alpha'}(a)\mathbf{C}_j^{\alpha\alpha';\alpha''w} = \sum_v B_{vw}(a)\sum_k Z_{kj}^{\alpha''}(a)\mathbf{C}_k^{\alpha\alpha';\alpha''v} \quad \text{for all } a \in A(G)^{\alpha\alpha'\alpha''} \tag{4.5}$$

since $a(\alpha) = \alpha, a(\alpha') = \alpha'$, and $a(\alpha'') = \alpha''$ must be valid for all $a \in A(G)^{\alpha\alpha'\alpha''}$. Assuming not only that the set $\{\mathbf{C}_j^{\alpha\alpha';\alpha''w}: w = 1,2,\ldots,m(\alpha,\alpha';\alpha'');j = 1,2,\ldots,n_{\alpha''}\}$ of CG vectors of G are known, but also that the unitary matrices $Z^{\alpha}(a)$, $Z^{\alpha'}(a)$, and $Z^{\alpha''}(a)$, $a \in A(G)^{\alpha\alpha'\alpha''}$, are given, the $m(\alpha,\alpha';\alpha'')$-dimensional unitary matrices $B(a)$, $a \in A(G)^{\alpha\alpha'\alpha''}$, are then *uniquely* determined. Therefore, CG vectors of G likewise transform according to the direct matrix product representation $B(a) \otimes Z^{\alpha''}(a)$, $a \in A(G)^{\alpha\alpha'\alpha''}$. As $Z^{\alpha\alpha'}(a)$, $a \in A(G)^{\alpha\alpha'\alpha''}$, can be interpreted as the Kronecker product of the projective unirreps $Z^{\alpha}(a)$ and $Z^{\alpha'}(a)$, $a \in A(G)^{\alpha\alpha'\alpha''}$, the matrices $B(a)$, $a \in A(G)^{\alpha\alpha'\alpha''}$, likewise constitute a projective

representation of $A(G)^{\alpha\alpha'\alpha''}$. A simple calculation yields

$$B(a)B(a') = P(a,a')B(aa') \quad \text{for all } a,a' \in A(G)^{\alpha\alpha'\alpha''}, \tag{4.6}$$

$$P(a,a') = Q^{\alpha}(a,a')Q^{\alpha'}(a,a')Q^{\alpha''}(a,a')^* \quad \text{for all } a,a' \in A(G)^{\alpha\alpha'\alpha''}, \tag{4.7}$$

where the standard factor system P is composed of the factor systems Q^{α}, $Q^{\alpha'}$, and $Q^{\alpha''}$. On account of

$$B(a_{G_j}) = \mathbf{1}_m \quad \text{for all } a_{G_j} \in A(G)_i \tag{4.8}$$

it follows immediately that

$$P(a_{G_j}, a_{G_k}) = 1 \quad \text{for all } a_{G_j}, a_{G_k} \in A(G)_i \tag{4.9}$$

when (1.13) and (1.14) are used. Since $A(G)_i$ is a normal subgroup of $A(G)^{\alpha\alpha'\alpha''}$, it suggests that $B(a)$, $a \in A(G)^{\alpha\alpha'\alpha''}$, forms a projective representation of the factor group $A(G)^{\alpha\alpha'\alpha''}/A(G)_i$. This requires that the factor system P is constant on left cosets of $A(G)^{\alpha\alpha'\alpha''}$ with respect to $A(G)_i$. The proof of this conjecture is simple and therefore omitted.

Utilizing the nonuniqueness of CG vectors, Eqs. (4.5) may be transformed into

$$Z^{\alpha\alpha'}(a)\mathbf{C}^{\alpha\alpha';\alpha''w}(\beta) = \sum_v (\beta^+ B(a)\beta)_{vw} \sum_k Z^{\alpha''}_{kj}(a)\mathbf{C}_k^{\alpha\alpha';\alpha''v}(\beta), \tag{4.10}$$

where the new CG vectors $\mathbf{C}^{\alpha\alpha';\alpha''w}(\beta)$, $\beta \in U(m)$ [= unitary group in $m(\alpha,\alpha';\alpha'')$ dimensions], are defined by

$$\mathbf{C}_j^{\alpha\alpha';\alpha''w}(\beta) = \sum_v \beta_{vw}\mathbf{C}_j^{\alpha\alpha';\alpha''v}. \tag{4.11}$$

Hence for suitably chosen $\beta = \beta_0$ the projective representation $B(a)$, $a \in A(G)^{\alpha\alpha'\alpha''}$ decomposes into a direct sum of its irreducible constituents. Therefore, also $(\beta_0^+ B(a)\beta_0) \otimes Z^{\alpha''}(a)$, $a \in A(G)^{\alpha\alpha'\alpha''}$, decomposes into a direct sum. Clearly, if $B(a)$, $a \in A(G)^{\alpha\alpha'\alpha''}$, contains a one-dimensional projective unirrep $\mathbb{D}^{\gamma}$ of $A(G)^{\alpha\alpha'\alpha''}$, the corresponding representation $\mathbb{D}^{\gamma}(a) \otimes Z^{\alpha''}(a)$, $a \in A(G)^{\alpha\alpha'\alpha''}$, forms a projective unirrep of the triple intersection group. Observe that this representation belongs to the factor system $Q^{\alpha}(a,a')Q^{\alpha'}(a,a')$, $a,a' \in A(G)^{\alpha\alpha'\alpha''}$, because the third factor $Q^{\alpha''}(a,a')$ cancels. Hence in this case the vectors $\mathbf{C}_j^{\alpha\alpha';\alpha''w}(\beta_0)$ are CG vectors of G and $A(G)^{\alpha\alpha'\alpha''}$ simultaneously.

Another interesting feature of $B(a)$, $a \in A(G)^{\alpha\alpha'\alpha''}$, may be derived when complex conjugation is taken into account. Provided that $\alpha,\alpha',\alpha'' \in A_{\mathrm{I}(*)} \cup A_{\mathrm{II}(*)}$ is satisfied, we derived in Ref. 9 the following identities:

$$\mathbf{C}_j^{\alpha\alpha';\alpha''w} = \sum_v f_{vw} U^{\alpha\alpha'} \sum_k (U_{jk}^{\alpha''}\mathbf{C}_k^{\alpha\alpha';\alpha''v})^*, \tag{4.12}$$

where f is a uniquely fixed $m(\alpha,\alpha';\alpha'')$-dimensional unitary matrix. f is either symmetric or skew-symmetric in accordance with

$$ff^* = (-1)^{c(\alpha)+c(\alpha')+c(\alpha'')}\mathbb{1}_m, \qquad c(\gamma) = \begin{cases} 0, & \gamma \in A_{\mathrm{I}(*)}, \\ 1, & \gamma \in A_{\mathrm{II}(*)}. \end{cases} \tag{4.13}$$

A simple manipulation of (4.5) yields

$$B(a)^* = fB(a)f^+ \quad \text{for all } a \in A(G)^{\alpha\alpha'\alpha''}, \tag{4.14}$$

where (4.12) is used. Therefrom it follows that the factor system (4.7) must be *real*. Combining (4.14) with such a unitary matrix β_0 that decomposes $B(a)$, $a \in A(G)^{\alpha\alpha'\alpha''}$, into a direct sum of its irreducible consitituents, Eqs. (4.14) transform into

$$B'(a)^* = f_0 B'(a) f_0^+, \quad B'(a) = \beta_0^+ B(a)\beta_0, \quad f_0 = \beta_0^t f \beta_0. \tag{4.15}$$

Note that this is in accordance with (IV.16) of Ref. 9. Therefore, we obtain as a by-product that only for $(-1)^{c(\alpha)+c(\alpha')+c(\alpha'')} = +1$ one-dimensional unirreps $B(a)$, $a \in A(G)^{\alpha\alpha'\alpha''}$, are possible. However, such ones are excluded, if the phase factor is -1 since type II(*) unirreps must be even-dimensional.

Finally let us turn to the second case, where at least one of the alphas of a given triplet $(\alpha,\alpha',\alpha'')$ belongs to $A_{\mathrm{III}(*)}$. It has been shown in Ref. 9 that complex conjugation correlates CG vectors of G that belong to the triplets $(\alpha,\alpha',\alpha'')$ and $(\alpha^*,\alpha^{*\prime},\alpha^{*\prime\prime})$. On account of (1.27) it follows that

$$A(G)^{\alpha^*\alpha^{*\prime}\alpha^{*\prime\prime}} = A(G)^{\alpha\alpha'\alpha''}. \tag{4.16}$$

Moreover, we may derive

$$Z^{\alpha^*\alpha^{*\prime}}(a)\mathbf{C}_j^{\alpha^*\alpha^{*\prime};\alpha^{*\prime\prime}w} = \sum_v E_{vw}(a) \sum_k Z^{\alpha^{*\prime\prime}}_{kj}(a)\mathbf{C}_k^{\alpha^*\alpha^{*\prime};\alpha^{*\prime\prime}v} \quad \text{for all } a \in A(G)^{\alpha\alpha'\alpha''}, \tag{4.17}$$

where $E(a)$, $a \in A(G)^{\alpha\alpha'\alpha''}$, are well defined $m(\alpha,\alpha';\alpha'')$-dimensional unitary matrices. Now it follows immediately by means of (II.11) of Ref. 9 and (1.28) [or (1.25) if necessary] that we can choose $E(a) = B(a)^*$ for all $a \in A(G)^{\alpha\alpha'\alpha''}$ without any loss of generality. Note that for such cases we can define

$$\mathbf{C}_j^{\alpha^*\alpha^{*\prime};\alpha^{*\prime\prime}w} = (U^{\alpha\alpha'} \sum_k U_{jk}^{\alpha''}\mathbf{C}_k^{\alpha\alpha';\alpha''w})^*, \tag{4.18}$$

as has been extensively discussed in Refs. 7 and 9 and in many other places. According to our definitions, Eqs. (4.12) are symmetry and (4.18) generating relations for CG vectors of G that are generated by complex conjugation.

5. GENERATING RELATIONS FOR CG VECTORS OF G

Let us start from Eqs. (4.4), which can be rewritten as follows:

$$\mathbf{C}_j^{a(\alpha),a(\alpha');a(\alpha''),v} = \sum_w B_{wv}(a) \sum_k Z^{\alpha''}_{kj}(a) \times Z^{\alpha\alpha'}(a)^+ \mathbf{C}_k^{\alpha\alpha';\alpha''w}. \tag{5.1}$$

In order to avoid conflicting relations and redundancies between CG vectors of G, we restrict the elements of $A(G)$ to left coset representatives $a_t \in A(G):A(G)^{\alpha\alpha'\alpha''}$. When now, for convenience, we choose $B(a_t) = \mathbb{1}_m$ for all elements a_t of the set $A(G):A(G)^{\alpha\alpha'\alpha''}$, it does not imply a loss of generality. Certainly, we can do this, if only the CG vectors, which occur on the right-hand side of (5.1) are given. By definition we call

$$\mathbf{C}_j^{a_t(\alpha),a_t(\alpha');a_t(\alpha''),v} = \sum_k Z^{\alpha''}_{kj}(a_t) Z^{\alpha\alpha'}(a_t)^+ \mathbf{C}_k^{\alpha\alpha';\alpha''v} \quad \text{for all } a_t \in A(G):A(G)^{\alpha\alpha'\alpha''}, \tag{5.2}$$

"generating" relations for CG vectors of G. The reason why we can choose $B(a_t) = 1_m$ for all $a_t \in A(G):A(G)^{\alpha\alpha'\alpha''}$ is that

$$(a_t(\alpha),a_t(\alpha'),a_t(\alpha'')) \neq (a_s(\alpha),a_s(\alpha'),a_s(\alpha'')) \Leftrightarrow a_t \neq a_s. \quad (5.3)$$

When CG vectors for a given triplet $(\alpha,\alpha',\alpha'')$ are known, CG vectors that belong to $(a_t(\alpha), a_t(\alpha'),a_t(\alpha''))$ immediately can be calculated by means of (5.2). This assumes that the matrices $Z^{\alpha}(a_t)$, $Z^{\alpha'}(a_t)$, and $Z^{\alpha''}(a_t)$ are known. On the other hand, when CG vectors that belong to $(\alpha,\alpha',\alpha'')$ and $(a_t(\alpha),a_t(\alpha'),a_t(\alpha''))$ have been determined independently the unitary matrix $B(a_t)$ is uniquely fixed. Observe that then $B(a_t)$ not necessarily is the $m(\alpha,\alpha';\alpha'')$-dimensional unit matrix. Irrespective of this problem, the generating relations (5.2) constitute a useful and practical tool for the calculation of CG vectors. For, if CG vectors, which belong to a fixed triplet $(\alpha,\alpha',\alpha'')$ are known [i.e., $\{C_j^{\alpha\alpha';\alpha''w}: w = 1,2,\ldots,m(\alpha,\alpha',\alpha'');\ j = 1,2,\ldots,n_{\alpha''}\}$ are taken for granted], corresponding CG vectors $\{C_j^{a_t(\alpha),a_t(\alpha');a_t(\alpha''),w}: a_t \in A(G):A(G)^{\alpha\alpha'\alpha''};\ w = 1,2,\ldots,m(\alpha,\alpha';\alpha'');\ j = 1,2,\ldots,n_{\alpha''}\}$ are obtained from (5.2) by varying a_t over the whole set $A(G):A(G)^{\alpha\alpha'\alpha''}$. Certainly this approach is only then practical if for a given group G the groups $A(G)$, $A(G)^{\alpha\alpha'\alpha''}$ and the matrices $Z^{\alpha}(a_t), Z^{\alpha'}(a_t), Z^{\alpha''}(a_t)$, $a_t \in A(G):A(G)^{\alpha\alpha'\alpha''}$, can be determined easily. Nevertheless, the generating relations (5.2) are [together with the symmetry relations (4.5)] important, since they reveal additional symmetries of CG coefficients. These relations can be used to calculate CG vectors out of known ones and to save space when CG coefficients are tabulated. Finally they can be utilized for additional crosschecks, provided that CG vectors have been calculated independently. Assuming now that $A(G)$, $A(G)^{\alpha\alpha'\alpha''}$ and the respective matrices $Z^{\gamma}(a_t)$, $a_t \in A(G)^{\alpha\alpha'\alpha''}$, with $\gamma = \alpha,\alpha',\alpha''$ are explicitly known, the computation of CG vectors by means of (5.2) works as follows: At first compute for a given triplet $(\alpha,\alpha',\alpha'')$ CG vectors and determine then by means of (5.2) for each element $(a_t(\alpha),a_t(\alpha'),a_t(\alpha''))$, $a_t \in A(G):A(G)^{\alpha\alpha'\alpha''}$, of the triplet orbit $\Delta(\alpha\alpha'\alpha'')$ the CG vectors. Provided that this has been done, choose a further triplet (β,β',β''), which does not belong to $\Delta(\alpha\alpha'\alpha'')$. Compute CG vectors, which refer to (β,β',β'') as usual, and proceed then as before by calculating for each element $(a_s(\beta),a_s(\beta'),a_s(\beta''))$, $a_s \in A(G):A(G)^{\beta\beta'\beta''}$, the corresponding CG vectors. Continue in the same way until each element of the product set $A_G \times A_G \times A_G$ is taken into account.

Let us now discuss briefly generating relations which refer to (3.9) and (3.12), respectively. By virtue of the definitions (3.8), Eq. (5.1) become

$$C_j^{\alpha\alpha';b_t(\alpha''),v} = \sum_k Z_{kj}^{\alpha''}(b_t) Z^{\alpha\alpha'}(b_t)^{+} C_k^{\alpha\alpha';\alpha''v}$$

$$\text{for all } b_t \in A(G)^{\alpha\alpha'}:A(G)^{\alpha\alpha'\alpha''}, \quad (5.4)$$

where only the CG vectors, which occur on the right-hand side are assumed to be known. The remaining CG vectors $\{C_j^{\alpha\alpha';b_t(\alpha'')v}: b_t \in A(G)^{\alpha\alpha'}:A(G)^{\alpha\alpha'\alpha''};\ v = 1,2,\ldots,m(\alpha,\alpha';\alpha'');\ j = 1,2,\ldots,n_{\alpha''}\}$ are defined by (5.4), since otherwise $B(b_t)$ cannot be chosen as the unit matrix 1_m. Hence nontrivial left coset representatives $b_t \in A(G)^{\alpha\alpha'}:A(G)^{\alpha\alpha'\alpha''}$ correlate CG vectors of a given Kronecker product $D^{\alpha\alpha'}$, which belong to inequivalent unirreps $D^{b_t(\alpha'')}$.

For the other case, which is characterized by (3.12), Eqs.(5.1) become

$$C_j^{c_t(\alpha),c_t(\alpha');\alpha'',v} = \sum_k Z_{kj}^{\alpha''}(c_t) Z^{\alpha\alpha'}(c_t)^{+} C_k^{\alpha\alpha';\alpha''v}$$

$$\text{for all } c_t \in A(G)^{\alpha''}:A(G)^{\alpha\alpha'\alpha''}, \quad (5.5)$$

where we have chosen once again $B(c_t) = 1_m$, $c_t \in A(G)^{\alpha''}:A(G)^{\alpha\alpha'\alpha''}$. This choice is allowed, if the CG vectors on the right-hand side of (5.5) are taken for granted and the remaining CG vectors are defined by (5.5). Accordingly, Eqs. (5.5) correlate CG vectors, which belong to inequivalent Kronecker products $D^{\alpha\alpha'}$ and $D^{c_t(\alpha),c_t(\alpha')}$, but transform according to the same unirrep, namely $D^{\alpha''}$.

Finally, one has to note that the three different sets of generating relations, which are defined by (5.2), (5.4), and (5.5), respectively, cannot be used in general simultaneously. The reason for it is that the special choices $B(a_t) = 1_m$, $a_t \in A(G):A(G)^{\alpha\alpha'\alpha''}$ and $B(b_t) = 1_m$, $b_t \in A(G)^{\alpha\alpha'}:A(G)^{\alpha\alpha'\alpha''}$ and $B(c_t) = 1_m$, $c_t \in A(G)^{\alpha''}:A(G)^{\alpha\alpha'\alpha''}$ would lead in general to conflicting definitions of the corresponding CG vectors.

6. SPECIAL SYMMETRIES FOR CG VECTORS OF G

The results of the preceding sections suggest the question: Are the special symmetries of $3jm$ factors of Ref. 8 involved in our automorphism approach? In order to get a better grip of these symmetries, let us start from equivalence relations of the type

$$D^{\gamma}(g) D^{\alpha}(g) = W^{\gamma}(\alpha) D^{\gamma(\alpha)}(g) W^{\gamma}(\alpha)^{+} \quad \text{for all } g \in G, \quad (6.1)$$

where γ labels a nontrivial one-dimensional unirrep of G and $W^{\gamma}(\alpha)$ is an n_{α}-dimensional unitary matrix. $W^{\gamma}(\alpha)$ is uniquely defined (up to a phase factor) by $D^{\gamma}(g), D^{\alpha}(g)$, and $D^{\gamma(\alpha)}(g)$, $g \in G$. Obviously two cases may occur, namely either $\alpha = \gamma(\alpha) \in A_G$, or $\alpha \neq \gamma(\alpha) \in A_G$. Since

$$D^{\gamma}(g) D^{\gamma}(g) = D^{\gamma''}(g) \quad \text{for all } g \in G \quad (6.2)$$

always holds for one-dimensional unirreps D^{γ}, $D^{\gamma'}$, and $D^{\gamma''}$ of G, where $D^{\gamma''}$ is uniquely determined by D^{γ} and $D^{\gamma'}$, we extend an arbitrary Kronecker product $D^{\alpha\alpha'}(g)$, $g \in G$ by $D^{\gamma''}(g)$, $g \in G$:

$$D^{\gamma''}(g) D^{\alpha\alpha'}(g) = D^{\gamma}(g) D^{\alpha}(g) \otimes D^{\gamma'}(g) D^{\alpha'}(g)$$

$$= \sum_{\alpha''} \oplus\, m(\alpha,\alpha';\alpha'') D^{\gamma''}(g) D^{\alpha''}(g). \quad (6.3)$$

Provided that $D^{\gamma''}$ is the trivial unirrep $D^{1}(g) = 1, g \in G$, of G, then $D^{\gamma'}(g) = D^{\gamma}(g)^{*}$, $g \in G$. Irrespective of this special case, four different situations may occur on the left-hand side of (6.3), namely,

$$D^{\gamma(\alpha)}(g) \sim D^{\alpha}(g) \quad \text{and} \quad D^{\gamma'(\alpha')}(g) \sim D^{\alpha'}(g), \quad (6.4)$$

$$D^{\gamma(\alpha)}(g) \sim D^{\alpha}(g) \quad \text{and} \quad D^{\gamma'(\alpha')}(g) \not\sim D^{\alpha'}(g), \quad (6.5)$$

$$D^{\gamma(\alpha)}(g) \not\sim D^{\alpha}(g) \quad \text{and} \quad D^{\gamma'(\alpha')}(g) \sim D^{\alpha'}(g), \quad (6.6)$$

$$D^{\gamma(\alpha)}(g) \not\sim D^{\alpha}(g) \quad \text{and} \quad D^{\gamma'(\alpha')}(g) \not\sim D^{\alpha'}(g), \quad (6.7)$$

whereas on the right-hand side either $D^{\gamma''(\alpha'')}(g) \sim D^{\alpha''}(g)$ or

$D^{\gamma''(\alpha'')}(g) \downarrow D^{\alpha''}(g)$, $g \in G$, is realized. By using Schur's lemma with respect to G two different symmetries for CG vectors of G can be derived, which correspond to the special symmetries of $3jm$ factors discussed in Ref. 8:

$$W^{\gamma\gamma'}(\alpha\alpha')\mathbf{C}_j^{\alpha\alpha';\alpha''w} = \sum_v e_{vw} \sum_k W_{kj}^{\gamma''}(\alpha'')\mathbf{C}_k^{\alpha\alpha';\alpha''v}, \tag{6.8}$$

$$\mathbf{C}_j^{\gamma(\alpha),\gamma'(\alpha');\gamma''(\alpha''),w} = \sum_v h_{vw} \sum_k W_{kj}^{\gamma''}(\alpha'') W^{\gamma\gamma'}(\alpha\alpha')^{+} \mathbf{C}_k^{\alpha\alpha';\alpha''v}. \tag{6.9}$$

Note that $W^{\gamma\gamma'}(\alpha\alpha') = W^{\gamma}(\alpha) \otimes W^{\gamma'}(\alpha')$ and that CG vectors of G are denoted correspondingly. The $m(\alpha,\alpha';\alpha'')$-dimensional unitary matrix e is unique [for given $(\alpha,\alpha',\alpha'')$] if $\gamma(\alpha) = \alpha$, $\gamma'(\alpha') = \alpha'$, and $v''(\alpha'') = \alpha''$, whereas the matrix h can be chosen arbitrary, if at least one $\gamma(\alpha)$ is different from α and if only the CG vectors on one side of (6.9) are given. Finally observe that in accordance with our definitions Eqs. (6.8) are symmetry and Eqs. (6.9) generating relations for CG vectors of G.

The earlier raised question may now be asked as to whether the symmetries (6.8) and (6.9) for CG vectors of G can be explained by a particular $a \in A(G)^{\alpha\alpha'\alpha''}$ or $a_t \in A(G) : A(G)^{\alpha\alpha'\alpha''}$. We show the answer is *no* after arguing generally and after inspecting some striking examples at the end of Sec. 7. One general argument for this conjecture is that the trivial unirrep D^1 of G is mapped by D^{γ} on D^{γ}, whereas each automorphism can only map D^1 on D^1. A second argument is that the number of nontrivial one-dimensional unirreps of a given group G is in general different to the number of outer automorphisms. We conclude that (6.3) generates symmetries that are distinct from the automorphism symmetries.

7. APPLICATION TO SPACE GROUPS

In order to show the utility of the automorphism approach, we apply it to space group CG vectors in which we are primarily interested. Hereafter we denote a space group by G, its normal subgroup of translations by T, and its point group by P, which is isomorphic to the factor group G/T. It is well known[10,11] that the automorphism group $A(G)$ of a given space group G is isomorphic to the factor group $N(G)/C(G)$ of the affine normalizer $N(G)$ with respect to the affine centralizer $C(G)$. In order to simplify the following discussion, we assume in addition that $Z(G)$ and $C(G)$ are trivial. This means that $A(G) = N(G)$ and that G is a normal subgroup of $A(G)$. Fortunately, this situation is realized for many space groups, as can be seen from Table 3 of Ref. 11. Accordingly, for such cases we are not forced to embed G in the corresponding holomorph. But note, on the other hand, that $N(G) = A(G)$ need not be a semidirect product of G with a group that is isomorphic to the factor group $A(G)/G$.

By virtue of our assumptions, each automorphism $a \in A(G)$ can be identified with a particular element $(S_a|\mathbf{v}_a)$ of the affine group $A(\mathbb{R}^3)$ of $\mathbb{R}^3$. Adopting the notation $g = (R|\mathbf{n}(R) + \mathbf{t})$ for space group elements, where R denotes a point group element, $\mathbf{t}$ a primitive translation, and $\mathbf{n}(R)$ a nonprimitive translation (characterizing screw axes or glide planes), we have

$$\begin{aligned} a(g) &= a[(R|\mathbf{n}(R) + \mathbf{t})] = (S_a|\mathbf{v}_a)(R|\mathbf{n}(R) + \mathbf{t})(S_a|v_a)^{-1} \\ &= (S_a R S_a^{-1}|S_a[\mathbf{n}(R) + \mathbf{t}] + \mathbf{v}_a - S_a R S_a^{-1}\mathbf{v}_a). \end{aligned} \tag{7.1}$$

The presence of $\mathbf{n}(R)$ is typical for nonsymmorphic space groups, whereas they can be chosen as zero vectors for symmorphic space groups. Since each mapping $a = (S_a|\mathbf{v}_a)$ of G onto G requires that $a(g) = (S_a|\mathbf{v}_a)(R|\mathbf{n}(R) + \mathbf{t})(S_a|\mathbf{v}_a)^{-1} \in G$, we infer that $S_a R S_a^{-1}$ must belong to the point group P. Moreover, $S_a\mathbf{t}$ and $\mathbf{v}_a - S_a R S_a^{-1}\mathbf{v}_a$ must belong to the translation group T and $S_a\mathbf{n}(R)$ must be equal to $\mathbf{n}(S_a R S_a^{-1})$ modulus a primitive lattice translation. Note in particular that the nonsingular matrices S_a and the vectors $\mathbf{v}_a$ must not depend on the respective space group elements. In many cases the $A(G)$ are isomorphic to certain group of motions, whereas for some space groups their automorphism groups cannot be identified with groups of motions.[10]

In order to be able to go into further details, we recall that space group unirreps $D^K(g), g \in G$, are labeled by pairs $K = (\mathbf{q},\alpha)$. Thereby the vector $\mathbf{q}$ is an element of the "representation domain" Δ BZ of the Brillouin zone BZ of G and α labels unirreps of the corresponding "little cogroup" $P(\mathbf{q})$. The unirreps of $P(\mathbf{q})$ are projective if G is nonsymmorphic.[2] By definition $P(\mathbf{q})$ consists of all elements of P that leave $\mathbf{q}$ invariant (up to reciprocal lattice vectors).[2] In order to be able to apply the automorphism approach, we need for each $a \in A(G)$ not only the mapping $a: G \to G$, but also the mapping $a: A_G \to A_G$ and the matrices $Z^K(a)$, which satisfy

$$\begin{aligned} D^K(a(g)) &= D^K((S_a|\mathbf{v}_a)(R|\mathbf{n}(R) + \mathbf{t})(S_a|\mathbf{v}_a)^{-1}) \\ &= Z^K(S_a|\mathbf{v}_a) D^{a(K)}(R|\mathbf{n}(R) + \mathbf{t}) Z^K(S_a|\mathbf{v}_a)^{+}. \end{aligned} \tag{7.2}$$

Finally we need the corresponding little groups $A(G)^K$ which contain G as a normal subgroup, i.e., $G \lhd A(G)^K \subseteq A(G)$. Provided that these tasks have been solved, one is in the position to apply the automorphism approach to space group CG vectors.

Example: In the following we discuss briefly an example which demonstrates among others that the special symmetries of $3jm$ factors and automorphism properties of CG coefficients are different in general. A detailed discussion of this and therewith related subjects will be published elsewhere.[12]

We consider the primitive cubic space group $G = P23$, whose automorphism group is given by $A(G) = Im3m$ [see Table 3 of Ref. 11], which contains G as normal subgroup, since $Z(G)$ is trivial. Both space groups are symmorphic and therefore can be written as a semidirect product of their translation and point groups. In our case we have

$$G = T_{\text{pc}} \circledS \mathcal{T}, \tag{7.3}$$

$$A(G) = T_{\text{bcc}} \circledS \mathcal{O}_h, \tag{7.4}$$

where T_{pc} denotes the primitive cubic translation group, T_{bcc} the body-centered cubic translation group, $\mathcal{T}$ the tetrahedral group, and $\mathcal{O}_h$ the octahedral group including the inversion. By virtue of the left coset decompositions

$$T_{\text{bcc}} = T_{\text{pc}} + (\mathbf{E}|\mathbf{b}_0)T_{\text{pc}}, \tag{7.5}$$

$$\mathcal{O}_h = \mathcal{T} + C_{2a}\mathcal{T} + \sigma_z\mathcal{T} + \sigma_{db}\mathcal{T}, \tag{7.6}$$

we arrive at

$$A(G) = G + (C_{2a}|\mathbf{0})G + (\sigma_z|\mathbf{0})G + (\sigma_{db}|\mathbf{0})G + (E|\mathbf{b}_0)G + (C_{2a}|\mathbf{b}_0)G + (\sigma_z|\mathbf{b}_0)G + (\sigma_{db}|\mathbf{b}_0)G, \tag{7.7}$$

where $\mathbf{b}_0$ is a special element of $T_{\rm bcc}$ and $C_{2a},\sigma_z,\sigma_{db}$ are special elements of $\mathcal{O}_h$. By definition, G must be a normal subgroup of $A(G)$, since $Z(G)$ is trivial. But note that the set

$$A(G):G = \{(E|\mathbf{0}),(C_{2a}|\mathbf{0}),(\sigma_z|\mathbf{0}),(\sigma_{db}|\mathbf{0}), (E|\mathbf{b}_0),(C_{2a}|\mathbf{b}_0),(\sigma_z|\mathbf{b}_0),(\sigma_{db}|\mathbf{b}_0)\} \tag{7.8}$$

of left coset representatives does not form a group, i.e., $A(G)$ cannot be written as a semidirect product of G with a group that is isomorphic to the factor group $A(G)/G$. Besides this, the index of G with respect to $A(G)$ is eight, i.e., seven nontrivial outer automorphisms (mod G) exist. The representation domain Δ BZ(G) of $G = P23$ consists roughly speaking of all elements of the union set[13] $\Omega\cup\sigma_z\Omega\cup\sigma_{db}\Omega\cup C_{2a}\Omega$, when Ω denotes the representation domain of the simple cubic space group $\mathcal{O}_h^1 = Pm3m$. A "general" point $\mathbf{q}$ of Δ BZ(G) is defined by $P(\mathbf{q}) = \{E\}$, when $P(\mathbf{q})$ denotes the corresponding little cogroup. For each such case the corresponding space group unirrep D^K must be 12-dimensional, since $|G|/|T_{\rm pc}| = 12$. Their matrix elements take the form[2]

$$D^{(\mathbf{q},1)}_{R',R''}(R|\mathbf{t}) = \delta_{R',RR''}\,e^{-iR'\mathbf{q}\cdot\mathbf{t}}, \quad R',R''\in\mathcal{T}, R\in\mathcal{T}, \mathbf{t}\in T_{\rm pc}, \tag{7.9}$$

where the superscript 1 denotes the trivial unirrep of $P(\mathbf{q})$ and $\delta_{R',RR''}$ the usual Kronecker delta. Specializing (7.1) to the left coset representatives $(S|\mathbf{0}),(S|\mathbf{b}_0)$ with $S\in\{E,C_{2a},\sigma_z,\sigma_{db}\}$, we obtain

$$(S|\mathbf{0})(R|\mathbf{t})(S|\mathbf{0})^{-1} = (SRS^{-1}|S\mathbf{t}), \tag{7.10}$$

$$(E|\mathbf{b}_0)(R|\mathbf{t})(E|\mathbf{b}_0)^{-1} = (R|\mathbf{t}+\mathbf{b}_0-R\mathbf{b}_0), \tag{7.11}$$

where the remaining conjugations are readily calculated by means of $(E|\mathbf{b}_0)(S|\mathbf{0}) = (S|\mathbf{b}_0)$. Inserting (7.10) and (7.11) into (7.9), we arrive immediately at the result that $(E|\mathbf{b}_0)$ maps $(\mathbf{q},1)$ onto $(\mathbf{q},1)$, whereas $(S|\mathbf{0})$ maps $(\mathbf{q},1)$ onto $(S^{-1}\mathbf{q},1)$. This means that

$$a((\mathbf{q},1)) = (\mathbf{q},1) \quad \text{for } a = (E|\mathbf{b}_0), \tag{7.12}$$

$$a((\mathbf{q},1)) = (S^{-1}\mathbf{q},1) \quad \text{for } a = (S|\mathbf{0}), \tag{7.13}$$

$$A(G)^K = G + (E|\mathbf{b}_0)G = I23, \quad K = (\mathbf{q},1). \tag{7.14}$$

Accordingly if $\mathbf{q}$ is a general point of Δ BZ(G) the corresponding little group $A(G)^K$ is the body-centered cubic space group $I23$. This means that there exists a unitary matrix $Z^K(E|\mathbf{b}_0)$, which satisfies

$$D^K(a(g)) = D^K(R|\mathbf{t}+\mathbf{b}_0-R\mathbf{b}_0) = Z^K(E|\mathbf{b}_0)D^K(R|\mathbf{t})Z^K(E|\mathbf{b}_0)^+. \tag{7.15}$$

A simple manipulation yields for the matrix elements

$$Z^K_{R',R''}(E|\mathbf{b}_0) = \delta_{R',R''}\,e^{-iR'\mathbf{q}\cdot\mathbf{b}_0}. \tag{7.16}$$

$Z^K((E|\mathbf{b}_0)$ forms, together with the 12-dimensional unit matrix, a projective representation of the factor group $A(G)^K/G\sim S_2$. On the other hand, when we take $a = (S|\mathbf{0})$, Eq. (7.2) becomes

$$D^K(a(g)) = D^K(SRS^{-1}|S\mathbf{t}) = Z^K(S|\mathbf{0})D^{(S^{-1}\mathbf{q},1)}(R|\mathbf{t})Z^K(S|\mathbf{0}), \tag{7.17}$$

where the matrix elements of $Z^K(S|\mathbf{0})$ turn out to be

$$Z^K_{R',R''}(S|\mathbf{0}) = \delta_{R',SR''S^{-1}}, \quad S\in\{E,C_{2a},\sigma_z,\sigma_{db}\}. \tag{7.18}$$

These four matrices do not depend on a particular $\mathbf{q}$, provided that $\mathbf{q}$'s of the boundary of Δ BZ(G) are excluded. Because of

$$A(G) = A(G)^K\circledS\{(E|\mathbf{0}),(C_{2a}|\mathbf{0}),(\sigma_z|\mathbf{0}),(\sigma_{db}|\mathbf{0})\} \tag{7.19}$$

the four 12-dimensional matrices $Z^K(S|\mathbf{0})$, $S\in\{E,C_{2a},\sigma_z,\sigma_{db}\}$, form an ordinary representation of the group $\{E,C_{2a},\sigma_z,\sigma_{db}\}$, as can be readily verified by means of (7.18).

Since we want to compare the automorphism properties of CG coefficients with the special symmetries of corresponding $3jm$ factors, it sufficies to investigate the mappings $L:A_G\to A_G$ that are generated by nontrivial one-dimensional unirreps D^L of G. Adopting the notation D^L, $L = (\mathbf{q},\gamma)$, for one-dimensional unirreps of G, the corresponding equivalence relations become

$$D^L(R|\mathbf{t})D^K(R|\mathbf{t}) = W^L(K)D^{L(K)}(R|\mathbf{t})W^L(K)^+, \tag{7.20}$$

where $L(K)\in A_G$ is the image of $K\in A_G$. A straightforward study of Table 5.7 (on p. 374) and Table 5.1 (on p. 230) of Ref. 2 yields that $G = P23$ possesses five nontrivial one-dimensional unirreps. Two belong to the Γ-point (i.e., $\mathbf{q} = \mathbf{0}$) and three belong to the R-point [i.e., $\mathbf{q}_R = (\pi/a)(1,1,1)$, when a denotes the lattice constant]. Four of them are pairwise of type III(*), which means that they are pairwise complex conjugate. The fifth unirrep is real and belongs to $\mathbf{q}_R$ and to the trivial unirrep D^1 of $P(\mathbf{q}_R) = \mathcal{T}$. Adopting the same notation as in Ref. 2, these unirreps take the form

$$D^{(0,j)}(R|\mathbf{t}) = D^j(R), \quad j = 2,3, \tag{7.21}$$

$$D^{(\mathbf{q}_R,j)}(R|\mathbf{t}) = e^{-i\mathbf{q}_R\cdot\mathbf{t}}D^j(R), \quad j = 1,2,3, \tag{7.22}$$

where $e^{-i\mathbf{q}_R\cdot\mathbf{t}}$ is real due to the definition of $\mathbf{q}_R$ and $\mathbf{t}\in T_{\rm pc}$. Moreover, it holds that $D^2(R) = D^3(R)^*$, $R\in\mathcal{T}$. Besides this, be aware that the index R of $\mathbf{q}_R$ never should be confused with the point group elements $R\in\mathcal{T}$ and that $P(\mathbf{q}_\Gamma) = P(\mathbf{q}_R) = \mathcal{T}$. According to our intentions, L has to vary over the set $\{(\mathbf{0},2),(\mathbf{0},3),(\mathbf{q}_R,1),(\mathbf{q}_R,2),(\mathbf{q}_R,3)\}$.

We assume in the following that K belongs to a general point $\mathbf{q}$ [i.e., $K = (\mathbf{q},1)$, since $P(\mathbf{q}) = \{E\}$]. The character test shows that

$$L(K) = K \quad \text{for } L = (\mathbf{0},2),(\mathbf{0},3), \quad K = (\mathbf{q},1), \tag{7.23}$$

where the similarity matrices $W^L(K)$ are given by

$$W^{(\mathbf{0},j)}_{R',R''}(K) = \delta_{R',R''}D^j(R'), \quad j = 2,3. \tag{7.24}$$

The remaining three cases $[L\in\{(\mathbf{q}_R,1),(\mathbf{q}_R,2),(\mathbf{q}_R,3)\}]$ lead to

$$L((\mathbf{q},1)) = (S_0\mathbf{q}+\mathbf{q}_R+\mathbf{Q},1), \tag{7.25}$$

where $S_0\in\mathcal{T}$ and $\mathbf{Q}$ (= reciprocal lattice vector) are uniquely defined by the constraint that $S_0\mathbf{q}+\mathbf{q}_R+\mathbf{Q}$ must belong to Δ BZ(G). Observe that $S_0\mathbf{q}+\mathbf{q}_R+\mathbf{Q}$ is a general point if $\mathbf{q}$

is a general point and that they never coincide. Note in addition that S_0 does not depend on the particular j of $(\mathbf{q}_R,j)$, whereas $W^{(\mathbf{q}_R,j)}(K)$, of course, depends on the corresponding j:

$$W^{(\mathbf{q}_R,j)}_{R',R''}(\mathbf{q},1)=\delta_{R',R''S_0}D^{j}(R'), \quad j=1,2,3. \tag{7.26}$$

Hence only two different situations may occur, namely either (7.23) or (7.25). In contrast thereto we obtained for the automorphism approach four different cases, namely (7.12) and (7.13). Therefrom, it follows that $a=(E\,|\mathbf{b}_0)$ and $L\in\{(\mathbf{0},2),(\mathbf{0},3)\}$ define the same mapping, namely $(\mathbf{q},1)$ onto $(\mathbf{q},1)$, but are presented by different similarity matrices $Z^K(E\,|\mathbf{b}_0)$ or $W^L(K)$, respectively. Therefrom, we infer that *different* symmetry relations for corresponding CG vectors of G are obtained, if either $a=(E\,|\mathbf{b}_0)\in A(G)$ or $L\in\{(\mathbf{0},2),(\mathbf{0},3)\}$ is taken. More drastic is the difference between (7.13) and (7.25), since for (7.13) three different nontrivial mappings are realized, whereas for (7.25) only one nontrivial mapping of A_G onto A_G occurs. This means for the latter case that it suffices to consider, e.g., $L=(\mathbf{q}_R,1)$, since the remaining $L\in\{(\mathbf{q}_R,2),(\mathbf{q}_R,3)\}$ are redundant insofar as they lead to the same mapping. Besides this, none of the vectors $S^{-1}\mathbf{q}$, $S\in\{E,C_{2a},\sigma_z,\sigma_{db}\}$, coincides with $S_0\mathbf{q}+\mathbf{q}_R+\mathbf{Q}$, if only $\mathbf{q}$ is a general point of $\Delta\,\mathrm{BZ}(G)$. Therefrom, it follows that $a=(S\,|\mathbf{0})$, $S\in\{C_{2a},\sigma_z,\sigma_{db}\}$ and $L=(\mathbf{q}_R,1)$ give rise to *different* generating relations for corresponding CG vectors of G. Hence both approaches can be used simultaneously, since $S^{-1}\mathbf{q}$, $S\in\{C_{2a},\sigma_z,\sigma_{db}\}$, and $S_0\mathbf{q}+\mathbf{q}_R+\mathbf{Q}$ are four different elements of $\Delta\,\mathrm{BZ}(G)$, which characterize four inequivalent unirreps of G.

Accordingly, both approaches present a useful tool for the computation of CG vectors and should therefore be employed for practical calculations (e.g., crosschecks). The present example shows obviously the utility of both approaches. The automorphism approach allows one to restrict the calculations of CG vectors to $\mathbf{q}$'s of the subset $\Delta\,\mathrm{BZ}(\mathscr{O}_h^1)=\Omega$ of $\Delta\,\mathrm{BZ}(G)$, since the remaining CG vectors are obtained by means of the corresponding generating relations that are associated with $a=(S\,|\mathbf{0})$, $S\in\{C_{2a},\sigma_z,\sigma_{db}\}$. The other approach leads to a similar reduction of the calculations, since the mapping $L=(\mathbf{q}_R,1):(\mathbf{q},1)\to(S_0\mathbf{q}+\mathbf{q}_R+\mathbf{Q},1)$ also allows one to restrict to $\mathbf{q}$'s, which belong to a corresponding subset of $\Delta\,\mathrm{BZ}(G)$.

The preceding discussion also suggests consideration of complex conjugation in comparison to the other two approaches. For simplicity, we assume once again that the unirrep D^K of $G=P23$ belongs to a general point of $\Delta\,\mathrm{BZ}(G)$. Complex conjugation requires

$$D^K(R\,|\mathbf{t})^*=U^KD^{K*}(R\,|\mathbf{t})U^{K+}, \tag{7.27}$$

where $K^*=(\mathbf{q},1)\in A_G$ labels the complex conjugate unirrep. Provided that $\mathbf{q}$ does not belong to the "surface" of $\Delta\,\mathrm{BZ}(G)$, Eq. (7.9) yields to

$$\mathbf{q}^*=-S^*\mathbf{q}, \tag{7.28}$$

where $\mathbf{q}^*\in\Delta\,\mathrm{BZ}(G)$ and $S^*\in\mathscr{T}$ are uniquely defined. Provided that $\mathbf{q}^*$ does not coincide with a particular $S^{-1}\mathbf{q}$, $S\in\{E,C_{2a},\sigma_z,\sigma_{db}\}$, the corresponding generating relation (4.18) is compatible with the generating relations that are generated by the automorphisms $a=(S\,|\mathbf{0})$, $S\in\{E,C_{2a},\sigma_z,\sigma_{db}\}$, and by $L=(\mathbf{q}_R,1)$. Obviously, the vectors $\mathbf{q}^*$ and $S_0\mathbf{q}+\mathbf{q}_R+\mathbf{Q}$ are always different, since their magnitude are different. Certainly, if, on the other hand, $\mathbf{q}$ would be identical with a particular $S^{-1}\mathbf{q}$, $S\in\{E,C_{2a},\sigma_z,\sigma_{db}\}$, the corresponding generating relations cannot be used simultaneously in the special form (4.18) and (5.2), since they would lead to conflicting defintions of CG vectors.

CONCLUDING REMARKS

This paper introduced a new class of symmetry and generating relations for CG vectors. These arose from the full automorphism group. We argued the utility of the automorphism approach for a practical and efficient computation of CG coefficients with a particular interest in applications to space groups. We showed, on hand of some striking examples, that symmetry and generating relations are *different* from those associated with automorphisms, from the multiplication by one-dimensional nontrivial unirreps, and from complex conjugation. Symmetry relations defined by (4.5), (4.12), and (6.8) can be used simultaneously, but one should be aware that the matrices $B(a)$, $a\in A(G)^{\alpha\alpha'\alpha''}$, f, and e are (up to phase factors) *unique* for given unirreps and given CG vectors. They only can transformed simultaneously into $\beta B(a)\beta^+$, $\beta f\beta^+$, and $\beta e\beta^+$ in accordance with (4.11). Certainly this is the only freedom when the unirreps of G are given. Hence one cannot choose independently $B(a)$, $a\in A(G)^{\alpha\alpha'\alpha''}$, f, and e in case of symmetry relations. Nevertheless, symmetry relations are very useful and can be employed to resolve the multiplicity problem, for instance, when $B(a)$, $a\in A(G)^{\alpha\alpha'\alpha''}$, either defines a unirrep of the factor group $A(G)^{\alpha\alpha'\alpha''}/A(G)_i$ or is a reducible representation that decomposes into inequivalent unirreps only. For both cases the multiplicity index can be identified with the irrep label and row index of those unirreps of the factor group that are contained in $B(a)$, $a\in A(G)^{\alpha\alpha'\alpha''}$. Generating relations, which are defined by outer automorphisms $[(\alpha,\alpha',\alpha'')\to(a_t(\alpha),a_t(\alpha'),a_t(\alpha'')),\ a_t\in A(G):A(G)^{\alpha\alpha'\alpha''}]$ were discussed and compared with those associated with complex conjugation $[(\alpha,\alpha',\alpha'')\to(\alpha^*,\alpha^{*\prime},\alpha^{*\prime\prime})]$ and with the multiplication by one-dimensional nontrivial unirreps $[(\alpha,\alpha',\alpha'')\to(\gamma(\alpha),\gamma'(\alpha'),\gamma''(\alpha''))]$, where (6.2) is now abbreviated as $\gamma''=\gamma\cdot\gamma'$]. In contrast to symmetry relations, it is admissible for generating relations to choose simultaneously $B(a_t)=\mathbf{1}_m$, $a_t\in A(G):A(G)^{\alpha\alpha'\alpha''}$, $f=\mathbf{1}_m$, and $h=\mathbf{1}_m$, if the triplets $(a_t(\alpha),a_t(\alpha'),a_t(\alpha''))$, $a_t\in A(G):A(G)^{\alpha\alpha'\alpha''}$, $(\alpha^*,\alpha^{*\prime},\alpha^{*\prime\prime})$, and $(\gamma(\alpha),\gamma'(\alpha'),\gamma''(\alpha''))$, are *different*, since otherwise conflicting CG vectors of G would be defined by (5.2), (4.18), and (6.9) (with $h_{vw}=\delta_{vw}$). Of course, it is also allowable to combine the different sets of generating relations in a generic way in order to obtain further generating relations. In any case the automorphism approach presents a new way of computing and correlating sets of CG coefficients and therefore should be used in combination with the other approaches for an efficient computation of CG coefficients.

ACKNOWLEDGMENTS

The author is indebted to Professor H. Wondratschek for drawing his attention to Ref. 10. Valuable discussions and correspondence with Professor J. Neubüser and Professor W. Opechowski are gratefully acknowledged. Useful critizisms and valuable amendments of the referee are also gratefully acknowledged.

[1]L. Jansen and M. Boon, *The Theory of Finite Groups. Applications in Physics* (North-Holland, Amsterdam, 1972).
[2]C. J. Bradley and A. P. Cracknell, *The Mathematical Theory of Symmetry in Solids* (Clarendon, Oxford, 1972).
[3]Ref. 1, p. 158.
[4]G. W. Mackey, Acta Math. **99**, 265 (1958).
[5]R. Dirl, J. Math. Phys. **20**, 659 (1979).
[6]M. Hamermesh, *Group Theory and Its Applications to Physical Problems* (Addison-Wesley, Reading, MA, 1964).
[7]P. H. Butler, Trans. R. Soc. London Ser. A **277**, 545 (1975).
[8]P. H. Butler and A. M. Ford, J. Phys. A **12**, 1357 (1979).
[9]R. Dirl, "On the uniqueness and reality of Clebsch–Gordan coefficients. I. Ordinary representations," J. Math. Phys. (to be published).
[10]E. Koch and W. Fischer, Acta Cryst. A **31**, 88 (1975).
[11]H. Burzlaff and H. Zimmerman, Z. Krist. **153**, 151 (1980).
[12]R. Dirl, "Automorphism properties of space group representations and space group selection rules" (in preparation).
[13]A. P. Cracknell, B. L. Davies, S. C. Miller, and W. F. Love, *Kronecker Product Tables* (Plenum, New York, 1979), Vol. 1.

On the quaternion representation of the proper Lorentz group SO(3,1)

K. N. Srinivasa Rao,[a] A. V. Gopala Rao, and B. S. Narahari
Department of Physics, University of Mysore, Mysore 570 006, India

(Received 1 July 1982; accepted for publication 7 October 1982)

Complex quaternions are investigated in detail, bringing out some new aspects of the relationship of the multiplicative group of unit complex quaternions (UCQ) with the proper Lorentz group SO(3,1). Constructing the proper Lorentz transformation (PLT) corresponding to a given UCQ, the quaternion parameters of a PLT are determined explicitly in terms of its element L_{ij}, and this quaternion parametrization is then utilized to obtain an interesting geometrical interpretation of SO(3,1) as the intersection of a hyperboloid with a cone in a real eight-dimensional Euclidean space E_8. The UCQ components are then related to the Lie–Cartan parameters of SO(3,1), leading to an identification of complex quantities which may be interpreted as the complex axis and angle of rotation. It is shown that any PLT admits a special type of Euler resolution which is at the same time a resolution into three Lorentz–Synge screws the two angles of which combine to form a complex Euler angle (or Euler–Brauer angle). It is also shown that on taking the rotation parameters in the formula for the D^j representation of SO(3) to be complex, one obtains the D^{j0} representation of SO(3,1), leading at once to its $D^{ij'}$ representation. Similarly, a formula for the character χ^{j0} of the D^{j0} representation, having a complete analogy to the character formula for SO(3), but in terms of a complex angle ω is obtained and this in turn yields a formula for the character $\chi^{ij'}$ in the $D^{ij'}$ representation of SO(3,1).

PACS numbers: 02.20. + b

I. INTRODUCTION

In an earlier paper,[1] PLT's were classified according to (i) the self-representation of the proper Lorentz group SO(3,1), (ii) the three-dimensional complex orthogonal representation by the group SO(3,C) (iii) the two-dimensional complex unimodular representation by the group SL(2,C), and lastly (iv) the representation by the multiplicative group Q of UCQ's. It was shown that a proper Lorentz transformation $L = (L_{ij})$ is characterized, in general, by two invariants θ_r and θ_b (to be denoted by ω_r and ω_b in this paper) each of which is expressed explicitly in terms of L_{ij} and a PLT for which both θ_r and θ_b are nonzero was called *screwlike* because such a transformation is *equivalent*, by a proper Lorentz transformation, to what Synge[2] has called a "4-screw." Since a screw with $\theta_b = 0$ is a pure rotation and one with $\theta_r = 0$ is a pure boost, PLT's with $\theta_b = 0$, $\theta_r \neq 0$ were called *rotationlike* and those for which $\theta_r = 0$, $\theta_b \neq 0$ were called *boostlike*. A PLT for which $\theta_r = \theta_b = 0$ (with the exception of the trivial identity) is known as a *null* or *singular* transformation, and such a PLT is not equivalent to a 4-screw. A rotationlike PLT with $\theta_r = \pi$ was called *exceptional* and is the same as the *involution* of Wigner[3] as it satisfies the algebraic relation $L^2 = 1$. The character $\chi(L)$ in each of the three representations (i), (ii), and (iii) and the constant term q_4 in the UCQ representation $q = \mathbf{q}\cdot\mathbf{e} + q_4$ essentially determine the classification. The constant q_4 in (iv) and the characters $\chi_A(L)$ and $\chi_C(L)$ in (ii) and (iii) are, in general, complex and have two degrees of freedom to provide a complete classification while, the $\chi(L)$ in (i) being real, it is necessary to invoke another real invariant ξ, which turns out to be the sum of all principal minors of the second order in the determinant of L. The four classification schemes are displayed in Table I.

We explore in this paper this last UCQ representation Q of SO(3,1) more fully and show that it has all the advantages of the (real) unit quaternion representation of the rotation group SO(3) and establishes a complete parallelism between the results for the nonnull transformations of SO(3,1) and the whole of SO(3) by means of an appropriate complexification, in addition to yielding other results for the null transformations of SO(3,1) which are the only ones that do not have SO(3) analogs.

In Sec. II, a PLT is expressed explicitly in terms of the UCQ components q_i, yielding a relationship between the invariants of L and the UCQ components. Conversely, the q_i are also expressed in terms of L_{ij} for all types of L listed in Table I. A geometrical representation of SO(3,1) in a real eight-dimensional Euclidean space E_8, which includes, as a particular case, the well-known interpretation of SO(3) as a unit sphere in a E_4 is also presented.

In Sec. III, the parameters q_i are related to the Lie–Cartan parameters of SO(3,1) obtaining two complex quantities ω and $\hat{\mathbf{n}}$, which are interpreted as the complex angle and axis corresponding to an L, using the three-dimensional complex orthogonal representation SO(3,C).

Section IV is devoted to a discussion of special types of Euler resolution of a PLT which are also resolutions of the PLT into a product of three Lorentz–Synge screws.

In Sec. V, using the known result that the D^{j0} representation of SL(2,C) may be obtained by replacing the Cayley–Klein parameters in the D^j representation of SU(2) by the generalized Cayley–Klein parameters, we obtain the $D^{ij'}$ representations of SO(3,1) in terms of complex quantities such as $(\omega,\hat{\mathbf{n}})$ introduced in Sec. III. By employing the Clebsch–Gordon theorem, the character formula for the $D^{ij'}$ representation of SO(3,1) is then obtained in terms of the complex angle ω.

[a] Formerly Professor of Theoretical Physics, now retired from service.

 0022-2488/83/081945-10$02.50

TABLE I. Classification of proper Lorentz transformations in the various representations.

Representation \ L	Nonplanar transformations	Planar transformations					Characterizing interger (dimension of the representation)
		General condition	Boostlike	Null	Rotationlike	Exceptional	
SO(3, 1)	$1+\frac{1}{2}\xi<\chi$	$1+\frac{1}{2}\xi=\chi$	$\infty>\chi>4$	$\chi=4$	$0<\chi<4$	$\chi=0$	4
SO(3, C)	χ_A complex or χ_A real and <-1	χ_A real and $\geqslant -1$	$\infty>\chi_A>3$	$\chi_A=3$	$-1<\chi_A<3$	$\chi_A=-1$	3
SL(2, C)	χ_C complex	χ_C real	$\infty>\chi_C>2$ or $-\infty<\chi_C<-2$	$\chi_C=\pm 2$	$0<\chi_C<2$ or $0>\chi_C>-2$	$\chi_C=0$	2
Q	q_4 complex	q_4 real	$\infty>q_4>1$ or $-\infty<q_4<-1$	$q_4=\pm 1$	$0<q_4<1$ or $0>q_4>-1$	$q_4=0$	1

II. EXPLICIT CONSTRUCTION OF THE UNIT COMPLEX QUATERNION CORRESPONDING TO A PROPER LORENTZ TRANSFORMATION

A *complex quaternion* a is defined by

$$a=a_\alpha e_\alpha+a_4=\mathbf{a}\cdot\mathbf{e}+a_4, \tag{2.1}$$

where a_i are any four complex numbers and e_α are the *quaternion units* satisfying

$$e_\alpha e_\beta=-e_\beta e_\alpha=e_\gamma, \quad \alpha,\beta,\gamma \text{ cyclic}, \quad \text{and } e_\alpha^2=-1. \tag{2.2}$$

(Throughout this paper, we use the Greek suffixes α,β,γ, etc., for the range 1,2,3 and the Latin suffixes i,j,k etc., for the range 1,2,3,4. Unless otherwise stated, we employ the summation convention of summing over repeated suffixes. In the Minkowski space–time, we use coordinates with $x_4\equiv ict$). The *Hamiltonian conjugate* $\bar{a}$ of a complex quaternion a is defined as

$$\bar{a}=-a_\alpha e_\alpha+a_4=(-\mathbf{a})\cdot\mathbf{e}+a_4, \tag{2.3}$$

and is to be contrasted with the *ordinary complex conjugate*

$$a^*=a_\alpha^* e_\alpha+a_4^*=\mathbf{a}^*\cdot\mathbf{e}^*+a_4^*, \tag{2.4}$$

where a_i^* is the complex conjugate of a_i. The (*quaternion*) *product* and the *scalar product* of two quaternions a and b are respectively given by

$$ab=(a_4b_4-\mathbf{a}\cdot\mathbf{b})+(a_4\mathbf{b}+b_4\mathbf{a}+\mathbf{a}\times\mathbf{b})\cdot\mathbf{e}\neq ba \tag{2.5}$$

and

$$\tfrac{1}{2}(a\bar{b}+b\bar{a})=\tfrac{1}{2}(\bar{a}b+\bar{b}a)=(\mathbf{a}\cdot\mathbf{b}+a_4b_4). \tag{2.6}$$

From the scalar product, it follows that the norm $\|a\|$ of a quaternion is given by

$$\|a\|=a\bar{a}=\bar{a}a=a_1^2+a_2^2+a_3^2+a_4^2, \tag{2.7}$$

and quaternions with unit norm are called *unit complex quaternions* (UCQ).

In a paper by Synge[4] dedicated to the memory of A. W. Conway (1875–1950), who first introduced quaternions into special relativity, it is shown that, with each proper Lorentz transformation (PLT), one can associate a pair of UCQ's which differ only in an overall sign. In that paper, each 4-vector $X\equiv(X_i)$, belonging to the Minkowski world, is mapped into the quaternion

$$x=\mathbf{X}\cdot\mathbf{e}+X_4, \tag{2.8}$$

called a *minquat* which satisfies

$$x+\bar{x}^*=0, \tag{2.9}$$

in view of the fact that X_α are real and X_4 is pure imaginary. Then it is shown that under the quaternion transformation

$$x\to x'=ax\bar{a}^*, \quad a\bar{a}=1, \tag{2.10}$$

minquats are transformed into minquats, with norms unaltered, i.e.,

$$x+\bar{x}^*=0\Leftrightarrow x'+\bar{x}'^*=0, \tag{2.11}$$

and

$$x\bar{x}=x'\bar{x}'. \tag{2.12}$$

From these two properties (2.11) and (2.12), it is then inferred that Eq. (2.10) implies the proper Lorentz transformation

$$X_i'=L_{ij}X_j, \tag{2.13}$$

and in turn the mapping

$$X\leftrightarrow x, \quad L\leftrightarrow\pm a, \quad L\equiv(L_{ij}), \tag{2.14}$$

between 4-vectors and minquats on the one hand and proper Lorentz transformations and UCQ's on the other.

It is to be noticed, however, that if one considers the complex conjugate of Eq. (2.10), i.e.,

$$x^*\to x'^*=a^*x^*\bar{a}, \quad a^*\bar{a}^*=1, \tag{2.15}$$

then one is led exactly to the same PLT as in Eq. (2.13), but with the *different mapping*

$$X\leftrightarrow x^*, \quad L\leftrightarrow\pm a^*. \tag{2.16}$$

This, together with Eq. (2.14), shows that, with each proper Lorentz transformation L, one can associate *any one* of the pairs of unit complex quaternions $\pm a$ or $\pm a^*$. Thus, the UCQ corresponding to a given PLT is indeterminate not only with regard to an overall sign, but also with regard to complex conjugation. In fact, as we shall see in Sec. V, the two different mappings yield the mutually *inequivalent* representations $D^{0\,1/2}$ and $D^{1/2\,0}$ of SO(3,1). Therefore, for the

purpose of constructing the PLT corresponding to a UCQ, one may use either of the two mappings in Eqs. (2.14) and (2.16); if Eq. (2.14) yields an L expressed in terms of a, Eq. (2.16) yields the *same* L expressed in terms of a^*. In this paper, we choose for convenience Eq. (2.16) with the changed notation

$$a^* \equiv q, \quad x \equiv \mathbf{X}\cdot\mathbf{e} - X_4,$$

so that Eqs. (2.15) and (2.16) become

$$x' = qx\bar{q}^*, \quad X' = LX, \tag{2.17}$$
$$X \leftrightarrow x, \quad L \leftrightarrow \pm q.$$

On writing $q = \mathbf{q}\cdot\mathbf{e} + q_4$ in the above equation and expanding, we obtain an equation corresponding to $X' = LX$ from which the elements of the matrix L may be easily identified as

$$\begin{aligned} L_{\alpha\beta} &= (q_4 q_4^* - \mathbf{q}\cdot\mathbf{q}^*)\delta_{\alpha\beta} + (q_\alpha q_\beta^* + q_\beta q_\alpha^*) \\ &\quad - \epsilon_{\alpha\beta\gamma}(q_\gamma q_4^* + q_4 q_\gamma^*), \\ L_{\alpha 4} &= (q_4 q_\alpha^* - q_\alpha q_4^*) + \epsilon_{\alpha\beta\gamma} q_\beta q_\gamma^*, \\ L_{4\alpha} &= (q_\alpha q_4^* - q_4 q_\alpha^*) + \epsilon_{\alpha\beta\gamma} q_\beta q_\gamma^*, \\ L_{44} &= q_i q_i^*, \end{aligned} \tag{2.18}$$

where $\epsilon_{\alpha\beta\gamma}$ is the usual antisymmetric permutation symbol. These equations give L_{ij} as explicit functions of the UCQ components q_i. When the q_i are real, Eq. (2.18) reduces to the rotation matrix formula (see, for example, Synge[5]) in the sense that

$$\begin{aligned} L_{\alpha\beta} \equiv R_{\alpha\beta} &= 2q_\alpha q_\beta - \delta_{\alpha\beta}(1 - 2q_4^2) - 2\epsilon_{\alpha\beta\gamma} q_\gamma q_4, \\ L_{\alpha 4} &= L_{4\alpha} = 0, \quad L_{44} = 1, \end{aligned} \tag{2.19}$$

where we have used $q_i q_i = 1$, in rewriting the coefficient of $\delta_{\alpha\beta}$ in $L_{\alpha\beta}$. This result is also evident from the prescriptions in Eqs. (2.10) and (2.15) as they both reduce to the rotation–quaternion relation[5]

$$R \leftrightarrow \pm q = \pm(\mathbf{q}\cdot\mathbf{e} + q_4), \quad q_i q_i = 1, \tag{2.20}$$

when the q_i are real.

Equations (2.18) may be inverted to obtain q_i in terms of L_{ij} also. To do this, we first express the two (real) invariants χ and ξ of the matrix L appearing in its charcteristic equation,[1] i.e.,

$$\lambda^4 - \chi\lambda^3 + \xi\lambda^2 - \chi\lambda + 1 = 0 \tag{2.21}$$

in terms of q_i. We may note that the invariants χ and ξ are, respectively, the trace and the sum of all second-order principal minors of L. A direct evaluation using Eq. (2.18) yields

$$\chi = 4q_4 q_4^*, \tag{2.22}$$
$$1 + \tfrac{1}{2}\xi = 2(q_4^2 + q_4^{*2}). \tag{2.23}$$

On introducing the two 3-vectors $\mathscr{E}$ and $\mathscr{H}$ of the antisymmetric part of L, i.e.,

$$\begin{aligned} \mathscr{E}_\alpha &\equiv -i(L_{\alpha 4} - L_{4\alpha})/2 = i(q_\alpha q_4^* - q_4 q_\alpha^*), \\ \mathscr{H}_\alpha &\equiv \epsilon_{\alpha\beta\gamma} L_{\beta\gamma}/2 = -(q_\alpha q_4^* + q_4 q_\alpha^*), \end{aligned} \tag{2.24}$$

we obtain another useful invariant of L, namely

$$\mathscr{E}\cdot\mathscr{H} = i(q_4^2 - q_4^{*2}). \tag{2.25}$$

This invariant vanishes for all planar PLT's as q_4 is then real (see Table I). For pure rotations, i.e., when the q_i are all real, $\mathscr{E} = 0$. Also, when either $q_\alpha = 0\,(\alpha = 1,2,3)$, or $q_4 = 0$, both the vectors $\mathscr{E}$ and $\mathscr{H}$ vanish. The former corresponds to the trivial identity transformation whereas the latter, i.e., $q_4 = 0$, corresponds to the exceptional PLT (see Table I).

Now, on solving Eqs. (2.23) and (2.25) for q_4, we get

$$q_4 = \pm[(\xi + 2 - i4\mathscr{E}\cdot\mathscr{H})/8]^{1/2}, \tag{2.26}$$

thus yielding q_4 in terms of the invariants of L. On eliminating q_4 between Eqs. (2.22) and (2.26), we obtain, incidentally,

$$\chi = \tfrac{1}{2}[(\xi + 2)^2 + 16(\mathscr{E}\cdot\mathscr{H})^2]^{1/2}, \tag{2.27}$$

which is a relation connecting the three invariants of L introduced above. This is to be expected since L has only two algebraically independent invariants. In order to obtain the remaining UCQ components q_α, we use the relation

$$\mathscr{H}_\alpha + i\mathscr{E}_\alpha = -2q_4^* q_\alpha, \tag{2.28}$$

implied by Eq. (2.24). This, together with Eq. (2.26) yields, when $q_4 \neq 0$, the UCQ components corresponding to nonexceptional PLT as

$$\begin{aligned} q_\alpha &\mp [(\xi + 2 + i4\mathscr{E}\cdot\mathscr{H})/2]^{-1/2}(\mathscr{H}_\alpha + i\mathscr{E}_\alpha), \\ q_4 &= \pm[(\xi + 2 - i4\mathscr{E}\cdot\mathscr{H})/8]^{1/2}. \end{aligned} \tag{2.29}$$

In the case of (the nonexceptional) planar transformations, the above formulas may also be expressed as

$$q_\alpha = \mp(\chi)^{-1/2}(\mathscr{H}_\alpha + i\mathscr{E}_\alpha), \quad q_4 = \pm(\chi/4)^{1/2}, \tag{2.30}$$

as then $\mathscr{E}\cdot\mathscr{H} = 0$ and $\xi + 2 = 2\chi$ (see Table I). In particular, for the (planar) null PLT's for which $\chi = 4$, the above formula further simplifies to

$$q_\alpha = \mp(\mathscr{H}_\alpha + i\mathscr{E}_\alpha)/2, \quad q_4 = \pm 1. \tag{2.31}$$

Equation (2.28) evidently fails to determine q_α in the case of the exceptional PLT's for which $q_4 = 0$, $\mathscr{E} = \mathscr{H} = 0$. However, in this case, we obtain, on using Eq. (2.18), the relation

$$q_\alpha q_\beta^* = \tfrac{1}{2}(L_{\alpha\beta} + L_{44}\delta_{\alpha\beta} + \epsilon_{\alpha\beta\gamma} L_{\gamma 4}). \tag{2.32}$$

As $q_i q_i = 1$ and $q_4 = 0$, at least one of the three q_α, say q_μ must be nonzero for an excpetional PLT, and we may use it to determine the other two q_α from Eq. (2.32). Thus we obtain

$$\begin{aligned} &q_4 = 0, \quad q_\alpha = (L_{\alpha\mu} + L_{44}\delta_{\alpha\mu} + \epsilon_{\alpha\mu\nu} L_{\nu 4})/2q_\mu^*, \\ &\text{and} \\ &q_\mu = \pm[(1 + L_{\mu\mu} L_{44} + L_{\mu\nu} L_{\rho 4} - L_{\mu\rho} L_{\nu 4} - L_{\mu 4}^2)/2]^{1/2}, \end{aligned} \tag{2.33}$$

where $L_{\mu\mu}$ is the element of L occurring in the μth row and the μth column (the repeated index μ *does not* indicate summation) and μ, ν, ρ are cyclic.

The above formulas yielding the UCQ components corresponding to a PLT reduce to the corresponding ones for the rotation group on restricting the q_i to be real. On recalling that $\mathscr{E} = 0$ when q_i are real, we obtain from Eq. (2.30)

$$\begin{aligned} q_\alpha &= \mp\tfrac{1}{2}(1 + \chi_R)^{-1/2}\epsilon_{\alpha\beta\gamma} R_{\beta\gamma}, \\ q_4 &= \pm\tfrac{1}{2}(1 + \chi_R)^{1/2}, \quad \chi \equiv 1 + \chi_R, \end{aligned} \tag{2.34}$$

which corresponds to nonexceptional rotations. Here, we have written $\chi = 1 + \chi_R$, where χ_R is the trace of the rota-

tion matrix $R_{\alpha\beta} \equiv L_{\alpha\beta}$ [see also Eq. (2.19)]. When $q_4 = 0$, i.e., $\chi = 0$, and $\chi_R = -1$, which corresponds to the case of exceptional rotations,[1,6] we obtain from Eq. (2.33)

$$q_4 = 0, \quad q_\alpha = (R_{\mu\alpha} + \delta_{\mu\alpha})/2q_\mu, \tag{2.35}$$

and

$$q_\mu = \pm [(1 + R_{\mu\mu})/2]^{1/2},$$

where $R_{\mu\mu} \neq -1$ is the element of R occurring in the μth row and μth column and the repeated index μ does not indicate summation.

Here, it is appropriate to mention the relation of the UCQ components q_i to the complex parameters q_i of SL(2,C) defined by[7-9]

$$g = g_i \sigma_i, \quad g \in \mathrm{SL}(2,C), \tag{2.36}$$

where σ_α are the Pauli matrices

$$\sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \tag{2.37}$$

and σ_4 is the 2×2 unit matrix. The condition $\det(g) = \pm 1$ then implies

$$\eta_{ij}\, g_i\, g_j = 1, \tag{2.38}$$

where $\eta \equiv (\eta_{ij})$ is the diagonal matrix diag$(-1, -1, -1, 1)$. It is to be observed that there is a basic difference between the UCQ components q_i and the similar looking g_i, whereas the q_i are defined with respect to the abstract hypercomplex system of the quaternions and hence are unique (except for the inherent ambiguity in sign), the g_i are defined with respect to the specific representation chosen for the e_α in terms of the Pauli matrices and hence are representation dependent. In fact, a change from a set σ_i of Pauli matrices to another equivalent set $\sigma_i' \equiv \tau^{-1} \sigma_i \tau$, where τ is any nonsingular 2×2 matrix, is equivalent to changing from one set of parameters g_i to another set g_i', so that

$$g = g_i \sigma_i = g_i' \sigma_i' = g_i' (\tau^{-1} \sigma_i \tau) = (M_{ji}\, g_i') \sigma_j,$$

where we have expressed $(\tau^{-1} \sigma_i \tau)$ in the basis σ_i as

$$\tau^{-1} \sigma_i \tau = \tilde{M}_{ij} \sigma_j.$$

Thus we obtain the transformation

$$g_j = M_{ji}\, g_i', \tag{2.39}$$

where the complex 4×4 matrix $M \equiv (M_{ij})$ has to satisfy, in view of Eq. (2.38),

$$\tilde{M} \eta M = \eta, \tag{2.40}$$

so that $\eta_{ij}\, g_i'\, g_j' = 1$.

We will now show that the UCQ representation leads to an interesting geometrical interpretation of the proper Lorentz group in an eight-dimensional real Euclidean space E_8. It is well known[10] that the rotation group SO(3) is topologically equivalent to a unit sphere with diametrically opposite points identified, in a four-dimensional real Euclidean space E_4. This also follows immediately from the fact (see Synge[5]) that the UCQ components $q_i \equiv u_i$ of a pure rotation are all real and satisfy the relation

$$u_1^2 + u_2^2 + u_3^2 + u_4^2 = 1. \tag{2.41}$$

On the other hand, for a pure boost (see Møller[11]) with velocity $\mathbf{w} = (w_x, w_y, w_z)$, the UCQ components [obtained by using Eq. (2.30)] are

$$\begin{aligned} q_4 \equiv u_4 &= [\tfrac{1}{2}(1 + \gamma)]^{1/2}, \\ q_\alpha \equiv i v_\alpha &= -i[2(\gamma + 1)]^{-1/2} \gamma w_\alpha / c \\ &= i[\tfrac{1}{2}(\gamma - 1)]^{1/2} w_\alpha / w, \end{aligned} \tag{2.42}$$

where $\gamma \equiv (1 - w^2/c^2)^{-1/2}$ and $w \equiv (w_x^2 + w_y^2 + w_z^2)^{1/2}$. These evidently satisfy

$$u_4^2 - v_1^2 - v_2^2 - v_3^2 = 1. \tag{2.43}$$

showing that the set of all pure boosts lie on a hyperboloid of two sheets[12] in an E_4, again with diametrically opposite points identified since $\pm q$ yield the same PLT.

On writing

$$q_k = u_k + i v_k, \tag{2.44}$$

the UCQ corresponding to a general PLT is split into two real quaternions u qnd v as

$$\begin{aligned} L \to \pm q &= \pm [(u_\alpha + i v_\alpha) e_\alpha + (u_4 + i v_4)] \\ &= \pm (u + iv), \end{aligned} \tag{2.45}$$

which satisfy the relations

$$u_k u_k - v_k v_k + 2 i u_k v_k = 1 \tag{2.46}$$

in view of $q_i q_i = 1$. This implies, in turn,

$$u_1^2 + u_2^2 + u_3^2 + u_4^2 - v_1^2 - v_2^2 - v_3^2 - v_4^2 = 1 \tag{2.47}$$

and

$$u_1 v_1 + u_2 v_2 + u_3 v_3 + u_4 v_4 = 0. \tag{2.48}$$

In a real E_8, Eq. (2.47) represents a central hyperboloid referred to its principal axes and Eq. (2.48) a central cone[12] so that a general PLT represented as a point in E_8 with coordinates (u_i, v_i) lies on the intersection of the two surfaces given by Eqs. (2.47) and (2.48). Thus it follows that *the proper Lorentz group SO(3,1) is geometrically represented in E_8 by the surface of intersection of the hyperboloid of Eq. (2.47) with the cone in Eq. (2.48) with diametrically opposite points identified.*

It may also be observed that only the general nonplanar PLT's need an E_8 for a geometrical representation and the planar PLT's form surfaces in the subspaces of E_8. As already seen, the subgroup of pure rotations is confined to a sphere in one four-dimensional subspace of E_8 and the set of all pure boosts is confined to a hyperboloid in another four-dimensional subspace. Other planar transformations with $v_4 = 0$ and $u_4 \gtrless 1$ are confined to a seven-dimensional subspace of E_8. Exceptional transformation with $u_4 = v_4 = 0$ need only a six-dimensional subspace. Lastly, null transformations with $u_4 = 1$ and $v_4 = 0$ lie on a six-dimensional subspace which is the intersection of the two surfaces

$$u_1^2 + u_2^2 + u_3^2 - v_1^2 - v_2^2 - v_3^2 = 0$$

and

$$u_1 v_1 + u_2 v_2 + u_3 v_3 = 0,$$

which are *both* central cones.[12]

III. RELATION BETWEEN THE UCQ COMPONENTS AND THE LIE–CARTAN PARAMETERS OF A PROPER LORENTZ TRANSFORMATION

The proper Lorentz group SO(3,1) being a Lie group, we know that every $L\in$SO(3,1) may be expressed in a sufficiently small neighborhood of the identity in the form

$$L = \exp(I), \tag{3.1}$$

where

$$I_{\alpha\beta} = \epsilon_{\alpha\beta\gamma}h_\gamma, \quad I_{\alpha 4} = -I_{4\alpha} = ie_\alpha \tag{3.2}$$

are the elements of the antisymmetric matrix I. The six real parameters (h_α, e_α) may be called the Lie–Cartan parameters of L. If $I^{(\alpha\beta)}$ and $I^{(\alpha 4)}$ denote the matrices of the infinitesimal transformations (in Minkowski coordinates) in the respective coordinate planes, then we have evidently

$$I = \sum_\alpha (e_\alpha I^{(\alpha 4)} + h_\alpha I^{(\beta\gamma)}), \quad \alpha,\beta,\gamma \text{ cyclic}. \tag{3.3}$$

By introducing the two sets of matrices

$$J_\alpha = \tfrac{1}{2}(I^{(\beta\gamma)} - iI^{(\alpha 4)}),$$
$$K_\alpha = \tfrac{1}{2}(I^{(\beta\gamma)} + iI^{(\alpha 4)}), \quad \alpha,\beta,\gamma, \text{ cyclic}, \tag{3.4}$$

as is done while showing[13] that the rotation group SO(4) is a direct product of two rotation groups each of which is isomorphic to SO(3), we may also write

$$I = J_\alpha\omega_\alpha + K_\alpha\omega_\alpha^* \equiv \Omega + \Omega', \tag{3.5}$$

where the parameters ω_α are defined by

$$\omega_\alpha \equiv h_\alpha + ie_\alpha \quad (\omega_\alpha^* \equiv h_\alpha - ie_\alpha), \tag{3.6}$$

and may be called the complex Lie–Cartan parameters of L. The elements of the skew-symmetric matrices Ω and Ω' are given by

$$\Omega_{\alpha\beta} = \tfrac{1}{2}\epsilon_{\alpha\beta\gamma}\omega_\gamma = \Omega'^*_{\alpha\beta}, \quad \Omega_{\alpha 4} = \tfrac{1}{2}\omega_\alpha = -\Omega'^*_{\alpha 4}, \tag{3.7}$$

and it is easy to verify that these matrices satisfy

$$\Omega\Omega' = \Omega'\Omega, \quad \Omega^2 = -\tfrac{1}{4}\omega^2, \quad \Omega'^2 = -\tfrac{1}{4}\omega^{*2}, \tag{3.8}$$

where

$$\omega^2 \equiv \omega_\alpha\omega_\alpha. \tag{3.9}$$

Thus we have, when $\omega^2 \neq 0$,

$$\mathscr{L} \equiv \exp\Omega = \cos(\omega/2) + (2\Omega/\omega)\sin(\omega/2) \tag{3.10}$$

and

$$\mathscr{L}' \equiv \exp\Omega' = \cos(\omega^*/2) + (2\Omega'/\omega^*)\sin(\omega^*/2), \tag{3.11}$$

where $\omega = \omega_r + i\omega_b$ is a root of ω^2. Hence,

$$L = \exp(\Omega + \Omega') = \mathscr{L}\mathscr{L}' = \mathscr{L}'\mathscr{L}, \tag{3.12}$$

from which we obtain on using Eqs. (3.7)–(3.11),

$$\begin{aligned}
L_{\alpha\beta} &= \delta_{\alpha\beta}(cc^* - \boldsymbol{\omega}\cdot\boldsymbol{\omega}^* ss^*/\omega\omega^*) \\
&\quad + \epsilon_{\alpha\beta\gamma}(\omega_\gamma c^* s/\omega + \omega_\gamma^* cs^*/\omega^*) \\
&\quad + ss^*(\omega_\alpha\omega_\beta^*/\omega\omega^* + \omega_\alpha^*\omega_\beta/\omega^*\omega), \\
L_{\alpha 4} &= \epsilon_{\alpha\beta\gamma}(\omega_\beta\omega_\gamma^*/\omega\omega^*)ss^* - cs^*\omega_\alpha^*/\omega^* + c^* s\omega_\alpha/\omega, \\
L_{4\alpha} &= \epsilon_{\alpha\beta\gamma}(\omega_\beta\omega_\gamma^*/\omega\omega^*)ss^* + cs^*\omega_\alpha^*/\omega^* - c^* s\omega_\alpha/\omega, \\
L_{44} &= cc^* + ss^*(\boldsymbol{\omega}\cdot\boldsymbol{\omega}^*/\omega\omega^*),
\end{aligned} \tag{3.13a}$$

where $c \equiv \cos(\omega/2)$ and $s \equiv \sin(\omega/2)$. Comparing this with Eq. (2.18), we obtain

$$\cos(\omega/2) = \pm q_4, \quad (\omega_\alpha/\omega)\sin(\omega/2) = \mp q_\alpha. \tag{3.14a}$$

On the other hand, if $\omega^2 = 0$, we have

$$\mathscr{L} = E_4 + \Omega \quad \text{and} \quad \mathscr{L}' = E_4 + \Omega',$$

where E_4 is the 4×4 unit matrix. Thus we obtain, in this case,

$$\begin{aligned}
L_{\alpha\beta} &= \delta_{\alpha\beta}(1 - \tfrac{1}{4}\boldsymbol{\omega}\cdot\boldsymbol{\omega}^*) + \tfrac{1}{2}\epsilon_{\alpha\beta\gamma}(\omega_\gamma + \omega_\gamma^*) \\
&\quad + \tfrac{1}{4}(\omega_\alpha\omega_\beta^* + \omega_\alpha^*\omega_\beta), \\
L_{\alpha 4} &= \tfrac{1}{4}\epsilon_{\alpha\beta\gamma}\omega_\beta\omega_\gamma^* + \tfrac{1}{2}(\omega_\alpha - \omega_\alpha^*), \\
L_{4\alpha} &= \tfrac{1}{4}\epsilon_{\alpha\beta\gamma}\omega_\beta\omega_\gamma^* - \tfrac{1}{2}(\omega_\alpha - \omega_\alpha^*), \\
L_{44} &= 1 + \tfrac{1}{4}(\boldsymbol{\omega}\cdot\boldsymbol{\omega}^*).
\end{aligned} \tag{3.13b}$$

On comparing this with Eqs. (2.18) as before, we get

$$q_4 = \pm 1, \quad q_\alpha = \mp\omega_\alpha/2. \tag{3.14b}$$

It may be observed that Eqs. (3.13b) and (3.14b) result respectively from Eqs. (3.13a) and (3.14a) on formally passing to the limit $\omega\to 0$. Also, from Eqs. (3.14a,b) and Table I, it follows that a *a general complex ω corresponds to a nonplanar PLT whereas a ω which is real or purely imaginary corresponds to a planar one. Further, a planar PLT is seen to be rotationlike, null, boostlike, or exceptional according as ω is real, zero, purely imaginary, or equal to $\pm\pi$.*

The formulas of Eqs. (3.14a,b) may be used together with Eqs. (2.26), (2.29)–(2.33) to express the Lie–Cartan parameters ω of a PLT directly in terms of the invariants of L as follows: In the general nonplanar case, we get

$$\begin{aligned}
&\cos(\omega/2) = [(\xi + 2 - i4\mathscr{E}\cdot\mathscr{H})/8]^{-1/2}, \\
&(\omega_\alpha/\omega)\sin(\omega/2) = [(\xi + 2 + i4\mathscr{E}\cdot\mathscr{H})/2]^{-1/2}(\mathscr{H}_\alpha + i\mathscr{E}_\alpha).
\end{aligned} \tag{3.15a}$$

Although there is an apparent ambiguity of sign in determining the ω_α from these equations, we may note that the ω_α are, nevertheless, uniquely determined by the PLT as it should be, since ω and $\sin(\omega/2)$ change sign simultaneously. For planar PLT's, Eq. (3.15a) becomes

$$\begin{aligned}
&\cos(\omega/2) = \tfrac{1}{2}(\chi)^{1/2}, \\
&(\omega_\alpha/\omega)\sin(\omega/2) = (\chi)^{-1/2}(\mathscr{H}_\alpha + i\mathscr{E}_\alpha),
\end{aligned} \tag{3.15b}$$

which contains as a particular case the well-known relations for pure rotations, i.e.,

$$\begin{aligned}
&\cos(\omega/2) = \tfrac{1}{2}(1 + \chi_R)^{1/2}, \\
&(\omega_\alpha/\omega)\sin(\omega/2) = (1 + \chi_R)^{-1/2}\mathscr{H}_\alpha
\end{aligned} \tag{3.15c}$$

obtained by taking $\chi = (1 + \chi_R)$, $\mathscr{E} = 0$, and all the quantities to be real. In the case of the exceptional PLT's, we get

$$\omega_\alpha = \pm\pi q_\alpha, \quad \omega^2 = \omega_\alpha\omega_\alpha = \pi^2, \tag{3.15d}$$

where the q_α are determined (up to an ambiguity in sign) by Eq. (2.33). In this case, however, the ambiguity in the sign of ω_α is *nonremovable* as, even for the subgroup SO(3) itself, the rotations corresponding to $\pm\pi$ lie at the (diametrically) opposite ponits of the three-dimensional unit sphere in the well-known topological representation of SO(3). This would also mean that the exceptional PLT's do not lie in a "suffi-

ciently small" neighborhood of the identity and the one–one mapping implied by $L = \exp(I)$ is no longer valid. Lastly, Eqs. (3.14b) and (2.31) imply that, in the null case,

$$\omega_\alpha = (\mathscr{H}_\alpha + i\mathscr{E}_\alpha), \quad \omega^2 = \omega_\alpha \omega_\alpha = 0, \tag{3.15e}$$

where we may note that the Lie–Cartan parameters ω_α are again uniquely determined by the null PLT.

It may be observed that Eq. (3.12) incidentally shows that every PLT may be resolved into a commuting product of two unimodular complex orthogonal matrices. This may be contrasted with a similar resolution of a PLT into a commuting product of two Lorentz transformations discussed elsewhere.[1]

We now proceed to give a geometrical interpretation of the complex quantities ω and $\hat{\mathbf{n}}$, where the complex unit vector $\hat{\mathbf{n}}$ is defined by

$$n_\alpha = \omega_\alpha/\omega, \quad \omega \neq 0. \tag{3.16}$$

It is easy to see that the transformation

$$x' = qx\bar{q} \tag{3.17}$$

[and not $x' = qx\bar{q}^*$ as in Eq. (2.17)] sends the complex "pure quaternion"

$$x = \mathbf{x}\cdot\mathbf{e},$$

into another complex pure quaternion x' and hence leads to the following 3×3 complex orthogonal matrix

$$A_{\alpha\beta} = 2q_\alpha q_\beta - \delta_{\alpha\beta}(1 - 2q_4^2) - 2\epsilon_{\alpha\beta\gamma} q_\gamma q_4, \tag{3.18}$$

which corresponds to the given unit complex quaternion q, i.e.,

$$L \leftrightarrow \pm q \leftrightarrow A. \tag{3.19}$$

Substituting for q_i in Eq. (3.18) from Eq. (3.14a), we get (in the nonnull case)

$$A_{\alpha\beta} = n_\alpha n_\beta(1 - \cos\omega) + \delta_{\alpha\beta}\cos\omega + \epsilon_{\alpha\beta\gamma} n_\gamma \sin\omega, \tag{3.20}$$

which has precisely the same structure as a rotation matrix[1,9] with, however, complex elements. Thus the complex unit vector $\hat{\mathbf{n}}$ (which is an eigenvector of A belonging to the eigenvalue 1) may be interpreted as the "*complex axis*" of the "complex rotation" represented by A and the number ω may be interpreted as the "*complex angle of rotation*." One can also define, *ab initio* the angle of rotation θ in the usual manner by

$$\cos\theta = (\hat{\mathbf{x}}, A\hat{\mathbf{x}}) = A_{\alpha\beta} x_\alpha x_\beta, \tag{3.21}$$

where $\hat{\mathbf{x}} \equiv (x_\alpha)$ is a unit vector orthogonal to the axis $\hat{\mathbf{n}}$, and from Eq. (3.20) it follows that $\theta = \omega$. When $\omega = \pm\pi$, Eq. (3.20) reduces to

$$A_{\alpha\beta} = 2n_\alpha n_\beta - \delta_{\alpha\beta}, \tag{3.22}$$

corresponding to the (symmetric) exceptional complex rotation. The "null" complex rotation matrix is obtained by substituting for q_i from Eq. (3.14b) in Eq. (3.18); we get[1]

$$A_{\alpha\beta} = \delta_{\alpha\beta} + \tfrac{1}{2}\omega_\alpha\omega_\beta + \epsilon_{\alpha\beta\gamma}\omega_\gamma. \tag{3.23}$$

This (null) complex orthogonal matrix is characterized completely by the complex null vector $\boldsymbol{\omega}$, which is also an eigenvector of A belonging to the eigenvalue $+1$ [all the three eigenvalues of A given by Eq. (3.23) are equal to $+1$]. Hence we may interpret $\boldsymbol{\omega}$ as the (complex-null) axis of this null rotation. From Eqs. (3.21) and (3.23) we get the "angle of rotation" to be

$$\omega = 0. \tag{3.24}$$

The relations in Eqs. (3.20), (3.22), and (3.23), on inversion, respectively, yield the angle and axis of A as

$$\omega = \arccos\tfrac{1}{2}(\chi_A - 1), \quad n_\alpha = \tfrac{1}{2}\epsilon_{\alpha\beta\gamma} A_{\beta\gamma}/\sin\omega,$$
$$\chi_A \equiv \mathrm{tr}(A), \tag{3.25a}$$
$$\omega = \pm\pi, \quad n_\alpha = (A_{\mu\alpha} + \delta_{\mu\alpha})/2n_\mu,$$

$$n_\mu = \pm[(A_{\mu\mu} + 1)/2]^{1/2}, \tag{3.25b}$$

$$\omega = 0, \quad \omega_\alpha = \tfrac{1}{2}\epsilon_{\alpha\beta\gamma} A_{\beta\gamma}, \tag{3.25c}$$

where in (3.25b) μ is a fixed index not to be summed upon and $A_{\mu\alpha}$ are the elements of some row (the μth row) of A along which $A_{\mu\mu} \neq -1$. If so desired, the "angle" ω in Eq. (3.25a) may be chosen such that $\omega_r = \mathrm{Re}(\omega)$ lies in the range $0 \leqslant \omega_r < \pi$ and this fixes a particular "direction" for the (complex) axis $\hat{\mathbf{n}}$. Also, this choice evidently conforms to the one usually adopted[5] in the case of real nonexceptional rotations. In Eq. (3.25b), the ambiguous signs for the axis and angle is essential as already remarked. Regarding Eq. (3.25c), we may note that both the identity and null PLT's have $\omega = 0$. However, while the identity corresponds to $\omega_\alpha \equiv 0$, a null PLT has $\omega_\alpha \neq 0$. The formulas in Eqs. (3.25a,b) are complexified versions of the corresponding rotation group formulas and reduce to the same when A is taken to be real. The formula in Eq. (3.25c) for null complex rotations has no analog in SO(3).

Equation (3.18) may similarly be inverted to yield q_i in terms of $A_{\alpha\beta}$; we obtain

$$q_\alpha = -\tfrac{1}{4}\epsilon_{\alpha\beta\gamma} A_{\beta\gamma}/q_4, \quad q_4 = \tfrac{1}{2}(\chi_A + 1)^{1/2}, \tag{3.26a}$$

in the nonexceptional case and in the exceptional case,

$$q_\alpha = (A_{\mu\alpha} + \delta_{\mu\alpha})/2q_\mu, \quad q_\mu = \pm[(A_{\mu\mu} + 1)/2]^{1/2}, \tag{3.26b}$$

where μ is not a dummy and $A_{\mu\alpha}$ are the elements of any row of A along which the diagonal element $A_{\mu\mu}$ is not equal to -1. It may similarly be seen that Eqs. (3.26a,b) are complexified versions of the rotation group formulas in Eqs. (2.34) and (2.35) (with the new notation $R_{\alpha\beta} \equiv A_{\alpha\beta}$).

In passing, we consider one interesting application of the representation $L \to A$, or more directly the pure-quaternion transformation of Eq. (3.16), to the Lorentz transformation of an electromagnetic field. The electromagnetic field tensor F_{ij} may be represented by the (pure) field quaternion

$$f = (\mathbf{H} + i\mathbf{E})\cdot\mathbf{e} \equiv \mathbf{F}\cdot\mathbf{e}, \tag{3.27}$$

where $\mathbf{H}$ and $\mathbf{E}$ are, respectively, the magnetic and electric field intensities. A general pure boost q represented by Eq. (2.42) sends f to f', where

$$f' = \mathbf{F}'\cdot\mathbf{e} = qf\bar{q}$$
$$= [\gamma\mathbf{F} + w^{-2}(1 - \gamma)(\mathbf{w}\cdot\mathbf{F})\mathbf{w} - (i\gamma/c)(\mathbf{w}\times\mathbf{F})]\cdot\mathbf{e},$$

which yields immediately the well-known[11,14] transformation formulas

$$\mathbf{E}' = \gamma\mathbf{E} + w^{-2}(1-\gamma)(\mathbf{w}\cdot\mathbf{E})\mathbf{w} - \gamma c^{-1}(\mathbf{w}\times\mathbf{H}),$$
$$\mathbf{H}' = \gamma\mathbf{H} + w^{-2}(1-\gamma)(\mathbf{w}\cdot\mathbf{H})\mathbf{w} + \gamma c^{-1}(\mathbf{w}\times\mathbf{E}) \tag{3.28}$$

for the field vectors under an arbitrary boost.

IV. THE RESOLUTION OF A PROPER LORENTZ TRANSFORMATION INTO A PRODUCT OF THREE LORENTZ-SYNGE SCREWS AND COMPLEX EULER-BRAUER ANGLES

The complex orthogonal 3×3 matrix A corresponding to a proper Lorentz transformation L may be resolved into a product of three complex rotations in the coordinate planes as in the case of real rotations.[5] Here we may note, for future reference, that complex rotations in the three coordinate planes are given by

$$A_1(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix},$$
$$A_2(\theta) = \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix},$$

and

$$A_3(\theta) = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{4.1}$$

However, unlike real rotations, it is not possible (this would be evident from the discussion to follow) to cover the entire group SO(3,C) by a single factorization scheme like the familiar Euler product $A_3(\alpha)A_2(\beta)A_3(\gamma)$, and we need a minimum of three different factorization schemes to cover the entire group. In view of its obvious advantage in yielding a complexified version of the Wigner formula[13] for a subclass of Lorentz transformations (to be derived in the next section), we retain the above-mentioned Euler product and use in addition the Euler–Brauer products (see, for example, Murnaghan[15]) $A_3(\alpha)A_2(\beta)A_1(\gamma)$ and $A_3(\alpha)A_1(\beta)A_2(\gamma)$ for covering the entire group.

By an extension of the well-known procedure[15] adopted for real rotations to complex A, we observe that it is always possible to choose three complex angles (α,β,γ) such that A is factorizable into at least one of the three products mentioned above. More explicitly, we have the following results expressing (α,β,γ) in terms of the elements $A_{\mu\nu}$ of the matrix A:

Case (i) $A_{33}^2 \neq 1$ or $q_3^2 + q_4^2 \neq 1$ or 0

In this case, the Euler product resolution

$$A = A_3(\alpha)A_2(\beta)A_3(\gamma) \tag{4.2}$$

is certainly possible, and by a straightforward calculation we obtain

$$\cos\beta = A_{33}, \quad \sin\beta = (1 - A_{33}^2)^{1/2},$$
$$\cos\alpha = A_{13}(1-A_{33}^2)^{-1/2}, \quad \sin\alpha = -A_{23}(1-A_{33}^2)^{-1/2}, \tag{4.3}$$
$$\cos\gamma = A_{31}(1-A_{33}^2)^{-1/2}, \quad \sin\gamma = A_{32}(1-A_{33}^2)^{-1/2},$$

where any one particular branch of the function $(1-A_{33}^2)^{1/2}$ may be used. Using these formulas, it is not difficult to see that it is always possible to choose the complex Euler angles $\alpha = \alpha_r + i\alpha_b$, $\beta = \beta_r + i\beta_b$, and $\gamma = \gamma_r + i\gamma_b$ to lie in the ranges

$$\begin{aligned} &0 \leqslant \alpha_r < 2\pi, \quad 0 \leqslant \alpha_b < \infty, \\ &0 \leqslant \beta_r \leqslant \pi, \quad -\infty \leqslant \beta_b < \infty, \\ &0 \leqslant \gamma_r < 2\pi, \quad 0 < \gamma_b < \infty. \end{aligned} \tag{4.4}$$

These particular ranges have been chosen for (α,β,γ) so that when they are real, i.e., for real rotations, they reduce to the corresponding ranges of the real Euler angles. We also note for future reference that Eqs. (3.26) and (4.3) together imply the following relation between the UCQ components q_i and the Euler angles (α,β,γ):

$$\begin{aligned} q_4 &= \cos(\beta/2)\cos[(\alpha+\gamma)/2], \\ q_3 &= -\cos(\beta/2)\sin[(\alpha+\gamma)/2], \\ q_2 &= -\sin(\beta/2)\cos[(\alpha-\gamma)/2], \\ q_1 &= -\sin(\beta/2)\sin[(\alpha-\gamma)/2]. \end{aligned} \tag{4.5}$$

When (α,β,γ) are real, these are precisely the rotation-group formulas.[5] However, in the case of Lorentz transformations it must be observed that the above representation of q_i in terms of Euler angles is possible only for such PLT's for which $(q_3^2 + q_4^2)$ is neither 0 nor 1. Secondly, given (α,β,γ), Eq. (4.5) determines q_i to within an overall sign reflecting the nature of the correspondence $L \to \pm q$. However, in the exceptional case corresponding to $q_4 = 0$ (i.e., $\beta = \pi$ or $\alpha + \gamma = \pi$), the sign of q_α is undetermined and both signs are to be admitted for q_α in view of the fact that in the topological representation of the rotation group by a sphere of radius π in E_3 diametrically opposite points are to be identified.[13]

Evidently the Euler angles are undefined when $A_{33} = \pm 1$ [see Eq. (4.3)]. In the case of real orthogonal A, $A_{33} = \pm 1$, and hence $\beta = 0$, would imply $A_{13} = A_{23} = A_{32} = 0$, and an Euler resolution is possible, although the angles α and γ are not determined uniquely.[16] On the other hand, for a complex orthogonal A, $A_{33}^2 = 1$ does not necessarily imply the vanishing of the elements A_{13}, A_{31}, A_{23}, and A_{32}. For example,

$$A = \begin{bmatrix} 1 & i-1 & 1+i \\ 1+i & -1 & i-1 \\ 1-i & i+1 & 1 \end{bmatrix}$$

is one such complex orthogonal matrix and in such cases an Euler resolution in the form $A = A_3A_2A_3$ is *impossible*.

However, when $A_{33}^2 = 1$, at least one of the two complex numbers A_{32}^2 and A_{31}^2 is not equal to 1 in view of the relation $A_{31}^2 + A_{32}^2 + A_{33}^2 = 1$, and we thus consider separately the following two cases:

Case (iia): $A_{33}^2 = 1$, $A_{32}^2 \neq 1$; or $(q_3q_2 + q_1q_4)^2 \neq \frac{1}{4}$, $q_3^2 + q_4^2 = 1$, or 0;

Case (iib): $A_{33}^2 = 1$, $A_{31}^2 \neq 1$; or $(q_3q_1 - q_2q_4)^2 \neq \frac{1}{4}$, $q_3^2 + q_4^2 = 1$, or 0

In these cases we have to adopt other resolution schemes.

In case (iia), we use the Euler–Brauer scheme[15]

$$A = A_3(\alpha)A_1(\beta)A_2(\gamma), \tag{4.6}$$

and, as before, we find after some simple algebra

$$\begin{aligned}
&\sin\beta = -A_{32}, \quad \cos\beta = (1-A_{32}^2)^{1/2},\\
&\sin\alpha = A_{12}(1-A_{32}^2)^{-1/2}, \quad \cos\alpha = A_{22}(1-A_{32}^2)^{-1/2}\\
&\sin\gamma = -A_{31}(1-A_{32}^2)^{-1/2} \quad \cos\gamma = A_{33}(1-A_{32}^2)^{-1/2},
\end{aligned} \tag{4.7}$$

and it is always possible to determine the branches of the square roots such that the complex Euler–Brauer angles (α,β,γ) lie in the ranges

$$\begin{aligned}
&0\leqslant\alpha_r\leqslant\pi, \quad -\infty<\alpha_b<\infty,\\
&0\leqslant\beta_r<2\pi, \quad 0\leqslant\beta_b<\infty,\\
&0\leqslant\gamma_r<2\pi, \quad 0\leqslant\gamma_b<\infty.
\end{aligned} \tag{4.8}$$

In case (iib), we may use the other Euler–Brauer scheme

$$A = A_3(\alpha)A_2(\beta)A_1(\gamma) \tag{4.9}$$

with the complex Euler–Brauer angles (α,β,γ) determined from

$$\begin{aligned}
&\sin\beta = A_{31}, \quad \cos\beta = (1-A_{31}^2)^{1/2},\\
&\sin\alpha = -A_{21}(1-A_{31}^2)^{-1/2}, \quad \cos\alpha = A_{11}(1-A_{31}^2)^{-1/2},\\
&\sin\gamma = -A_{32}(1-A_{31}^2)^{-1/2}, \quad \cos\gamma = A_{33}(1-A_{31}^2)^{-1/2},
\end{aligned} \tag{4.10}$$

where, as before, it is always possible to choose the branches of the square-roots involved such that the angles (α,β,γ) lie in the ranges given in Eq. (4.8).

Having shown that it is always possible to resolve a 3×3 complex orthogonal matrix A in at least one of the three factorization schemes discussed above, we now interpret these factorizations directly in terms of the (4×4) Lorentz matrix L that corresponds to A. The UCQ's corresponding to the special complex rotations of Eq. (4.1) are [see Eq. (3.26a)] given by

$$\begin{aligned}
&q_4 = \cos(\theta/2), \quad q_1 = -\sin(\theta/2), \quad q_2 = q_3 = 0,\\
&q_4 = \cos(\theta/2), \quad q_2 = -\sin(\theta/2), \quad q_1 = q_3 = 0,\\
&q_4 = \cos(\theta/2), \quad q_3 = -\sin(\theta/2), \quad q_1 = q_2 = 0.
\end{aligned} \tag{4.11}$$

Using these in Eq. (2.18), we get the corresponding PLT's to be, respectively,

$$\begin{aligned}
&S_1(\theta_r,\theta_b) = L^{(14)}(\theta_b)L^{(23)}(\theta_r) = L^{(23)}(\theta_r)L^{(14)}(\theta_b),\\
&S_2(\theta_r,\theta_b) = L^{(24)}(\theta_b)L^{(31)}(\theta_r) = L^{(31)}(\theta_r)L^{(24)}(\theta_b),\\
&S_3(\theta_r,\theta_b) = L^{(34)}(\theta_b)L^{(12)}(\theta_r) = L^{(12)}(\theta_r)L^{(34)}(\theta_b),
\end{aligned} \tag{4.12}$$

where $L^{(\alpha\beta)}(\theta_r)$ is a pure rotation [see Eq. (2.19)] through θ_r in the x^α–x^β plane and $L^{(\alpha4)}(\theta_b)$ is a pure boost along the x^α direction with velocity $w = c\tanh(\theta_b)$. The proper Lorentz transformation $S_\alpha(\theta_r,\theta_b)$ is a Lorentz–Synge 4-screw along the x^α direction (see Synge[2]). Equations (3.26), (4.11), and (4.12) imply, for example, for the screw

$$S_3(\theta_r,\theta_b) = \begin{bmatrix} \cos\theta_r & \sin\theta_r & 0 & 0\\ -\sin\theta_r & \cos\theta_r & 0 & 0\\ 0 & 0 & \cosh\theta_b & i\sinh\theta_b\\ 0 & 0 & -i\sinh\theta_b & \cosh\theta_b \end{bmatrix}, \tag{4.13}$$

the mappings

$$S_3(\theta_r,\theta_b)\leftrightarrow\pm[\cos(\theta/2) - e_3\sin(\theta/2)]\leftrightarrow A_3(\theta_r+i\theta_b), \tag{4.14}$$

where

$$A_3(\theta_r+i\theta_b) = \begin{bmatrix} \cos(\theta_r+i\theta_b) & \sin(\theta_r+i\theta_b) & 0\\ -\sin(\theta_r+i\theta_b) & \cos(\theta_r+i\theta_b) & 0\\ 0 & 0 & 1 \end{bmatrix}. \tag{4.15}$$

Thus the three factorization schemes of A given in Eq. (4.2), (4.6), and (4.9) correspond to the following screw-resolution schemes of the corresponding proper Lorentz transformation L.

(*i*) *Screw resolution in the Euler scheme*:

$$L = S_3(\alpha_r,\alpha_b)S_2(\beta_r,\beta_b)S_3(\gamma_r,\gamma_b) \tag{4.16a}$$

$$= L^{(34)}(\alpha_b)L^{(12)}(\alpha_r)L^{(24)}(\beta_b)L^{(31)}(\beta_r)L^{(34)}(\gamma_b)L^{(12)}(\gamma_r). \tag{4.16b}$$

(*ii*) *Screw resolutions in the Euler–Brauer scheme*:

$$L = S_3(\alpha_r,\alpha_b)S_1(\beta_r,\beta_b)S_2(\gamma_r,\gamma_b) \tag{4.17a}$$

$$= L^{(34)}(\alpha_b)L^{(12)}(\alpha_r)L^{(14)}(\beta_b)L^{(23)}(\beta_r)L^{(24)}(\gamma_b)L^{(31)}(\gamma_r), \tag{4.17b}$$

$$L = S_3(\alpha_r,\alpha_b)S_2(\beta_r,\beta_b)S_1(\gamma_r,\gamma_b) \tag{4.18a}$$

$$= L^{(34)}(\alpha_b)L^{(12)}(\alpha_r)L^{(24)}(\beta_b)L^{(31)}(\beta_r)L^{(14)}(\gamma_b)L^{(23)}(\gamma_r). \tag{4.18b}$$

We may note that the screw resolution in Eq. (4.16a) involves screws in *two* coordinate directions only whereas those in Eqs. (4.17a) and (4.18a) involve screws in all the three coordinate directions. It is also interesting to note the particular ordering of the six planar Lorentz transformations in Eqs. (4.16b), (4.17b), and (4.18b). A direct proof of these Euler resolutions appears to be difficult.

V. THE CHARACTER FORMULA FOR SO(3,1) IN ITS $D^{j'}$ REPRESENTATION IN TERMS OF THE COMPLEX ANGLE ω

It is well known[9,17] that the D^{j0} repreentation of the complex unimodular group SL(2,C) can be obtained formally from the D^j representation of the special unitary group SU(2), namely,[13]

$$\begin{aligned}
D^j_{m'm}(a,b) &= \sum_n \frac{[(j+m)!(j-m)!(j+m')!(j-m')!]^{1/2}}{(j+m-n)!(j-m'-n)!(m'-m+n)!}\\
&\quad\times a^{j+m-n}a^{*j-m'-n}b^n(-b^*)^{m'-m-n},
\end{aligned} \tag{5.1}$$

by simply replacing the SU(2) matrix $\begin{pmatrix} a & b\\ -b^* & a^* \end{pmatrix}$ by the SL(2,C) matrix $\begin{pmatrix} a & b\\ c & d \end{pmatrix}$. We thus have

$$\begin{aligned}
D^{j0}_{m'm}(a,b,c,d) &= \sum_n \frac{[(j+m)!(j-m)!(j+m')!(j-m')!]^{1/2}}{(j+m-n)!(j-m'-n)!(m'-m+n)!}\\
&\quad\times a^{j+m-n}b^n c^{m'-m-n}d^{j-m'-n}
\end{aligned} \tag{5.2}$$

in terms of the four complex SL(2,C) parameters a, b, c and d [satisfying $(ad-bc)=1$] also known as the generalized Cayley–Klein parameters.[18] We also know that, on expressing the Cayley–Klein parameters (a,b) in terms of the (real) Euler angles (α,β,γ) by

$$a = \cos(\beta/2)\exp[i(\alpha+\gamma)/2],$$
$$b = \sin(\beta/2)\exp[i(\alpha-\gamma)/2], \tag{5.3}$$

Eq. (5.1) becomes the *Wigner formula*[13] for the D^j representation of SO(3). Similarly by expressing (a,b) in terms of the axis $\hat{\mathbf{n}}$ and angle ω of rotation by

$$\begin{aligned} a &= \cos(\omega/2) + in_3\sin(\omega/2) \\ &= \cos(\omega/2) + i\cos\theta\sin(\omega/2), \\ b &= (n_2 + in_1)\sin(\omega/2) = i\sin\theta\sin(\omega/2)\exp(-i\phi), \end{aligned} \tag{5.4}$$

Eq. (5.1) becomes the *Carmeli formula*[9] for the D^j representation of SO(3), where $\hat{\mathbf{n}} = (n_1,n_2,n_3) = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$. These results suggest that we may similarly obtain formulas analogous to those of Wigner and Carmeli for the proper Lorentz group SO(3,1), by expressing the generalized Cayley–Klein parameters (a,b,c,d) in Eq. (5.2) in terms of the complex Euler angles and the complex angle-axis parameters obtained in the previous sections.

On using the two-dimensional irreducible representation $e_\alpha \to -i\sigma_\alpha$, for the quaternion units e_α (generating the quaternion ring) in terms of the Pauli matrices given by Eq. (2.37), we obtain the $D^{1/2\,0}$ or SL(2,C) representation of SO(3,1), namely,

$$\begin{aligned} &L \leftrightarrow \pm q \to \pm(-i\sigma_\alpha q_\alpha + q_4) \\ &\equiv \pm M = \pm\begin{bmatrix} q_4 - iq_3 & -q_2 - iq_1 \\ q_2 - iq_1 & q_4 + iq_3 \end{bmatrix}. \end{aligned} \tag{5.5}$$

We also observe that the mapping $L \to \pm q^*$; on the other hand, yields the matrix

$$M' = (-i\sigma_\alpha q_\alpha^* + q_4^*) = \sigma_2 M^* \sigma_2^{-1}$$

so that the latter mapping yields the in equivalent complex conjugate $D^{0\,1/2}$ representation. From Eq. (5.5) we may now identity the generalized Cayley–Klein parameters as

$$\begin{aligned} &a = q_4 - iq_3, \quad b = -q_2 - iq_1, \quad c = q_2 - iq_1, \\ &d = q_4 + iq_3. \end{aligned} \tag{5.6}$$

Using the expressions for q_i given by Eqs. (3.14a) and (3.16) in these, we obtain for *nonnull* PLT's the following relations between (a,b,c,d) and the complex axis-angle parameters $(\hat{\mathbf{n}},\omega)$:

$$\begin{aligned} &a = \cos(\omega/2) + in_3\sin(\omega/2), \quad b = (n_2 + in_1)\sin(\omega/2), \\ &c = -(n_2 - in_1)\sin(\omega/2), \quad d = \cos(\omega/2) - in_3\sin(\omega/2). \end{aligned} \tag{5.7}$$

For null PLT's, we have similarly, from Eqs. (3.14b) and (5.6),

$$\begin{aligned} &a = (1 + \tfrac{1}{2}i\omega_3), \quad b = \tfrac{1}{2}(\omega_2 + i\omega_1), \\ &c = -\tfrac{1}{2}(\omega_2 - i\omega_1), \quad d = (1 - \tfrac{1}{2}i\omega_3), \end{aligned} \tag{5.8}$$

where we recall that $\boldsymbol{\omega}$ is a null vector, i.e., $\boldsymbol{\omega}\cdot\boldsymbol{\omega} = 0$. We may also note here that when $(\hat{\mathbf{n}},\omega)$ are real, i.e., for pure rotations, Eq. (5.8) reduces to $a = 1 = d$, $b = c = 0$ corresponding to the identity whereas Eq. (5.7) reduces to Eq. (5.4) of pure rotations.

Since the D^{0j_1} representation of SL(2,C) is equivalent to the complex conjugate representation $D^{j_1 0^*}$ of $D^{j_1 0}$ and since $D^{j_1 j_2} = D^{j_1 0} \times D^{0 j_2}$, we have (see, for example, Schwartz[19])

$$D^{j_1 j_2}_{m_1' m_2'; m_1 m_2} = D^{j_1 0}_{m_1' m_1} D^{j_2 0^*}_{m_2' m_2}, \tag{5.9}$$

where the rows and columns of the (Konecker product) matrix $D^{j_1 j_2}$ are labelled by pairs of indices so that, for example, $D^{j_1 j_2}_{m_1' m_2'; m_1 m_2}$ is the element of $D^{j_1 j_2}$ occurring in the $(m_1' m_2')$th row and the $(m_1 m_2)$th column. Substituting Eqs. (5.7) and (5.8) in Eq. (5.2) yields the $D^{j_1 0}$ (and $D^{j_2 0}$) representation of SO(3,1) respectively for the nonnull and null PLT's. Using the $D^{j_1 0}$ and $D^{j_2 0}$ so obtained in Eq. (5.9) then yields the desired $D^{j_1 j_2}$ representation of SO(3,1) expressed in terms of the complex axis-angle parameter $(\hat{\mathbf{n}},\omega)$ for nonnull PLT's and in terms of the complex null axis $\boldsymbol{\omega}$ for null PLT's. While the representation formula for nonnull PLT's obtained in this manner involves a D^{j0} given by the complexified version of the Carmeli formula of SO(3), the formula for null PLT's has no analog in SO(3). Further, as is done with real rotations (see Carmeli[9]), we may introduce a pair of complex angles (θ,ϕ) to represent the complex unit vector $\hat{\mathbf{n}}$ by $\hat{\mathbf{n}} = (n_1,n_2,n_3) = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$ thus bringing the formula for the D^{j0} representation for nonnull PLT's to a form which is identical to the D^j representation formula of Carmeli[9] except that now the angles (θ,ϕ) are complex instead of being real. Similarly, it is possible to represent a null $\boldsymbol{\omega}$ by a pair of complex angles, but this representation is not very convenient as a single set of complex angles giving all the null complex vectors $\boldsymbol{\omega}$ does not exist.

Similarly, on expressing (a,b,c,d) in terms of the complex Euler angles (α,β,γ), using Eqs. (4.5) and (5.6), we obtain

$$\begin{aligned} &a = \cos(\beta/2)\exp[i(\alpha+\gamma)/2], \\ &b = \sin(\beta/2)\exp[i(\alpha-\gamma)/2], \\ &c = -\sin(\beta/2)\exp[-i(\alpha-\gamma)/2], \\ &d = \cos(\beta/2)\exp[-i(\alpha+\gamma)/2], \end{aligned} \tag{5.10}$$

which through Eqs. (5.2) and (5.9) leads to a formula for the $D^{j_1 j_2}$ representation of SO(3,1) involving only a complexified version of the Wigner formula for the D^j representation of SO(3). Evidently, *this formula alone does not cover the entire group SO* (3,1) as the complex Euler angles given by Eq. (4.5) can be defined only for a subclass of PLT's. Moreover, this formula, unlike the one involving the complexified Carmeli formula of SO(3), does not completely cover even the nonnull PLT's *in toto*. And, as it is necessary to use the other two Euler–Brauer schemes in addition, to cover the entire group (see Sec. IV), the advantage of using the diagonal representation matrices for two Z rotations is now not available. It thus appears that a *single* formula for the $D^{jj'}$ representation covering the entire group SO(3,1) would only be in terms of the quaternion parameters which are explicitly expressed in terms of the elements L_{ij} of L as in Sec. II.

We next proceed to obtain the character formula for the proper Lorentz group in its $D^{jj'}$ representation. First, we prove that in the case of nonnull PLT's

$$\chi^{jj'}(L) = \chi^{jj'}(\omega) = \frac{\sin[(j+\frac{1}{2})\omega]\,\sin[(j'+\frac{1}{2})\omega^*]}{\sin(\omega/2)\sin(\omega^*/2)}. \tag{5.11}$$

The proof is by induction. In the first place, we note that the

above formula yields $\chi^{00} = 1$, as it ought to, for the trivial identity representation D^{00}. Secondly, it gives $\chi^{1/2\,0} = \sin\omega/\sin(\omega/2) = 2\cos(\omega/2)$, which we know is also correct by virtue of Eq. (5.7) yielding the $D^{1/2\,0}$ representation $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. We now assume the formula (5.11) to be true for the $D^{j-1/2,0}$ and D^{j0} representations and prove its validity for the $D^{j+1/2,0}$ representation. From the Clebsch–Gordon theorem,[10] we have

$$D^{j0} \times D^{1/2\,0} = D^{j+1/2,0} + D^{j-1/2,0}, \tag{5.12}$$

so that we must have

$$\begin{aligned} \chi^{j+1/2,0} &= \chi^{j0}\chi^{1/2\,0} - \chi^{j-1/2,0} \\ &= \frac{\sin[(j+\tfrac{1}{2})\omega]\,2\cos(\omega/2)}{\sin(\omega/2)} - \frac{\sin(j\omega)}{\sin(\omega/2)} \\ &= \frac{\sin[(j+1)\omega]}{\sin(\omega/2)}. \end{aligned} \tag{5.13}$$

The result of Eq. (5.11) is thus true for $D^{j+1/2,0}$ and by induction the assertion is proved. Since $D^{0j'}$ is merely equivalent to the complex conjugate of $D^{j'0}$ and $D^{jj'} = D^{j0} \times D^{0j'}$, the result of Eq. (5.11) now follows for all nonnull PLT's. It may be observed that the character formula χ^{j0} for the D^{j0} representation of the nonnull PLT's is obtained by a formal complexification of the character formula for the D^{j} representation of SO(3).

Clearly the character formula of Eq. (5.11) yields on passing to the limit $\omega \to 0$

$$\chi^{jj'}(L) = (2j+1)(2j'+1), \tag{5.14}$$

which is the character formula for null PLT's in the $D^{jj'}$ representation. This result, as in the nonnull case, may also be proved independently by induction.

Note added in proof: To avoid misunderstanding, we wish to state that the proper Lorentz group SO(3,1) is identical with the *proper orthochronous Lorentz group*.

ACKNOWLEDGMENTS

One of us (B.S.N.) is grateful to the UGC (India) for the award of a fellowship under the Faculty Improvement Programme and to Professor B. Sanjeevaiah for facilities and encouragement.

[1] K. N. Srinivasa Rao, D. Saroja, and A. V. Gopala Rao, J. Math. Phys. **22**, 2167–79 (1981).
[2] J. L. Synge, *Relativity: The Special Theory*, 2nd ed. (North-Holland, Amsterdam, 1972), pp. 89–94.
[3] E. P. Wigner, Ann. Math. **40**, 149–204 (1939).
[4] J. L. Synge, *Quaternions, Lorentz Transformations and the Conway–Dirac–Eddington Matrices* (Communications of the Dublin Institute for Advanced Studies, Dublin, 1972), Ser. A, No. 21, pp. 1–67.
[5] J. L. Synge, "Classical Dynamics" in *Encyclopaedia of Physics*, edited by S. Flugge (Springer, Berlin, 1960), Vol. III/1, pp. 15–25.
[6] H. Weyl, *The Classical Groups* (Princeton U.P., Princeton, New Jersey, 1939), p. 56.
[7] A. O. Barut, *Electrodynamics and Classical Theory of Fields and Particles* (MacMillan, New York, 1964), p. 27.
[8] W. Ruhl, *The Lorentz Group and Harmonic Analysis* (Benjamin, New York, 1970), as quoted in Ref. 9.
[9] M. Carmeli, *Group Theory and General Relativity* (McGraw-Hill, New York, 1977), pp. 7–12, 34–8, 49–50.
[10] I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and Their Applications* (Pergamon, New York, 1963), p. 13.
[11] C. Møller, *The Theory of Relativity* (Oxford U.P., London, 1972), pp. 110, 120.
[12] N. V. Efimov and E. R. Rozendorn, *Linear Algebra and Multidimensional Geometry* (Mir Publishers, Moscow, 1975), pp. 402–16.
[13] M. Hameresh, *Group Theory* (Addison-Wesely, Reading, Mass., 1962), pp. 308, 328, 354–5.
[14] J. L. Anderson, *Principles of Relativitiy Physics* (Academic, New York, 1967), p. 262.
[15] F. D. Murnaghan, *The Theory of Group Representations* (The Johns Hopkins Press, Baltimore, 1949), pp. 297, 320.
[16] E. P. Wigner, *Group Theory* (Academic, New York, 1959), p. 152.
[17] V. I. Smirnov, *Linear Algebra* (Pergamon, New York, 1964), p. 286.
[18] V. A. Fock, *Fundamentals of Quantum Mechanics* (Mir Publishers, Moscow, 1978), p. 293.
[19] H. M. Schwartz, *Introduction to Special Relativity* (McGraw-Hill, New York, 1968), p. 381.

Clebsch–Gordan coefficients for $SU(5) \supset SU(3) \times SU(2) \times U(1)$ theories[a)]

In-Gyu Koh,[b)] J. Patera,[c)] and C. Rousseau[d)]
Centre de Recherche de Mathématiques Appliquées, Université de Montréal, Montréal H3C 3J7, Québec, Canada

(Received 7 June 1982; accepted for publication 14 January 1983)

The Clebsch–Gordan coefficients are calculated for the following tensor products of SU(5) representations: $5 \otimes 5$, $5 \otimes 10$, $5 \otimes \overline{10}$, $15 \otimes 5$, $15 \otimes \bar{5}$, $10 \otimes 10$, $5 \otimes \bar{5}$, $10 \otimes \overline{10}$, $5 \otimes 24$, and $10 \otimes 24$. Each case is calculated twice: once in a weight vector basis independent of any semisimple subgroup, the second time in a basis which refers to $SU(3) \times SU(2) \times U(1) \subset SU(5)$.

PACS numbers: 02.20. + b, 11.30.Ly, 12.90. + b

1. INTRODUCTION

The purpose of this paper is to provide the Clebsch–Gordan coefficients (CGC) frequently needed in models of unification of the weak and electromagnetic interactions with the strong ones based on the group-subgroup pair $SU(5) \supset SU(3) \times SU(2) \times U(1)$. More precisely, we consider the products of SU(5) representations shown in Table I. Standard notations for the representations and their reductions (branching rules) to representations of the subgroup are shown in Table II.

During the last 20 years many papers were devoted to CGC for finite-dimensional representations of semisimple Lie groups (cf. Ref. 1 and references therein). In connection with developing the $SU(5) \supset SU(3) \times SU(2) \times U(1)$ model a few CGC were calculated in the original papers.[2] A critical review of the literature is a major task, which we do not undertake here. Nevertheless, one hardly avoids the conclusion that there are no satisfactory methods for the needs, for instance, of particle physics with its ever changing emphasis on different groups, subgroups, and representations.[3] Our aim here is more than just to recalculate the old CGC and to find some new ones. By this computation we would like to demonstrate a method that is both practical and versatile. It hinges on the known fact that each representation space can be generated from the highest weight vector by successive application of generators of the Lie algebra.

In this article we choose an important particular case instead of dealing with the method in its full generality. It allows us to point out general properties as observations made on specific cases without providing abstract proofs. The general properties underlying our computations are described elsewhere.[4,5] An earlier version of the method is in Ref. 6.

There are three problems one faces when computing CGC. The first is the choice of the basis for each space. Although we consider only a particular case, our method is applicable to representations of any semisimple or reductive Lie group–subgroup pair, which is clear already from the context of this paper and which will be illustrated also by subsequent papers. The choice made here involves a standard set of generators for each Lie algebra related to the roots of the algebra. Each representation space is viewed as a direct sum of mutually orthogonal subspaces labeled by weights (additive quantum numbers) of the representation. There exists a unique highest weight for each irreducible representation. It is used to identify the representation. The weight subspaces of the simplest representations are all one-dimensional so that weights specify orthonormal basis vectors. This is the case, for instance, for all irreducible representations of SU(2) or the fundamental representations of SU(5) of dimension 5 and 10 and also for all SU(5) representations with the highest weight $(k,0,0,0)$, $k = 1,2,\ldots$. Starting from products of these representations, we *define* complete bases in spaces of higher representations as products of basis vectors from the simplest ones and their linear combinations. It turns out to be a satisfactory procedure for all cases of practical interest at present, not only those considered here. Practically, we rarely need to use this construction explicitly; its purpose is to provide a well-defined basis in each space. The overall phases in irreducible subspaces are chosen by simplicity only.

The second problem one faces is due to the frequent preference in physics given to a particular subgroup of the

TABLE I. Identification of tables of SU(5) Clebsch–Gordan coefficients and tensor products of representations. For further properties of these representations, see also Table II.

Tensor product of SU(5) representations	Reduction of the product	Table of CG coefficients SU(5)-basis	Subgroup basis
$5 \otimes 5$	$15 \oplus 10$	Table IV	Table IV′
$5 \otimes 10$	$40 \oplus \overline{10}$	Table V	Table V′
$5 \otimes \overline{10}$	$45 \oplus \bar{5}$	Table VI	Table VI′
$15 \otimes 5$	$35 \oplus 40$	Table VII	Table VII′
$15 \otimes \bar{5}$	$70 \oplus 5$	Table VIII	Table VIII′
$10 \otimes 10$	$50 \oplus 45 \oplus \bar{5}$	Table IX	Table IX′
$5 \otimes \bar{5}$	$24 \oplus 1$	Table X	Table X′
$10 \otimes \overline{10}$	$75 \oplus 24 \oplus 1$	Table XI	Table XI′
$5 \otimes 24$	$70 \oplus \overline{45} \oplus 5$	Table XII	Table XII′
$10 \otimes 24$	$175 \oplus \overline{40} \oplus 15 \oplus 10$	Table XIII	Table XIII′

[a)] Work supported in part by the National Research Council of Canada, Ministère de l'Education du Québec and by the Korea Trader's Foundation and the Korea Science and Engineering Foundation.

[b)] Permanent address: Dept. of Physics, Sogang University, Seoul, C.P.O. Box 1142, Korea.
[c)] Present address: Caltech, Pasadena, California 91125.
[d)] Dept. de Mathématiques, Université de Montréal, Montréal, Québec, Canada.

 0022-2488/83/081955-13$02.50

TABLE II. Relevant irreducible representations of SU(5) and their reduction to representations of SU(3)×SU(2)×U(1). Analogous results for SU(5) representations $\bar{5}$, $\overline{10}$, $\overline{15}$, $\overline{35}$, $\overline{40}$, $\overline{45}$, $\overline{50}$, $\overline{70}$, and $\overline{175}$ are obtained by (i) permuting the SU(5) representation labels $(a\ b\ c\ d) \rightarrow (d\ c\ b\ a)$, (ii) permuting the SU(3) labels $(p\ q) \rightarrow (q\ p)$, and (iii) changing the sign of the U(1) label.

Dimension	Highest weight	Reduction to SU(3)×SU(2)×U(1) representation.
5	(1000)	$(10)(0)(\bar{2}) \oplus (00)(1)(\bar{3})$
10	(0100)	$(10)(1)(\bar{1}) \oplus (01)(0)(4) \oplus (00)(0)(\bar{6})$
15	(2000)	$(20)(0)(4) \oplus (10)(1)(\bar{1}) \oplus (00)(2)(\bar{6})$
24	(1001)	$(11)(0)(0) \oplus (10)(1)(1) \oplus (01)(1)(\bar{1}) \oplus (00)(2)(0) \oplus (00)(0)(0)$
35	(3000)	$(30)(0)(6) \oplus (20)(1)(1) \oplus (10)(2)(\bar{4}) \oplus (00)(3)(\bar{9})$
40	(1100)	$(20)(1)(1) \oplus (11)(0)(6) \oplus (10)(2)(\bar{4}) \oplus (01)(1)(1) \oplus (10)(0)(\bar{4}) \oplus (00)(1)(\bar{9})$
45	(1010)	$(11)(1)(3) \oplus (20)(0)(\bar{2}) \oplus (01)(2)(\bar{2}) \oplus (10)(1)(\bar{7}) \oplus (10)(0)(8) \oplus (01)(0)(\bar{2}) \oplus (00)(1)(3)$
50	(0200)	$(20)(2)(\bar{2}) \oplus (11)(1)(3) \oplus (02)(0)(8) \oplus (10)(1)(\bar{7}) \oplus (01)(0)(\bar{2}) \oplus (00)(0)(\overline{12})$
70	(2001)	$(21)(0)(2) \oplus (20)(1)(7) \oplus (11)(1)(\bar{3}) \oplus (01)(2)(\bar{8}) \oplus (10)(2)(\bar{2}) \oplus (00)(3)(\bar{3}) \oplus (10)(0)(2) \oplus (00)(1)(\bar{3})$
75	(0110)	$(11)(2)(0) \oplus (02)(1)(5) \oplus (20)(1)(\bar{5}) \oplus (11)(0)(0) \oplus (01)(1)(\bar{5}) \oplus (10)(1)(5) \oplus (10)(0)(\overline{10}) \oplus (01)(0)(10) \oplus (00)(0)(0)$
175	(1101)	$(21)(1)(\bar{1}) \oplus (20)(2)(4) \oplus (12)(0)(4) \oplus (11)(2)(\bar{6}) \oplus (11)(1)(9) \oplus (10)(3)(\bar{1}) \oplus (02)(1)(\bar{1}) \oplus (20)(0)(4) \oplus (01)(2)(4) \oplus (11)(0)(\bar{6}) \oplus (01)(1)(\overline{11}) \oplus 2(10)(1)(\bar{1}) \oplus (00)(2)(\bar{6}) \oplus (01)(0)(4) \oplus (00)(0)(\bar{6})$

given group, for instance, SU(3)×SU(2)×U(1) in SU(5). The weights (quantum numbers) used to specify basis vectors should be those related to the subgroup rather than the group. In principle, a subgroup can be inserted in the group in many equivalent ways. However, for simplicity of our computation, a particular insertion is an imperative. Namely, all generators of the subgroup which lower (raise) the weights of a weight vector must be linear combinations of group generators which lower (raise) to group weights. Practically, the insertion is specified by a projection matrix transforming any group weight into a subgroup weight. A construction of these matrices is described in Ref. 7, and many of them are listed in Table IV of Ref. 8.

The last (not least) problem is that of the large number of CGC. The CGC corresponding to the tensor product

$$\Lambda \otimes \Lambda' = \bigotimes_{i=1}^{s} \Lambda_i \tag{1}$$

of representations Λ and Λ' are merely transformation coefficients from the product basis to the basis in the reduced form. Instead of calculating and tabulating all the coefficients, it suffices[5] to find and list a small subset of them which is representative of all the others. More precisely, we use a finite subgroup N of SU(5) or SU(3)×SU(2)×U(1) (depending on the actual basis) which permutes the weight subspaces, suitably chosen basis vectors, and also CGC, and which allows us to produce any other basis vector, or CGC, from those listed here.

Section 2 contains preliminaries indispensable for the SU(5)⊃SU(3)×SU(2)×U(1) case. For a general case one should consult Refs. 4 and 5. In Sec. 3, CGC tables are presented. The coefficients here are calculated relative to a basis independent of the subgroup. Section 4 contains CGC calculated in the subgroup basis. It is explained how to relate any particular CGC one may need to an entry in the tables.

For convenience we adopt the convention of writing a minus sign as a bar over the corresponding integer in all matrixlike symbols.

2. MATHEMATICAL PRELIMINARIES

The immediate motivation for our own computation is the standard assignment of one family of 15 fermions to a reducible representation of SU(5). It is summarized in Table

TABLE III. Assignment of basic fermions to SU(5) and SU(3)×SU(2)×U(1) representation spaces. Each irreducible representation is denoted by its highest weight. The entries are ordered according to SU(5) weights. For convenience negative signs are shown as bars over the cipher.

Fermion	Electric charge	SU(3)×SU(2)×U(1) Weight	SU(3)×SU(2)×U(1) Representation	SU(5) Weight	SU(5) Representation
d_y^c	1/3	$(01)(0)(\bar{2})$	$(01)(0)(\bar{2})$	(0001)	(0001)
ν_c	0	$(00)(1)(3)$	$(00)(1)(3)$	$(001\bar{1})$	
d_b^c	1/3	$(1\bar{1})(0)(\bar{2})$	$(01)(0)(\bar{2})$	$(01\bar{1}0)$	
e	−1	$(00)(\bar{1})(3)$	$(00)(1)(3)$	$(1\bar{1}00)$	
d_r^c	1/3	$(\bar{1}0)(0)(\bar{2})$	$(01)(0)(\bar{2})$	$(\bar{1}000)$	
u_r	2/3	$(10)(1)(\bar{1})$	$(10)(1)(\bar{1})$	(0100)	(0100)
u_y^c	−2/3	$(01)(0)(4)$	$(01)(0)(4)$	$(1\bar{1}10)$	
u_b	2/3	$(\bar{1}1)(1)(\bar{1})$	$(10)(1)(\bar{1})$	$(\bar{1}010)$	
d_r	−1/3	$(10)(\bar{1})(\bar{1})$	$(10)(1)(\bar{1})$	$(10\bar{1}1)$	
e^c	1	$(00)(0)(\bar{6})$	$(00)(0)(\bar{6})$	$(\bar{1}1\bar{1}1)$	
u_b^c	−2/3	$(1\bar{1})(0)(4)$	$(01)(0)(4)$	$(100\bar{1})$	
u_y	2/3	$(0\bar{1})(1)(\bar{1})$	$(10)(1)(\bar{1})$	$(\bar{1}10\bar{1})$	
d_b	−1/3	$(\bar{1}1)(\bar{1})(\bar{1})$	$(10)(1)(\bar{1})$	$(010\bar{1})$	
u_r^c	−2/3	$(\bar{1}0)(0)(4)$	$(01)(0)(4)$	$(0\bar{1}1\bar{1})$	
d_y	−1/3	$(0\bar{1})(\bar{1})(\bar{1})$	$(10)(1)(\bar{1})$	$(00\bar{1}0)$	

III. Note the one-to-one correspondence between weights and the fermions.

A. SU(n) Lie algebras and roots

Let us describe the Lie algebras of SU(n), $n = 2,3,5$, involved in our computation and also their action on the fermions of Table III.

An SU(n) has $n^2 - 1$ generators. For our purpose the following choice is by far the most convenient:

$$\begin{aligned} &e_{\pm\alpha_i},\ h_{\alpha_i}, \quad i = 1,2,...,n-1, \\ &e_{\pm(\alpha_i+\alpha_{i+1})}, \quad i = 1,2,...,n-2, \\ &\vdots \\ &e_{\pm(\alpha_1+\cdots+\alpha_{n-1})}. \end{aligned} \tag{2}$$

Here α_i, $i = 1,...,n-1$, are vectors (simple roots of the algebra) of a real Euclidean space. Their lengths and relative angles are given by the scalar products

$$(\alpha_i,\alpha_j) = 2\delta_{ij} - \delta_{1,|j-i|}, \quad i,j = 1,2,...,n-1. \tag{3}$$

The subscripts ζ of $e_{\pm\zeta}$ in (2) taken with positive sign form the set Π of positive roots of the algebra. The $n-1$ generators h_{α_i} have the following linearity property:

$$h_{a\alpha_i+b\alpha_j} = ah_{\alpha_i} + bh_{\alpha_j}, \quad a,b\in R. \tag{4}$$

The nonzero commutation relations of SU(n) generators are:

$$\begin{aligned} &[e_\alpha,e_\beta] = N_{\alpha\beta}e_{\alpha+\beta} \quad \text{if } \alpha,\beta, \text{ and } \alpha+\beta\in\Pi \text{ or} \\ &\qquad -\alpha-\beta\in\Pi, \\ &[e_\alpha,e_{-\alpha}] = h_\alpha \quad \text{if } \alpha\in\Pi, \\ &[h_\alpha,e_{\pm\alpha}] = \pm 2e_{\pm\alpha} \quad \text{if } \alpha\in\Pi, \end{aligned} \tag{5}$$

where $N_{\alpha\beta}$ are constants usually chosen equal to ± 1. We fix them subsequently only as far as some of the generators $e_{\alpha+\beta}$ are needed. One notices that from the $2n-2$ generators $e_{\pm\alpha_i}$ one obtains all the others by commutation. Therefore, it suffices to consider only the action of those generators in any representation.

Let us point out that the generators (2) can be associated with the $n\times n$ matrices E_{ij} with 1 at the intersection of ith row and jth column and 0 elsewhere as follows:

$$\left.\begin{aligned} e_{\alpha_i+\cdots+\alpha_k} &= E_{i,k+1} \\ e_{-\alpha_i-\cdots-\alpha_k} &= E_{k+1,i} \end{aligned}\right\}, \quad i\leqslant k, \quad i,k = 1,...,n,$$

$$h_{\alpha_i} = E_{ii} - E_{i+1,i+1}, \quad i = 1,2,...,n-1.$$

B. Weights and representations

Every irreducible representation space we consider here is built out of the weight vectors (fermions) of Table III. Denoting each fermion by its SU(5) weight $w = (w_1,w_2,w_3,w_4)$ and h_{α_i} and $e_{\pm\alpha_i}$, respectively, by h_i and $e_{\pm i}$, one can write the action of the SU(5) generators on the fermions as follows:

$$h_i(w_1,w_2,w_3,w_4) = w_i(w_1,w_2,w_3,w_4), \quad i = 1,2,3,4, \tag{6}$$

$$e_{\pm 1}(w) = \begin{cases} 0 & \text{if } w_1 = 0 \text{ or } w_1\gtrless 0 \\ (w_1 \pm 2, w_2 \mp 1, w_3, w_4), \end{cases}$$

$$e_{\pm 2}(w) = \begin{cases} 0 & \text{if } w_2 = 0 \text{ or } w_2\gtrless 0 \\ (w_1 \mp 1, w_2 \pm 2, w_3 \mp 1, w_4), \end{cases}$$

$$e_{\pm 3}(w) = \begin{cases} 0 & \text{if } w_3 = 0 \text{ or } w_3\gtrless 0 \\ (w_1, w_2 \mp 1, w_3 \pm 2, w_4 \mp 1), \end{cases}$$

$$e_{\pm 4}(w) = \begin{cases} 0 & \text{if } w_4 = 0 \text{ or } w_4\gtrless 0 \\ (w_1, w_2, w_3 \mp 1, w_4 \pm 2). \end{cases} \tag{7}$$

Here all upper (lower) signs and inequalities have to be taken simultaneously in each equality. Let us point out that an element X of the SU(5) Lie algebra acts on a product of two fermions as a derivation, i.e.,

$$X(w)(w') = (X(w))(w') + (w)(X(w')). \tag{8}$$

The coordinates w_i of a weight w are equal to the integers

$$w_i = 2(w,\alpha_i)/(\alpha_i,\alpha_i), \quad i = 1,2,...,n-1. \tag{9}$$

In particular, taking for w one of the positive roots, one finds its coordinates (9) using (3). An irreducible representation is specified conveniently by the highest weight. Coordinates (9) of a highest weight are nonnegative integers. For examples, see Table II.

There exists a standard algorithm for computing all weights of a representation starting from the highest one (cf. Refs. 4 and 8 and references therein).

Let us underline the difference between a weight w and the weight vector (w). The weight vector belongs to a representation space, i.e., it is transformed by elements of the representation of the group, weight w is just the set of integer labels ("quantum numbers") used to denote the weight vector. The group representation does not act on it.

C. SU(3)×SU(2)×U(1) in SU(5)

Generators of the subgroup SU(3)×SU(2)×U(1) are linear combinations of SU(5)-generators. The projection matrix

$$P = \begin{pmatrix} 1 & 0 & 0 & 2 \\ 1 & 0 & 1 & \bar{1} \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & \bar{2} \end{pmatrix}, \quad \bar{1} = -1,\ \bar{2} = -2, \tag{10}$$

acting from the *right* on every SU(5)-weight w transforms it into wP, a weight of SU(3)×SU(2)×U(1). In particular, it projects the SU(5) simple roots into those of the subgroup. That can be used to find the subgroup generators compatible with the projection matrix (10). Thus one finds the SU(3) generators as

$$\begin{aligned} e_{\pm\beta_1} &= e_{\pm(\alpha_1+\alpha_2)} = [e_{\pm 1},e_{\pm 2}], \\ e_{\pm\beta_2} &= e_{\pm(\alpha_3+\alpha_4)} = [e_{\pm 3},e_{\pm 4}], \end{aligned} \tag{11}$$

those of SU(2) as

$$e_{\pm\gamma} = e_{\pm(\alpha_2+\alpha_3)} = [e_{\pm 2},e_{\pm 3}], \tag{12}$$

and the one of U(1) as

$$h = 2h_1 - h_2 + h_3 - 2h_4. \tag{13}$$

Note that the electric charge operator Q, whose eigenvalues are listed in the second column of Table III, is

$$Q = \tfrac{1}{2}(h_2 + h_3) - \tfrac{1}{6}h. \tag{14}$$

The U(1) component of an SU(3)×SU(2)×U(1) weight does not vary within an irreducible representation. There-

TABLES IV–X. SU(5) Clebsch–Gordan coefficients.

5x5

		(2000)	(2000)	(0100)
(1000)(1000)		1	(0100)	(0100)
(1000)($\bar{1}$100)		$1/\sqrt{2}$	$1/\sqrt{2}$	
($\bar{1}$100)(1000)		$1/\sqrt{2}$	$-1/\sqrt{2}$	

5x10

		(1100)	(1100)	(0010)
(1000)(0100)		1	$(0010)_1$	$(0010)_2$	(0010)
(1000)($\bar{1}$010)		$1/\sqrt{2}$		$1/\sqrt{3}$
($\bar{1}$100)(1$\bar{1}$10)		$1/\sqrt{2}$	$1/\sqrt{2}$	$-1/\sqrt{3}$
(0$\bar{1}$10)(0100)			$1/\sqrt{2}$	$1/\sqrt{3}$

5x$\bar{10}$

		(1010)	(1010)		(0001)
(1000)(0010)		1	$(0001)_1$	$(0001)_2$	$(0001)_3$	(0001)
(1000)($\bar{1}$001)		$1/\sqrt{2}$			1/2
($\bar{1}$100)(1$\bar{1}$01)		$1/\sqrt{2}$	$1/\sqrt{2}$		-1/2
(0$\bar{1}$10)(01$\bar{1}$1)			$1/\sqrt{2}$	$1/\sqrt{2}$	1/2
(00$\bar{1}$1)(0010)				$1/\sqrt{2}$	-1/2

5x$\bar{5}$

		(1001)	(1001)			(0000)
(1000)(0001)		1	$(0000)_1$	$(0000)_2$	$(0000)_3$	$(0000)_4$	(0000)
(1000)($\bar{1}$000)		$1/\sqrt{2}$				$1/\sqrt{5}$
($\bar{1}$100)(1$\bar{1}$00)		$1/\sqrt{2}$	$1/\sqrt{2}$			$-1/\sqrt{5}$
(0$\bar{1}$10)(01$\bar{1}$0)			$1/\sqrt{2}$	$1/\sqrt{2}$		$1/\sqrt{5}$
(00$\bar{1}$1)(001$\bar{1}$)				$1/\sqrt{2}$	$1/\sqrt{2}$	$-1/\sqrt{5}$
(000$\bar{1}$)(0001)					$1/\sqrt{2}$	$1/\sqrt{5}$

15x5

		(3000)	(3000)	(1100)			
(2000)(1000)		1	(1100)	(1100)			
(0100)(1000)		$\sqrt{2}/\sqrt{3}$	$1/\sqrt{3}$	(3000)	(1100)	
(2000)($\bar{1}$100)		$1/\sqrt{3}$	$-\sqrt{2}/\sqrt{3}$	(0010)	$(0010)_1$	$(0010)_2$
($\bar{1}$010)(1000)				$1/\sqrt{3}$	$1/\sqrt{6}$	$\sqrt{2}/\sqrt{3}$
(1$\bar{1}$10)($\bar{1}$100)				$1/\sqrt{3}$	$1/\sqrt{6}$	$-1/\sqrt{6}$
(0100)(0$\bar{1}$10)				$1/\sqrt{3}$	$-\sqrt{2}/\sqrt{3}$	$-1/\sqrt{6}$

15x$\bar{5}$

		(2001)	(2001)	(0101)	(2001)				(1000)
(2000)(0001)		1	(0100)(0001)	1	$(1000)_1$	$(1000)_2$	$(1000)_3$	$(1000)_4$	(1000)
(100$\bar{1}$)(0001)								$1/\sqrt{2}$	$1/\sqrt{6}$
(10$\bar{1}$1)(001$\bar{1}$)							$1/\sqrt{2}$	$1/\sqrt{2}$	$-1/\sqrt{6}$
(1$\bar{1}$10)(01$\bar{1}$0)						$1/\sqrt{2}$	$1/\sqrt{2}$		$1/\sqrt{6}$
(0100)(1$\bar{1}$00)					$\sqrt{2}/\sqrt{3}$	$1/\sqrt{2}$			$-1/\sqrt{6}$
(2000)($\bar{1}$000)					$1/\sqrt{3}$				$1/\sqrt{3}$

10x10

		(0200)	(0200)	(1010)							
(0100)(0100)		1	(1010)	(1010)							
(0100)(1$\bar{1}$10)		$1/\sqrt{2}$	$1/\sqrt{2}$	(1010)			(0200)		(0001)
(1$\bar{1}$10)(0100)		$1/\sqrt{2}$	$-1/\sqrt{2}$	$(0001)_1$	$(0001)_2$	$(0001)_3$	$(0001)_1$	$(0001)_2$	(0001)
(1$\bar{1}$10)($\bar{1}$1$\bar{1}$1)				1/2	1/2	1/2	1/2	1/2	$1/\sqrt{6}$
($\bar{1}$1$\bar{1}$1)(1$\bar{1}$10)				-1/2	-1/2	-1/2	1/2	1/2	$1/\sqrt{6}$
($\bar{1}$010)(10$\bar{1}$1)				1/2		-1/2	1/2		$-1/\sqrt{6}$
(10$\bar{1}$1)($\bar{1}$010)				-1/2		1/2	1/2		$-1/\sqrt{6}$
(0100)(0$\bar{1}$01)					1/2			1/2	$-1/\sqrt{6}$
(0$\bar{1}$01)(0100)					-1/2			1/2	$-1/\sqrt{6}$

fore, we do not need to use it in the computation of CGC besides what is shown in Table II. Consequently, it suffices to consider only the left three columns of the matrix P, i.e., the projection matrix for $SU(5) \supset SU(3) \times SU(2)$. The U(1) weights associated with any representation other than those in Table II are readily found from Ref. 9.

D. Finite groups W and N

In order to find a desired CGC or a basis vector from those listed in our tables below, one has to use two finite groups related to the problem. The group W permutes the weights of a representation, without changing the whole weight system. It is generated by reflections r_i in planes orthogonal to simple roots α_i. An SU(n) weight $w = (w_1,\ldots,w_{n-1})$ is transformed as follows:

$$r_i w = w - w_i \alpha_i. \tag{15}$$

All weights which are obtained from a given one by the action of the Weyl group belong to the same W-orbit. It is known that each W-orbit contains precisely one weight with all components nonnegative (dominant weight). Thus, for instance, all weights of the SU(5) representation 5 are in the same W-orbit. The same is true for the weights of the representation 10.

The Weyl group acts in the space spanned by simple roots. Similar action in the representation space is provided by the group N generated by the operators[5]

$$\begin{aligned} R_\xi &= \exp(e_{-\xi}) \exp(-e_\xi) \exp(e_{-\xi}) \\ &= (1 + e_{-\xi} + \cdots)(1 - e_\xi + \cdots)(1 + e_{-\xi} + \cdots). \end{aligned} \tag{16}$$

Applications of (15) and (16) are shown in subsequent sections.

The group W acts in the weight space. The group $N \subset SU(5)$; therefore, it acts in the representation space. W and N are not isomorphic as groups: If (w) denotes a weight vector of weight w, and $r_\xi w$ is given by (15), then[5]

$$R_\xi(w) = \pm (r_\xi w), \quad \text{when } r_\xi w \neq w. \tag{17}$$

The relation (17) allows one to reduce the number of CGC which have to be listed in our tables. Suppose that w is the dominant weight on an orbit of W (such a weight is always unique) and that the weight subspace V^Λ_w of a representation space V^Λ has dim $V^\Lambda_w > 1$. Suppose further that a complete basis in the subspace V^Λ_w has been already constructed in some way. Then (17) allows one to translate that basis to any other subspace $V^\Lambda_{r_j w}$, $r_j \in W$. Hence it suffices to construct the basis only once in V^Λ_w instead of in each subspace corresponding to the same W-orbit.

The actual sign in (17) has to be determined either by direct computation or from the general prescription of Ref. 5. For the SU(5) weight vectors from representation spaces of (0001) and (0100) shown in Table III, the action (17) of R_{α_i}, where α_i is any of the simple roots of SU(5), is given by

$$\begin{aligned} &R_{\alpha_i}(w) = (w) \quad \text{if } r_i w = w, \\ &R_{\alpha_i}(w) = (w,\alpha_i)(r_i w) \quad \text{if } r_i w \neq w, \quad i = 1,2,3, \text{ and } 4, \end{aligned} \tag{18}$$

where (w,ξ) is the scalar product of the weight w with the positive root ξ. In particular, $R_{\alpha_1}(0001) = R_{\alpha_2}(0001)$ $= R_{\alpha_3}(0001) = (0001)$, $R_{\alpha_4}(0001) = (001\bar{1})$, $R_{\alpha_1}(001\bar{1})$ $= R_{\alpha_2}(001\bar{1}) = (001\bar{1})$, and $R_{\alpha_3}(001\bar{1}) = (01\bar{1}0)$, $R_{\alpha_4}(001\bar{1})$ $= -(0001)$, etc. Let us get the last equality in all details:

$$\begin{aligned} &R_{\alpha_4}(001\bar{1}) \\ &\quad = (1 + e_{-4} + \cdots)(1 - e_4 + \cdots)(1 + e_{-4} + \cdots)(001\bar{1}) \\ &\quad = (1 + e_{-4} + \cdots)(1 - e_4 + \cdots)(001\bar{1}) \\ &\quad = (1 + e_{-4} + \cdots)\{(001\bar{1}) - (0001)\} \\ &\quad = (001\bar{1}) - (0001) - (001\bar{1}) \\ &\quad = -(0001). \end{aligned}$$

Here we used repeatedly Eqs. (7).

The operators R_ξ represent elements of the group SU(n); hence, they act on the tensor product of two weight vectors according to

$$R_\xi(w)(w') = (R_\xi(w))(R_\xi(w')). \tag{19}$$

Subsequently, we consider only weight vectors, not the weights.

3. SU(5) CLEBSCH–GORDAN COEFFICIENTS

In this section we describe CGC related to the tensor products listed in Table I. No reference is made here to any subgroup. We comment separately on each table starting from the simplest one, introducing some conventions, providing examples, and pointing out important aspects of the derivation as they arise.

A. CGC for $5 \otimes 5$, Table IV

Multiplicity of each weight in (1000) is one. Hence we can use the weights of this representation to denote orthonormal basis vectors of the representation space without ambiguity. Relevant transformation properties are given in (6) and (7). As an illustration let us rewrite the content of Table IV as follows:

$$\binom{2000}{2000} = (1000)(1000), \tag{20}$$

$$\binom{2000}{0100} = \frac{1}{\sqrt{2}}(1000)(\bar{1}100) + \frac{1}{\sqrt{2}}(\bar{1}100)(1000), \tag{21}$$

$$\binom{0100}{0100} = \frac{1}{\sqrt{2}}(1000)(\bar{1}100) - \frac{1}{\sqrt{2}}(\bar{1}100)(1000). \tag{22}$$

Here on the left we write the highest weight above the weight, thus underlining the irreducible subspace to which the vector belongs. On the right side the weights of the representation (1000) denote the basis vectors of the product space. Using Table III, the right side can be read as a combination of two-fermion states.

Equation (21) is obtained from (20) by the action of e_{-1} accordiong to (7) and (8). One could also describe independently the action of the generators on the left side. However, for our limited purposes it is simpler to define the left side by the right one. Namely,

$$e_{-1}\begin{pmatrix}2000\\2000\end{pmatrix}$$
$$= (e_{-1}(1000))(1000) + (1000)(e_{-1}(1000)) \sim \begin{pmatrix}2000\\0100\end{pmatrix}$$
$$= \frac{1}{\sqrt{2}}(\bar{1}100)(1000) + \frac{1}{\sqrt{2}}(1000)(\bar{1}100).$$

Here $1/\sqrt{2}$ is just the normalization coefficient. The highest weight vector $\binom{0100}{0100}$ does not belong to the subspace which is generated from $\binom{2000}{2000}$; hence, it is orthogonal to $\binom{2000}{0100}$. Choosing conveniently its phase, one gets (22).

One can continue applying generators e_{-i} and thus construct further basis vectors for both irreducible subspaces. Some examples are:

$$e_{-1}e_{-1}\begin{pmatrix}2000\\2000\end{pmatrix} \sim \begin{pmatrix}2000\\\bar{2}200\end{pmatrix} = (\bar{1}100)(\bar{1}100),$$
$$e_{-2}\begin{pmatrix}2000\\0100\end{pmatrix} \sim \begin{pmatrix}2000\\1\bar{1}10\end{pmatrix}$$
$$= \frac{1}{\sqrt{2}}(0\bar{1}10)(1000) + \frac{1}{\sqrt{2}}(1000)(0\bar{1}10),$$
$$e_{-1}\begin{pmatrix}2000\\1\bar{1}10\end{pmatrix} \sim \begin{pmatrix}2000\\\bar{1}010\end{pmatrix}$$
$$= \frac{1}{\sqrt{2}}(0\bar{1}10)(\bar{1}100) + \frac{1}{\sqrt{2}}(\bar{1}100)(0\bar{1}10) \sim e_{-2}\begin{pmatrix}2000\\\bar{2}200\end{pmatrix},$$
$$e_{-3}\begin{pmatrix}2000\\1\bar{1}10\end{pmatrix} \sim \begin{pmatrix}2000\\10\bar{1}1\end{pmatrix}$$
$$= \frac{1}{\sqrt{2}}(00\bar{1}1)(1000) + \frac{1}{\sqrt{2}}(1000)(00\bar{1}1), \tag{23}$$
$$e_{-2}\begin{pmatrix}0100\\0100\end{pmatrix} \sim \begin{pmatrix}0100\\1\bar{1}00\end{pmatrix}$$
$$= \frac{1}{\sqrt{2}}(1000)(0\bar{1}10) - \frac{1}{\sqrt{2}}(0\bar{1}10)(1000),$$

etc. These computations are easy in a simple case as this one. They quickly become prohibitively long when considering higher representations and/or higher rank groups.

An important observation can already be made here: There is considerable repetition of CGC values. The coefficients take only two nonzero values, 1 and $1/\sqrt{2}$, in the representation space (2000), and $\pm 1/\sqrt{2}$ in (0100). The explanation lies in relation (17). Indeed, one has

$$\begin{pmatrix}2000\\\bar{2}200\end{pmatrix} = R_{\alpha_1}\begin{pmatrix}2000\\2000\end{pmatrix}$$
$$= R_{\alpha_1}((1000)(1000)) = (\bar{1}100)(\bar{1}100),$$
$$\begin{pmatrix}2000\\0\bar{2}20\end{pmatrix} = R_{\alpha_2}R_{\alpha_1}\begin{pmatrix}2000\\2000\end{pmatrix}, \tag{24}$$
$$\begin{pmatrix}2000\\00\bar{2}2\end{pmatrix} = R_{\alpha_3}R_{\alpha_2}R_{\alpha_1}\begin{pmatrix}2000\\2000\end{pmatrix},$$
$$\begin{pmatrix}2000\\000\bar{2}\end{pmatrix} = R_{\alpha_4}R_{\alpha_3}R_{\alpha_4}R_{\alpha_1}\begin{pmatrix}2000\\2000\end{pmatrix}.$$

Similarly,

$$\begin{pmatrix}2000\\1\bar{1}10\end{pmatrix} = R_{\alpha_2}\begin{pmatrix}2000\\0100\end{pmatrix},$$
$$\begin{pmatrix}2000\\\bar{1}010\end{pmatrix} = R_{\alpha_1}R_{\alpha_2}\begin{pmatrix}2000\\0100\end{pmatrix}, \tag{25}$$
$$\begin{pmatrix}2000\\10\bar{1}1\end{pmatrix} = R_{\alpha_3}R_{\alpha_2}\begin{pmatrix}2000\\0100\end{pmatrix}, \quad \text{etc.}$$

The application of the operators R_{α_i} according to (17) using also (16) is much faster than the computation as in (23). The CGC corresponding to all nonzero vectors

$$R_{\alpha_{i_1}}\cdots R_{\alpha_{i_k}}\begin{pmatrix}2000\\2000\end{pmatrix} \tag{26}$$

are equal to 1, those in

$$R_{\alpha_{i_r}}\cdots R_{\alpha_{i_t}}\begin{pmatrix}2000\\0100\end{pmatrix} \tag{27}$$

are all $1/\sqrt{2}$, and in

$$R_{\alpha_{i_s}}\cdots R_{\alpha_{i_w}}\begin{pmatrix}0100\\0100\end{pmatrix} \tag{28}$$

are $\pm 1/\sqrt{2}$, where the sign is decided using (22).

B. CGC for 5×10, Table V

Comments similar to those in the previous case can be made about Table V. The new feature here is the presence of two weight vectors $\binom{1100}{0010}_1$ and $\binom{1100}{0010}_2$ in the (1100)-representation space. They are obtained from the highest one as follows:

$$e_{-1}e_{-2}\begin{pmatrix}1100\\1100\end{pmatrix} \sim \begin{pmatrix}1100\\0010\end{pmatrix}_1$$
$$= \frac{1}{\sqrt{2}}(1000)(\bar{1}010) + \frac{1}{\sqrt{2}}(\bar{1}100)(1\bar{1}10) \tag{29}$$
$$e_{-2}e_{-1}\begin{pmatrix}1100\\1100\end{pmatrix} \sim \begin{pmatrix}1100\\0010\end{pmatrix}_2$$
$$= \frac{1}{\sqrt{2}}(\bar{1}100)(1\bar{1}10) + \frac{1}{\sqrt{2}}(0\bar{1}10)(0100). \tag{30}$$

The product basis vectors in the first column are pairwise orthogonal. Their linear combinations (29) and (30) are not. The vector $\binom{0010}{0010}$ of the last column is orthogonal to (29) and (30). An application of the operator R_{α_i} to the four-vectors of Table V according to (17) and (16) produces the rest of the basis in the reduced space and thus also all the CGC.

C. CGC for $5\times\overline{10}$, Table VI

Table VI is similar to Table V. There are five dominant vectors, four in the (1010)-space: $\binom{1010}{1010}$ and three others of the same weight: $\binom{1010}{0001}_1$, $\binom{1010}{0001}_2$, $\binom{1010}{0001}_3$. In the (0001)-space there is only one dominant vector, $\binom{0001}{0001}$. The subscript i at $\binom{1010}{0001}_i$ indicates that e_{-i} was the last generator of the three needed to get to the (0001)-subspace from the highest weight one.

TABLES XI–XIII. SU(5) Clebsch–Gordan coefficients.

5x24

	(2001)	(2001)	(0101)
(1000)(1001)	1	(0101)	(0101)
(1000)($\bar{1}$101)	1/$\sqrt{2}$	1/$\sqrt{2}$	
($\bar{1}$100)(1001)	1/$\sqrt{2}$	-1/$\sqrt{2}$	

	(0101)			(2001)				(1000)
	$(1000)_2$	$(1000)_3$	$(1000)_4$	$(1000)_1$	$(1000)_2$	$(1000)_3$	$(1000)_4$	(1000)
$(1000)(0000)_1$				$\sqrt{2}/\sqrt{3}$				- 2/$\sqrt{15}$
$(1000)(0000)_2$	1/$\sqrt{2}$				1/$\sqrt{2}$			3/2$\sqrt{15}$
$(1000)(0000)_3$		1/$\sqrt{2}$				1/$\sqrt{2}$		- 1/$\sqrt{15}$
$(1000)(0000)_4$			1/$\sqrt{2}$				1/$\sqrt{2}$	1/2$\sqrt{15}$
($\bar{1}$100)(2$\bar{1}$00)	-1/2			1/$\sqrt{3}$	1/2			$\sqrt{5}$/2$\sqrt{6}$
(0$\bar{1}$10)(11$\bar{1}$0)	-1/2	-1/2			1/2	1/2		-$\sqrt{5}$/2$\sqrt{6}$
(00$\bar{1}$1)(101$\bar{1}$)		-1/2	-1/2			1/2	1/2	$\sqrt{5}$/2$\sqrt{6}$
(000$\bar{1}$)(1001)			-1/2				1/2	-$\sqrt{5}$/2$\sqrt{6}$

10x$\bar{10}$

	(0110)	(0110)		(1001)
(0100)(0010)	1	$(1001)_2$	$(1001)_3$	(1001)
(0100)(1$\bar{1}$01)	1/$\sqrt{2}$			-1/$\sqrt{3}$
(1$\bar{1}$10)(01$\bar{1}$1)	1/$\sqrt{2}$	1/$\sqrt{2}$		1/$\sqrt{3}$
(10$\bar{1}$1)(0010)		1/$\sqrt{2}$		-1/$\sqrt{3}$

	(0110)					(1001)				(0000)
	$(0000)_a$	$(0000)_b$	$(0000)_c$	$(0000)_d$	$(0000)_e$	$(0000)_1$	$(0000)_2$	$(0000)_3$	$(0000)_4$	(0000)
(0100)(0$\bar{1}$00)			1/2				-1/$\sqrt{6}$			1/$\sqrt{10}$
(1$\bar{1}$10)($\bar{1}$1$\bar{1}$0)			1/2	1/2		-1/$\sqrt{6}$	-1/$\sqrt{6}$	-1/$\sqrt{6}$		-1/$\sqrt{10}$
($\bar{1}$010)(10$\bar{1}$0)	1/2			1/2		-1/$\sqrt{6}$		1/$\sqrt{6}$		1/$\sqrt{10}$
(10$\bar{1}$1)($\bar{1}$01$\bar{1}$)	1/2			1/2	1/2	1/$\sqrt{6}$		-1/$\sqrt{6}$	-1/$\sqrt{6}$	1/$\sqrt{10}$
($\bar{1}$1$\bar{1}$1)(1$\bar{1}$1$\bar{1}$)	1/2	1/2	1/2	1/2	1/2	1/$\sqrt{6}$	1/$\sqrt{6}$	1/$\sqrt{6}$	1/$\sqrt{6}$	-1/$\sqrt{10}$
(100$\bar{1}$)($\bar{1}$001)					1/2	-1/$\sqrt{6}$			-1/$\sqrt{6}$	-1/$\sqrt{10}$
($\bar{1}$10$\bar{1}$)(1$\bar{1}$01)		1/2			1/2	-1/$\sqrt{6}$	-1/$\sqrt{6}$		1/$\sqrt{6}$	1/$\sqrt{10}$
(0$\bar{1}$01)(010$\bar{1}$)		1/2	1/2				1/$\sqrt{6}$		-1/$\sqrt{6}$	1/$\sqrt{10}$
(0$\bar{1}$1$\bar{1}$)(01$\bar{1}$1)	1/2	1/2					-1/$\sqrt{6}$	-1/$\sqrt{6}$	-1/$\sqrt{6}$	-1/$\sqrt{10}$
(00$\bar{1}$0)(0010)	1/2							-1/$\sqrt{6}$		1/$\sqrt{10}$

10x24

	(1101)	(1101)		(0011)
(0100)(1001)	1	$(0011)_1$	$(0911)_2$	(0011)
(1$\bar{1}$10)($\bar{1}$101)	1/$\sqrt{2}$	1/$\sqrt{2}$	1/$\sqrt{3}$	
($\bar{1}$010)(1001)	1/$\sqrt{2}$		-1/$\sqrt{3}$	
(0100)(0$\bar{1}$11)		1/$\sqrt{2}$	-1/$\sqrt{3}$	

	(1101)			(2000)
	$(2000)_2$	$(2000)_3$	$(2000)_4$	(2000)
(0100)(2$\bar{1}$00)	1/$\sqrt{2}$			1/2
(1$\bar{1}$10)(11$\bar{1}$0)	1/$\sqrt{2}$	1/$\sqrt{2}$		-1/2
(100$\bar{1}$)(1001)			1/$\sqrt{2}$	-1/2
(10$\bar{1}$1)(101$\bar{1}$)		1/$\sqrt{2}$	1/$\sqrt{2}$	1/2

	(2000)	(0011)		(1101)						(0100)
	(0100)	$(0100)_3$	$(0100)_4$	$(0100)_{21}$	$(0100)_{31}$	$(0100)_{41}$	$(0100)_2$	$(0100)_3$	$(0100)_4$	(0100)
$(0100)(0000)_1$	1/2			1/$\sqrt{2}$						- 1/$\sqrt{10}$
$(0100)(0000)_2$							$\sqrt{2}/\sqrt{3}$			$\sqrt{2}/\sqrt{5}$
$(0100)(0000)_3$		-1/$\sqrt{3}$						1/$\sqrt{2}$		- 2$\sqrt{2}$/3$\sqrt{5}$
$(0100)(0000)_4$			-1/$\sqrt{3}$						1/$\sqrt{2}$	$\sqrt{2}$/3$\sqrt{5}$
($\bar{1}$010)(11$\bar{1}$0)	-1/$\sqrt{8}$	-1/$\sqrt{6}$		1/2	1/2					$\sqrt{5}$/6
($\bar{1}$10$\bar{1}$)(1001)	-1/$\sqrt{8}$		-1/$\sqrt{6}$			1/2				$\sqrt{5}$/6
(1$\bar{1}$10)($\bar{1}$2$\bar{1}$0)	-1/$\sqrt{8}$	1/$\sqrt{6}$		1/2	1/2		1/$\sqrt{3}$	1/2		-$\sqrt{5}$/6
(100$\bar{1}$)($\bar{1}$101)	-1/$\sqrt{8}$		1/$\sqrt{6}$			1/2			1/2	-$\sqrt{5}$/6
(10$\bar{1}$1)($\bar{1}$11$\bar{1}$)	1/$\sqrt{8}$	1/$\sqrt{6}$	1/$\sqrt{6}$		1/2	1/2		1/2	1/2	$\sqrt{5}$/6
($\bar{1}$1$\bar{1}$1)(101$\bar{1}$)	1/$\sqrt{8}$	-1/$\sqrt{6}$	-1/$\sqrt{6}$		1/2	1/2				-$\sqrt{5}$/6

D. CGC in Tables VII–XI

These tables are similar to the previous ones, Table XI being the most complicated of them all. Each of the five linearly independent vectors of the weight (0000) in (0110) in this Table can be obtained in many different ways. For instance, one way is as follows:

$$\begin{pmatrix}0110\\0000\end{pmatrix}_a \sim e_{-3}e_{-2}e_{-1}e_{-4}e_{-3}e_{-2}\begin{pmatrix}0110\\0110\end{pmatrix},$$
$$\begin{pmatrix}0110\\0000\end{pmatrix}_b \sim e_{-2}e_{-1}e_{-3}e_{-4}e_{-3}e_{-2}\begin{pmatrix}0110\\0110\end{pmatrix},$$
$$\begin{pmatrix}0110\\0000\end{pmatrix}_c \sim e_{-2}e_{-1}e_{-3}e_{-2}e_{-4}e_{-3}\begin{pmatrix}0110\\0110\end{pmatrix},$$
$$\begin{pmatrix}0110\\0000\end{pmatrix}_d \sim e_{-1}e_{-2}e_{-3}e_{-2}e_{-4}e_{-3}\begin{pmatrix}0110\\0110\end{pmatrix},$$
$$\begin{pmatrix}0110\\0000\end{pmatrix}_e \sim e_{-4}e_{-3}e_{-2}e_{-1}e_{-2}e_{-3}\begin{pmatrix}0110\\0110\end{pmatrix}.$$

E. CGC in Tables XII and XIII

There is a new phenomenon in these tables. At least one of the representations in the product has some weights of multiplicity >1. consequently, in the product there occur weight vectors with subscripts. In order to know correctly all the properties of such a basis vector, one has to use a corresponding earlier table, where such a vector is built out of the simple ones.

Consider the vector

$$\begin{pmatrix}0101\\1000\end{pmatrix}_2$$
$$= \tfrac{1}{2}(\bar{1}100)(2\bar{1}00) + \tfrac{1}{2}(0\bar{1}10)(11\bar{1}0) - \frac{1}{\sqrt{2}}(1000)(0000)_2$$

of Table XII. Here $(0000)_2$ stands for the vector $\binom{1001}{0000}_2$ given in Table X, or also for the same vector of the 12th column of Table XI. If some detailed properties of $(0000)_2$ are needed one has to refer to one of these tables. Clearly, it is simpler to use Table X than Table XI.

4. CLEBSCH–GORDAN COEFFICIENTS IN THE SUBGROUP BASIS

In this section we require that each basis vector has a definite SU(3)×SU(2)×U(1)-weight and that it belongs entirely to one subspace irreducible with respect to the subgroup. These requirements are obviously satisfied by the fermions of Table III. Their subgroup weights are given in column 3, and the subgroup multiplets are identified in column 4. However, for other SU(5) representations the situation is generally more complicated.

Every SU(5)-weight $w = (w_1\ w_2\ w_3\ w_4)$ corresponds to a definite subgroup weight $v = (v_1\ v_2)(v_3)(v_4)$. The correspondence is made explicit by means of (10):

$$v = wP = (w_1 + w_2, w_3 + w_4)(w_2 + w_3) \times (2w_1 - w_2 + w_3 - 2w_4), \tag{31}$$

where the parentheses indicate the SU(3), SU(2), and U(1) weights, respectively.

In order to assure that a subgroup weight vector belongs to just one irreducible subspace, it has to be generated from the highest weight, as before, using the generators (11) and (12) only. Also the operators r_i and R_i act in the same way, but now i stands for a root of SU(3) or SU(2). The finite groups W and N are smaller in the case of the subgroup than for SU(5). The subgroup weights which are dominant (all components $\geqslant 0$) are more numerous. Consequently, the tables of previous section need not only reinterpretation of the weights using (31), but also refinement (they get bigger).

The product space reduces first according to SU(5) and then each of the spaces reduces further according to the subgroup. The weight vectors with repeated subgroup weights referring to different SU(5) subspaces are automatically orthogonal, those referring to the same SU(5) subspace but different subgroup subspaces must be constructed to be orthogonal.

For simplicity of notation the U(1) part of each subgroup weight is omitted. It does not affect the CGC, and it is given in Table II.

Example: CGC for 5×5 in the subgroup basis, Table IV′

The Tables IV and IV′ refer to the same tensor product of SU(5) representations. The product basis in both cases is the same. Table IV′ is considerably bigger than Table IV because there are more subgroup weights which are dominant. The first line of each column of Table IV′ indicates the SU(5)-irreducible subspace. The second line gives the highest weight of each subspace irreducible with respect to the subgroup. The third line gives the actual subgroup weight vector. The three linear combinations of Table IV can easily be identified. For instance, $\binom{2000}{2000}$ of (20) becomes the highest weight vector $\binom{20}{20}\binom{0}{0}$ of the (2000) space, etc. The vector $\binom{20}{01}\binom{0}{0}$ is obtained as follows:

$$\begin{aligned} e_{-\beta_1}\begin{pmatrix}20\\20\end{pmatrix}\begin{pmatrix}0\\0\end{pmatrix} &= (e_{-1}e_{-2} - e_{-2}e_{-1})\begin{pmatrix}2000\\2000\end{pmatrix} \\ &= (e_{-1}e_{-2} - e_{-2}e_{-1})(1000)(1000) \\ &= -e_{-2}(\bar{1}100)(1000) - e_{-2}(1000)(\bar{1}100) \\ &= -(0\bar{1}10)(1000) - (1000)(0\bar{1}10). \end{aligned}$$

Thus one has

$$\begin{pmatrix}20\\01\end{pmatrix}\begin{pmatrix}0\\0\end{pmatrix} = -\frac{1}{\sqrt{2}}(0\bar{1}10)(1000) - \frac{1}{\sqrt{2}}(1000)(0\bar{1}10). \tag{32}$$

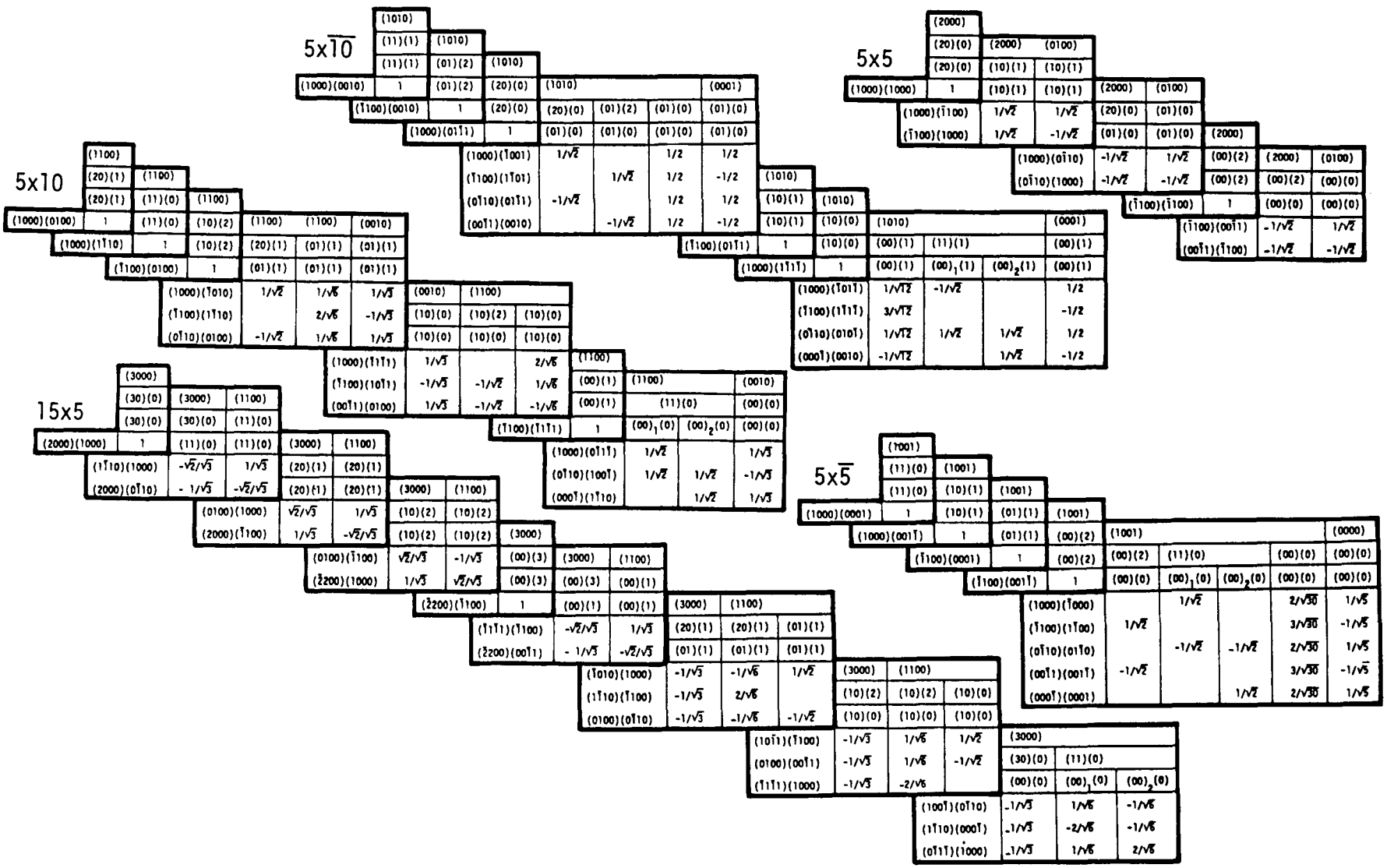
The highest weight vector $\binom{01}{01}\binom{0}{0}$ from the SU(5)-subspace (0100) is found by requiring its orthogonality to $\binom{20}{01}\binom{0}{0}$ of (32):

$$\begin{pmatrix}01\\01\end{pmatrix}\begin{pmatrix}0\\0\end{pmatrix} = \frac{1}{\sqrt{2}}(1000)(0\bar{1}10) - \frac{1}{\sqrt{2}}(0\bar{1}10)(1000), \tag{33}$$

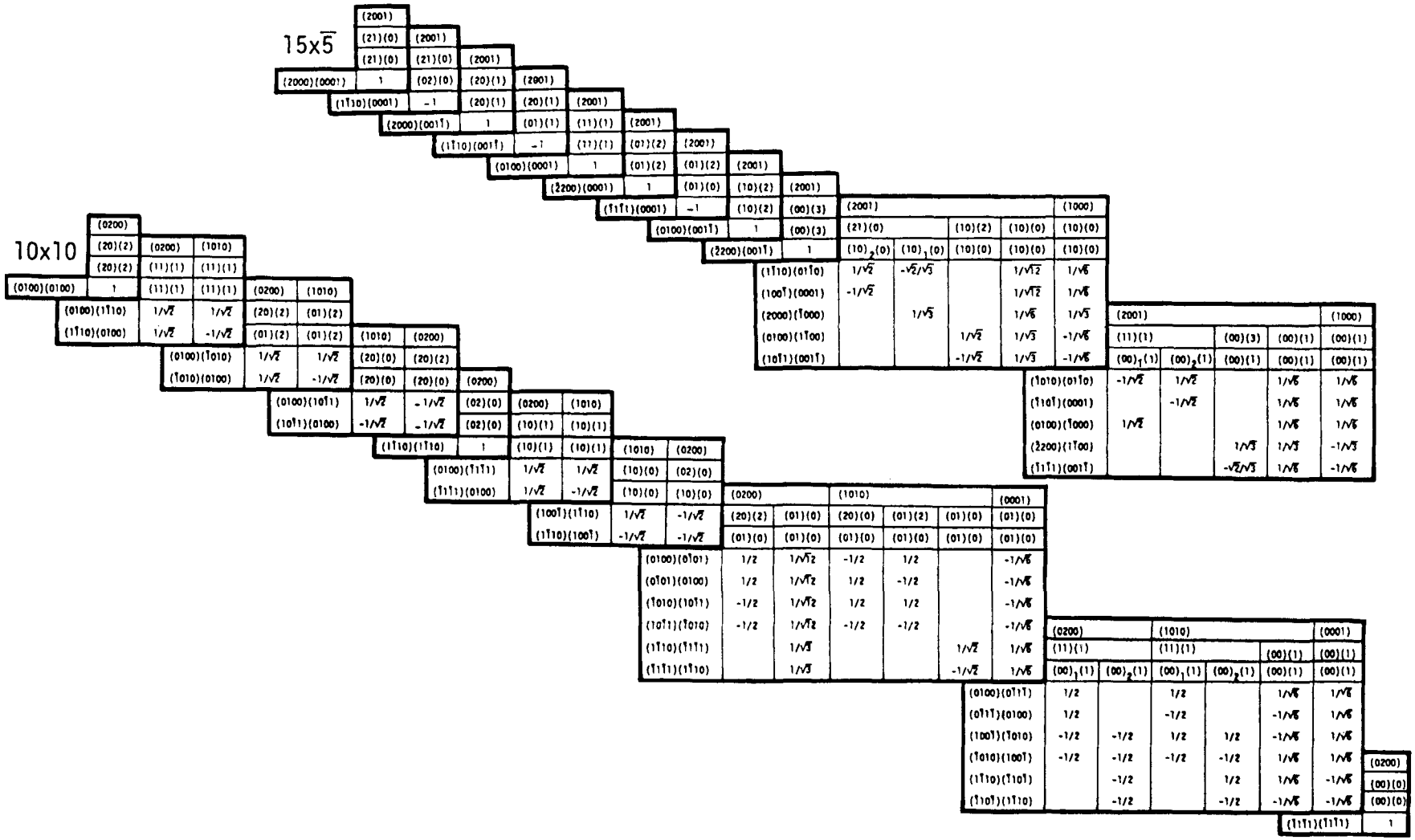
where again the phase is chosen by convenience only.

Since the subgroup weight (00)(2) is unique in the whole system, the highest weight vector $\binom{00}{00}\binom{2}{2}$ of the (2000)-space is found in a different way. Using the projection matrix P of (10), one has to find among the product basis the one basis

TABLES IV′–VII′ and X′. Clebsch–Gordan coefficients in the subgroup basis.



TABLES VIII′ and IX′. SU(5) Clebsch–Gordan coefficients in the subgroup basis.



TABLES XI′ and XII′. SU(5) Clebsch–Gordan coefficients in the subgroup basis.

5x24

	(2001)		
	(21)(0)	(2001)	(0101)
	(21)(0)	(21)(0)	(02)(0)
(1000)(1001)	1	(02)(0)	(02)(0)

		(2001)	(0101)
	(1000)(0$\bar{1}$11)	$-1/\sqrt{2}$	$1/\sqrt{2}$
	(0$\bar{1}$10)(1001)	$-1/\sqrt{2}$	$-1/\sqrt{2}$

			(2001)	(0101)
		(1000)(10$\bar{1}$1)	1	(01)(1)
			(20)(1)	(01)(1)
			(20)(1)	(01)(1)

				(2001)	(0101)
			(1000)(0$\bar{1}$2$\bar{1}$)	$-1/\sqrt{2}$	$1/\sqrt{2}$
			(0$\bar{1}$10)(101$\bar{1}$)	$-1/\sqrt{2}$	$-1/\sqrt{2}$
			(11)(1)	(11)(1)	
			(11)(1)	(11)(1)	

					(2001)	(0101)
				(1000)($\bar{1}$101)	$1/\sqrt{2}$	$1/\sqrt{2}$
				($\bar{1}$100)(1001)	$1/\sqrt{2}$	$-1/\sqrt{2}$
				(01)(2)	(2001)	(0101)
				(01)(2)	(10)(2)	(10)(2)

						(2001)
					($\bar{1}$100)($\bar{1}$101)	1
					(10)(2)	(10)(2)
					(10)(2)	(10)(2)

							(2001)	(0101)
						(1000)($\bar{1}$11$\bar{1}$)	$1/\sqrt{2}$	$1/\sqrt{2}$
						($\bar{1}$100)(101$\bar{1}$)	$1/\sqrt{2}$	$-1/\sqrt{2}$
						(00)(3)	(2001)	(0101)
						(00)(3)	(01)(2)	(01)(0)

								(2001)	(0101)
							($\bar{1}$100)($\bar{1}$11$\bar{1}$)	1	(01)(0)
								(01)(0)	(01)(0)

									(2001)	(0101)
								($\bar{1}$100)(00$\bar{1}$2)	$-1/\sqrt{2}$	$1/\sqrt{2}$
								(00$\bar{1}$1)($\bar{1}$101)	$-1/\sqrt{2}$	$-1/\sqrt{2}$

10x$\overline{10}$

	(0110)			
	(11)(2)	(0110)		(1001)
	(11)(2)	(11)(2)	(11)(0)	(11)(0)
(0100)(0010)	1	(11)(0)	(11)(0)	(11)(0)

					(0110)
	(0100)(1$\bar{1}$01)	$1/\sqrt{2}$	$1/\sqrt{6}$	$1/\sqrt{3}$	(02)(1)
	(1$\bar{1}$10)(01$\bar{1}$1)		$2/\sqrt{6}$	$-1/\sqrt{3}$	(02)(1)
	(10$\bar{1}$1)(0010)	$-1/\sqrt{2}$	$1/\sqrt{6}$	$1/\sqrt{3}$	(20)(1)

						(0110)
		(1$\bar{1}$10)(0010)	1	(20)(1)	(10)(0)	(10)(0)
		(0100)(01$\bar{1}$1)	1	(10)(0)	(01)(0)	(01)(0)

							(0110)		(1001)
			(1$\bar{1}$1$\bar{1}$)(01$\bar{1}$1)	1	(01)(0)	(11)(2)	(11)(2)	(00)(2)
			(1$\bar{1}$10)(1$\bar{1}$1$\bar{1}$)	1	(01)(0)	$(00)_1(2)$	$(00)_2(2)$	(00)(2)

										(1001)
						(0100)($\bar{1}$01$\bar{1}$)	$-1/\sqrt{2}$		$1/\sqrt{3}$	(00)(2)
						($\bar{1}$010)(010$\bar{1}$)	$-1/\sqrt{2}$	$-1/\sqrt{2}$	$-1/\sqrt{3}$	(00)(0)
						($\bar{1}$10$\bar{1}$)(0010)		$-1/\sqrt{2}$	$1/\sqrt{3}$	(00)(0)

10x$\overline{10}$ cont.

	(0110)		(1001)
	(02)(1)	(10)(1)	(10)(1)
	(10)(1)	(10)(1)	(10)(1)
(0100)(1$\bar{1}$1$\bar{1}$)	$-1/\sqrt{2}$	$1/\sqrt{6}$	$1/\sqrt{3}$
(1$\bar{1}$10)(010$\bar{1}$)	$-1/\sqrt{2}$	$-1/\sqrt{6}$	$-1/\sqrt{3}$
(100$\bar{1}$)(0010)		$-2/\sqrt{6}$	$1/\sqrt{3}$

				(0110)		(1001)
			(0100)($\bar{1}$001)	$1/\sqrt{2}$	$1/\sqrt{6}$	$1/\sqrt{3}$
			($\bar{1}$010)(01$\bar{1}$1)	$1/\sqrt{2}$	$-1/\sqrt{6}$	$-1/\sqrt{3}$
			($\bar{1}$1$\bar{1}$1)(0010)		$-2/\sqrt{6}$	$1/\sqrt{3}$

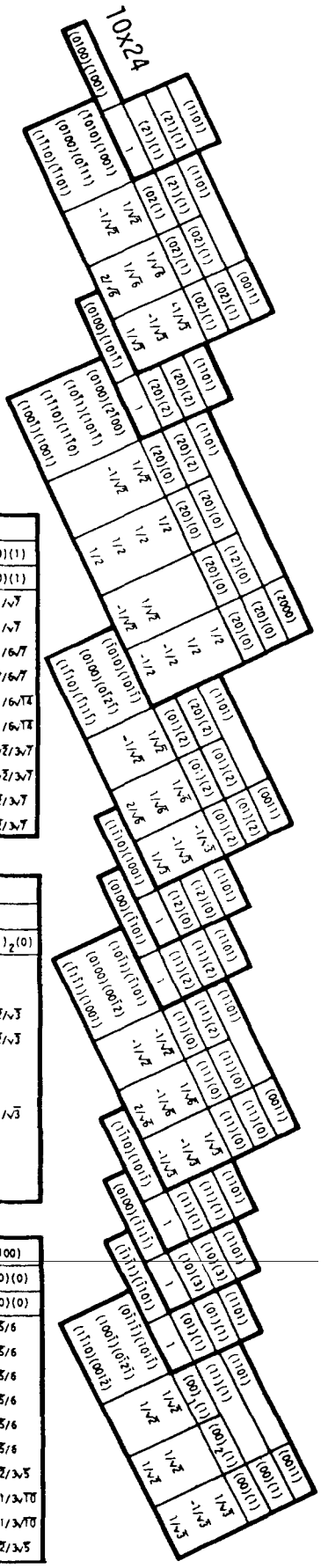
5x24 cont.

	(2001)		(0101)		(2001)		(0101)	(1000)
	(11)(1)		(11)(1)		(00)(3)	(00)(1)	(00)(1)	(00)(1)
	$(00)_1(1)$	$(00)_2(1)$	$(00)_1(1)$	$(00)_2(1)$	(00)(1)	(00)(1)	(00)(1)	(00)(1)
($\bar{1}$100)(0000)$_1$	$1/\sqrt{2}$		$-1/\sqrt{2}$			$1/\sqrt{6}$	$-1/2\sqrt{3}$	$1/2\sqrt{15}$
($\bar{1}$100)(0000)$_2$	$-1/\sqrt{2}$		$1/\sqrt{2}$		$\sqrt{2}/\sqrt{3}$	$1/\sqrt{6}$	$1/2\sqrt{3}$	$\sqrt{3}/2\sqrt{5}$
($\bar{1}$100)(0000)$_3$		$-1/\sqrt{2}$		$1/\sqrt{2}$	$-\sqrt{2}/\sqrt{3}$		$-1/\sqrt{3}$	$-1/\sqrt{15}$
($\bar{1}$100)(0000)$_4$		$1/\sqrt{2}$		$-1/\sqrt{2}$		$1/\sqrt{6}$	$-1/2\sqrt{3}$	$1/2\sqrt{15}$
(0$\bar{1}$10)($\bar{1}$2$\bar{1}$0)	$-1/2$	$-1/2$	$-1/2$	$-1/2$		$1/2\sqrt{3}$	$1/\sqrt{24}$	$-\sqrt{5}/2\sqrt{6}$
(000$\bar{1}$)($\bar{1}$101)		$1/2$		$1/2$		$1/2\sqrt{3}$	$1/\sqrt{24}$	$-\sqrt{5}/2\sqrt{6}$
(00$\bar{1}$1)($\bar{1}$11$\bar{1}$)					$-1/\sqrt{3}$	$1/2\sqrt{3}$	$\sqrt{3}/2\sqrt{2}$	$\sqrt{5}/2\sqrt{6}$
(1000)($\bar{2}$100)	$1/2$		$1/2$			$1/2\sqrt{3}$	$1/\sqrt{24}$	$-\sqrt{5}/2\sqrt{6}$

	(2001)		(0101)		(2001)		(0101)	(1000)
	(11)(1)		(11)(1)		(00)(3)	(00)(1)	(00)(1)	(00)(1)
	$(00)_1(1)$	$(00)_2(1)$	$(00)_1(1)$	$(00)_2(1)$	(00)(1)	(00)(1)	(00)(1)	(00)(1)
(1000)(0000)$_1$	$\sqrt{2}/\sqrt{3}$			$1/\sqrt{3}$				$-2/\sqrt{15}$
(1000)(0000)$_2$	$-\sqrt{2}/\sqrt{3}$		$1/\sqrt{2}$			$1/\sqrt{2}$	$-1/2$	$\sqrt{3}/2\sqrt{5}$
(1000)(0000)$_3$		$-1/\sqrt{2}$	$-1/\sqrt{2}$	$1/2\sqrt{3}$	$1/\sqrt{2}$	$-1/\sqrt{2}$		$-1/\sqrt{15}$
(1000)(0000)$_4$		$1/\sqrt{2}$		$1/2\sqrt{3}$	$-1/\sqrt{2}$		$-1/2$	$1/2\sqrt{15}$
(0$\bar{1}$10)(11$\bar{1}$0)	$-1/\sqrt{3}$	$-1/2$		$1/\sqrt{24}$	$-1/2$		$1/2\sqrt{2}$	$-\sqrt{5}/2\sqrt{6}$
(000$\bar{1}$)(1001)		$1/2$		$1/\sqrt{24}$	$1/2$		$1/2\sqrt{2}$	$-\sqrt{5}/2\sqrt{6}$
($\bar{1}$100)(2$\bar{1}$00)			$1/2$	$1/\sqrt{6}$		$-1/2$	$1/2\sqrt{2}$	$\sqrt{5}/2\sqrt{6}$
(00$\bar{1}$1)(101$\bar{1}$)			$-1/2$	$1/\sqrt{6}$		$1/2$	$1/2\sqrt{2}$	$-\sqrt{5}/2\sqrt{6}$

	(1001)				(0110)				(0000)
	(00)(2)	(11)(0)		(00)(0)	(11)(2)		(11)(0)		(00)(0)	(00)(0)
	(00)(0)	$(00)_1(0)$	$(00)_2(0)$	(00)(0)	$(00)_1(0)$	$(00)_2(0)$	$(00)_1(0)$	$(00)_2(0)$	(00)(0)	(00)(0)
(0100)(0$\bar{1}$00)	$1/\sqrt{6}$	$-1/\sqrt{6}$		$1/\sqrt{90}$	$-1/2$		$-1/\sqrt{12}$		$1/\sqrt{18}$	$1/\sqrt{10}$
(1$\bar{1}$10)($\bar{1}$1$\bar{1}$0)			$-1/\sqrt{6}$	$4/\sqrt{90}$				$1/\sqrt{3}$	$1/\sqrt{18}$	$-1/\sqrt{10}$
($\bar{1}$010)(10$\bar{1}$0)	$1/\sqrt{6}$	$1/\sqrt{6}$	$1/\sqrt{6}$	$1/\sqrt{90}$	$1/2$	$1/2$	$1/\sqrt{12}$	$1/\sqrt{12}$	$1/\sqrt{18}$	$1/\sqrt{10}$
(10$\bar{1}$1)($\bar{1}$01$\bar{1}$)	$-1/\sqrt{6}$	$-1/\sqrt{6}$		$1/\sqrt{90}$	$1/2$		$-1/\sqrt{12}$		$1/\sqrt{18}$	$1/\sqrt{10}$
($\bar{1}$1$\bar{1}$1)(1$\bar{1}$1$\bar{1}$)				$-6/\sqrt{90}$					$1/\sqrt{2}$	$-1/\sqrt{10}$
(100$\bar{1}$)($\bar{1}$001)		$1/\sqrt{6}$	$1/\sqrt{6}$	$4/\sqrt{90}$			$-1/\sqrt{3}$	$-1/\sqrt{3}$	$1/\sqrt{18}$	$-1/\sqrt{10}$
($\bar{1}$10$\bar{1}$)(1$\bar{1}$01)	$1/\sqrt{6}$		$-1/\sqrt{6}$	$1/\sqrt{90}$		$-1/2$		$-1/\sqrt{12}$	$1/\sqrt{18}$	$1/\sqrt{10}$
(0$\bar{1}$01)(010$\bar{1}$)	$-1/\sqrt{6}$	$1/\sqrt{6}$	$1/\sqrt{6}$	$1/\sqrt{90}$	$-1/2$	$-1/2$	$1/\sqrt{12}$	$1/\sqrt{12}$	$1/\sqrt{18}$	$1/\sqrt{10}$
(0$\bar{1}$1$\bar{1}$)(01$\bar{1}$1)		$-1/\sqrt{6}$		$4/\sqrt{90}$			$1/\sqrt{3}$		$1/\sqrt{18}$	$-1/\sqrt{10}$
(00$\bar{1}$0)(0010)	$-1/\sqrt{6}$		$-1/\sqrt{6}$	$1/\sqrt{90}$		$1/2$		$-1/\sqrt{12}$	$1/\sqrt{18}$	$1/\sqrt{10}$

10x24



10x24 cont.

	(1101)			(2000)	(1101)		(0011)
	(11)(2)		(00)(2)	(00)(2)	(11)(1)		(00)(1)
	$(00)_1(2)$	$(00)_2(2)$	(00)(2)	(00)(2)	$(00)_1(1)$	$(00)_2(1)$	(00)(1)
$(\bar{1}0\bar{1}0)(\bar{1}2\bar{1}0)$	$1/\sqrt{2}$	$1/\sqrt{2}$	$1/2\sqrt{3}$	-1/2			
$(0\bar{1}00)(2\bar{1}00)$	$1/\sqrt{2}$		$-1/2\sqrt{3}$	1/2			
$(\bar{1}1\bar{1}0)(\bar{1}1\bar{1}1)$		$-1/\sqrt{2}$	$1/2\sqrt{3}$	-1/2			
$(\bar{1}1\bar{1}1)(\bar{1}1\bar{1}\bar{1})$			$\sqrt{3}/2$	1/2			
$(0\bar{1}1\bar{1})(1011)$					$1/\sqrt{2}$		$1/\sqrt{3}$
$(100\bar{1})(0\bar{1}2\bar{1})$					$1/\sqrt{2}$	$1/\sqrt{2}$	$-1/\sqrt{3}$
$(1\bar{1}10)(001\bar{2})$						$1/\sqrt{2}$	$1/\sqrt{3}$

10x24 cont.

	(1101)			(0011)	(1101)	(0100)	(2000)	(0011)	(1101)	
	(21)(1)		(02)(1)	(02)(1)	(10)(3)	(10)(1)	(10)(1)	(10)(1)	(10)(1)	(10)(1)
	$(10)_1(1)$	$(10)_2(1)$	(10)(1)	(10)(1)	(10)(1)	(10)(1)	(10)(1)	(10)(1)	(10)(1)	(10)(1)
$(0100)(0000)_1$	$\sqrt{2}/\sqrt{3}$					$-1/\sqrt{10}$	1/2		$\sqrt{3}/2\sqrt{7}$	$1/\sqrt{7}$
$(0100)(0000)_2$	$-\sqrt{2}/\sqrt{3}$				$\sqrt{2}/\sqrt{3}$	$\sqrt{2}/\sqrt{5}$			$-2/\sqrt{21}$	$1/\sqrt{7}$
$(0100)(0000)_3$		$-1/\sqrt{2}$	$1/\sqrt{6}$	$-1/\sqrt{3}$	$-\sqrt{2}/\sqrt{3}$	$-2\sqrt{2}/3\sqrt{5}$		1/3	$2/\sqrt{21}$	$1/6\sqrt{7}$
$(0100)(0000)_4$		$1/\sqrt{2}$	$-1/\sqrt{6}$	$1/\sqrt{3}$		$\sqrt{2}/3\sqrt{5}$		1/3	0	$7/6\sqrt{7}$
$(\bar{1}010)(11\bar{1}0)$	$1/\sqrt{3}$	1/2	$1/2\sqrt{3}$	$-1/\sqrt{6}$		$\sqrt{5}/6$	$-1/\sqrt{8}$	$1/3\sqrt{2}$	$-\sqrt{3}/2\sqrt{14}$	$1/6\sqrt{14}$
$(\bar{1}10\bar{1})(1001)$		-1/2	$-1/2\sqrt{3}$	$1/\sqrt{6}$		$\sqrt{5}/6$	$-1/\sqrt{8}$	$1/3\sqrt{2}$	$-\sqrt{3}/2\sqrt{14}$	$1/6\sqrt{14}$
$(1\bar{1}10)(\bar{1}2\bar{1}0)$			$1/\sqrt{3}$	$1/\sqrt{6}$		$-\sqrt{5}/6$	$-1/\sqrt{8}$	$-1/3\sqrt{2}$	$-\sqrt{3}/2\sqrt{14}$	$2\sqrt{2}/3\sqrt{7}$
$(100\bar{1})(\bar{1}101)$			$-1/\sqrt{3}$	$-1/\sqrt{6}$		$-\sqrt{5}/6$	$-1/\sqrt{8}$	$-1/3\sqrt{2}$	$-\sqrt{3}/2\sqrt{14}$	$2\sqrt{2}/3\sqrt{7}$
$(10\bar{1}1)(\bar{1}11\bar{1})$					$-1/\sqrt{3}$	$\sqrt{5}/6$	$1/\sqrt{8}$	$-\sqrt{2}/3$	$-5/2\sqrt{42}$	$\sqrt{2}/3\sqrt{7}$
$(\bar{1}1\bar{1}1)(101\bar{1})$						$-\sqrt{5}/6$	$1/\sqrt{8}$	$\sqrt{2}/3$	$-9/\sqrt{42}$	$-\sqrt{2}/3\sqrt{7}$

10x24 cont.

	(1101)			(2000)	(1101)	(0100)	(0011)		(1101)	
	(20)(2)	(01)(2)	(01)(0)	(20)(0)	(20)(0)	(01)(0)	(01)(2)	(01)(0)	(12)(0)	
	(01)(0)	(01)(0)	(01)(0)	(01)(0)	(01)(0)	(01)(0)	(01)(0)	(01)(0)	$(01)_1(0)$	$(01)_2(0)$
$(1\bar{1}10)(0000)_1$			1/2	-1/2	1/2	$-1/\sqrt{10}$			$1/\sqrt{2}$	
$(1\bar{1}10)(0000)_2$		$\sqrt{2}/\sqrt{3}$	1/6	1/2	-1/2	$1/3\sqrt{10}$	$1/\sqrt{3}$	1/3	$-1/\sqrt{2}$	
$(1\bar{1}10)(0000)_3$		$-\sqrt{2}/\sqrt{3}$	1/3			$-2\sqrt{2}/3\sqrt{5}$	$-1/\sqrt{3}$	-1/3		$-\sqrt{2}/\sqrt{3}$
$(1\bar{1}10)(0000)_4$			1/3			$\sqrt{2}/3\sqrt{5}$		2/3		$\sqrt{2}/\sqrt{3}$
$(\bar{1}010)(2\bar{1}00)$	1/2	$1/2\sqrt{3}$	$1/3\sqrt{2}$	$1/2\sqrt{2}$	$1/2\sqrt{2}$	$\sqrt{5}/6$	$-1/\sqrt{8}$	$-1/3\sqrt{2}$		
$(0\bar{1}0\bar{1})(1001)$			$-1/6\sqrt{2}$	$1/2\sqrt{2}$	$-1/2\sqrt{2}$	$\sqrt{5}/6$		$\sqrt{2}/3$	1/2	
$(100\bar{1})(0\bar{1}11)$			$1/6\sqrt{2}$	$1/2\sqrt{2}$	$-1/2\sqrt{2}$	$-\sqrt{5}/6$		$-\sqrt{2}/3$	1/2	$1/\sqrt{3}$
$(10\bar{1}1)(0\bar{1}2\bar{1})$	1/2	$-1/2\sqrt{3}$	$1/3\sqrt{2}$	$-1/2\sqrt{2}$	$-1/2\sqrt{2}$	$\sqrt{5}/6$	$1/\sqrt{8}$	$-1/3\sqrt{2}$		
$(0\bar{1}01)(101\bar{1})$	1/2	$1/2\sqrt{3}$	$-1/3\sqrt{2}$	$-1/2\sqrt{2}$	$-1/2\sqrt{2}$	$-\sqrt{5}/6$	$-1/\sqrt{8}$	$1/3\sqrt{2}$		
$(0100)(1\bar{2}10)$	-1/2	$1/2\sqrt{3}$	$1/3\sqrt{2}$	$-1/2\sqrt{2}$	$-1/2\sqrt{2}$	$\sqrt{5}/6$	$-1/\sqrt{8}$	$-1/3\sqrt{2}$		

10x24 cont.

	(1101)						(2000)	(0011)		(0100)
	(11)(2)		(11)(0)		(00)(2)	(00)(0)	(00)(2)	(11)(0)		(00)(0)
	$(00)_1(0)$	$(00)_2(0)$	$(00)_1(0)$	$(00)_2(0)$	(00)(0)	(00)(0)	(00)(0)	$(00)_1(0)$	$(00)_2(0)$	(00)(0)
$(10\bar{1}1)(2\bar{1}00)$	-1/2		$1/2\sqrt{3}$		$1/\sqrt{24}$	1/6	$-1/2\sqrt{2}$	$1/\sqrt{8}$		$-\sqrt{5}/6$
$(\bar{1}010)(01\bar{2}1)$	-1/2	-1/2	$-1/2\sqrt{3}$	$-1/2\sqrt{3}$	$-1/\sqrt{24}$	1/6	$1/2\sqrt{2}$	$-1/\sqrt{8}$	$-1/\sqrt{8}$	$-\sqrt{5}/6$
$(0\bar{1}00)(\bar{1}0\bar{1}1)$	-1/2		$-1/2\sqrt{3}$		$1/\sqrt{24}$	-1/6	$-1/2\sqrt{2}$	$-1/\sqrt{8}$		$\sqrt{3}/6$
$(0\bar{1}01)(1\bar{2}\bar{1}0)$	1/2	1/2	$-1/2\sqrt{3}$	$-1/2\sqrt{3}$	$1/\sqrt{24}$	1/6	$-1/2\sqrt{2}$	$-1/\sqrt{8}$	$-1/\sqrt{8}$	$-\sqrt{5}/6$
$(00\bar{1}0)(\bar{1}\bar{1}01)$		1/2		$-1/2\sqrt{3}$	$-1/\sqrt{24}$	-1/6	$1/2\sqrt{2}$		$-1/\sqrt{8}$	$\sqrt{3}/6$
$(\bar{1}\bar{1}0\bar{1})(00\bar{1}2)$		1/2		$1/2\sqrt{3}$	$-1/\sqrt{24}$	1/6	$1/2\sqrt{2}$		$1/\sqrt{8}$	$-\sqrt{3}/6$
$(\bar{1}\bar{1}\bar{1}1)(0000)_1$			$\sqrt{2}/\sqrt{3}$			$\sqrt{2}/3$		$-1/\sqrt{3}$		$\sqrt{2}/3\sqrt{5}$
$(\bar{1}\bar{1}\bar{1}1)(0000)_2$			$-\sqrt{2}/\sqrt{3}$		$\sqrt{3}/2$	$1/3\sqrt{2}$	1/2	$1/\sqrt{3}$		$1/3\sqrt{10}$
$(\bar{1}\bar{1}\bar{1}1)(0000)_3$				$-\sqrt{2}/\sqrt{3}$	$-\sqrt{3}/2$	$1/3\sqrt{2}$	-1/2		$1/\sqrt{3}$	$1/3\sqrt{10}$
$(\bar{1}\bar{1}\bar{1}1)(0000)_4$				$\sqrt{2}/\sqrt{3}$		$\sqrt{2}/3$			$-1/\sqrt{3}$	$\sqrt{2}/3\sqrt{5}$

vector whose SU(5) weight is transformed into (00)(2). Clearly one has $(00)(2) = (\bar{2}200)P$; hence

$$\binom{00}{00}\binom{2}{2} = (\bar{1}100)(\bar{1}100). \tag{34}$$

Similar comments apply to the rest of Table IV'.

The coefficients in the eight linear combinations given in Table IV' represent all CGC in this problem. Applying the operators R_{β_1}, R_{β_2}, and R_γ as presented by (17) to the entries in the table, one produces the rest of the space. For instance,

$$R_{\beta_1}\binom{20}{20}\binom{0}{0} = \binom{20}{\bar{2}2}\binom{0}{0}. \tag{35}$$

Using (18), one has also

$$R_{\beta_1}(1000)(1000)$$
$$= (R_{\beta_1}(1000))(R_{\beta_1}(1000)) = (0\bar{1}10)(0\bar{1}10). \tag{36}$$

Combining (35) and (36), we get

$$\binom{20}{\bar{2}2}\binom{0}{0} = (0\bar{1}10)(0\bar{1}10). \tag{37}$$

5. CONCLUDING REMARKS

Let us point out several features of the method used in this paper. The first is its algorithmic nature: It allows one to calculate relatively easily by hand almost any CGC one may want. However, it provides no formulas for them. The second is the systematic use of relation (17) which generalizes the well-known symmetry relation of the SU(2) Clebsch–Gordan coefficients:

$$C(l_1,l_2,l;m_1,m_2,m_1+m_2)$$
$$= (-1)^{l-l_1-l_2}C(l_1,l_2,l;-m_1,-m_2,-m_1-m_2).$$

The third distinctive feature is the systematic use of weight systems of representations. It provides the generality of the method with respect to the choice of the group, its representation, and also the subgroup; Eqs. (16) and (17) are valid for any simple Lie group.

Finally, let us point out that the economy in computing CGC using the property (17) of the operators defined by (16) is approximately equal to the order $|W|$ of the Weyl group W. One has $|W| = n!$ for SU(n).

ACKNOWLEDGMENTS

We are grateful to Dr. R. V. Moody, Dr. R. Slansky, and Dr. R. T. Sharp for helpful discussions.

[1] J. J. de Swart, Rev. Mod. Phys. **35**, 916 (1963); L. C. Biedenharn, J. Math. Phys. **4**, 436 (1963); G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 1449 (1963); J. C. Carter, J. J. Coyne, and S. Meshkov, Phys. Rev. Lett. **14**, 523, 850 (1965);R. T. Sharp and H. von Baeyer, J. Math. Phys. **7**, 1105 (1966); C. Chew and R. T. Sharp, Nucl. Phys. B **2**, 697 (1967); J. Louck, Am. J. Phys. **38**, 3 (1970); V. Vanagas, *Algebraic Methods in the Nuclear Theory* (Mintis, Vilnius, 1971); M. Hongoh and R. T. Sharp, J. Math. Phys. **15**, 782 (1974); B. G. Wybourne, *Classical Groups for Physicists* (Wiley, New York, 1974); V. Rabl, G. Campbell, Jr., and K. C. Wali, J. Math. Phys. **16**, 2494 (1975); V. A. Karasev and L. A. Shelepin, in *Group Theoretical Methods in Physics* (Consultants Bureau, New York, 1975); E. M. Haacke, J. W. Moffat, and P. Savaria, J. Math. Phys. **17**, 2041 (1976); S. I. So and D. Strottman, J. Math. Phys. **20**, 153 (1979); R. Anderson and G. C. Joshi, J. Math. Phys. **20**, 1015 (1979); A. V. Klimyk and A. M. Garvilik, J. Math. Phys. **20**, 1624 (1979); D. Strottman, J. Math. Phys. **20**, 1643 (1979); A. V. Klimyk, *Matrix Elements and Clebsch–Gordan Coefficients of Group Representations* (Naukova Dumka, Kiev, 1979); U. Abresch, A. Bovier, O. Lechtenfeld, M. Lüliug, V. Rittenberg, and G. Weymans, J. Math. Phys. **21**, 2481 (1980).

[2] M. Gell-Mann, P. Ramond, and R. Slansky, Rev. Mod. Phys. **50**, 721 (1978); R. Slansky, Phys. Rep. **79**, 1 (1981); G. Feldman, T. Fulton, and P. T. Matthews, J. Phys. G: Nucl. Phys. **8**, 295 (1982).

[3] H. Georgi and C. Jarlskog, Phys. Lett. B **86**, 297 (1979); H. Georgi and D. Nanopulos, Nucl. Phys. B **159**, 16 (1979).

[4] R. V. Moody and J. Patera, Bull. Am. Math. Soc. **7**, 237 (1982).

[5] R. V. Moody and J. Patera, "Charge conjugation operators in simple Lie groups," Preprint CRMA-1097, 1982.

[6] J. Patera, "A universal method for computing Clebsch–Gordan coefficients," Preprint CRMA-58, 1970, unpublished.

[7] A. Navon and J. Patera, J. Math. Phys. **8**, 489 (1967).

[8] W. McKay, J. Patera, and D. Sankoff, in *Computers in Nonassociative Rings and Algebras*, edited by R. Beck and B. Kolman (Academic, New York, 1977).

[9] J. Patera and R. T. Sharp, Phys. Rev. D **25**, 1141 (1982).

Tensorial analysis of the hyperfine interaction operator by extensive use of Racah algebra and of translational invariance

Lionel Bréhamet
IRF DB/SBPh, Centre d'Etudes Nucléaires de Saclay, 91191 Gif-sur-Yvette Cedex, France

(Received 30 April 1982; accepted for publication 25 February 1983)

Strict application of Racah algebra gives an efficient way of deriving the tensorial form of the hyperfine interaction one-electron operator involved in NMR spin coupling effects. In particular, this procedure avoids the use of classic explicit arguments of integration around the origin to find the Fermi-contact term. The linear momentum is retained in its explicit translationally invariant form as long as possible during the calculations. This allows the tensorial expansion of the operator to be obtained rapidly at any origin of coordinates. The reduced matrix elements of interest are given by new general closed expressions in the $j-j$ coupling scheme.

PACS numbers: 02.20. + b, 33.10.Cs, 33.25.Fs

I. INTRODUCTION

Standard Racah algebra methods[1] have long been used to solve problems of electronic spin-spin or spin-other-orbit interactions.[2–7] This procedure is suitable for studying interactions between nuclear and electronic moments in a molecular system. We are going here to make systematic use of Racah algebra for an examination of the hyperfine interaction operator H involved in the coupling between nuclear spins.[8] According to the usual rules and notations of vectorial calculus, the scalar interaction H may be written[8–10] as the sum of a spin-dipolar term H_s and of an orbital term H_l

$$H = H_s + H_l, \tag{1}$$

where

$$H_s = -\mathbf{s}\cdot(\nabla\times(\nabla\times\boldsymbol{\mu}/r)), \tag{2}$$

$$H_l = \mathbf{p}\cdot(\nabla\times\boldsymbol{\mu}/r). \tag{3}$$

In tensorial notations, the resulting equations have been given[1,11,12] as follows:

$$H = \{\mathsf{H}\cdot\mu\}, \tag{4}$$

where

$$\mathsf{H} = \mathsf{H}_s + \mathsf{H}_l, \tag{5}$$

$$\mathsf{H}_s = (8\pi/3)\,\mathsf{s}\delta(\mathsf{r}) - \sqrt{10}r^{-3}(\mathsf{sC}^{(2)})^{(1)}, \tag{6}$$

with the prescription of deleting the r^{-3} term for S-states, and

$$\mathsf{H}_l = r^{-3}\mathsf{l}. \tag{7}$$

The factor 2β (β being the Bohr magneton) has been omitted. In these equations, s stands for the electronic spin, p for the linear momentum, and ∇ for the usual gradient operator. The electronic radius-vector r, the electronic angular momentum l, and the spherical tensor of rank 2, $C^{(2)}$, are assumed to be relative to the same origin of coordinates, denoted by 0, where the nuclear spin μ is localized. The tridimensional Dirac delta distribution $\delta(\mathsf{r})$ describes the Fermi-contact term[13] [so that the use of the term "spin-dipolar" in Eq. (2) is somewhat improper].

The systematic use of tensor algebra provides a most effective method for deducing Eq. (6) from Eq. (2), for example, and the Fermi-contact term may be obtained without classic explicit argument of infinitesimal integration around the origin. By further retaining the translationally invariant explicit form p as long as possible during the calculations, we have an immediate means of expanding H_s as the sum of operators relative to another arbitrary origin. The only expansion required will be that of $1/r$ in Legendre polynomials. Some developments are formally equivalent to those found in two-body studies,[5,7] but, in this case, such an extensive use of the p operator invariance is excluded.

IIA. THE H_s OPERATOR

Equation (2) may be written in tensorial notations as

$$H_s = 2\sqrt{3}\,\{\mathsf{s}\{\mathsf{p}\{\mathsf{p}\mu\}^{(1)}\}^{(1)}\}^{(0)}\left(\frac{1}{r}\right). \tag{8}$$

According to Racah,[1] simple recouplings with $6-j$ Wigner coefficients allow the nuclear spin part to be separated from the rest and give the scalar product $H_s = \{\mathsf{H}_s\cdot\mu\}$ with

$$\mathsf{H}_s = \left[-\frac{2\sqrt{3}}{3}\,\mathsf{s}(\mathsf{pp})^{(0)} + \{\mathsf{s}(\mathsf{pp})^{(1)}\}^{(1)} + \sqrt{\frac{5}{3}}\,\{\mathsf{s}(\mathsf{pp})^{(2)}\}^{(1)}\right]\left(\frac{1}{r}\right). \tag{9}$$

The general expression for $(\mathsf{pp})^{(k)}$ is given in Appendix A by Eq. (A6). We may now inspect the cases $k = 0$ or 1 and especially the nontrivial case $k = 2$. The tables of Rotenberg *et al.*[14] and Edmonds' formulas[12] are used to evaluate some $n-j$ Wigner coefficients occuring in Eq. (A6) for the different cases.

1. The $(\mathsf{pp})^{(0)}$ term

The coefficients of $(1/r)(\partial/\partial r)$ and of $-\partial^2/\partial r^2$ are easily calculated from Eq. (A6). Using the identity (B2) and the l-commutation relation to evaluate the coefficient of $-1/r^2$, we find

$$(\mathsf{pp})^{(0)} = -\frac{1}{\sqrt{3}}\left[\mathsf{l}^2 - \frac{\partial^2}{\partial r^2} - \frac{2}{r}\frac{\partial}{\partial r}\right]. \tag{10}$$

As expected, this expression gives the Laplacian in spherical coordinates, thus

$$(\mathsf{pp})^{(0)} = (1/\sqrt{3})\Delta. \tag{11}$$

 0022-2488/83/081968-05$02.50

2. The $(pp)^{(1)}$ term

The coefficient of interest here is that of $-1/r^2$. Using the identity (B1) and the value of a special $9-j$ Wigner coefficient,

$$\begin{Bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 2 & 1 & 1 \end{Bmatrix} = \frac{1}{18},$$

we find the final result $(pp)^{(1)}$ to be zero, as expected from $\nabla \times \nabla = 0$.

3. The $(pp)^{(2)}$ term

With the aid of the equality

$$\begin{Bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 2 & 2 & 2 \end{Bmatrix} = \frac{1}{30}\sqrt{\frac{7}{3}},$$

we obtain a preliminary expression:

$$\begin{aligned}(pp)^{(2)} = &-\frac{1}{r^2}\left[\frac{1}{3}\mathsf{I}^{(2)} + \frac{\sqrt{14}}{3}(\mathsf{C}^{(2)}\mathsf{I}^{(2)})^{(2)}\right. \\ &\left. - \frac{\sqrt{2}}{3}\mathsf{C}^{(2)}(\mathsf{II})^{(0)} + 2(\mathsf{C}^{(2)}\mathsf{I})^{(2)}\right] \\ &+ \frac{1}{r}\frac{\partial}{\partial r}\left[2(\mathsf{C}^{(2)}\mathsf{I})^{(2)} + \frac{2}{\sqrt{6}}\mathsf{C}^{(2)}\right] \\ &- \frac{\partial^2}{\partial r^2}\frac{2}{\sqrt{6}}\mathsf{C}^{(2)}, \end{aligned} \tag{12}$$

where the $\mathsf{I}^{(2)}$ notation is that given in Appendix B. If we insert $-(4/\sqrt{6})\mathsf{C}^{(2)}$ as a multiplicative coefficient of $(1/r)(\partial/\partial r)$, the Laplacian appears again, but associated this time with the spherical tensor $\mathsf{C}^{(2)}$. Using the identity (B4), we obtain the final result

$$\begin{aligned}(pp)^{(2)} = &-\frac{\sqrt{6}}{3}\mathsf{C}^{(2)}\Delta + [\sqrt{6}\mathsf{C}^{(2)} + 2(\mathsf{C}^{(2)}\mathsf{I})^{(2)}]\frac{1}{r}\frac{\partial}{\partial r} \\ &- \left[\mathsf{I}^{(2)} + (\mathsf{C}^{(2)}\mathsf{I})^{(2)} + 2\sqrt{\frac{2}{3}}\mathsf{C}^{(2)}\mathsf{I}^2\right]\frac{1}{r^2}. \end{aligned} \tag{13}$$

The operator H_s is thus given by

$$\begin{aligned}\mathsf{H}_s = \left\{\mathsf{s}\left[\left[-\frac{2}{3}\Delta - \sqrt{\frac{10}{3}}\mathsf{C}^{(2)}\Delta\right] + \left[\sqrt{10}\mathsf{C}^{(2)}\right.\right.\right. \\ \left. + 2\sqrt{\frac{5}{3}}(\mathsf{C}^{(2)}\mathsf{I})^{(2)}\right]\frac{1}{r}\frac{\partial}{\partial r} - \sqrt{\frac{5}{3}}\left[\mathsf{I}^{(2)} + (\mathsf{C}^{(2)}\mathsf{I})^{(2)}\right. \\ \left.\left.\left. + 2\sqrt{\frac{2}{3}}\mathsf{C}^{(2)}\mathsf{I}^2\right]\frac{1}{r^2}\right]\right\}^{(1)}\left(\frac{1}{r}\right). \end{aligned} \tag{14}$$

If all the terms of this equation, which are defined from a given origin, are assumed to be relative to the same one, i.e., 0, those including I give exactly zero when operating on $1/r$.

Using now the relation $\Delta(1/r) \equiv -4\pi\delta(\mathsf{r})$ (this identity implicitly contains a result of integration theory), we find that the remaining terms in Eq. (14) are $(8\pi/3)\delta(\mathsf{r})$ and $-\sqrt{10}[-(4\pi/3)C^{(2)}\delta(\mathsf{r}) + \mathsf{C}^{(2)}r^{-3}]$. It is important here to discuss these terms because they have formed the subject of lengthy analyses,[9,15,16] in connection with the theory of hyperfine structure splitting. Complicated arguments involving integration of infinitesimal spheres have been used, necessitating prescriptions to indicate the path for integrating around the origin. With our formulation, no *explicit* argument of this type is necessary. First, the term $(8\pi/3)\delta(\mathsf{r})$ appears straightforwardly, as being the Fermi-contact term.[13] Secondly, if spherically symmetrical wave functions (S-state electron) are involved, the problem of the indeterminate integral $\langle S|\mathsf{C}^{(2)}r^{-3}|S\rangle$, which has the form $0\times\infty$, finds an immediate solution. Mere inspection of the second term (multiplied by $-\sqrt{10}$) will give the appropriate mathematical prescription. Indeed, no experimental data have yet required the use of an additional contact term; therefore,

$$\langle S|\mathsf{C}_q^{(2)}r^{-3}|S\rangle = (4\pi/3)\delta_{q0}\langle S|\delta(\mathsf{r})|S\rangle, \tag{15}$$

a result which previously demanded lengthy proof.[16]

IIB. THE H'_s DISPLACED OPERATOR

Let us now suppose that the nuclear moment μ' is given relative to an origin 0′ different from 0. We wish to expand the corresponding H'_s operator as a sum of operators relative to the first origin 0 (primed quantities are taken with respect to 0′). Thus

$$\mathsf{H}'_s = \sum_{k=0}^{\infty}\mathsf{h}_s^k. \tag{16}$$

The advantage of Eq. (14) is as follows. The complicated but translationally invariant tensor of rank 1 operating on $1/r'$ may be expressed in relation to any origin, especially that of interest, 0. It only remains to expand $1/r'$ in Legendre polynomials. There is no need to expand higher powers $1/r'^k$, with $k \geqslant 2$, as used in standard references.[11]

The notations used now will be more or less classic, but require some explanation. If 1 stands for the electron, we have $0'1 = r'$, $00' = d$, $r_< = \inf(r',d)$, $r_> = \sup(r',d)$. Two spherical tensors will appear: $C_0^{(k)}(1)$ denoted by $\mathsf{C}^{(k)}$, and $C_0^{(k)}(0')$ by $C'^{(k)}$, with $C_q'^{(k)} = \delta_{q0}$ if $0 = 0'$. $C'^{(k)}$ is a kind of geometrical constant, but the meaning of $\{\mathsf{C}^{(s)}C'^{(t)}\}^{(k)}$ must be understood as a classic tensorial form. Expansion of $1/r'$ in Legendre polynomials may be given as

$$\frac{1}{r'} = \sum_k \frac{r_<{}^k}{r_>{}^{k+1}}(-1)^k(2k+1)^{1/2}\{\mathsf{C}^{(k)}C'^{(k)}\}^{(0)}. \tag{17}$$

Now, if we take Eqs. (14) and (17), we must keep in mind the fact that the operator I acts on C. In terms of the form $\{\mathsf{C}\{\mathsf{C}C'\}\}$, the usual recouplings between C have to be performed. Replacing Eq. (17) in (14) and using the result

$$\begin{aligned}&(\mathsf{II})^{(t)}\{\mathsf{C}^{(k)}C'^{(k)}\}^{(0)} \\ &= (-1)^t(2k+1)^{1/2}[k(k+1)]\begin{Bmatrix} k & k & t \\ 1 & 1 & k \end{Bmatrix} \\ &\quad \{\mathsf{C}^{(k)}C'^{(k)}\}^{(t)}, \end{aligned} \tag{18}$$

we see that most of the detailed recouplings are formally equivalent to those of Horie,[5] although we are concerned here with monoelectronic operators. As a result, a sum of two complementary operators is found, acting for radial integration from 0 to d (when r^k is in numerator) and from d to ∞ (when r^k is in denominator). We shall use this integration convention later. Thus

$$H'_s = \left[\frac{8\pi}{3}s + \frac{4\pi}{3}\sqrt{10}(sC^{(2)})^{(1)}\right]\delta(r')$$
$$- \sum_{k=0}^{\infty} \frac{1}{2\sqrt{3}}(-1)^k \left[\frac{(2k+5)!}{(2k)!}\right]^{1/2}$$
$$\times\left[\frac{r^k}{d^{k+3}}\{(sC^{(k)})^{(k+1)}C'^{(k+2)}\}^{(1)}\right.$$
$$\left. + \frac{d^k}{r^{k+3}}\{(sC^{(k+2)})^{(k+1)}C'^{(k)}\}^{(1)}\right]. \quad (19)$$

Equation (19) gives a quite clear means to understand the nil translation limit ($d \to 0$). Firstly, the angular integration (in $\sin\theta\, d\theta\, d\phi$) must be performed. This operation sets the $k = 0$ value if the apparently problematic S-states case is involved. Secondly, the radial integration $\int_0^d$ (in $r^2\, dr$) is performed on the only nonvanishing $(1/d^3)\{sC'^{(2)}\}^{(1)}$ term and, thirdly, d is set equal to zero. The result cancels exactly with the $(sC^{(2)})^{(1)}\delta(r')$ term. This correct procedure is not conflicting with Eq. (15), but agrees perfectly with it.

The Dirac distribution can also be expanded using the relation

$$\delta(r') = \frac{1}{4\pi}\sum_k (2k+1)\frac{\delta(r-d)}{r^2}P_k(\cos\omega), \quad (20)$$

where P_k is a Legendre polynomial and ω the angle between r and r'. Standard recouplings and some algebraic rearrangments lead to the final form for h_s^k, defined in Eq. (16):

$$h_s^k = \frac{1}{\sqrt{3}}(-1)^k \frac{\delta(r-d)}{r^2}$$
$$\times[\{-(2k+1)^{3/2}(sC^{(k)})^{(k)}C'^{(k)}$$
$$+ [(k+2)(2k+1)](2k+3)^{-1/2}(sC^{(k)})^{(k+1)}C'^{(k)}$$
$$+ [(k-1)(2k+1)](2k-1)^{-1/2}$$
$$\times(sC^{(k)})^{(k-1)}C'^{(k)}\}^{(1)}]$$
$$+ \frac{1}{2\sqrt{3}}(-1)^k\left[\frac{(2k+5)!}{(2k)!}\right]^{1/2}$$
$$\times\left[\left[(2k+3)^{-1}\frac{\delta(r-d)}{r^2} - \frac{r^k}{d^{k+3}}\right]\right.$$
$$\times\{(sC^{(k)})^{(k+1)}C'^{(k+2)}\}^{(1)}$$
$$+ \left[(2k+3)^{-1}\frac{\delta(r-d)}{r^2} - \frac{d^k}{r^{k+3}}\right]$$
$$\left.\times\{(sC^{(k+2)})^{(k+1)}C'^{(k)}\}^{(1)}\right], \quad (21)$$

where if $k = 0$, some terms obviously have to be deleted. Given the usual normalized S-state wave functions, it can be further deduced from Eq. (21) (when $k = 0$ and $d = 0$) that the classic Fermi-contact term can be written for practical use as $\frac{2}{3}(\delta(r)/r^2)s$.

IIIA. THE H_l OPERATOR

In order to obtain the full displaced H' hyperfine operator, it is desirable to establish a similar formula for the orbital part H_l. The presence of the angular momentum l precludes the crucial problems found in Sec. II for limiting cases. In contrast to Sec. II, we note that the translational invariance property can no longer be used. In fact, in Eq. (3) the term $1/r$ cannot be extracted and replaced at the right of a translationally invariant operator. As for Eq. (16), we write

$$H'_l = \sum_{k=0}^{\infty} h_l^k. \quad (22)$$

From systematic recouplings and with the radial integration convention used in Eq. (19), we obtain

$$h_l^k = \frac{1}{\sqrt{3}}(-1)^k$$
$$\times\left[\left\{[k+2](2k+3)^{1/2}\frac{r^{k-1}}{d^{k+2}}(C^{(k+1)}l)^{(k+1)}C'^{(k+1)}\right.\right.$$
$$+ [2k+3](2k+1)^{1/2}\frac{r^{k-1}}{d^{k+2}}(C^{(k+1)}l)^{(k)}C'^{(k+1)}$$
$$\left.+ [(k+1)(k+2)(2k+3)]^{1/2}\frac{r^k}{d^{k+2}}C^{(k+1)}C'^{(k+1)}\frac{\partial}{\partial r}\right\}$$
$$+ \left\{k(2k+1)^{1/2}\frac{d^k}{r^{k+3}}(C^{(k)}l)^{(k)}C'^{(k)}\right.$$
$$+ [2k+1](2k+3)^{1/2}\frac{d^k}{r^{k+3}}(C^{(k)}l)^{(k+1)}C'^{(k)}$$
$$\left.\left.- [k(k+1)(2k+1)]^{1/2}\frac{d^k}{r^{k+2}}C^{(k)}C'^{(k)}\frac{\partial}{\partial r}\right\}^{(1)}\right]. \quad (23)$$

Some terms clearly vanish if $k = 0$. As expected, the nil translation limit ($d = 0$) gives Eq. (7):

$$H_l = r^{-3}l.$$

We shall now be able to show that the reduced matrix elements of interest can be expressed in a closed form if one special basis of eigenfunctions is used.

IVA. GENERAL REDUCED MATRIX ELEMENTS IN THE $J-J$ COUPLING SCHEME

Formulas such as (19) or (21), and (23) are useful for the evaluation of molecular integrals, as already emphasized[16] in the molecular physics domain. Moreover, the electronic operators $(sC^{(s)})^{(t)}$ and $(C^{(s)}l)^{(t)}$ have explicit reduced matrix elements if we make use of the $j-j$ coupling scheme.[1] This scheme is a useful framework for carrying out relativistic calculations,[17-19] where each of the four components of the one-electron Dirac wave function includes standard coupled functions such as $|l\,\tfrac{1}{2}\,JM\rangle$ (the total angular momentum is denoted by $J = l + s$, and M is a value of J_z). Thus for electrons we need formulas for the reduced matrix elements $(l\,\tfrac{1}{2}J\,\|(sC^{(s)})^{(t)}\|\,l'\,\tfrac{1}{2}J')$ and $(l\,\tfrac{1}{2}J\,\|(C^{(s)}l)^{(t)}\|\,l'\,\tfrac{1}{2}J')$.

Separate formulas have already been given[20] for $(sC^{(s)})^{(t)}$, but none for $(C^{(s)}l)^{(t)}$. However, the formulas unfortunately conceal a general structure, which enables the twelve possible cases for l, J, s, t, l', J' to be grouped together and exhibits some quality of symmetry. Such new formulas will be given below. If we use standard techniques for decoupling and recoupling, with the basic intention of exhibiting the fundamental selection rule by means of the $3-j$ Wigner coefficient

$$\begin{pmatrix} J & t & J' \\ \frac{1}{2} & 0 & -\frac{1}{2} \end{pmatrix}$$

occurring in nuclear physics,[21] the formulas can be obtained by straightforward deduction.[22] It should be pointed out here that, in the course of calculations, an additional error[23,24] was found in the Tables of Rotenberg *et al.*[14] The recursion relation on the magnetic numbers (Formula 1.48, p. 10) must be read as

$$
\begin{aligned}
&-[(j_3+m_1+m_2+1)(j_3-m_1-m_2)]^{1/2}\\
&\quad\times\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & -m_3+1 \end{pmatrix}\\
&=[(j_2+m_2+1)(j_2-m_2)]^{1/2}\\
&\quad\times\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2+1 & -m_3 \end{pmatrix}\\
&\quad+[(j_1+m_1+1)(j_1-m_1)]^{1/2}\\
&\quad\times\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1+1 & m_2 & -m_3 \end{pmatrix}.
\end{aligned}
\tag{24}
$$

Our results giving the reduced matrix elements are the following for $s+t>0$:

$$
\begin{aligned}
&(l\tfrac{1}{2}J\,\|(\mathsf{sC}^{(s)})^{(t)}\|\,l'\tfrac{1}{2}J')\\
&\quad=[8(2s+1)(s+t+1)]^{-1/2}(-1)^{J+1/2}\\
&\quad\times[(2J+1)(2J'+1)]^{1/2}\\
&\quad\times\tfrac{1}{2}[1+(-1)^{l+s+l'}][(-1)^{J+1/2+l}(2J+1)\\
&\quad+(-1)^{J'+1/2+l'}(2J'+1)+s(s+1)-t(t+1)]\\
&\quad\times\begin{pmatrix} J & t & J' \\ \frac{1}{2} & 0 & -\frac{1}{2} \end{pmatrix}.
\end{aligned}
\tag{25}
$$

$$
\begin{aligned}
&(l\tfrac{1}{2}J\,\|(\mathsf{C}^{(s)}\|)^{(t)})\|\,l'\tfrac{1}{2}J')\\
&\quad=(-1)^{s+t}[(1-t+s)!(1+t-s)!]^{-1/2}(2t+1)^{1/2}\\
&\quad\times\left[\frac{(s+t-1)!}{(s+t+2)!}\right]^{1/2}(-1)^{J+1/2}[(2J+1)(2J'+1)]^{1/2}\\
&\quad\times\tfrac{1}{2}[1+(-1)^{l+s+l'}]\begin{pmatrix} J & t & J' \\ \frac{1}{2} & 0 & -\frac{1}{2} \end{pmatrix}\\
&\quad\times\tfrac{1}{4}[[(-1)^{J+1/2+l}(2J+1)-(-1)^{s+t}\\
&\quad\times(-1)^{J'+1/2+l'}(2J'+1)]\\
&\quad\times[(-1)^{J+1/2+l}(2J+1)+(-1)^{J'+1/2+l'}\\
&\quad\times(2J'+1)+2]-4t(t+1)].
\end{aligned}
\tag{26}
$$

As initially anticipated, Formulas (25) and (26) display some analogy with the well-known equality

$$
\begin{aligned}
&(l\tfrac{1}{2}J\,\|\mathsf{C}^{(s)}\|\,l'\tfrac{1}{2}J')\\
&\quad=(-1)^{J+1/2}[(2J+1)(2J'+1)]^{1/2}\tfrac{1}{2}[1+(-1)^{l+s+l'}]\\
&\quad\times\begin{pmatrix} J & s & J' \\ \frac{1}{2} & 0 & -\frac{1}{2} \end{pmatrix}.
\end{aligned}
\tag{27}
$$

In particular, Eq. (26) immediately leads to previously obtained results,[7] e.g., the "proportionality" between the tensors $(\mathsf{C}^{(k-1)}\|)^{(k)}$ and $(\mathsf{C}^{(k+1)}\|)^{(k)}$, and between the reduced matrix elements of $(\mathsf{C}^{(k)}\|)^{(k)}$ and $\mathsf{C}^{(k)}$.

V. CONCLUSION

In a somewhat unconventional but systematic manner, we have used Racah algebra to inspect the usual hyperfine interaction one-electron operator. The method offers a most effective means not only straightforwardly obtaining tensorial formulas at any origin of coordinates, but even of solving problems of infinitesimal analysis. This report concludes with new general formulas for reduced matrix elements of interest in the $j-j$ coupling scheme.

APPENDIX A: GENERAL EXPRESSION OF THE TENSOR $(\mathsf{pp})^{(k)}$

According to Innes and Ufford,[7] the linear momentum p can be written

$$
\mathsf{p}=i\frac{\sqrt{2}}{r}(\mathsf{C}^{(1)}\|)^{(1)}-i\frac{\partial}{\partial r}\mathsf{C}^{(1)},
\tag{A1}
$$

where l stands for the angular momentum and where $\mathsf{C}^{(m)}$ is the spherical tensor of rank m. Thus formally (dropping out the k tensor rank superscript)

$$
\begin{aligned}
(\mathsf{pp})=&-\frac{1}{r^2}[2((\mathsf{C}^{(1)}\|)^{(1)}(\mathsf{C}^{(1)}\|)^{(1)})\\
&+\sqrt{2}(\mathsf{C}^{(1)}(\mathsf{C}^{(1)}\|)^{(1)})]\\
&+\frac{\sqrt{2}}{r}\frac{\partial}{\partial r}[(\mathsf{C}^{(1)}(\mathsf{C}^{(1)}\|)^{(1)})\\
&+((\mathsf{C}^{(1)}\|)^{(1)}\mathsf{C}^{(1)})]\\
&-\frac{\partial^2}{\partial r^2}(\mathsf{C}^{(1)}\mathsf{C}^{(1)}).
\end{aligned}
\tag{A2}
$$

Using standard recoupling techniques, and taking the necessary precautions when handling noncommutating operators, we find that all the tensors $(\ )^{(k)}$ can easily be expressed as function of $(\mathsf{C}^{(s)}\|)^{(k)}$ and $\mathsf{C}^{(k)}$, except one: $((\mathsf{C}^{(1)}\|)^{(1)}(\mathsf{C}^{(1)}\|)^{(1)})^{(k)}$, which is more difficult to reduce. If the operators $\mathsf{C}^{(1)}$ and l were commutating, the reduced matrix elements of the last mentioned tensor would be the following:

$$
\begin{aligned}
&(l\,\|\sum_{s,t}[(2s+1)(2t+1)]^{1/2}\begin{Bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ s & t & k \end{Bmatrix}\\
&\quad\times((\mathsf{C}^{(1)}\mathsf{C}^{(1)})^{(s)})(\|)^{(t)})^{(k)}\,\|\,l')\\
&\quad=\sum_{l''}(-1)^k(2k+1)^{1/2}\begin{Bmatrix} l'' & l' & 1 \\ 1 & 1 & l' \end{Bmatrix}\\
&\quad\times(l\,\|\mathsf{C}^{(1)}\|\,l'')(l''\|\mathsf{C}^{(1)}\|\,l')\\
&\quad\times(l'\|\|\|l')^2\begin{Bmatrix} 1 & 1 & 1 \\ l'' & l' & 1 \\ l & l' & k \end{Bmatrix}.
\end{aligned}
\tag{A3}
$$

To obtain Eq. (A3) we have used the identity of Arima *et al.*[25] for the $9-j$ coefficient on the left-hand side. Using a (corrected[23]) formula given by de-Shalit and Talmi,[26] we find that

$$
\begin{aligned}
&(l'\|\|\|l'\|\|l')\begin{Bmatrix} 1 & 1 & 1 \\ l'' & l' & 1 \\ l & l' & k \end{Bmatrix}\\
&\quad=(-1)^k\left[\begin{Bmatrix} 1 & 1 & k \\ l' & l & l'' \end{Bmatrix}\begin{Bmatrix} l & l'' & l' \\ 1 & 1 & l'' \end{Bmatrix}\right.\\
&\quad\left.-\frac{1}{2\sqrt{6}}[k(k+1)-4]\begin{Bmatrix} 1 & 1 & k \\ l' & l & l'' \end{Bmatrix}\right].
\end{aligned}
\tag{A4}
$$

Replacing Eq. (A4) in (A3) results in exhibiting expressions of the reduced matrix elements of $((C^{(1)}l)^{(1)}(C^{(1)}l)^{(1)})^{(k)}$ and of $(C^{(1)}(C^{(1)}l)^{(1)})^k$. The resulting equation may be written

$$3\sum_{s,t}[(2s+1)(2t+1)]^{1/2}\begin{Bmatrix}1&1&1\\1&1&1\\s&t&k\end{Bmatrix}\times((C^{(1)}C^{(1)})^{(s)}(ll)^{(t)})^{(k)}$$
$$=(-1)^k\left[((C^{(1)}l)^{(1)}(C^{(1)}l)^{(1)})^{(k)}-\frac{\sqrt{2}}{4}[k(k+1)-4](C^{(1)}(C^{(1)}l)^{(1)})^{(k)}\right]. \quad (A5)$$

Simple standard recouplings between spherical tensors and the use of Appendix B lead to the final result

$$(pp)^{(k)}=-\frac{1}{r^2}\left[(-1)^k\sum_s 6(2s+1)\begin{pmatrix}1&1&s\\0&0&0\end{pmatrix}\times\sum_t(2t+1)^{1/2}\begin{Bmatrix}1&1&1\\1&1&1\\s&t&k\end{Bmatrix}(C^{(s)}(ll)^{(t)})^{(k)}\right.$$
$$-(-1)^k\tfrac{1}{2}(k-1)(k+2)\sum_s\sqrt{6}(2s+1)\times\begin{pmatrix}1&1&s\\0&0&0\end{pmatrix}\begin{Bmatrix}k&1&s\\1&1&1\end{Bmatrix}(C^{(s)}l)^{(k)}\Big]$$
$$+\frac{1}{r}\frac{\partial}{\partial r}\left[(2k+1)^{1/2}\begin{pmatrix}1&1&k\\0&0&0\end{pmatrix}[k(k+1)]^{1/2}\times(C^{(k)}l)^{(k)}+\tfrac{1}{2}[k(k+1)-4]C^{(k)}\right]$$
$$-\frac{\partial^2}{\partial r^2}(2k+1)^{1/2}\begin{pmatrix}1&1&k\\0&0&0\end{pmatrix}C^{(k)}. \quad (A6)$$

APPENDIX B: IDENTITIES CONNECTING THE SPHERICAL TENSOR $C^{(2)}$ AND THE ANGULAR MOMENTUM l

Starting from the relation defining l as the vector product of r and p, $l=-\sqrt{2}r(C^{(1)}p)^{(1)}$, and using Eq. (A1) as well as the commutation relation $l=-\sqrt{2}(ll)^{(1)}$, we obtain

$$\sqrt{10}(C^{(2)}l)^{(1)}=l. \quad (B1)$$

From this first identity, we can deduce three other relations by operating with l on the right of Eq. (B1) and considering the parts of rank 0, 1, and 2, respectively. Recoupling the orbital terms leads to

$$\sqrt{10}(C^{(2)}l^{(2)})^{(0)}=(ll)^{(0)}, \quad (B2)$$

where the definition of Rudzikas *et al.*[27] has been used for the "iterated" angular momentum of rank 2: $(ll)^{(2)}=l^{(2)}$. Expressed as a scalar product, Eq. (B2) becomes

$$-\sqrt{6}(C^{(2)}\cdot l^{(2)})=l^2. \quad (B2')$$

The tensors of rank 1 and 2 give us two more identities:

$$-2\sqrt{5}(C^{(2)}l^{(2)})^{(1)}=\sqrt{3}l, \quad (B3)$$
$$\sqrt{42}(C^{(2)}l^{(2)})^{(2)}+3\sqrt{3}(C^{(2)}l)^{(2)}-2\sqrt{2}C^{(2)}l^2=2\sqrt{3}l^{(2)}. \quad (B4)$$

[1]U. Fano and G. Racah, *Irreducible Tensorial Sets* (Academic, New York, 1959).
[2]I. Talmi, Phys. Rev. **89**, 1065 (1953).
[3]F. R. Innes, Phys. Rev. **91**, 31 (1953).
[4]R. E. Trees, Phys. Rev. **92**, 308 (1953).
[5]H. Horie, Prog. Theoret. Phys. Jpn. **10**, 296 (1953).
[6]J. Hope and L. W. Longdon, Phys Rev. **102**, 1124 (1956).
[7]F. R. Innes and C. W. Ufford, Phys. Rev. **111**, 194 (1958).
[8]N. F. Ramsey, Phys. Rev. **91**, 303 (1953).
[9]S. M. Blinder, J. Mol. Spectrosc. **5**, 17 (1960).
[10]J. D. Memory, *Quantum Theory of Magnetic Resonance Parameters* (McGraw-Hill, New York, 1968), Chap. 4.
[11]B. R. Judd, *Operator Techniques in Atomic Spectroscopy* (McGraw-Hill, New York, 1963).
[12]A. R. Edmonds, *Angular Momentum in Quantum Mechanics*, revised edition (Princeton U. P., Princeton, NJ, 1974).
[13]E. Fermi, Z. Phys. **60**, 320 (1930).
[14]M. Rotenberg, R. Bivins, N. Metropolis, and J. K. Wooten, Jr., *The 3−j and 6−j Symbols* (MIT Press, Cambridge, MA, 1959).
[15]H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One- and Two-Electrons Atoms* (Plenum, New York, 1957), 1st ed., Secs. 39,40; (Plenum, New York, 1977), 2nd ed., Sec. 22.
[16]R. M. Pitzer, C. W. Kern, and W. N. Lipscomb, J. Chem. Phys. **37**, 267 (1962).
[17]M. E. Rose, *Relativistic Electron Theory* (Wiley, New York, 1961).
[18]P. O. Lipas, P. Pyykkö, and G. Pajanne, J. Chem. Phys. **58**, 3248 (1973).
[19]P. Pyykkö, Chem. Phys. **22**, 289 (1976).
[20]D. M. Brink and G. R.Satchler, *Angular Momentum*, 2nd ed. (Clarendon, Oxford, 1971), Appendix VI.
[21]J. Raynal, Nucl. Phys. A **97**, 572 (1967).
[22]L. Bréhamet, thesis, University of Pierre et Marie Curie, Paris VI, 1977.
[23]C. Stassis and S. A. Williams, J. Math. Phys. **17**, 480 (1976).
[24]R. Chatterjee and J. M. Dixon, J. Math. Phys. **20**, 1622 (1979).
[25]A. Arima, H. Horie, and Y. Tanabe, Prog. Theoret. Phys. Jpn. **11**, 143 (1954).
[26]A. de-Shalit and I.Talmi, *Nuclear Shell Theory*, (Academic, New York, 1963).
[27]Z. B. Rudzikas, A. A. Bandzaitis, Ya. I. Vizbaraite, and A. P. Jucys, Liet. Fiz. Rinkinys **4**, 5 (1964).

Maximal abelian subalgebras of real and complex symplectic Lie algebras[a),b)]

J. Patera, P. Winternitz, and H. Zassenhaus[c)]
Centre de recherche de mathématiques appliquées, Université de Montréal, Montréal, Québec, Canada

(Received 7 June 1982; accepted for publication 15 October 1982)

We provide guidelines for classifying maximal abelian subalgebras (MASA's) of the symplectic Lie algebras $sp(2n, \mathbb{R})$ and $sp(2n, \mathbb{C})$ into conjugacy classes under the Lie groups $Sp(2n, \mathbb{R})$ and $Sp(2n, \mathbb{C})$, respectively. The task of classifying all MASA's is reduced to the classification of orthogonally indecomposable (OID) MASA's. Two types of orthogonally indecomposable MASA's of $sp(2n, \mathbb{C})$ exist: 1. Indecomposable maximal abelian nilpotent subalgebras (MANS's). 2. Decomposable MASA's [their classification reduces to a classification of MANS's of $sl(n, \mathbb{C})$]. Four types of orthogonally indecomposable MASA's of $sp(2n, \mathbb{R})$ exist: 1. Absolutely indecomposable MASA's (MANS's). 2. Relatively indecomposable MASA's [their classification reduces to a classification of MANS's of $su(p, q)$ for $p + q = n$]. 3. Decomposable absolutely OID MASA's [involving MANS's of $sl(n,\mathbb{R})$]. 4. Decomposable relatively OID MASA's [involving MANS's of $sl(n/2, \mathbb{C})$, for n even]. Low-dimensional cases of $sp(2n, F)$ ($n = 1, 2, 3, F = \mathbb{R}$ or $\mathbb{C}$) are treated exhaustively. The algebras $sp(2, \mathbb{R})$, $sp(4, \mathbb{R})$, and $sp(6, \mathbb{R})$ have 3, 10, and 30 classes of MASA's, respectively; $sp(2, \mathbb{C})$, $sp(4, \mathbb{C})$, and $sp(6, \mathbb{C})$ have 2, 5, and 14 classes of MASA's, respectively. For $n \geqslant 4$ infinitely many classes of MASA's exist.

PACS numbers: 02.20.Sv

1. INTRODUCTION

The purpose of this paper is to provide a classification of the maximal abelian subalgebras (MASA's) of the real and complex symplectic Lie algebras $sp(2n, \mathbb{R})$ and $sp(2n, \mathbb{C})$ into conjugacy classes under the action of the corresponding classical Lie groups $Sp(2n, \mathbb{R})$ and $Sp(2n, \mathbb{C})$, respectively.

This study is part of a general program, the aim of which is to construct all Lie subalgebras of Lie algebras of interest in physics and, similarly, to construct all closed subgroups of the corresponding Lie groups. Earlier articles have been devoted to all maximal solvable subgroups of semisimple Lie groups,[1,2] to all subgroups of the conformal group of space-time and some of its interesting subgroups.[3–6] For the motivation of our interest in subgroup classification, we refer to our previous articles: further references are listed, e.g., in Refs. 5 and 7.

Applications include the systematic study of spontaneous or explicit symmetry breaking,[7–9] the construction of symmetry adapted wave functions,[10] the separation of variables in partial differential equations,[11–14] the construction of solutions of Yang–Mills and other nonlinear equations invariant under specific subgroups of the conformal group,[9,15] the derivation of nonlinear superposition principles for certain systems of nonlinear equations[16] and many others.

The problem of finding all maximal abelian subgroups of a given Lie group is of particular interest. They provide maximal sets of additive quantum numbers for quantum mechanical systems, and the simplest integrals of motion in involution for classical ones. MASA's are related to the simplest types of separable coordinates for partial differential equations ("ignorable coordinates" for which the corresponding solutions are exponentials).[12–14] Different MASA's of, e.g., $sl(n, \mathbb{R})$ provide different types of systems of nonlinear equaitons with superposition principles (matrix Riccati equations)[16] and have many other applications.

Mathematically, the classification of MASA's of the classical Lie algebras is the natural extension of the classification of individual elements into orbits and strata under the action of the corresponding classical Lie group—a matter of prime theoretical and practical importance.[17–21] The two classification problems do indeed have much in common.

A large amount of literature has been devoted to the classification of MASA's of semisimple Lie algebras. A very important type of MASA's are *Cartan subalgebras*, which are maximal abelian and self-normalizing algebras, or equivalently MASA's having the property that for each of their elements X, $\mathrm{ad}X$ is a semisimple linear matrix (in the adjoint representation). Over the field of complex numbers all Cartan subalgebras of a given semisimple Lie algebra are conjugate. Over the field of real numbers the number of conjugacy classes of Cartan subalgebras of a semisimple Lie algebra is finite, but not necessarily equal to one. The Cartan subalgebras of the real simple Lie algebras have been classified by several authors.[22–24]

A further type of abelian subalgebras that has received much attention are *maximal abelian nilpotent algebras* (MANS's) of the associative algebras of all complex or real matrices of a given dimension. This work was largely done by Kravchuk and has been reviewed and further developed by Suprunenko and Tyshkevich.[25] A MANS of a semisimple Lie algebra is defined as a subalgebra that is maximal among the commutative subalgebras consisting entirely of elements with nilpotent adjoint action. A MANS consists entirely of nilpotent matrices in any representation of finite degree. Maltsev[26] has constructed a certain subclass of the MANS's for all simple Lie algebras of finite dimension over the com-

[a)] Supported in part by the National Research Council of Canada and the Ministère de l'Education du Québec.
[b)] Published with the help of a grant from the "Fonds FCAC pour L'aide et le soutien à la recherche."
[c)] Permanent address: Department of Mathematics, Ohio State University, Columbus, Ohio, United States of America.

 0022-2488/83/081973-13$02.50

plex number field, namely the MANS's of maximal dimension.

In two recent research announcements[27] and in a larger publication in preparation[28] we have generalized Kravchuk's results in such a manner as to provide methods of constructing MANS's and all other MASA's of an aribtrary classical Lie algebra. In this paper we specialize to the fields of real and complex numbers and apply the general methods[27,28] to construct MASA's of $sp(2n, \mathbb{R})$ and $sp(2n, \mathbb{C})$.

The symplectic Lie algebras have been singled out for two reasons in this article. The first is because of their importance in applications. Indeed the group $\mathrm{Sp}(2n, \mathbb{R})$ is the group of real linear canonical transformations of a classical or quantum mechanical system with n degrees of freedom.[29] Symplectic Lie groups and algebras underlie the Hamiltonian formulations of the problems in the theory of dynamical systems, ranging from nuclear physics[30,31] to optimal control theory.[32] Hamiltonians and other integrals of motion that are quadratic polynomials in some dynamical variables (e.g., many particle coordinates and the conjugate momenta or creation and annihilation operators, etc.) play an important role in various fields of physics. They generate a $sp(2n, \mathbb{R})$ algebra and the $\mathrm{Sp}(2n, \mathbb{R})$ group of canonical transformations can be used to classify such Hamiltonians into orbits.[17] Our study of MASA's of $sp(2n, \mathbb{R})$ provides a classification of complete sets of commuting quadratic integrals of motion into orbits. The results have already been used to construct wave functions for arbitrary quadratic Hamiltonians.[33]

The symplectic Lie groups and Lie algebras also make their appearance in particle physics in the context of supersymmetric field theories and supergravity, where the orthosymplectic supergroups appear, e.g., as gauge groups.[34–37]

The second reason for our interest in the MASA's of symplectic Lie algebras is a mathematical one, namely that in particular $sp(2n, \mathbb{R})$ is actually one of the most complicated cases and as such provides a good illustration of the general situation. Thus, $sp(2n, \mathbb{R})$ has four different types of orthogonally indecomposable MASA's, matching the four different types of orthogonally indecomposable elements.[21] The algebra $sp(2n, \mathbb{C})$, on the other hand, has only two types of orthogonally indecomposable MASA's, again matching the two types of orthogonally indecomposable elements.[21] Each type will be treated separately below.

In Sec. 2 we provide some general information on the symplectic Lie groups and Lie algebras. In Sec. 3 we present some decomposition theorems and general results on the MASA's of $sp(2n, \mathbb{R})$ and $sp(2n, \mathbb{C})$. Section 4 is devoted more specifically to MANS's and Sec. 5 to low dimensional examples ($n = 1, 2, 3$). Section 6 contains the conclusions and a future outlook.

2. THE SYMPLECTIC LIE ALGEBRAS AND LIE GROUPS

We shall consider the Lie algebra $sp(2n, F)$, where F is either $\mathbb{R}$ or $\mathbb{C}$, of $2n \times 2n$ matrices X over F, satisfying

$$XK + KX^T = 0, \tag{1}$$

Where K is an antisymmetric nondegenerate matrix

$$K = -K^T, \quad \det K \neq 0 \tag{2}$$

(the superscript T throughout denotes transposition). The symplectic group $\mathrm{Sp}(2n, F)$ is then realized by $2n \times 2n$ matrices G over F satisfying

$$GKG^T = K. \tag{3}$$

When classifying either elements of $sp(2n, F)$ or subalgebras, in particular MASA's, of $sp(2n, F)$, it is useful to leave K flexible. Rather than fixing K and classifying matrices X under symplectic transformations G satisfying (3), we shall classify pairs of matrices (X, K) under general linear transformations. The pairs (X, K) and (X', K'), both satisfying (1) belong to the same conjugacy class if and only if there exists a matrix $G_0 \in \mathrm{GL}(2n, F)$ such that

$$G_0 X G_0^{-1} = X', \quad G_0 K G_0^T = K'. \tag{4}$$

When studying MANS's we find the following realizations of K particularly useful:

$$K_{\lambda\mu} = \begin{pmatrix} & & & I_\lambda \\ & & H_\mu & \\ & -H_\mu & & \\ -I_\lambda & & & \end{pmatrix},$$

$$\lambda + \mu = n, \quad 0 \leqslant \lambda \leqslant n, \quad 0 \leqslant \mu \leqslant n, \tag{5}$$

where λ and μ are nonnegative integers, $I_\lambda \in F^{\lambda \times \lambda}$ is a unit matrix and $H_\mu \in F^{\mu \times \mu}$ is

$$H_\mu = \begin{pmatrix} & & 1 \\ & \iddots & \\ 1 & & \end{pmatrix}. \tag{6}$$

In particular we shall use

$$K_\lambda \equiv K_{\lambda 0} = \begin{pmatrix} 0 & I_\lambda \\ -I_\lambda & 0 \end{pmatrix}. \tag{7}$$

When dealing with orthogonally decomposable elements of $sp(2n, F)$ a useful realization is

$$K = \begin{pmatrix} K_{\lambda_1\mu_1} & & \\ & \ddots & \\ & & K_{\lambda_k\mu_k} \end{pmatrix},$$

$$n_1 + \cdots + n_k = n, \tag{8}$$

where each $K_{\lambda_i\mu_i}$ has the form (5).

When studying (absolutely) indecomposable elements we use

$$K = F_{2n} = \begin{pmatrix} & & & & 1 \\ & & & -1 & \\ & & \iddots & & \\ & 1 & & & \\ -1 & & & & \end{pmatrix} \tag{9}$$

and for relatively indecomposable elements (see below) we put

$$K = K_0 \otimes F_n, \quad K_0 = \begin{cases} I_2 & \text{for } n = 2k, \\ F_2 & \text{for } n = 2k + 1. \end{cases} \tag{10}$$

Other realizations of K also occur and will be specified below.

Let us now list without proof several known properties of symplectic Lie algebras $sp(2n, F)$, and their elements $x \in sp(2n, F)$.

1. The real (complex) dimension of $sp(2n, R)$ [$sp(2n, C)$] is $d = n(2n+1)$.

2. The matrices X act on a $2n$-dimensional real (complex) vector space V. A matrix X is *decomposable* if two non-zero subspaces V_1 and V_2 of V exist such that

$$XV_1 \subseteq V_1 \quad XV_2 \subseteq V_2, \quad V_1 + V_2 = V, \quad V_1 \cap V_2 = 0. \tag{11}$$

The matrix X is *orthogonally decomposable* if in addition to (11) we have

$$V_1 \perp V_2, \quad \text{i.e., } f_1^T K f_2 = 0 \tag{12}$$

for all $f_1 \in V_1, f_2 \in V_2$.

3. Every matrix $X \in sp(2n, F)$ is conjugate under $\mathrm{Sp}(2n, F)$ to the block diagonal sum of orthogonally indecomposable elements. The classification of elements X into conjugacy classes under $\mathrm{Sp}(2n, F)$ thus reduces to a classification of orthogonally indecomposable elements.

4. Two types of orthogonally indecomposable elements of $sp(2n, \mathbb{C})$ exist.

a. *Indecomposable elements*; X has a single eigenvalue 0 and (X, K) can be represented by

$$X = J_{2n}, \quad K = F_{2n}, \tag{13a}$$

where

$$J_{2n} = \begin{pmatrix} 0 & & & \\ 1 & 0 & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{pmatrix} \tag{13b}$$

is a Jordan matrix and F_{2n} is given in (9).

b. *Decomposable elements*; X has two complex eigenvalues $\pm\alpha$ and (X, K) can be represented by

$$X = \begin{pmatrix} \alpha I_n + J_n & 0 \\ 0 & -\alpha I_n - J_n^T \end{pmatrix}, \quad K = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix} \tag{14}$$

with $\alpha \in \mathbb{C}$; for n even we must require $\alpha \neq 0$.

5. Four types of orthogonally indecomposable elements of $sp(2n, \mathbb{R})$ exist:

a1. *Absolutely indecomposable elements*; X has a single eigenvalue 0 and (X, K) can be represented by

$$X = J_{2n}, \quad K = \epsilon F_{2n}, \quad \epsilon = \pm 1. \tag{15}$$

a2. *Relatively indecomposable elements*; X has a pair of pure imaginary latent eigenvalues $\pm ib$ and (X, K) can be represented by

$$X = \begin{pmatrix} 0 & b \\ -b & 0 \end{pmatrix} \otimes I_n + I_2 \otimes J_n,$$
$$K = \epsilon K_0 \otimes F_n, \quad b > 0, \quad \epsilon = \pm 1 \tag{16}$$

with K_0 as in (10).

b1. *Decomposable and absolutely orthogonally indecomposable elements*; X has a pair of real eigenvalues $\pm a$ and (X, K) can be represented by

$$X = \begin{pmatrix} aI_n + J_n & 0 \\ 0 & -aI_n - J_n^T \end{pmatrix}, \quad K = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix} \tag{17}$$

with $a \in \mathbb{R}$, $a \neq 0$ for n even.

b2. *Decomposable relatively orthogonally indecomposable elements*; these exist only for n even, X has four latent complex eigenvalues $\pm(a \pm ib)$ and (X, K) can be represented by

$$X = \begin{pmatrix} A & 0 \\ 0 & -A^T \end{pmatrix}, \quad K = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}, \tag{18}$$
$$A = \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \otimes I_{n/2} + I_2 \otimes J_{n/2}, \quad a, b \in \mathbb{R}, a \neq 0, b > 0.$$

Absolutely indecomposable elements remain indecomposable after a field extension from $\mathbb{R}$ to $\mathbb{C}$. Relatively indecomposable elements become decomposable after such an extension. Relatively orthogonally indecomposable elements become orthogonally decomposable after a field extension; they are orthogonally indecomposable over $\mathbb{R}$.

When classifying MANS's of $sp(2n, F)$ we shall fix K in the form (5) and classify up to conjugacy under $\mathrm{Sp}(2n, F)$ realized as the Lie group of $2n \times 2n$ matrices G over F satisfying (3) with $K = K_{\lambda\mu}$. We can write an element $G \in \mathrm{Sp}(2n, F)$ as

$$G = \begin{pmatrix} g_{11} & g_{12} & g_{13} & g_{14} \\ g_{21} & g_{22} & g_{23} & g_{24} \\ g_{31} & g_{32} & g_{33} & g_{34} \\ g_{41} & g_{42} & g_{43} & g_{44} \end{pmatrix}, \tag{19}$$

where g_{ik} are rectangular matrices:

$$\begin{aligned} &g_{11}, g_{14}, g_{41}, g_{44} \in F^{\lambda \times \lambda}, \\ &g_{22}, g_{23}, g_{32}, g_{33} \in F^{\mu \times \mu}, \\ &g_{12}, g_{13}, g_{42}, g_{43} \in F^{\lambda \times \mu}, \\ &g_{21}, g_{31}, g_{24}, g_{34} \in F^{\mu \times \lambda}. \end{aligned} \tag{20}$$

Equation (3) could be solved to yield all constraints on the matrices g_{ik}, however we shall only need to realize specific types of $\mathrm{Sp}(2n, F)$ matrices. They are

$$G_{\mathrm{I}} = \begin{pmatrix} g_{11} & & & \\ & g_{22} & & \\ & & H g_{22}^{-T} H & \\ & & & g_{11}^{-T} \end{pmatrix},$$
$$g_{11} \in \mathrm{GL}(\lambda, F), g_{22} \in \mathrm{GL}(\mu, F), \tag{21}$$

$$G_{\mathrm{II}} = \begin{pmatrix} I & g_{12} & & \\ & I & & \\ & & I & -H g_{12}^T \\ & & & I \end{pmatrix}, \quad g_{12} \in F^{\lambda \times \mu}, \tag{22}$$

$$G_{\mathrm{III}} = \begin{pmatrix} I & 0 & g_{13} & 0 \\ & I & 0 & H g_{13}^T \\ & & I & 0 \\ & & & I \end{pmatrix}, \quad g_{13} \in F^{\lambda \times \mu}, \tag{23}$$

$$G_{\mathrm{IV}} = \begin{pmatrix} I & & & \\ & I & g_{23} & \\ & & I & \\ & & & I \end{pmatrix}, \quad g_{23} = H g_{23}^T H, \; g_{23} \in F^{\mu \times \mu}, \tag{24}$$

$$G_V = \begin{pmatrix} I & & \\ & I & \\ & g_{32} \quad I & \\ & & I \end{pmatrix}, \quad g_{32} = H g_{32}^T H, \; g_{32} \in F^{\mu \times \mu} \tag{25}$$

[we have put $H \equiv H_\mu$ as in (6)].

3. GENERAL PROPERTIES OF MAXIMAL ABELIAN SUBALGEBRAS OF $sp(2n, F)$

In this section we shall state several decomposition and classification theorems that reduce the task of classifying MASA's of $sp(2n, F)$ to that of classifying MANS of $sp(2n, F)$, $sl(n, F)$ and $su(p, q)$. The theorems are special cases or adaptations of more general theorems, valid for arbitrary classical Lie algebras over division rings.[27,28] In this article we omit the proofs.

Let us start with some necessary definitions. Similarly as an element of $sp(2n, F)$, a MASA of $sp(2n, F)$ will be called *orthogonally decomposable* if it can be represented by matrices $X \in F^{2n \times 2n}$ satisfying (1) that are all simultaneously decomposable and correspond to the same orthogonal decomposition of the vector space V. Otherwise a MASA will be called *orthogonally indecomposable* (OID). An orthogonally indecomposable MASA can be *decomposable*, if all of its elements are simultaneously decomposable and correspond to the same (nonorthogonal) decomposition of V. Otherwise it is *indecomposable*. We shall call an indecomposable MASA of $sp(2n, R)$ *absolutely(relatively) indecomposable* if it remains (does not remain) indecomposable after an extension of the ground field from $\mathbb{R}$ to $\mathbb{C}$. A decomposable MASA of $sp(2n, \mathbb{R})$ will be called a *relatively orthogonally indecomposable* MASA if it becomes orthogonally decomposable after a field extension from $\mathbb{R}$ to $\mathbb{C}$.

The first theorem reduces the task of classifying all MASA's to that of classifying orthogonally indecomposable ones.[27,28]

Theorem 1: An arbitrary orthogonally decomposable MASA of $sp(2n, F)$ can be represented by a set of matrix pairs (X, K)

$$X = \begin{pmatrix} X_1 & & & \\ & X_2 & & \\ & & \ddots & \\ & & & X_k \end{pmatrix}, \quad K = \begin{pmatrix} K_{\lambda_1 \mu_1} & & & \\ & \ddots & & \\ & & & K_{\lambda_k \mu_k} \end{pmatrix}, \tag{26}$$

$$X_j K_{\lambda_j \mu_j} + K_{\lambda_j \mu_j} X_j^T = 0, \quad 1 \leqslant j \leqslant k, \; 2 \leqslant k \leqslant n,$$
$$\lambda_j + \mu_j = n_j, \quad 1 \leqslant n_1 \leqslant \cdots \leqslant n_k \leqslant n - 1,$$
$$n_1 + n_2 + \cdots + n_k = n,$$

where

(i) For each i the component matrices X_i form an orthogonally indecomposable MASA, say S_i, of $sp(2n, F)$. The entries in S_i run through all possible values independently for $i = 1, \ldots, k$.

(ii) At most one of the algebras S_i is a MANS.

Conversely every such matrix pair (X, K) represents a conjugacy class of orthogonally decomposable MASA's of $sp(2n, F)$. □

Let us now turn to orthogonally indecomposable MASA's and consider the fields $F = \mathbb{C}$ and $F = \mathbb{R}$ separately.

Theorem 2: Two types of orthogonally indecomposable MASA's of $sp(2n, \mathbb{C})$ exist:

(i) Indecomposable MASA's. They are by necessity MANS's, i.e., they consist entirely of nilpotent matrices. Conversely, every MANS is an indecomposable MASA.

(ii) Decomposable MASA's. They can be represented by a set of matrix pairs (X, K)

$$X = \begin{pmatrix} A & 0 \\ 0 & -A^T \end{pmatrix}, \quad K = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}, \tag{27}$$

where the matrices $A \in gl(n, \mathbb{C})$ form an indecomposable MASA of $gl(n, \mathbb{C})$, i.e.,

$$A = \mathbb{C} I_n \oplus \text{MANS of } sl(n, \mathbb{C}). \tag{28}$$

Conversely, every such set of pairs represents an OID decomposable MASA. □

The MANS's of $sp(2n, \mathbb{C})$ occurring in the application of this theorem, are discussed in Sec. 4 below. The MANS's of $sl(n, \mathbb{C})$ are treated elsewhere[25,27,28] and in the present context this is a lower-dimensional problem.

Theorem 3: Four types of orthogonally indecomposable MASA's of $sp(2n, \mathbb{R})$ exist

(i) *Absolutely indecomposable* MASA's. These are MANS's of $sp(2n, \mathbb{R})$ and upon field extension from $\mathbb{R}$ to $\mathbb{C}$ they become MANS's of $sp(2n, \mathbb{C})$, i.e., they remain indecomposable.

(ii) *Relatively indecomposable* MASA's. Such MASA's must contain an element representing the imaginary unit i, i.e., a matrix $X_0 \in sp(2n, \mathbb{R})$ satisfying

$$X_0^2 + 1 = 0, \tag{29}$$

such that after complexification X_0 is diagonalizable and has n eigenvalues equal to i and n equal to $(-i)$. There exist $[n/2] + 1$ conjugacy classes with respect to $\mathrm{Sp}(2n, \mathbb{R})$ of such elements[21] of $sp(2n, \mathbb{R})$ and they can be represented by the sets of matrix pairs (X_0, K) with

$$X_0 = K_1 \otimes I_n, \quad K = K_1 \otimes (I_k \oplus -I_{n-k}) \tag{30}$$
$$(n + 1)/2 \leqslant k \leqslant n, \quad k \in \mathbb{Z},$$

where K_1 is as in (7). The centralizer of X_0 in $sp(2n, \mathbb{R})$ is isomorphic to $u(k, n - k)$. All relatively indecomposable MASA's of $sp(2n, \mathbb{R})$ can be written in the form

$$S = \begin{pmatrix} 0 & a \\ -a & 0 \end{pmatrix} \otimes I_n \oplus \text{MANS of } su(k, n - k), \; (n + 1)/2 \leqslant k \leqslant n - 1. \tag{31}$$

Upon field extension these MASA's become decomposable.

(iii) *Decomposable and absolutely* OID MASA's. These can be represented by a set of matrix pairs (X, K)

$$X = \begin{pmatrix} A & 0 \\ 0 & -A^T \end{pmatrix}, \quad K = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}, \tag{32}$$

where the matrices A together form the MASA of $sl(n, \mathbb{R})$:

$$\{A\} = \mathbb{R} I_n \oplus \text{MANS of } sl(n, \mathbb{R}). \tag{33}$$

Upon field extension these MASA's remain OID.

(iv) *Decomposable and relatively* OID MASA's. These exist only for n even and they can be represented by a set of pairs (32) where the matrices A form the MASA of $gl(n/2, \mathbb{C})$:

$$\{A\} = \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \otimes I_{n/2} \oplus \text{MANS of } sl(n/2, C). \quad (34)$$

Conversely, any matrix pair of the type listed in (i)–(iv), represents an OID MASA of $sp(2n, \mathbb{R})$. □

Comments on the proof and application of Theorem 3. Much of this theorem follows directly from the general theory, developed elsewhere[27,28] for all real classical quadratic Lie algebras [$su(p, q)$, $o(p, q)$, $o^*(2n)$, $sp(2n, \mathbb{R})$ and $sp(p, q)$]. Thus, absolutely indecomposable MASA's are always MANS's, relatively indecomposable MASA's always contain an element X_0 satisfying (29) and their classification reduces to a classification of the MANS's of the centralizers of X_0 in the corresponding classical Lie algebra. Decomposable OID MASA's of both types are always decomposable into precisely two blocks, as in (32).

In the $sp(2n, \mathbb{R})$ case under consideration the centralizer of X_0 in $gl(2n, \mathbb{R})$ is isomorphic to $gl(n, \mathbb{C})$ and is realized by $n \times n$ matrices with real 2×2 matrices as entries

$$\text{cent}(X_0, gl(2n, \mathbb{R})) = \begin{pmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & & \vdots \\ A_{n1} & \cdots & A_{nn} \end{pmatrix},$$

$$A_{ik} = \begin{pmatrix} a_{ik} & b_{ik} \\ -b_{ik} & a_{ik} \end{pmatrix}, \quad a_{ik}, b_{ik} \in \mathbb{R}, \ i, k = 1,\ldots,n. \quad (35)$$

The matrices A_{ik} provide a realization of the complex number field $\mathbb{C}$. Now let us restrict to the centralizer of X_0 in $sp(2n, \mathbb{R})$. For elements $X \in \text{cent}(X_0, gl(2n, \mathbb{R}))$ the symplectic condition (1) with K as in (30) translates into a condition on the matrices $X \in gl(n, \mathbb{C})$:

$$\tilde{X}\tilde{K} + \tilde{K}\tilde{X}^+ = 0, \quad \tilde{K} = \begin{pmatrix} iI_k & 0 \\ 0 & -iI_{n-k} \end{pmatrix} = -\tilde{K}^+, \quad (36)$$

where the superscript $+$ denotes Hermitian conjugation. The matrices X are thus characterized as belonging to $u(k, n-k)$ and we find that the centralizer of X_0 in $sp(2n, \mathbb{R})$ is

$$\text{cent}(X_0, sp(2n, \mathbb{R})) = \text{cent}(X_0, gl(2n, \mathbb{R}) \cap sp(2n, \mathbb{R})) \approx u(k, n-k). \quad (37)$$

For $k = n$ (we assme $k \geqslant [(n+1)/2]$) we obtain $u(n)$, i.e., the Lie algebra of the compact group $U(n)$. Compact Lie algebras do not have any MANS's (up to conjugacy they each have a unique MASA, namely the Cartan subalgebra), hence the condition $k \leqslant n-1$ in (31).

The decomposable relatively OID MASA's (34) involve MANS's of $sl(n/2, \mathbb{C})$ (n even) realized by real matrices of dimension $n/2 \times n/2$, the entries of which are 2×2 matrices A_{ik} as in (35).

To summarize: a complete classification of the orthogonally indecomposable MASA's of $sp(2n, \mathbb{R})$ into conjugacy classes under $\text{Sp}(2n, \mathbb{R})$ has been reduced to the classification of four subtypes of MASA's. The first type involves a classification of MANS of $sp(2n, \mathbb{R})$ under $\text{Sp}(2n, \mathbb{R})$, the second a classification of MANS's of $su(k, n-k)$ under $\text{SU}(k, n-k)$ with $(n+1)/2 \leqslant k \leqslant n-1$. The third and fourth types involve classifications of MANS's of $sl(n, \mathbb{R})$ under $\text{SL}(n, \mathbb{R})$ and of $sl(n/2, \mathbb{C})$ under $\text{SL}(n/2, \mathbb{C})$ (for n even), respectively.

The MANS's of $sp(2n, \mathbb{R})$ and $sp(2n, \mathbb{C})$ are studied in Sec. 4, those of $sl(n, \mathbb{R})$, $sl(n, \mathbb{C})$ and $su(p, q)$ elsewhere.[25,27,28]

The Cartan subalgebras are a special case. All Cartan subalgebras of $sp(2n, \mathbb{C})$ are of course $\text{Sp}(2n, \mathbb{C})$ conjugate and are orthogonally decomposable according to the decomposition $n = 1 + 1 + \cdots + 1$.

The algebra $sp(2n, \mathbb{R})$ has $(n+2)^2/4$ conjugacy classes of Cartan subalgebras for n even and $(n+1)(n+3)/4$ for n odd.[24] For $n \geqslant 3$ they are all orthogonally decomposable. The algebra $sp(2, \mathbb{R})$ has two orthogonally indecomposable Cartan subalgebras, represented by

$$S_1 = \begin{pmatrix} a & 0 \\ 0 & -a \end{pmatrix} \text{ and } S_2 = \begin{pmatrix} 0 & a \\ -a & 0 \end{pmatrix}, \ K = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad (38a)$$

respectively. The algebra $sp(4, \mathbb{R})$ has just one class of orthogonally indecomposable (but decomposable) Cartan subalgebras, represented by

$$S = \begin{pmatrix} a & b & & \\ -b & a & & \\ & & -a & b \\ & & -b & -a \end{pmatrix} \quad K = \begin{pmatrix} 0 & I_2 \\ -I_2 & 0 \end{pmatrix}. \quad (38b)$$

4. MAXIMAL ABELIAN NILPOTENT SUBALGEBRAS OF SYMPLECTIC LIE ALGEBRAS

4.1 General theory

In order to classify all MANS's of $sp(2n, F)$ into $\text{Sp}(2n, F)$ conjugacy classes we make use of the Kravchuk signature.[25] For the algebra $sl(N, F)$ the Kravchuk signature is a triplet of nonnegative integers (λ, ρ, ν), satisfying

$$\lambda + \rho + \nu = N, \quad \lambda \geqslant 1, \quad \nu \geqslant 1. \quad (39)$$

Here λ is invariantly characterized as the F dimension of the linear subspace of all N-columns that are simultaneously annihilated upon multiplication from the left by all elements of the given MANS and ν is the codimension of the F linear subspace generated by the images of the MANS acting on the N-column space.

For $sp(2n, F)$ the condition (1) implies $\lambda = \nu$ and we adopt the Kravchuk signature to be the triplet

$$(\lambda\ 2\mu\ \lambda), \ \lambda + \mu = n, \ 1 \leqslant \lambda \leqslant n, \ 0 \leqslant \mu \leqslant n-1. \quad (40)$$

Let us first state some general results on the MANS of $sp(2n, F)$ that are proven elsewhere.[28]

Theorem 4: 1. The elements X of a MANS of $sp(2n, F)$ with Kravchuk signature $(\lambda, 2\mu, \lambda)$ can be simultaneously transformed into the form

$$X = \begin{pmatrix} 0 & B & D & Y \\ 0 & R & S & HD^T \\ 0 & 0 & -HR^TH & -HB^T \\ 0 & 0 & 0 & 0 \end{pmatrix},$$
$$B, D \in F^{\lambda \times \mu}, \ R, S \in F^{\mu \times \mu}, \ Y \in F^{\lambda \times \lambda}$$
$$Y = Y^T, \ S = HS^TH, \quad (41)$$

where $H \equiv H_\mu$ is given by (6), X satisfies (1) with K as in (5) and R is a nilpotent matrix.

2. The matrices D, R, and S all depend linearly on the elements of B. From here on we consider the special case when B, D, R, and S all depend linearly on the elements

$$b = (b_1, b_2, ..., b_\mu), \quad b_i \in F \tag{42}$$

of a single row of B; let this be the first row. The matrices in (41) then can be written as

$$B = \begin{pmatrix} b \\ bQ_2 \\ \vdots \\ bQ_\lambda \end{pmatrix}, \quad D = \begin{pmatrix} bP_1 \\ bP_2 \\ \vdots \\ bP_\lambda \end{pmatrix},$$

$$R = (R_1 b^T, ..., R_\mu b^T), \quad S = (U_\mu b^T, ..., U_1 b^T), \tag{43}$$

where Q_i, P_i, R_i and U_i are fixed $F^{\mu\times\mu}$ matrices and R_i are such that all entries in the rows $i, i+1,...,\mu$ vanish.

3. The elements $b_i (1 \leqslant i \leqslant \mu)$ of the row b are assumed to be free. The elements of the symmetric matrix $Y = Y^T$ are arbitrary. Therefore the F-dimension of a MANS of $sp(2n, F)$ of Kravchuk signature $(\lambda, 2\mu, \lambda)$ thus obtained is

$$d_{n\lambda} = n + \lambda(\lambda - 1)/2. \tag{44}$$

The maximal and minimal possible dimensions of these MANS's are obtained for $\lambda = n$ and $\lambda = 1$, respectively; and are equal to

$$d_{\max} = d_{nn} = n(n+1)/2, \quad d_{\min} = d_{n1} = n, \tag{45}$$

i.e., for $sp(2n, F)$ the minimal dimension of such a MANS is equal to the rank (the dimension of a Cartan MASA).

4. The commutativity relation $[X, X'] = 0$ for two matrices of the type (41) implies the following constraints on the matrices Q_i, P_i, R_i, and U_i $(Q_1 \equiv I)$:

$$P_k = HP_k^T H, \quad Q_j P_k H - P_k H Q_j^T = P_j H Q_k^T - Q_k P_j H,$$
$$R_a^T = R_a, \quad U_a^T = U, \quad 1 \leqslant j, k \leqslant \lambda, \quad 1 \leqslant a \leqslant \mu \tag{46}$$

and we can put

$$P_1 = 0, \quad \mathrm{Tr} Q_i = 0. \tag{47}$$

The conditions $U_a^T = U_a$ and $S = HS^T H$ together imply

$$(U_a)_{bc} = S_{bac}, \tag{48}$$

where S_{bac} is a completely symmetric three component tensor in μ variables. □

Comments on the proof of Theorem 4. Statement 1 is a direct consequence of the symplectic character of X, and of X being nilpotent with Kravchuk signature $(\lambda, 2\mu, \lambda)$. The special case considered from Eq. (42) on is the case when the matrix B is "one rowed" (depends linearly on one row only). This special case covers completely the Kravchuk signature $(\lambda, 2\mu, \lambda)$ with $u = 0$, 1, or 2. The form of D, R and S is then an elementary consequence of commutativity. Statement 3 is elementary. Relations (46) and (48) can be obtained in a straightforward manner from commutativity and (47) is a result of normalization (conjugacy) by suitably chosen $\mathrm{Sp}(2n, \mathbb{R})$ transformations of the type (24) and (21).

The general theory could be developed further, i.e., further consequences of commutativity could be analyzed and further simplifying transformations performed for arbitrary Kravchuk signatures. Such a study, involving an analysis of relations (46) would take us into the domain of algebraic geometry.

Instead we now turn to specific Kravchuk signatures.

4.2 The signature $(n, 0, n)$

In this case the result is trivial. There exists precisely one such MANS, its dimension is $d_{\max} = n(n+1)/2$ and it can be written as

$$X_1 = \begin{pmatrix} 0 & Y \\ 0 & 0 \end{pmatrix}, \quad Y = Y^T \in F^{n\times n}. \tag{49}$$

This is the type of MANS's studied by Maltsev for all classical Lie algebras over $\mathbb{C}$ (MANS's of maximal dimension).[26]

4.3 The signature $(n-1, 2, n-1)$

In this case $\mu = 1$, so that Q_i, P_i, R_i and S in (43), (46)–(48) are just numbers. Since the Q_i are traceless we have:

$$Q_j = 0, \quad P_j = \mu_j, \quad P_1 = 0, \quad 2 \leqslant j \leqslant n-1,$$
$$R = 0, \quad S = \kappa b_1, \quad b = b_1. \tag{50}$$

A transformation GxG^{-1} with G of type (21), can be found that will take κ into $\kappa = 0$ or $\kappa = 1$. Using transformations (21)–(23) we obtain precisely two representative MANS of signature $(n-2, 2, n-2)$. In the first case the matrices $X = X_1$ are obtained by substituting

$$Q_i = 0, \qquad 2 \leqslant i \leqslant n-1,$$
$$P_k = 0, \qquad 1 \leqslant k \leqslant n-1, \tag{51}$$
$$R = 0, \ S = b$$

into (43), and in the second case we substitute

$$Q_i = 0, \qquad 2 \leqslant i \leqslant n-1,$$
$$P_2 = 1, \ P_k = 0 \quad 3 \leqslant k \leqslant n-1, \tag{52}$$
$$R = S = 0$$

into (43) in order to obtain the matrices $X = X_2$. Finally we have

$$X_1 = \left(\begin{array}{c|cc|ccccc} & b & 0 & & & & & \\ & 0 & 0 & & & & & \\ 0_{n-1} & \vdots & \vdots & & & Y & & \\ & 0 & 0 & & & & & \\ \hline & 0 & b & 0 & 0 & \cdots & 0 & \\ & 0 & 0 & -b & 0 & \cdots & 0 & \\ \hline & & & & & 0_{n-1} & & \end{array}\right),$$

$$Y = Y^T \in F^{(n-1)\times(n-1)}, \ b \in F \tag{53}$$

and

$$X_2 = \left(\begin{array}{c|cc|ccccc} & b & 0 & & & & & \\ & 0 & b & & & & & \\ 0_{n-1} & 0 & 0 & & & Y & & \\ & \vdots & \vdots & & & & & \\ & 0 & 0 & & & & & \\ \hline & 0 & 0 & 0 & b & 0 & \cdots & \\ & 0 & 0 & -b & 0 & 0 & \cdots & \\ \hline & & & & & 0_{n-1} & & \end{array}\right),$$

$$Y = Y^T \in F^{(n-1)\times(n-1)}, \ b \in F \tag{54}$$

Both types of MANS's exist for $F = \mathbb{R}$ and $F = \mathbb{C}$.

4.4 The signature ($n-2, 4, n-2$)

A. General pattern

In this case we have B, D, R, and S as in (43) with

$$b=(b_1, b_2), \quad P_1=0, \quad Q_k=\begin{pmatrix}\alpha_k & \beta_k \\ \gamma_k & -\alpha_k\end{pmatrix},$$

$$P_k=\begin{pmatrix}\mu_k & \nu_k \\ \rho_k & \mu_k\end{pmatrix}, \quad 2\leqslant k\leqslant n-2 \tag{55}$$

$$R=\begin{pmatrix}0 & \omega b_1 \\ 0 & 0\end{pmatrix}, \quad S=(U_2 b^T, U_1 b^T),$$

where

$$U_1=\begin{pmatrix}s_{111} & s_{112} \\ s_{211} & s_{212}\end{pmatrix}, \quad U_2=\begin{pmatrix}s_{121} & s_{122} \\ s_{221} & s_{222}\end{pmatrix}, \tag{56}$$

$$s_{112}=s_{121}=s_{211}, \quad s_{122}=s_{212}=s_{221}$$

and ω can be chosen to be either $\omega=0$ or $\omega=1$.

In addition, commutativity implies that Q_k and P_j satisfy (46) and

$$Q_k U_2 - U_2 Q_k^T = 0, \tag{57}$$

$$Q_k U_1 - U_1 Q_k^T = \omega\begin{pmatrix}0 & -\rho_k \\ \rho_k & 0\end{pmatrix}. \tag{58}$$

Furthermore, if $\omega=1$, we have

$$Q=\begin{pmatrix}\alpha_k & \beta_k \\ 0 & -\alpha_k\end{pmatrix}, \quad s_{122}=s_{222}=0. \tag{59}$$

B. Simplification of the symmetric tensor s_{ijk}

We use the symplectic transformation (21) with $g_{11}=I$ to transform the general matrix X into $X'=GXG^{-1}$, having the same structure as X. The row vector b is replaced by $b'=bg_{22}^{-1}$, the matrices Q_i and P_i undergo some transformation (not changing their structure) and the matrix S changes to

$$S'=g_{22}(U_2 g_{22}^T b'^T, U_1 g_{22}^T b'^T) H g_{22}^T H \tag{60}$$

which implies

$$s'_{ijk}=(g_{22})_{ia}(g_{22})_{jb}(g_{22})_{kc}s_{abc}. \tag{61}$$

The problem is thus to classify symmetric three component tensors in two variables under the action of the general linear group GL(2, F). This is a well known problem[38] directly related to the classification of homogeneous cubic polynomials in two real or complex variables

$$C(x)=\sum_{i,j,k=1}^{2} s_{ijk}x_i x_j x_k. \tag{62}$$

The result is that a real polynomial (62) can be carried by a linear substitution to one of 5 standard forms:

$$0, x_1^3, x_1^2 x_2, x_1 x_2(x_1+x_2) \text{ or } x_1(x_1^2+x_2^2) \tag{63}$$

corresponding to the following forms of S

$$S_{\mathrm{I}}=\begin{pmatrix}0 & 0 \\ 0 & 0\end{pmatrix}, \qquad s_{ijk}=0,$$

$$S_{\mathrm{II}}=\begin{pmatrix}0 & b_1 \\ 0 & 0\end{pmatrix}, \qquad s_{111}=1, \qquad s_{122}=s_{112}=s_{222}=0,$$

$$S_{\mathrm{III}}=\begin{pmatrix}b_1 & b_2 \\ 0 & b_1\end{pmatrix}, \qquad s_{112}=1, \qquad s_{111}=s_{122}=s_{222}=0, \tag{64}$$

$$S_{\mathrm{IV}}=\begin{pmatrix}b_1+b_2 & b_2 \\ b_1 & b_1+b_2\end{pmatrix}, \quad s_{112}=s_{122}=1, \quad s_{111}=s_{222}=0,$$

$$S_{\mathrm{V}}=\begin{pmatrix}b_2 & b_1 \\ b_1 & b_2\end{pmatrix}, \qquad s_{111}=s_{122}=1, \quad s_{112}=s_{222}=0.$$

For $F=\mathbb{C}$ (complex polynomials) the last two are equivalent, so one of them, say S_{V}, must be omitted. For $\omega=0$ all of the above forms are allowed. For $\omega=1$, the commutativity relations imply $s_{122}=s_{222}=0$ so that only $S_{\mathrm{I}}, S_{\mathrm{II}}$ or S_{III} survive. A further simplification is possible for $\omega=1$. The transformation GXG^{-1} where G is of the form (25) will transform R into

$$R'=R-Sg_{32}=0 \text{ for } S=\begin{pmatrix}b_1 & b_2 \\ 0 & b_1\end{pmatrix}, \quad g_{32}=\begin{pmatrix}0 & 1 \\ 0 & 0\end{pmatrix}.$$

Thus for $S=S_{\mathrm{III}}$, the case $\omega=1$ can be disregarded.

Finally, the cases to be considered for $F=\mathbb{R}$ are

$$\begin{aligned}&\omega=0: \quad S_{\mathrm{I}},\ldots,S_{\mathrm{V}}\\ &\omega=1: \quad S_{\mathrm{I}}, S_{\mathrm{II}}.\end{aligned} \tag{65}$$

For $F=\mathbb{C}$, S_{V} should be omitted.

C. Simplification of the matrices P_j

So far we have completely classified and standardized the matrices R and S in the matrix (41). Now we wish to simplify further by transformations carrying X to GXG^{-1}, $G\in\mathrm{Sp}(2n, F)$, where G is so chosen as not to change the parts of X that we have already simplified.

Each of the cases (65) must be treated separately. Let us first consider the matrices P_j. They form a bundle of 2×2 real H symmetric matrices, (i.e., $P_j=HP_j^T H$). The dimension d_B of such a bundle can be $d_B=0,1,2$, or 3. A matrix of the form (21) with g_{22} so chosen as to stabilize R and S can be used to standardize this bundle of matrices P_j. Thus, in case I we have $S=S_{\mathrm{I}}=0$. For $\omega=0$, g_{22} is arbitrary. the case $d_B=0$ is excluded, since the algebra would have the wrong Kravchuk signature. If $d_B=1$ we can transform the bundle

into a form, where $P_k = 0$, $3 \leqslant k \leqslant n-2$ and

$$P_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ or } \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \text{ for } F = \mathbb{R} \tag{66a}$$

$$P_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ or } \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \text{ for } F = \mathbb{C}. \tag{66b}$$

In each case the last matrix corresponds to the wrong Kravchuk signature and must be omitted. If $d_B = 2$ we can arrange that $P_k = 0$ for $4 \leqslant k \leqslant n-2$ and the couple P_2, P_3 takes one of the following forms

$$\left\{\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\right\}, \quad \left\{\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right\},$$

or

$$\left\{\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right\} \quad \text{for } F = \mathbb{R} \tag{67a}$$

and

$$\left\{\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\right\} \text{ or } \left\{\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right\} \text{ for } F = \mathbb{C}. \tag{67b}$$

For $d_B = 3$ we have $P_k = 0$ for $5 \leqslant k \leqslant n-2$ and

$$\{P_2, P_3, P_4\} = \left\{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\right\}. \tag{68}$$

For $S = S_{\mathrm{I}} = 0$, $\omega = 1$ we have

$$P_k = \begin{pmatrix} \mu_k & \nu_k \\ 0 & \mu_k \end{pmatrix}, \quad g_{22} = \begin{pmatrix} x & y \\ 0 & x^2 \end{pmatrix}.$$

In this case we can have $d_B = 0$, 1, or 2. Again $d_B = 0$ leads to the wrong Kravchuk signature; $d_B = 1$ leads to $P_2 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ and $d_B = 2$ to $P_2 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, $P_3 = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$ (all other $P_k = 0$).

For $S = S_{\mathrm{II}}$ we first use G in the form (22) to transform all P_k to the form $P_k = \left(\begin{smallmatrix} \mu_k & 0 \\ \rho_k & \mu_k \end{smallmatrix}\right)$. Taking G as in (21) with $g_{11} = \mathrm{I}$, $g_{22} = \left(\begin{smallmatrix} 1 & x \\ 0 & y \end{smallmatrix}\right)$ we can change the basis for the bundle $\{P_k\}$. We then use (21) with $g_{11} \in \mathrm{GL}(n-2, F)$ to standardize the bundle to one of the forms

$$P_2 = \left\{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right\}, \quad P_2 = \left\{\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}\right\},$$

$$P_2 = \left\{\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\right\}, \quad P_3 = \left\{\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}\right\}$$

(all other $P_k = 0$).

Similarly, using (21) and (22) type transformations for S_{III}, S_{IV} and S_{V} we can standardize the bundles of matrices P_k to several specific forms. These are all listed in the second column of Table I.

D. Simplification of the matrices Q_j

Having already fixed the form of S, R, and P_j we now proceed to establish all mutually nonconjugate allowed forms of the set of matrices Q_j. We already know that they have the form

$$Q_k = \begin{pmatrix} \alpha_k & \beta_k \\ \gamma_k & -\alpha_k \end{pmatrix}, \quad 2 \leqslant k \leqslant n-2 \tag{69}$$

and that for $\omega = 1$ we have $\gamma_k = 0$. For each set of matrices S, R and P_k we must satisfy the commutativity relations (57), (58), and (46). The relations (57) and (58) are easy to satisfy.

TABLE I. Maximal abelian subalgebras of $sp(2n\mathbb{R})$ of nilpotent matrices with Kravchuk signature $(n-2\ 4\ n-2)$, classified into Sp $(2n, \mathbb{R})$ conjugacy classes. Each class is given in the form (41) as a set of $2n \times 2n$ matrices $X_i^{Z,\omega}$. A particular class is specified by a choice of standard form of matrices S, B, D, and R of (41). The superscript $Z \in$ I, II,..., V indicates the form S_Z of (64) of the matrix S; $\omega = 0$ or 1 indicates the form (55) of R; the matrices B and D are given by the entries P_k and Q_k of the corresponding row of the Table and by (43). Throughout $P_1 = 0$, $Q_1 = I_2$, $Q_k = P_k = 0$ for $4 \leqslant k \leqslant n-2$. The corresponding table for $sp(2n, \mathbb{C})$ is obtained by omitting the 10 entires $X_4, \ldots, X_8$, X_{11}, X_{12}, and X_{34}, X_{35}, X_{36}.

Algebras	P_2	P_3	P_4	Q_2	Q_3	Q_4
$x_1^{\mathrm{I},0}$	$\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$	0	0	0	0	0
$x_2^{\mathrm{I},0}$	$\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$	0	0	0	$\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$	0
$x_3^{\mathrm{I},0}$	$\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$	0	0	0	$\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$	$\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$
$x_4^{\mathrm{I},0}$	$\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$	0	0	0	0	0
$x_5^{\mathrm{I},0}$	$\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$	0	0	0	$\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$	0
$x_6^{\mathrm{I},0}$	$\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$	0	0	0	$\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$	0
$x_7^{\mathrm{I},0}$	$\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$	0	0	0	$\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$	0
$x_8^{\mathrm{I},0}$	$\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$	0	0	0	$\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$	$\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$
$x_9^{\mathrm{I},0}$	$\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$	$\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$	0	0	0	0
$x_{10}^{\mathrm{I},0}$	$\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$	$\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$	0	0	0	$\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$
$x_{11}^{\mathrm{I},0}$	$\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$	$\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$	0	0	0	0
$x_{12}^{\mathrm{I},0}$	$\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$	$\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$	0	0	0	$\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$
$x_{13}^{\mathrm{I},0}$	$\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$	$\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$	0	0	0	0
$x_{14}^{\mathrm{I},0}$	$\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$	$\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$	0	0	0	$\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$
$x_{15}^{\mathrm{I},0}$	$\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$	$\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$	$\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$	0	0	0
$x_{16}^{\mathrm{I},1}$	$\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$	0	0	0	0	0
$x_{17}^{\mathrm{I},1}$	$\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$	0	0	0	$\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$	0
$x_{18}^{\mathrm{I},1}$	$\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$	$\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$	0	0	0	0
$x_{19}^{\mathrm{I},1}$	$\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$	$\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$	0	0	$\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$	0
$x_{20}^{\mathrm{I},1}$	$\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$	$\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$	0	0	0	$\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$
$x_{21}^{\mathrm{II},0}$	$\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$	0	0	0	0	0
$x_{22}^{\mathrm{II},0}$	$\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$	0	0	0	$\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$	0
$x_{23}^{\mathrm{II},0}$	$\left(\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right)$	0	0	0	0	0
$x_{24}^{\mathrm{II},0}$	$\left(\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right)$	0	0	0	$\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$	0
$x_{25}^{\mathrm{II},0}$	$\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$	$\left(\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right)$	0	0	0	0

TABLE I (*continued*).

$X_{26}^{II,1}$	$\begin{pmatrix}1&0\\0&1\end{pmatrix}$	0	0	0	0	0
$X_{27}^{II,1}$	$\begin{pmatrix}1&0\\0&1\end{pmatrix}$	0	0	0	$\begin{pmatrix}0&1\\0&0\end{pmatrix}$	0
$X_{28}^{III,0}$	0	0	0	0	0	0
$X_{29}^{III,0}$	0	0	0	$\begin{pmatrix}0&1\\0&0\end{pmatrix}$	0	0
$X_{30}^{III,0}$	$\begin{pmatrix}0&0\\1&0\end{pmatrix}$	0	0	0	0	0
$X_{31}^{IV,0}$	0	0	0	0	0	0
$X_{32}^{IV,0}$	0	0	0	$\begin{pmatrix}1&-2\\2&-1\end{pmatrix}$	0	0
$X_{33}^{IV,0}$	$\begin{pmatrix}0&1\\0&0\end{pmatrix}$	0	0	0	0	0
$X_{34}^{V,0}$	0	0	0	0	0	0
$X_{35}^{V,0}$	0	0	0	$\begin{pmatrix}0&1\\1&0\end{pmatrix}$	0	0
$X_{36}^{V,0}$	$\begin{pmatrix}0&1\\0&0\end{pmatrix}$	0	0	0	0	0

Indeed, for $S = S_I$ they are always satisfied. For $S = S_{II}$ and $\omega = 0$ we find $\gamma_k = 0$. For $S = S_{II}$ and $\omega = 1$ relation (58) can only be satisfied if $P_2 = \left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$ and $P_k = 0$, $3 \leqslant k \leqslant n-2$. For $S = S_{III}$, we obtain $\alpha_k = \gamma_k = 0$, for $S = S_{IV}$ we have $\beta_k = -2\alpha_k$, $\gamma_k = 2\alpha_k$, and for $S = S_V$ we obtain $\alpha_k = 0$, $\gamma_k = \beta_k$.

Finally, we must impose the commutativity relations (46) and use all transformations stabilizing S, R, and P_j to simplify the bundle of matrices Q_k. As an example let us consider the case when

$$S = S_I = \begin{pmatrix}0&0\\0&0\end{pmatrix},\quad R = \begin{pmatrix}0&0\\0&0\end{pmatrix},$$
$$P_2 = \begin{pmatrix}0&1\\1&0\end{pmatrix},\quad P_k = 0,\ 3 \leqslant k \leqslant n-2. \tag{70}$$

Relation (46) for $j = 2$, $k \geqslant 2$ implies $Q_k = Q_k^T$. Now take G in the form (25). The matrix D, in (41), i.e., the bundle of matrices P_i, is not affected, however the choice

$$g_{32} = \begin{pmatrix}\alpha_2&\beta_2\\\beta_2&-\alpha_2\end{pmatrix}$$

transforms Q_2 into $Q_2' = 0$. Choosing a further G in the form (21) we can now rearrange the bundle of matrices Q_j without changing the matrices P_j. Three possibilities occur in the dimension d_Q of the bundle $\{Q_j\}$ which can be $d_Q = 0$, 1 or 2. Thus we obtain the bundles

$$\{Q_k = 0,\ 2 \leqslant k \leqslant n-2\},$$
$$\left\{Q_3 = \begin{pmatrix}1&0\\0&-1\end{pmatrix},\quad Q_k = 0,\ k = 2,\ 4 \leqslant k \leqslant n-2\right\} \tag{71}$$

and

$$\left\{Q_3 = \begin{pmatrix}1&0\\0&-1\end{pmatrix},\ Q_4 = \begin{pmatrix}0&1\\1&0\end{pmatrix},\right.$$
$$\left.Q_k = 0,\quad k = 2,\quad 5 \leqslant k \leqslant n-2\right\}, \tag{72}$$

respectively.

All other sets of S, R and P_i can be treated similarly. The results for $F = \mathbb{R}$ are presented in Table I. A symbol for each algebra is given in the first column: $X_i^{A,\omega}$. The subscript i simply enumerates all $(n-2,4,n-2)$ type MANS's and there are precisely 36 $\mathrm{Sp}(2n, \mathbb{R})$ conjugacy classes of such subalgebras. The superscript A can be I, II, III, IV or V and tells us the form of $S = S_A$ as listed in (64). The label $\omega = 0$ or 1 tells us whether we have

$$R = \begin{pmatrix}0&0\\0&0\end{pmatrix} \text{ or } R = \begin{pmatrix}0&b_1\\0&0\end{pmatrix}.$$

The remaining columns give us the values of P_2, P_3, P_4 and Q_2, Q_3, Q_4. The matrices P_k and Q_k for $5 \leqslant k \leqslant n-2$ always vanish. These 36 different subalgebras exist if $n \geqslant 6$. For $n = 5$ only 29 types exist, because all those with $P_4 \neq 0$ or $Q_4 \neq 0$ must be eliminated. For $n = 4$ only 14 survive, since those with $P_k \neq 0$ or $Q_k \neq 0$ $(k = 3, 4)$ must be eliminated. For $n = 3$ exactly 3 survive, namely those with $P_i = Q_i = 0$ for $i = 2$ ($X_{28}^{III,0}$, $X_{31}^{IV,0}$ and $X_{34}^{V,0}$). For $n = 2$ or 1 no such MANS exist.

For $F = \mathbb{C}$ the results are simpler, namely the algebras $X_4^{I,0}, \ldots, X_8^{I,0}, X_{11}^{I,0}, X_{12}^{I,0}, X_{34}^{V,0}, X_{35}^{V,0}$ and $X_{36}^{V,0}$ should be dropped. For $n \geqslant 6$, 26 MANS'S survive, for $n = 5$ just 21, for $n = 4$ precisely two ($X_{28}^{III,0}$ and $X_{31}^{IV,0}$).

We observe a phenomenon of *saturation*: all complications that can arise do arise for $n = 6$ and nothing new happens for $n \geqslant 7$.

Kravchuk signatures $(n-\mu, 2\mu, n-\mu)$ for $\mu \geqslant 3$ are much harder to treat and for high enough values of n and μ there will exist infinitely many different mutually nonconjugated MANS. Their classification will require the introduction of continuous parameters. The principles of classification, including "saturation" will however be the same as in this section.

5. LOW-DIMENSIONAL CASES

In this section we will consider all MASA's of $sp(2n, F)$ for $n = 1, 2, 3$ since the theory developed in Secs. 3 and 4 suffices completely in these cases. We shall denote the MASA's of $sp(2n, \mathbb{R})$ and $sp(2n, \mathbb{C})$ as R_k^{2n} and C_k^{2n}, respectively; $2n$ refers to the dimension, $k = 1, 2, 3, \cdots$ enumerates the individual MASA's. Each MASA is represented by a matrix pair (X, K); the letters in X denote real or complex parameters $(a, b, \cdots) \in F$ with $F = \mathbb{R}$ or $F = \mathbb{C}$.

5.1. The algebra $sp(2, F)$

In this case $n = 1$ which is too low for the general theory to be of much use and in any case, the results are well known.

For $sp(2, \mathbb{R})$ we obtain 3 MASA's, represented by

$$R_1^2 = \begin{pmatrix}a&0\\0&-a\end{pmatrix},\quad R_2^2 = \begin{pmatrix}0&a\\-a&0\end{pmatrix},\quad R_3^2 = \begin{pmatrix}0&a\\0&0\end{pmatrix} \tag{73}$$

with $K = \left(\begin{smallmatrix}0&1\\-1&0\end{smallmatrix}\right)$ in all 3 cases. Here R_1^2 and R_2^2 are the noncompact and compact Cartan subalgebras, R_3^2 is a MANS with Kravchuk signature (1 0 1). For $sp(2, \mathbb{C})$ the first two above cases are mutually conjugate, so we are left with

two Sp(2, $\mathbb{C}$) classes of MASA's:

$$C_1^2 \sim R_1^2, \quad C_2^2 \sim R_3^2, \tag{74}$$

the first being the Cartan subalgebra, the second a MANS; we have $K = K_2$.

5.2 The algebra *sp(4, F)*

In this case we apply the general theory and note that all types of MASA's discussed in Sec. 3 exist.

A. Orthogonally decomposable MASA's

The only admissible partition of $n = 4$ is $4 = 2 + 2$. For $F = \mathbb{C}$ we obtain just two such MASA's, for $F = R$ five.

$$F = \mathbb{C}\text{: } C_1^4 = C_1^2 \oplus C_1^2, \quad C_2^4 = C_1^2 \oplus C_2^2, \tag{75}$$

$$F = \mathbb{R}\text{: } R_1^4 = R_1^2 \oplus R_1^2, \quad R_2^4 = R_1^2 \oplus R_2^2,$$
$$R_3^4 = R_1^2 \oplus R_3^2, \quad R_4^4 = R_2^2 \oplus R_2^2, \quad R_5^4 = R_2^2 \oplus R_3^2. \tag{76}$$

In all cases we have $K = K_2 \oplus K_2$.

B. Orthogonally indecomposable MASA's

Let us consider the case $F = \mathbb{R}$ first.

1. Absolutely indecomposable MASA's, i.e., MANS's. For Kravchuk signature (2 0 2) (49) reduces to

$$R_6^4 = \begin{pmatrix} 0 & 0 & a & b \\ 0 & 0 & b & c \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad K = \begin{pmatrix} 0 & I_2 \\ -I_2 & 0 \end{pmatrix} \tag{77}$$

and for (1 2 1), (53) reduces to

$$R_7^4 = \begin{pmatrix} 0 & b & 0 & c \\ 0 & 0 & b & 0 \\ 0 & 0 & 0 & -b \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$K = \begin{pmatrix} & & & 1 \\ & & 1 & \\ & -1 & & \\ -1 & & & \end{pmatrix}, \tag{78}$$

and (54) is not realized in this low dimension.

2. Relatively indecomposable MASA's. The only relevant matrix pair (30) is obtained for $n = 2, k = 1$. The corresponding MASA's will consist of the element X_0 in (30) (for $n = 2$) plus a MANS of $su(1,1)$ where $u(1,1)$ is realized by the matrices

$$X = \begin{pmatrix} 0 & b_1 & a & c \\ -b_1 & 0 & -c & a \\ a & -c & 0 & b_2 \\ c & a & -b_2 & 0 \end{pmatrix},$$

$$K = \begin{pmatrix} 0 & 1 & & \\ -1 & 0 & & \\ & & 0 & -1 \\ & & 1 & 0 \end{pmatrix}.$$

The algebra $su(1,1)$ has a single MANS so we obtain

$$R_8^4 = \begin{pmatrix} 0 & b+c & c & 0 \\ -b-c & 0 & 0 & c \\ c & 0 & 0 & b-c \\ 0 & c & -b+c & 0 \end{pmatrix},$$

$$K = \begin{pmatrix} 0 & 1 & & \\ -1 & 0 & & \\ & & 0 & -1 \\ & & 1 & 0 \end{pmatrix}. \tag{79}$$

3. Decomposable MASA's. In (32) we have $A \in gl(2, \mathbb{R})$; $sl(2, \mathbb{R})$ has just one MANS, so we obtain

$$R_9^4 = \begin{pmatrix} a & b & & \\ 0 & a & & \\ & & -a & 0 \\ & & -b & -a \end{pmatrix}, \quad K = \begin{pmatrix} 0 & I_2 \\ -I_2 & 0 \end{pmatrix}. \tag{80}$$

4. Relatively orthogonally decomposable MASA's. Just one MASA of the type (34) can exist for $n = 2$, namely

$$R_{10}^4 = \begin{pmatrix} a & b & 0 & 0 \\ -b & a & 0 & 0 \\ 0 & 0 & -a & b \\ 0 & 0 & b & -a \end{pmatrix}, \quad K = \begin{pmatrix} 0 & I_2 \\ -I_2 & 0 \end{pmatrix}. \tag{81}$$

The case $F = \mathbb{C}$ is somewhat simpler since no field extensions are involved. The orthogonally indecomposable MASA's of $sp(4, \mathbb{C})$ are represented by two MANS's

$$C_3^4 \sim R_6^4, \quad C_4^4 \sim R_7^4 \tag{82}$$

and one decomposable MASA

$$C_5^4 \sim R_9^4. \tag{83}$$

To summarize: Ten Sp(4, $\mathbb{R}$) conjugacy classes of MASA's of $sp(4, \mathbb{R})$ exist, represented by R_i^4 $(1 \leqslant i \leqslant 10)$. Among them, four, namely R_1^4, R_2^4, R_4^4, and R_{10}^4, are Cartan subalgebras and two, namely, R_6^4 and R_7^4 are MANS's. Note that R_{10}^4 is an orthogonally indecomposable Cartan subalgebra and that orthogonally indecomposable Cartan subalgebras of $sp(2n, \mathbb{R})$ exist only for $n = 1$ and 2(R_1^2, R_2^2, and R_{10}^4). The dimensions of the MASA's are $d = 3$ for R_6^4, $d = 2$ for all other ones. Five Sp(4, $\mathbb{C}$) conjugacy classes of MASA's of $sp(4, \mathbb{C})$ exist, represented above by C_i^4 $(1 \leqslant i \leqslant 5)$. Among them C_1^4 is the Cartan subalgebra, two, namely C_3^4 and C_4^4 are MANS'S. The dimensions are $d = 3$ for C_3^4, $d = 2$ for the rest. It is of course well known that any complex semisimple Lie algebra has a single conjugacy class of Cartan subalgebras and that for $sp(2n, \mathbb{C})$, $n \geqslant 2$ these are orthogonally decomposable according to the pattern $n = 1 + 1 + \cdots + 1$.

5.3. The algebra *sp(6, F)*

In this case $n = 3$ is odd which eliminates certain types of MASA's.

A. Orthogonally decomposable MASA's

The possible partitions are $6 = 2 + 2 + 2$ and $6 = 2 + 4$. For $sp(2n, \mathbb{R})$ we have

$$R^6_1 = R^2_1 \oplus R^2_1 \oplus R^2_1,\ R^6_2 = R^2_1 \oplus R^2_1 \oplus R^2_2,$$
$$R^6_3 = R^2_1 \oplus R^2_1 \oplus R^2_3,\ R^6_4 = R^2_1 \oplus R^2_2 \oplus R^2_2,$$
$$R^6_5 = R^2_1 \oplus R^2_2 \oplus R^2_3\ R^6_6 = R^2_2 \oplus R^2_2 \oplus R^2_2,$$
$$R^6_7 = R^2_2 \oplus R^2_2 \oplus R^2_3 \tag{84}$$

with $K = K_2 \oplus K_2 \oplus K_2$ in all cases, and

$$R^6_{8+\alpha} = R^2_1 \oplus R^4_{6+\alpha},\quad \alpha = 0,1,\ldots,4,$$
$$R^6_{13+\alpha} = R^2_2 \oplus R^4_{6+\alpha},\quad \alpha = 0,1,\ldots,4,$$
$$R^6_{18} = R^2_3 \oplus R^4_8,$$
$$R^6_{20} = R^2_3 \oplus R^4_{10},\quad R^6_{19} = R^2_3 \oplus R^4_9, \tag{85}$$

with $K = K_2 \oplus K_4$ in all cases.

For $sp(2n, \mathbb{C})$ we have

$$C^6_1 = C^2_1 \oplus C^2_1 \oplus C^2_1,\quad C^6_2 = C^2_1 \oplus C^2_1 \oplus C^2_2 \tag{86}$$

with $K = K_2 \oplus K_2$, and

$$C^6_3 = C^2_1 \oplus C^4_3,\quad C^6_4 = C^2_1 \oplus C^4_4,$$
$$C^6_5 = C^2_1 \oplus C^4_5,\quad C^6_6 = C^2_2 \oplus C^4_5 \tag{87}$$

with $K = K_2 \oplus K_4$.

B. Orthogonally indecomposable MASA's

Let us again consider $F = \mathbb{R}$ first

1. Absolutely indecomposable MASA's, i.e., MANS's. The Kravchuk signature (3 0 3) provides a single MANS, namely

$$R^6_{21} = \begin{pmatrix} 0 & Y \\ 0 & 0 \end{pmatrix},\quad Y = Y^T \in R^{3\times 3},\quad K = \begin{pmatrix} 0 & I_3 \\ -I_3 & 0 \end{pmatrix}. \tag{88}$$

The signature (2 2 2) provides two further MANS's, corresponding to (53) and (54)

$$R^6_{22} = \begin{pmatrix} 0 & 0 & b & 0 & c & d \\ 0 & 0 & 0 & 0 & d & e \\ 0 & 0 & 0 & b & 0 & 0 \\ 0 & 0 & 0 & 0 & -b & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$R^6_{23} = \begin{pmatrix} 0 & 0 & b & 0 & c & d \\ 0 & 0 & 0 & b & d & e \\ 0 & 0 & 0 & 0 & 0 & b \\ 0 & 0 & 0 & 0 & -b & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{89}$$

with $K = K_{2,1}$ as in (5). The signature (1 4 1) leads to 3 more MANS's, namely those in Table I with $P_i = Q_i = 0$ for $i \geqslant 2$. (Entries $X^{\mathrm{III},0}_{28}$, $C^{\mathrm{IV},0}_{31}$, and $X^{\mathrm{V},0}_{34}$). We denote them

$$R^6_{24+\alpha} = \begin{pmatrix} 0 & B & 0 & C \\ 0 & 0 & S_\alpha & 0 \\ 0 & 0 & 0 & -H_2 B^T \\ 0 & 0 & 0 & 0 \end{pmatrix},$$
$$B = (b_1, b_2),\quad \alpha = 0, 1, 2,$$
$$b_1, b_2, c \in \mathbb{R}$$
$$S_0 \equiv S_{\mathrm{III}} = \begin{pmatrix} b_1 & b_2 \\ 0 & b_1 \end{pmatrix},$$
$$S_1 = S_{\mathrm{IV}} = \begin{pmatrix} b_1 + b_2 & b_2 \\ b_1 & b_1 + b_2 \end{pmatrix},$$
$$S_2 = S_{\mathrm{V}} = \begin{pmatrix} b_2 & b_1 \\ b_1 & b_2 \end{pmatrix} \tag{90}$$

and use $K = K_{1,2}$.

2. Relatively indecomposable MASA's. The only relevant matrix pair (30) in this case is obtained for $n = 3$, $k = 2$. The centralizer of the element X_0 (30) in this case is $u(2, 1)$ and we must find all MANS's of $su(2, 1)$. This is a simple task since $su(2, 1)$ allows only one Kravchuk signature (1 1 1) corresponding to one single MANS. Writing this MANS of $su(2, 1)$ as a 6×6 real symplectic matrix we obtain

$$R^6_{27} = \begin{pmatrix} 0 & a & b & 0 & 0 & c \\ -a & 0 & 0 & b & -c & 0 \\ & & 0 & a & 0 & b \\ & & -a & 0 & -b & 0 \\ & & & & 0 & a \\ & & & & -a & 0 \end{pmatrix},$$
$$K = \begin{pmatrix} & & I_2 \\ & K_2 & \\ -I_2 & & \end{pmatrix} \tag{91}$$

(we use a realization of $sp(6, \mathbb{R})$ adapted to the MANS's of $su(2, 1)$.

3. Decomposable MASA's. In (32) we have $A \in gl(3, \mathbb{R})$; we must find all MASA's of $sl(3, \mathbb{R})$. Just three of these exist, corresponding to Kravchuk signature (2 0 1), (1 0 2) and (1 1 1). We obtain 3 more MASA's of $sp(6, \mathbb{R})$

$$R^6_{28+\alpha} = \begin{pmatrix} A_{28+\alpha} & 0 \\ 0 & -A^T_{28+\alpha} \end{pmatrix},\quad \alpha = 0,1,2,$$
$$K = \begin{pmatrix} 0 & I_3 \\ -I_3 & 0 \end{pmatrix} \tag{92}$$

with

$$A_{28} = \begin{pmatrix} a & 0 & b \\ 0 & a & c \\ 0 & 0 & a \end{pmatrix},\quad A_{29} = \begin{pmatrix} a & b & c \\ 0 & a & 0 \\ 0 & 0 & a \end{pmatrix},$$
$$A_{30} = \begin{pmatrix} a & b & c \\ 0 & a & b \\ 0 & 0 & a \end{pmatrix}. \tag{93}$$

4. Relatively orthogonally decomposable MASA's. Since $n = 3$ is odd no such MASA's exist for $sp(6, \mathbb{R})$.

Let us now consider the simpler case of $sp(6, \mathbb{C})$. The orthogonally decomposable MASA's have already been listed in (86) and (87). Just 5 MANS's exist:

$$C_7^6 \sim R_{21}^6,\ C_8^6 \sim R_{22}^6,\ C_9^6 \sim R_{23}^6,$$
$$C_{10}^6 \sim R_{24}^6,\ C_{11}^6 \sim R_{25}^6 \tag{94}$$

as do three decomposable MASA's

$$C_{12}^6 \sim R_{28}^6,\quad C_{13}^6 \sim R_{29}^6,\quad C_{14}^6 \sim R_{30}^6. \tag{95}$$

To summarize: 30 Sp(6, $\mathbb{R}$) conjugacy classes of MASA's of $sp(6, \mathbb{R})$ exist, represented above by R_i^6, $1 \leqslant i \leqslant 30$. Among them six, namely, $R_1^6, R_2^6, R_4^6, R_6^6, R_{12}^6, R_{17}^6$ are Cartan subalgebras ($R_1^6, R_2^6, R_4^6, R_6^6$ correspond to the decomposition $6 = 2 + 2 + 2$, R_{12}^6 and R_{17}^6 to $6 = 4 + 2$). Six of the MASA's are MANS's, namely $R_{21}^6, \ldots, R_{26}^6$. The dimensions of the MASA's are $d = 6$ for R_{21}^6, $d = 4$ for R_8^6, $R_{13}^6, R_{12}^6, R_{23}^6$ and $d = 3$ for all the rest.

Altogether 14 Sp(6, $\mathbb{C}$) conjugacy classes of MASA's of $sp(6, \mathbb{C})$ exist. The only Cartan subalgebra is C_1^6, the MANS's are $C_7^6, \ldots, C_{11}^6$. The dimensions are $d = 6$ for C_7^6, $d = 4$ for C_3^6, C_8^6, C_9^6, and $d = 3$ for all the rest.

6. CONCLUSIONS

The main conclusion of this article is that it is possible to proceed quite far in the classification of maximal abelian subalgebras of $sp(2n, F)$ and that a close parallelism exists between a classification of these MASA's and the classification of individual elements of $sp(2n, F)$. This parallelism is brought out particularly clearly by the theorems of Sec. 3.

Theorem 1 reduces the task of classifying all MASA's to that of classifying the orthogonally indecomposable ones. The same holds for the classification of elements. Theorem 2 states that two types of orthogonally indecomposable MASA's of $sp(2n, \mathbb{C})$ exist: MANS's, corresponding to indecomposable (nilpotent) elements (13) and decomposable MASA's, corresponding to decomposable elements (14). Theorem 3 states that four types of orthogonally indecomposable MASA's of $sp(2n, \mathbb{R})$ exist: (i) MANS's, corresponding to absolutely indecomposable elements (15), (ii) Relatively indecomposable MASA's corresponding to relatively indecomposable elements (16). (iii) Decomposable MASA's, corresponding to decomposable elements (17), and (iv) relatively orthogonally decomposable MASA's, corresponding to relatively orthogonally decomposable elements (18). The crux of the matter is a classification of maximal abelian nilpotent subalgebras. To classify MASA's of $sp(2n, \mathbb{R})$ we need a classification of the MANS's of $sp(2n, \mathbb{R})$, but also of $su(p, q)$ ($p + q = n$), $sl(n, \mathbb{R})$ and $sl(n/2, \mathbb{C})$ (for n even). In Sec. 4 we have outlined a method for classifying MANS's of $sp(2n, F)$ for all Kravchuk signatures $(n - \mu, 2\mu, n - \mu)$. The method is developed in complete detail for $\mu = 0$, 1 and 2. The computational details for $\mu \geqslant 3$ remain a formidable task. The results for $\mu \geqslant 2$ have made it possible to classify all MASA's of $sp(2, F)$, $sp(4, F)$, *and* $sp(6, F)$. The algebra $sp(8, F)$ would already involve the Kravchuk signature (1 6 1).

In a forthcoming paper we shall supply the proof of Theorem 4 and a further analysis of other types of MANS's of $sp(2n, F)$ that exist for $n \geqslant 4$.

Work is in progress on similar studies for all other classical Lie algebras.

[1] J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **15**, 1378, 1932 (1974).
[2] M. Perroud, J. Math. Phys. **17**, 1028 (1976).
[3] J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **16**, 1597, 1615 (1975); **17**, 717 (1976).
[4] J. Patera, R. T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **17**, 977 (1976); Can. J. Phys. **54**, 950 (1976).
[5] J. Patera, R. T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **18**, 2259 (1977).
[6] G. Burdet, J. Patera, M. Perrin, and P. Winternitz, J. Math. Phys. **19**, 1758 (1978); Ann. Sci. Math. Québec **2**, 81 (1978).
[7] P. Winternitz, in "Group Theoretical Methods in Physics," Proceedings of the Fifth International Colloquium, edited by B. Kolman and R. T. Sharp, (Academic, New York, 1977).
[8] C. P. Boyer, R. T. Sharp, and P. Winternitz, J. Math. Phys. **17**, 1439 (1976); J. Beckers, J. Patera, M. Perroud, and P. Winternitz, J. Math. Phys. **18**, 72 (1977).
[9] J. Beckers, J. Harnad, M. Perroud, and P. Winternitz, J. Math. Phys. **19**, 2126 (1978).
[10] J. Patera and P. Winternitz, J. Chem. Phys. **65**, 2725 (1976).
[11] P. Winternitz and I. Friš, Yad. Fiz. **1**, 899 (1965). [Sov. J. Nucl. Phys. **1**, 636 (1965)].
[12] W. Miller Jr., *Symmetry and Separation of Variables* (Addison-Wesley, Reading, MA, 1977).
[13] C. P. Boyer, E. G. Kalnins, and W. Miller Jr., Commun. Math. Phys. **59**, 285 (1978); E. G. Kalnins and W. Miller Jr., SIAM J. Math. Anal. **11**, 1011 (1980).
[14] E. G. Kalnins, W. Miller Jr., and P. Winternitz, SIAM J. Appl. Math. **30**, 630 (1976); W. Miller Jr., J. Patera, and P. Winternitz, J. Math. Phys. **22**, 251 (1981); C. P. Boyer, E. G. Kalnins, and P. Winternitz, Preprint CRMA-1064, Montréal, 1981.
[15] J. Harnad, S. Shnider, and L. Vinet, J. Math. Phys. **20**, 931 (1979); **21**, 2719 (1980); J. Harnad and L. Vinet, Phys. Lett. B **6**, 589 (1978).
[16] R. L. Anderson, J. Harnad, and P. Winternitz, Lett. Math. Phys. **5**, 143 (1981); Physica D **4**, 164 (1982); J. Math. Phys. **24**, 1062 (1983).
[17] J. Williamson, Am J. Math. **58**, 141 (1936); H. Zassenhaus, Canad. Math. Bull. **1**, 31, 101, 183 (1958); G. E. Wall, J. Austral. Math. Soc. **3**, 1 (1963); I. K. Cibuna, Ukr. Math. Zh. **18**, 79, 122 (1966); J. G. F. Belinfante and P. Winternitz, J. Math. Phys. **12**, 1041 (1971); N. Burgoyne and R. Cushman, Celest. Mech. **8**, 435 (1974); A. J. Laub and K. Meyer, Celest. Mech. **9**, 213 (1974).
[18] T. A. Springer and R. Sternberg, *Lecture Notes in Mathematics*, Vol. 131 (Springer, New York, 1970); N. Burgoyne and R. Cushman, J. Algebra **44**, 339 (1977).
[19] V. I. Arnold, Russ. Math. Surveys **26**, 29 (1971); D. M. Galin, Uspekhi Mat. Nauk. **27**, 241 (1972); D. M. Galin, *Trudy Sem. Petrovs.*, **1**, 63 (1975).
[20] J. Patera and C. Rousseau, J. Math. Phys. **23**, 705 (1982). J. Patera, C. Rousseau, and D. Schlomiuk, J. Phys. A: Math. Gen. **15**, 1063 (1982).
[21] D. Djoković, J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **24**, 1363 (1983).
[22] Harish-Chandra, Trans. Amer. Math. Soc. **83**, 98 (1956).
[23] B. Kostant, Proc. Nat. Acad. Sci. USA **41**, 967 (1955).
[24] M. Sugiura, J. Math. Soc. Japan **11**, 374 (1959).
[25] D. A. Suprunenko and R. I. Tyshkevich, *Commutative Matrices* (Academic, New York, 1968).
[26] A. I. Maltsev, Izv. Akademii Nauk SSSR, Ser. Mat. **9**, 291 (1945) [Am. Math. Soc. Transl. Ser. 1, **9**, 214 (1962)].
[27] J. Patera, P. Winternitz, and H. Zassenhaus, C. R. Math. Rep. Acad. Sci. Canada **2**, 231, 237 (1980).
[28] J. Patera, P. Winternitz, and H. Zassenhaus (to be published).
[29] V. I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer, New York, 1978); R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Benjamin-Cummings, Reading, MA, 1978); J. M. Souriau, *Structures des systemes dynamiques* (Dunod, Paris, 1969).
[30] M. Moshinsky, Groups in Physics, *Collective Model of the Nucleus, Canonical Transformations in Quantum Mechanics* (Presses de l'Universite de Montreal, Montreal, 1979).

[31]P. Kramer, M. Moshinsky, and T. H. Seligman, in *Group Theory and its Applications*, edited by E. M. Loebl (Academic, New York, 1975) Vol. 3.
[32]A. E. Bryson, Jr. and Yu-Chi Ho, *Applied Optimal Control* (Blaisdell, Waltham, MA, 1969).
[33]M. Moshinsky and P. Winternitz, J. Math. Phys. **21**, 1667 (1980).
[34]P. G. O Freund and I. Kaplansky, J. Math. Phys. **17**, 228 (1976).
[35]S. W. Macdowell and F. Mansouri, Phys. Rev. Lett. **38**, 739 (1977).
[36]P. Nath and R. Arnowitt, Phys. Lett. B **65**, 73 (1976).
[37]Y. Ne'eman and T. Regge, Riv. Nuovo Cimento (3) **1**, 1 (1978); A. D'adda, R. D'Auria, P. Fre, and T. Regge, Riv. Nuovo Cimento (3) **3**, 1 (1980).
[38]G. Salmon, *Lessons Introductory to the Modern Higher Algebra* (Chelsea, New York, 1966).

Octonionic description of exceptional Lie superalgebras

A. Sudbery
Department of Mathematics, University of York, Heslington, York YO1 5DD, England

(Received 10 August 1982; accepted for publication 24 September 1982)

The structures of the exceptional Lie superalgebras **G**(3) and **F**(4) are expressed in terms of octonions.

PACS numbers: 02.20.Sv, 02.10.Vr, 11.30.Pb

1. INTRODUCTION AND SUMMARY OF OCTONION ALGEBRA

It is well known[1-4] that the five exceptional Lie algebras can all be described in terms of the octonions, which, indeed, seem to underlie all exceptional structures in algebra and geometry.[1] In this paper we show that this is true of Lie superalgebras also. The list of simple Lie superalgebras[5] contains ten countably infinite families, a one-parameter family, and two exceptional algebras **G**(3) and **F**(4). The last two have structures which appear very simple when expressed in terms of octonions; this description is the subject of this paper. An alternative description, without using octonions, has been given by DeWitt and van Nieuwenhuizen.[6]

First we summarize those properties of octonions (the Cayley division algebra over $\mathbb{R}$) that we will need; for a full treatment see Ref. 2 or Ref. 3. The octonions, which we will denote by $\mathbb{O}$, form an eight-dimensional real algebra with a multiplication which is not associative but *alternative*, i.e., the associator

$$[x,y,z] = x(yz) - (xy)z \tag{1}$$

is a totally antisymmetric function of $x, y, z \in \mathbb{O}$. The algebra has an identity, which enables us to regard $\mathbb{R}$ as a subspace of $\mathbb{O}$ (the subspace spanned by the identity). Thus we can write $\mathbb{O} = \mathbb{R} \oplus \mathbb{O}'$, where $\mathbb{O}$ is a seven-dimensional subspace whose elements are called pure octonions. $\mathbb{O}$ has a conjugation $x \mapsto \bar{x}$ defined by

$$x = \alpha + a \Rightarrow \bar{x} = \alpha - a \quad (\alpha \in \mathbb{R},\ a \in \mathbb{O}'). \tag{2}$$

This is an anti-involution, i.e., $\bar{\bar{x}} = x$ and

$$\overline{xy} = \bar{y}\bar{x}. \tag{3}$$

It follows that the function $\langle x, y\rangle$ defined by

$$\langle x,y\rangle = \frac{1}{2}(x\bar{y} + y\bar{x}) = \frac{1}{2}(\bar{x}y + \bar{y}x) \tag{4}$$

always takes values in $\mathbb{R}$; in fact this defines a positive definite inner product on $\mathbb{O}$. We denote the length of x in this inner product by $|x|$, so that

$$|x|^2 = x\bar{x}. \tag{5}$$

The real numbers in $\mathbb{O}$ commute and associate with all other elements, so it is only the pure octonion parts of elements that contribute to commutators and associators. Hence

$$[\bar{x},y] = -[x,y] \tag{6}$$

and

$$[\bar{x},y,z] = -[x,y,z]. \tag{7}$$

The automorphism group of $\mathbb{O}$ is a Lie group whose Lie algebra is the compact form of the exceptional Lie algebra G_2. This is the derivation algebra of $\mathbb{O}$, i.e., the set of maps $D:\mathbb{O}\to\mathbb{O}$ satisfying

$$D(xy) = x(Dy) + (Dx)y. \tag{8}$$

Every automorphism is orthogonal and every derivation is antisymmetric with respect to the inner product (4). The maps of left and right multiplication by elements of $\mathbb{O}'$ are also antisymmetric. Thus if we define L_a and R_a by

$$L_a(x) = ax, \quad R_a(x) = xa \quad (a \in \mathbb{O}', x \in \mathbb{O}) \tag{9}$$

and let L be the set of all L_a, R the set of all R_a, then L and R are seven-dimensional subspaces of the Lie algebra $o(8)$ of all antisymmetric maps of $\mathbb{O}$. In fact $o(8)$ is the direct sum of these two subspaces and the 14-dimensional derivation algebra G_2:

$$o(8) = G_2 \oplus L \oplus R. \tag{10}$$

Let $K_a = L_a - R_a$, so that $K_a(x) = [a,x]$, and let K be the set of all K_a with $a \in \mathbb{O}'$. Then each K_a maps $\mathbb{O}'$ to itself. The derivations in G_2 have the sampe property, so the Lie algebra $o(7)$ of antisymmetric maps of $\mathbb{O}'$ is made up as

$$o(7) = G_2 \oplus K. \tag{11}$$

Given $x, y \in \mathbb{O}$, let $A(x,y)$ be the map obtained by taking the associator with x and y:

$$A(x,y)z = [x,y,z]. \tag{12}$$

Then $A(x,y)$ is antisymmetric and maps $\mathbb{O}'$ to $\mathbb{O}'$, so according to (11) it can be written as

$$A(x,y) = D(x,y) + K_a, \tag{13}$$

where $D(x,y)$ is a derivation and a is some element of $\mathbb{O}'$. In fact $a = -\frac{1}{3}[x,y]$, so the derivation $D(x,y)$ is given by

$$D(x,y)z = [x,y,z] + \frac{1}{3}[[x,y],z]. \tag{14}$$

This map $D:\mathbb{O}\times\mathbb{O}\to G_2$ plays an important role in the structure of exceptional Lie algebras and superalgebras. It has the properties

$$D(y,x) = -D(x,y) = D(\bar{x},y), \tag{15}$$

$$[E,D(x,y)] = D(Ex,y) + D(x,Ey) \tag{16}$$

for any $E \in G_2$.

One of the most remarkable properties of the octonions emerges from considering a generalization of the derivation

 0022-2488/83/081986-03$02.50

equation (8) in the form

$$A(xy) = (A^{\#}x)y + x(A^{\flat}y), \tag{17}$$

where A, $A^{\#}$, $A^{\flat}$ are antisymmetric maps. The *principle of triality* asserts that if $A\epsilon o(8)$ is given, then there exist unique maps $A^{\#}$, $A^{\flat}\epsilon o(8)$ satisfying (17). It is easy to see that $[A_1, A_2]^{\#} = [A_1{}^{\#}, A_2{}^{\#}]$ and $[A_1, A_2]^{\flat} = [A_1{}^{\flat} A_2{}^{\flat}]$, so the correspondences $A \mapsto A^{\#}$, $A^{\flat}$ define representations of $o(8)$. These are not equivalent to the self-representation. A further representation, which is equivalent to the self-representation, is given by $A \mapsto \overline{A}$ where

$$\overline{A}x = \overline{(A\bar{x})}.$$

Another way of expressing this is to say that the correspondence $A \mapsto A^{\#}$, $A^{\flat}$, $\overline{A}$ are automorphisms of $o(8)$ of which only $A \mapsto \overline{A}$ is an inner automorphism. They generate a group of six automorphisms of $o(8)$ which is isomorphic to S_3.

Explicitly, $A^{\#}$, $A^{\flat}$, and $\overline{A}$ are given as follows:

If

$$A = D + L_a + R_b \quad (D\epsilon G_2, a, b\epsilon\mathbb{O}'), \tag{18}$$

then

$$A^{\#} = D + L_a + R_{a-b}, \tag{19}$$

$$A^{\flat} = D - L_{a-b} + R_b, \tag{20}$$

$$\overline{A} = D - L_b - R_a. \tag{21}$$

If $A\epsilon o(7)$, then $A = \overline{A}$ and $A^{\#} = \overline{A^{\flat}}$. In that case, on calculating $(A^{\#})^{\#}$ and $(A^{\#})^{\flat}$ we find the following:

$$A(xy) = (A^{\#}x)y + x(A^{\flat}y), \tag{22}$$

$$A^{\#}(xy) = (Ax)y + x(A^{\#}y), \tag{23}$$

$$A^{\flat}(xy) = (A^{\flat}x)y + x(Ay). \tag{24}$$

Note that the multiplication operators L_a form a representation of the Clifford algebra of the seven-dimensional orthogonal space $\mathbb{O}'$, for it follows from the alternative law (1) that

$$L_a L_b + L_b L_a = L_{ab+ba}, \tag{25}$$

and if $a, b\epsilon\mathbb{O}'$, (14) gives

$$ab + ba = -2\langle a,b\rangle. \tag{26}$$

Now from (23) we have

$$L_{Aa} = [A^{\#}, L_a]. \tag{27}$$

It follows that $A \mapsto A^{\#}$ is the spin representation of $o(7)$.

Most of this theory is also valid for the other alternative algebras $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{H}$; the only change that has to be made is that the statement of uniqueness does not hold in the principle of triality. [This can be taken into account by considering the triality algebra, i.e., the Lie algebra of triples $(A, A^{\#}, A^{\flat})$ satisfying (17), as the analog of $o(8)$ in the other cases; see Ref. 2.]

2. CONSTRUCTION OF THE LIE SUPERALGEBRAS G(3) AND F(4)

We recall that a Lie superalgebra L is a direct sum $L = L_0 \oplus L_1$, in which L_0 is a Lie algebra and L_1 is a vector space carrying a representation ρ of L_0, which gives the bracket between L_0 and L_1:

$$[X,v] = \rho(X)v \quad (X\epsilon L_0, v\epsilon L_1). \tag{28}$$

The bracket between elements of L_1 is a symmetric bilinear map from $L_1 \times L_1$ to L_0 which is covariant in the sense that

$$[X,[v,w]] = [\rho(X)v,w] + [v,\rho(X)w] \quad (X\epsilon L_0, v,w\epsilon L_1) \tag{29}$$

and which satisfies

$$\rho([u,v])w + \rho([v,w])u + \rho([w,u])v = 0. \tag{30}$$

The Lie algebra L_0 and the representation ρ for all simple Lie superalgebras are specified by Kac.[5] It remains to construct the bracket on L_1 satisfying (24) and (30).

In $\mathbf{G}(3)$ (more precisely, in a particular real form of it), $L_0 = sl(2,\mathbb{R}) \oplus G_2$ and $L_1 = \mathbb{R} \otimes V$ where V is a seven-dimensional space carrying the fundamental representation of G_2. Here, as in Sec. 3, G_2 denotes the compact real form of the algebra which can be identified with the algebra of derivations of $\mathbb{O}$; hence V can be identified with $\mathbb{O}'$. Thus L_1 can be regarded as the space of two-component vectors with entries from $\mathbb{O}'$, with $sl(2,\mathbb{R})$ acting as 2×2 matrices and G_2 acting componentwise as derivations.

Now suppose $\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$, $\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \epsilon L_1$. We define the bracket $[\mathbf{a}, \mathbf{b}]$ as

$$[\mathbf{a},\mathbf{b}] = (3D, -4M), \tag{31}$$

where $D\epsilon G_2$ and $M\epsilon sl(2,\mathbb{R})$ are given by

$$\begin{aligned} D &= D(a_1,b_2) - D(a_2,b_1) \\ &= \epsilon_{ij} D(a_i,b_j) \end{aligned} \tag{32}$$

where $D(x,y)$ is defined in (14) and ϵ_{ij} is the two-dimensional antisymmetric symbol; and

$$\begin{aligned} M &= M(\mathbf{a},\mathbf{b}) \\ &= \begin{pmatrix} \langle a_1,b_2\rangle + \langle a_2,b_1\rangle & 2\langle a_1,b_1\rangle \\ -2\langle a_2,b_2\rangle & -\langle a_1,b_2\rangle - \langle a_2,b_1\rangle \end{pmatrix} \\ &= \mathrm{Re}\,(\mathbf{a}\mathbf{b}^{\dagger} + \mathbf{b}\mathbf{a}^{\dagger})\epsilon, \end{aligned} \tag{33}$$

where the dagger denotes the octonion conjugate transpose, Re denotes the real part of an octonion, and $\epsilon = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$.

The covariance equation (29) is quickly verified. When $X = E\epsilon G_2$, (29) follows from (16) and the fact that E is antisymmetric, so that $\langle Ea_i,b_j\rangle + \langle a_i,Eb_j\rangle = 0$. When $X = N\epsilon sl(2,\mathbb{R})$, (29) follows from the equation

$$\epsilon N + N^T\epsilon = 0. \tag{34}$$

To verify (30), consider the ith component ($i = 1$ or 2) of $\rho([\mathbf{a},\mathbf{b}])\mathbf{c} + \rho([\mathbf{b},\mathbf{c}])\mathbf{a} + \rho([\mathbf{c},\mathbf{a}])\mathbf{b}$: this is

$$\begin{aligned} &3\epsilon_{jk} D(a_j,b_k)c_i - 4(\langle a_i,b_j\rangle + \langle b_i,a_j\rangle)\epsilon_{jk} c_k \\ &\quad + \text{terms obtained by cyclically permuting } a, b, c. \end{aligned} \tag{35}$$

Now (14) and (26) give

$$3D(a,b)c = [a,b,c] + 4(\langle a,c\rangle b - \langle b,c\rangle a) \tag{36}$$

for $a, b, c\epsilon\mathbb{O}'$. Using this, together with the identity

$$\epsilon_{jk}\delta_{il} = \epsilon_{ik}\delta_{jl} - \epsilon_{ij}\delta_{kl}, \tag{37}$$

(35) can be written as

$$\epsilon_{ij}([a_k,b_j,c_k] - [a_j,b_k,c_k])$$
$$+ 4\epsilon_{ij}(\langle a_k,c_k\rangle b_j + \langle b_k,c_k\rangle a_j - 2\langle a_k,b_k\rangle c_j)$$
$$+ 4\epsilon_{ij}(\langle a_j,b_k\rangle c_k - \langle a_j,c_k\rangle b_k + \langle b_j,a_k\rangle c_k - \langle b_j,c_k\rangle a_k)$$
$$+ \text{cyclic permutations}$$
$$= 0.$$

This completes the verification that Eqs. (21)–(23) define a Lie superalgebra.

In **F**(4) the Lie algebra L_0 is $sl(2,\mathbb{R}) \oplus o(7)$ and the representation ρ is the tensor product of the eponymous two-dimensional representation of $sl(2,\mathbb{R})$ with the spin representation of $o(7)$. As explained in Sec. 2, the latter can be taken to act on $\mathbb{O}$, so in this case $L_1 = \mathbb{O}^2$.

Given $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, $\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \epsilon\, \mathbb{O}^2$, we define the bracket $[\mathbf{x}, \mathbf{y}]$ as

$$[\mathbf{x},\mathbf{y}] = (C(\mathbf{x},\mathbf{y}), -3M(\mathbf{x},\mathbf{y})), \tag{38}$$

where $M(\mathbf{x},\mathbf{y})$ is as in (33) and $C(\mathbf{x},\mathbf{y}) \epsilon o(7)$ is given by

$$C(\mathbf{x},\mathbf{y})z = \epsilon_{ij}\{(zx_i)\bar{y}_j - x_i(\bar{y}_j z)\} \quad (z \epsilon \mathbb{O}). \tag{39}$$

That this defines an element of $o(7)$ can be seen by writing it as

$$C(\mathbf{x},\mathbf{y})z = \epsilon_{ij}\{2[x_i,y_j,z] - [x_i\bar{y}_j,z]\}. \tag{40}$$

With this definition of the bracket, (29) holds for $X \epsilon sl(2,\mathbb{R})$ in the same way as it does for **G**(3). For $X = A \epsilon o(7)$ we have

$$M(A^{\#}\mathbf{x},\mathbf{y}) + M(\mathbf{x},A^{\#}\mathbf{y}) = 0 \tag{41}$$

since $A^{\#}$ is antisymmetric; and applying Eqs. (22)–(24) to (40) gives

$$C(A^{\#}\mathbf{x},\mathbf{y}) + C(\mathbf{x},A^{\#}\mathbf{y}) = [A, C(\mathbf{x},\mathbf{y})]. \tag{42}$$

As $\rho(A) = A^{\#}$, (41) and (42) are the same as (24) for this algebra.

To verify (30), we need to find $C(\mathbf{x},\mathbf{y})^{\#}$. Use (14) to write (40) in the form (18) and apply (20): the result is

$$C(\mathbf{x},\mathbf{y})^{\#}z = -\epsilon_{ij}\{(x_i\bar{y}_j)z + 2(z\bar{y}_j)x_i\}. \tag{43}$$

Now from (40) it follows that $C(\mathbf{x},\mathbf{y}) = C(\mathbf{y},\mathbf{x})$; hence

$$C(\mathbf{x},\mathbf{y})^{\#}\mathbf{z} + C(\mathbf{y},\mathbf{z})^{\#}\mathbf{x} + C(\mathbf{z},\mathbf{y})^{\#}\mathbf{x} = \tfrac{1}{2}\{C(\mathbf{x},\mathbf{y})^{\#}\mathbf{z}$$
$$+ \text{terms obtained from all permutations of } x, y, z\}.$$

Let

$$P_k = \epsilon_{ij}(x_i\bar{y}_j)z_k$$
$$+ \text{terms obtained from all permutations of } x, y, z,$$
$$Q_k = \epsilon_{ij}(z_k\bar{y}_j)x_i$$
$$+ \text{terms obtained from all permutations of } x, y, z,$$

so that

$$C(\mathbf{x},\mathbf{y})^{\#}z_k + \text{cyclic permutation terms}$$
$$= -\tfrac{1}{2}P_k - Q_k. \tag{44}$$

Writing

$$\epsilon_{ij}\delta_{kl} = \tfrac{1}{2}(\epsilon_{il}\delta_{jk} + \epsilon_{ij}\delta_{ik} + \epsilon_{ij}\delta_{kl}),$$

we find

$$P_k = \epsilon_{ij}\{\langle y_j,z_k\rangle + \langle z_j,y_k\rangle\}x_i$$
$$+ \text{cyclic permutation terms} + \tfrac{1}{2}Q_k,$$
$$Q_k = \epsilon_{ij}\{\langle y_j,z_k\rangle + \langle z_j,y_k\rangle\}x_i$$
$$+ \text{cyclic permutation terms} + \tfrac{1}{2}P_k.$$

Hence, by (44),

$$C(\mathbf{x},\mathbf{y})^{\#}z_k + \text{cyclic permutation terms}$$
$$= 3\{\langle y_k,z_j\rangle + \langle z_k,y_j\rangle\}\epsilon_{ji}x_i + \text{cyclic permutation terms}$$
$$= 3M(\mathbf{y},\mathbf{z})_{ki}x_i + \text{cyclic permutation terms} \tag{45}$$

which is the form taken by (30) for the bracket (38). This completes the verification that Eqs. (38)–(39) define a Lie superalgebra.

[1] J. R. Faulkner and J. C. Ferrar, "Exceptional Lie algebras and related algebraic and geometric structures," Bull. London Math. Soc. **9**, 1 (1977).
[2] P. Ramond, "Introduction to exceptional Lie groups and algebras," Caltech Preprint CALT-68-577 (1976).
[3] R. D. Schafer, *Introduction to Nonassociative Algebras* (Academic, New York, 1966).
[4] N. Jacobson, *Exceptional Lie Algebras* (Dekker, New York, 1971).
[5] V. G. Kac, "Lie superalgebras," Adv. Math. **26**, 8 (1977).
[6] B. S. DeWitt and P. van Nieuwenhuizen, "Explicit construction of the exceptional superalgebras F(4) and G(3)", Santa Barbara Preprint ITP-SB-81-16 (1982).

On a new class of gradient formulas in the angular momentum theory

A. W. Niukkanen
Vernadsky Institute of Geochemistry and Analytical Chemistry, Academy of Sciences of the USSR, 117334, Moscow, USSR

(Received 18 August 1982; accepted for publication 6 October 1982)

Recently B. F. Bayman has derived the formula for $\mathscr{Y}_{lm}(\nabla)\phi(r)\mathscr{Y}_{LM}(\mathbf{r})$, where $\mathscr{Y}_{lm}(\mathbf{r})$ is a solid spherical harmonic, in terms of the derivatives of the function $\phi(r)$ by the scalar parameter r. This result is clarified and essentially generalized for the case of the tensor product $\{\mathscr{Y}^n_l(\nabla)\otimes\phi(r)\mathscr{Y}_L(\mathbf{r})\}_{\lambda\mu}$, where $\mathscr{Y}^n_{lm}(\nabla)=\Delta^n\mathscr{Y}_{lm}(\nabla)$, in this paper. Applications to the Taylor series in the three-dimensional Euclidean space, as well as some other expansions, are discussed briefly.

PACS numbers: 02.30.Bi, 03.65. — w

1. INTRODUCTION

In physical applications the necessity often arises of expanding the function $f(\mathbf{r})$, specified in the three-dimensional Euclidean space, into the Taylor series or some other type of expansion. The corresponding formulas can be easily written down by using Cartesian coordinates of the vector $\mathbf{r}$. The corresponding expression is, however, rather cumbersome and not appropriate for calculations and it also does not take into account the transformation properties which are usually inherent in the functions relevant to physical applications. It is desirable to introduce these properties into expansion formulas, in an explicit form.

The functions met in applications are usually transformed by irreducible representations of the rotation group. Such a function will be referred to as an irreducible spherical tensor (IST).

In the case of a function which is not IST the reduction of a problem to the IST case is reached via preliminary expansion of $f(\mathbf{r})$ over the complete set of spherical functions (the Laplace series). Though a three-dimensional problem is essentially distinguished from a one-dimensional one, the IST case bears some general resemblance with the one-dimensional problem, since the coefficients in the Taylor expansion are given here by differentiation of the "scalar part" of a tensor by one (radial) variable, the difference being that the differentiation operator is now not reduced to the n-order derivative, but it has more complicated structure.

It is essential, however (and this is the main result of the present work), that the IST expansion problem is closely connected with the classical orthogonal polynomials theory. In the case of the Taylor series, for example, the differential operator turns out to have a form of the associated Laguerre polynomial. This conclusion provides a new formulation of expansion formulas which might lead to more profound insight into a formal background of the expansion theory.

2. NOTATIONS

Let $Y_{lm}(\mathbf{r})$ be the spherical function (of angular variables θ,ϕ of vector $\mathbf{r}$) with the usual Condon and Shortley phase.[1,2] Let us introduce the uniform tensor polynomial (n is an integer)

$$\mathscr{Y}^n_{lm}(\mathbf{r})=r^{l+2n}Y_{lm}(\mathbf{r})=r^{2n}\mathscr{Y}^0_{lm}(\mathbf{r}),$$

where $\mathscr{Y}^0_{lm}(\mathbf{r})\equiv\mathscr{Y}_{lm}(\mathbf{r})$ is the harmonic polynomial (the regular solid harmonic). An arbitrary IST $f_{lm}(\mathbf{r})$ is written as

$$f_{lm}(\mathbf{r})=f(r)\mathscr{Y}^0_{lm}(\mathbf{r}).$$

The notation $f_l(\mathbf{r})$ means symbolically $2l+1$ components, $f_{lm}(\mathbf{r})$, $-l\leqslant m\leqslant l$. We use also the usual[1] definition of the irreducible tensor product (ITP):

$$\{f_l(\mathbf{r})\otimes\phi_\lambda(\mathbf{q})\}_{LM}=\sum_{m,\mu}\langle lm,\lambda\mu|LM\rangle f_{lm}(\mathbf{r})\phi_{\lambda\mu}(\mathbf{q}),$$

where $\langle lm,\lambda\mu|LM\rangle$ is the Clebsch–Gordan coefficient.[1]

3. TAYLOR SERIES

The Taylor series of an IST $f_{LM}(\mathbf{r})$ can be written as

$$f_{LM}(\mathbf{r}+\mathbf{R})=\sum_{n,l,\lambda}a(n,l)(-1)^l\frac{\pi(\lambda)}{\pi(l,L)}\times\{\mathscr{Y}^n_l(\mathbf{R})\otimes\{\mathscr{Y}^n_l(\nabla)\otimes f_L(\mathbf{r})\}_\lambda\}_{LM},\quad(1)$$

where $\pi(a,b,\cdots)=[(2a+1)(2b+1)\cdots]^{1/2}$, and $a(n,l)$ is the coefficient in the "plane wave" expansion

$$\exp(\mathbf{a}\cdot\mathbf{b})=\sum_{n,l}a(n,l)\{\mathscr{Y}^n_l(\mathbf{a})\otimes\mathscr{Y}^n_l(\mathbf{b})\}_{00},\quad(2)$$

$$a(n,l)=(-1)^l\pi^{3/2}\frac{\pi(l)}{2^{l+2n-1}n!\Gamma(n+l+3/2)},$$

where $0\leqslant n<\infty$ in Eq. (1) and the summation over l,λ is determined by usual triangle condition.[1] The dependence of the quantity $\mathscr{Y}^n_{lm}(\nabla)$ in Eq. (1) on Cartesian components $\partial/\partial x,\partial/\partial y,\partial/\partial z$ of the gradient operator $\nabla=\nabla(\mathbf{r})$, which *commute* one with another, is fairly similar as in the case of usual (nonoperator) vector argument. One should note that the quantities $\mathscr{Y}^n_{lm}(\nabla)$ and $\mathscr{Y}^n_{lm}(\mathbf{r})$ are transformed under rotations in a similar way. The functions $\mathscr{Y}^n_{lm}(\mathbf{R})$ in Eq. (1) play the part of generalized powers, and the quantities

$$f^{n,l,L}_{\lambda,\mu}(\mathbf{r})=\{\mathscr{Y}^n_l(\nabla)\otimes f_L(\mathbf{r})\}_{\lambda\mu}$$

have the meaning of generalized (tensor) derivatives of the function $f(\mathbf{r})$ if one bears in mind the analogy with one-dimensional Taylor series.

4. THE GRADIENT FORMULA

The following relations hold:

$$\{\mathscr{Y}^n_l(\nabla)\otimes f(r)\mathscr{Y}^0_L(\mathbf{r})\}_{\lambda\mu}=\mathscr{Y}^0_{\lambda\mu}(\mathbf{r})H(l,L,\lambda)2^\nu\nu!N^\lambda_{\nu,\omega}(r)f(r),\quad(3)$$

 0022-2488/83/081989-03$02.50

$$N^{\lambda}_{\nu,\omega}(r) = L^{\lambda+1/2}_{\nu}\left(-\frac{r^2}{2}\mathscr{D}\right)\mathscr{D}^{\omega}, \quad \mathscr{D} \equiv \frac{\partial}{\partial(r^2/2)}, \tag{4}$$

$$\nu = n + (l + L - \lambda)/2, \quad \omega = n + (l + \lambda - L)/2, \tag{5}$$

$$H(a,b,c) = (4\pi)^{-1/2}\pi(a,b)\langle ao,bo|co\rangle/\pi(c), \tag{6}$$

where $L^{\alpha}_{\nu}(x)$ is an associated Laguerre polynomial.[3] When using the explicit formula for the Laguerre polynomial in Eq. (4) or in any other algebraic transformation, the expression $x^k = (-1)^k K^k \mathscr{D}^k$ should be considered as the "power" (or quasipower) of the quantity $x = -K\mathscr{D}(K)$, $K \equiv r^2/2$, i.e., the quantity $\mathscr{D}(K)$ should be considered as the differentiation operator (by $r^2/2$) for function f only, and with respect to the multiplier $r^2/2$ the expression $\mathscr{D}(K)$ should be considered as a permutable symbol. Such a convention does not interfere, but, on the contrary, facilitates obtaining various analytical corollaries from the Eq. (3).

Note that, since the quantity $H(l,L,\lambda)$ in Eq. (3) is proportional to the Clebsch–Gordan coefficient with zero projections of momenta [cf. Eq. (6)], then, in virtue of the known selection rule, the quantities $l + L - \lambda$ and $l + \lambda - L$ may assume only even values. Hencewith, the parameters ν and ω in Eqs. (3)–(5) are integers.

5. AUXILIARY IDENTITIES

To prove the central formula (3), we need some auxiliary differential identities.

Any function of the scalar argument, $r = (x^2 + y^2 + z^2)^{1/2}$, has the following differentiation property (cf. Ref. 4):

$$d^n(x)f(r) = M_n(x\sqrt{\mathscr{D}})\sqrt{\mathscr{D}}^n f(r),$$

$$d(x) = \frac{\partial}{\partial x}, \quad \mathscr{D} = d\left(\frac{r^2}{2}\right), \tag{7}$$

where M_n is the modified Hermite polynomial

$$M_n(x) = (-i)^n H_n(ix),$$

$$H_n(x) = (-1)^n \exp(x^2/2) d^n(x) \exp(-x^2/2). \tag{8}$$

The Hermite polynomials H_n in Eq. (8) are normalized according to Ref. 5 (in a more usual convention the multipliers 1/2 in exponents are absent). The expression in the right-hand side of the Eq. (7) should be treated as an ordered one. Namely, when using the explicit polynomial expression for the differential operator in the right-hand side of Eq. (7), the operators $\mathscr{D}^k$ (only integer "powers" k take place) should be placed in the right position, and the multipliers x^q—in the left one [following Maslov[6] notations, the operators $\mathscr{D}$ in Eq. (7) should be marked by the ordering index 1, and operators x with index 2, that would correspond to explicit indication of operators' positions].

Similar differential rules take place for y and z variables. Thus, for any polynomial $P(\mathbf{r}) \equiv P(x,y,z)$ we have

$$P(\nabla)f(r) = P\left[M(x\sqrt{\mathscr{D}})\sqrt{\mathscr{D}}, M(y\sqrt{\mathscr{D}})\sqrt{\mathscr{D}}, \times M(z\sqrt{\mathscr{D}})\sqrt{\mathscr{D}}\right] f(r), \tag{9}$$

where

$$(M(x\sqrt{\mathscr{D}})\sqrt{\mathscr{D}})^i \equiv M_i(x\sqrt{\mathscr{D}})\sqrt{\mathscr{D}}^i.$$

Using the integral representation of H_n polynomials,[5] one may easily obtain the integral representation of polynomials M_n,

$$M_n(x) = (2\pi)^{-1/2}\int_{-\infty}^{\infty} dx'\,(x + x')^n \exp\left(-\frac{1}{2}x'^2\right). \tag{10}$$

Substituting Eq. (10) into Eq. (9), we have

$$P(\nabla)f(r) = (2\pi)^{-3/2}\int d\mathbf{r}' \times P(\mathbf{r}\mathscr{D} + \mathbf{r}'\sqrt{\mathscr{D}})\exp\left(-\frac{1}{2}r'^2\right) f(r). \tag{11}$$

In order to transit from the integral representation (11) to an algebraic form, one should specify the type of the polynomial P and to use some addition theorem either for the polynomials P, or for the Gaussian exponent (after the substitution $\mathbf{r}' = \mathbf{r}'' - \mathbf{r}\sqrt{\mathscr{D}}$) in the integrand (11).

In the case under consideration, the function $P(\nabla) = \mathscr{Y}^n_{lm}(\nabla)$ can be transformed with the help of the addition theorem for polynomials $\mathscr{Y}$ (cf. Ref. 7 or 8), or, equivalently, we can use the Kumar expansion[9] for the Gaussian exponent. After integration of the right-hand side of Eq. (11) in spherical coordinates we obtain

$$\mathscr{Y}^n_{lm}(\nabla)f(r) = 2^n n!\mathscr{Y}^0_{lm}(\mathbf{r}) \times L^{l+1/2}_n\left(-\frac{1}{2}r^2\mathscr{D}\right)\mathscr{D}^{n+l} f(r). \tag{12}$$

In the particular case $n = 0$, Eq. (12) gives

$$\mathscr{Y}^0_{lm}(\nabla)f(r) = \mathscr{Y}^0_{lm}(\mathbf{r})\mathscr{D}^l f(r). \tag{13}$$

6. PROOF OF FORMULA (3)

To prove the gradient formula (3), it is sufficient to use relations (12) and (13).

Let $\phi(r)$ be such a scalar function that $f(r) = \mathscr{D}^l\phi(r)$. Then, the IST $\mathscr{Y}^0_{LM}(\mathbf{r})f(r)$ appearing in the left-hand side of Eq. (3) can be written, by means of Eq. (13), as $\mathscr{Y}^0_{LM}(\nabla)\phi(r)$. As a result, the irreducible tensor product in the left-hand side of Eq. (3) can be transformed, by placing the function $\phi(r)$ outside of the sign of the tensor product, to the following form:

$$\{\mathscr{Y}^n_{lm}(\nabla) \otimes \mathscr{Y}^0_L(\mathbf{r})f(r)\}_{\lambda\mu} = \{\mathscr{Y}^n_l(\nabla) \otimes \mathscr{Y}^0_L(\nabla)\}_{\lambda\mu}\phi(r). \tag{14}$$

In the right-hand side of Eq. (14) we have now ITP of functions of the *same argument*. Applying the addition theorem for spherical functions[1]

$$\{Y_l(\mathbf{r}) \otimes Y_L(\mathbf{r})\}_{\lambda\mu} = H(l,L,\lambda)Y_{\lambda\mu}(\mathbf{r}),$$

we have

$$\{\mathscr{Y}^n_l(\nabla) \otimes \mathscr{Y}^0_L(\nabla)\}_{\lambda\mu} = H(l,L,\lambda)\mathscr{Y}^{n+(l+L-\lambda)/2}_{\lambda\mu}(\nabla). \tag{15}$$

Substituting Eq. (15) into Eq. (14) and applying relation (12) with due account of the identity $f(r) = \mathscr{D}^l\phi(r)$, we, finally, get at the formula (3).

7. NEW INTERPRETATION OF BAYMAN FORMULA

The particular case of the general formula (3) yields a clear and compact alternative to the recent Bayman's re-

sult.[2] To this end, one should merely let $n = 0$ (the case considered by Bayman) in the above formula. If we use now explicit expressions for the resulting Laguerre polynomial[3] and the Clebsch–Gordan coefficient[1] participating in the definition of $H(l,L,\lambda)$ then, applying an evident substitution of a summation variable, one may easily verify that the final expression for a tensor derivative in the case $n = 0$ is identical to Bayman's[2] result.

Letting $n = 0, l = 1$, one may easily transform Eq. (3) to the well-known Darwin's formulas.[1]

8. GENERALIZED HYPERGEOMETRIC SERIES (GHS)

In the case of IST, the scalar part of which may be represented by GHS[10] of the argument r^2 (for brevity we call this function IST-GHS),

$$f(r) = {}_pF_q\left(\begin{matrix}\alpha_s;ur^2/2\\ \rho_t\end{matrix}\right) = \sum_{k=0}^{\infty} {}_pC_q^k\left(\begin{matrix}\alpha_s\\ \rho_t\end{matrix}\right)\frac{1}{k!}(u(r^2/2))^k,$$

$${}_pC_q^k\left(\begin{matrix}\alpha_s\\ \rho_t\end{matrix}\right) = \frac{(\alpha_1)_k\cdots(\alpha_p)_k}{(\rho_1)_k\cdots(p_q)_k}$$

[$(\alpha)_n$ is the Pochhammer symbol], the following important relation

$$N^{\lambda}_{\nu,\omega}\,{}_pF_q\left(\begin{matrix}\alpha_s;ur^2/2\\ \rho_t\end{matrix}\right)$$
$$= \frac{u^{\omega}}{\nu!}\left(\lambda+\frac{3}{2}\right)_{\nu}\,{}_pC_q^{\omega}\left(\begin{matrix}\alpha_s\\ \rho_t\end{matrix}\right)$$
$$\times\,{}_{p+1}F_{q+1}\left(\begin{matrix}\alpha_s+\omega, L+\omega+3/2;ur^2/2\\ \rho_t+\omega,\lambda+3/2\end{matrix}\right) \qquad (16)$$

holds. This means that tensor differentiation of IST-GHS leads to the IST of the same class. If one of parameters in the initial GHS denominator is equal to $L + 3/2$ (the situation often met in applications), the differentiation does not lead to changing the number of parameters in the GHS. The straightforward substitution of relations (16) and (3) into Eq. (1) gives the general formula of the Taylor expansion for IST-GHS of arbitrary type. Since many physically important functions (such as $\mathscr{Y}^n_{lm}(\mathbf{r})$, spherical waves, isotropic oscillator eigenfunctions, etc.) have IST-GHS form, such an approach leads to unified treatment of different expansions (for example, of Friedman and Russak,[11] Nozawa,[12] Moshinsky,[13] Kay, Todd, and Silverstone,[7] Smirnov,[8] expansions, etc.) and provides a systematic background for obtaining new expansions as particular cases of the general formula.

9. OTHER EXPANSIONS

Formula (1) has been obtained by applying the plane wave expansion, Eq. (2), to the shift operator, $\exp(2\mathbf{R}\nabla)$ (with the subsequent reassociation of the triple ITP by standard formulas of the angular momentum theory[1]).

If other types of the shift operator expansions are used, this results in IST expansions different from the Taylor series. In this case, $\mathscr{Y}^n_{lm}(\nabla)$ in Eq. (1) is substituted by other differential operators which have, however, the form of IST of the gradient operator ∇. The "scalar part" of such operators is usually represented by an expansion in powers of the Laplace operator Δ, i.e., the corresponding IST may be represented as a series in gradient operators, $\mathscr{Y}^n_{lm}(\nabla)$. Formula (3) gives an action of such operators on an arbitrary IST, $f_{LM}(\mathbf{r})$. In many cases, an analytical form of the resulting expressions allows one to sum up the arising operator series, thus connecting it with some special functions different from the Laguerre polynomials. This, in turn, results in compact formulas for IST expansion not exclusively of Taylor type, but of different types as well.

ACKNOWLEDGMENTS

The author is grateful to Professor L. A. Gribov, Dr. B. K. Novosadov (Vernadsky Institute of Geochemistry and Analytical Chemistry), Professor L. A. Shelepin and Dr. V. P. Karasev (Physical Institute of Academy of Sciences of the USSR) for helpful discussions of the final results. The author is greatly indebted to Dr. B. K. Novosadov for drawing the author's attention to Bayman's paper. Professor Yu. F. Smirnov indicated the possible links of the technique described in this paper with nuclear shell model formalism and some group theoretical topics. These links will be, presumably, discussed in a separate communication.

[1] D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonsky, *Quantum theory of angular momentum* (Nauka, Leningrad, 1975) (in Russian).
[2] B. F. Bayman, J. Math. Phys. **19**, 2558 (1978).
[3] *Higher Transcendental Functions*, edited by Erdélyi (McGraw-Hill, New York, 1953), Vol. 2.
[4] A. W. Niukkanen, Inter. J. Quant. Chem. **XVIII**, 695 (1980).
[5] J. Kampé de Fériet, R. Campbell, G. Petiau, and T. Vogel, *Fonctions de la Physique Mathématique* (Dunod, Paris, 1957).
[6] V. P. Maslov, *Operator methods* (Nauka, Moscow, 1973).
[7] K. G. Kay, H. D. Todd, and H. J. Silverstone, J. Chem. Phys. **51**, 2359 (1969).
[8] A. W. Niukkanen, Chem. Phys. Lett. **69**, 174 (1980).
[9] K. Kumar, J. Math. Phys. **7**, 671 (1966).
[10] *Higher Transcendental Functions*, edited by Erdélyi (McGraw-Hill, New York, 1953), Vol. 1.
[11] B. Friedman and J. Russak, Quart. Appl. Math. **12**, 13 (1954).
[12] R. Nozawa, J. Math. Phys. **7**, 1841 (1966).
[13] M. Moshinsky, Nucl. Phys. **13**, 104 (1959).

On the higher orders of hyperspherical harmonics

E. Arribas and J. Navarro
Departamento de Física Teórica, Facultad de Ciencias Físicas, Burjassot-Valencia, Spain
M. Fabre de la Ripelle
Division de Physique Théorique,[a] Institut de Physique Nucléaire, F-91406 Orsay Cedex, France

(Received 20 July 1982; accepted for publication 22 October 1982)

We suggest a procedure to evaluate matrix elements between hyperspherical harmonics of any order. The method is based on the hyperspherical expansion of a Slater determinant constructed with oscillator wavefunctions. Explicit formulae are given for all matrix elements up to order $L_m + 2$.

PACS numbers: 02.30.Gp, 03.65.Fd

I. INTRODUCTION

The hyperspherical harmonic method (HHM) has been applied in recent years to nuclear bound states. However, except for some few-body problems, calculations have been only performed in the L_m approximation in which the hyperspherical expansion of the wavefunction is restricted to those HH having the minimal degree L_m deduced from the Pauli principle. A detailed description of the method, as well as some L_m calculations, can be found in Ref. 1, for instance.

Basically the HHM uses internal coordinates consisting of an hyperradius r and a set Ω of $3A - 4$ angles; the internal wavefunction is expanded on a complete basis of angular functions, the so-called hyperspherical harmonics. Then the nonrelativistic Schrödinger equation is transformed into a system of coupled differential equations, the solutions of which give the hyperradial functions of the wavefunction expansion and the energies of the bound states. A preliminary technical problem arises therefore with the evaluation of matrix elements of a translationally invariant operator on an HH basis. For instance, when this operator is the total potential energy, its matrix elements are the coupling coefficients for the system of differential equations. This has been already worked out in the first-order approximation, by following essentially two ways. In the first one,[1,2] the square of HH of order L_m is integrated over a set of suitable $3A - 6$ internal variables, thus getting one- and two-body coefficients, as an intermediate step of obtaining matrix elements of one- and two-body operators. In the second way[3] a connection is established between hyperspherical matrix elements and the matrix elements taken between Slater determinants constructed from harmonic oscillator wavefunctions (HO determinants henceforth), in which the order

$$L = \sum_{i=1}^{A} (2n_i + l_i) \tag{1.1}$$

is minimum (L_m). Here n_i and l_i are, respectively, the usual radial and orbital quantum numbers of the ith nucleon.

In this paper a generalization of this second way is given, based on the result than an HO determinant of order L, can be written as a linear combination of a known function of the hyperradius times an HH. The orders of the HH appearing in such a linear combination range from L_m to L. In Ref. 4 a generalization of the first way has been given. In the present work we use a direct connection with the shell model leading to a straightforward physical interpretation of the HH.

The plan of this work is as follows. We discuss in Sec. II the hyperspherical expansion of an HO determinant. The coefficients are explicitly calculated up to the order $L_m + 2$ for 1p–1h excitations over a closed-shell core. In Sec. III, a general procedure for evaluating hyperspherical matrix elements is suggested. In particular, we find the result of Baz' and Zhukov for the order L_m. Explicit formulae for the matrix elements needed up to the order $L_m + 2$ are given.

II. HYPERSPHERICAL EXPANSION OF AN HO DETERMINANT

Let us first define the notations and recall some results derived in Ref. 1.

Let $(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_A)$ be the coordinates of the nucleons. The center-of-mass

$$\mathbf{R} = \frac{1}{A} \sum_{i=1}^{A} \mathbf{r}_i \tag{2.1}$$

will be eliminated by choosing the Jacobi coordinates

$$\boldsymbol{\xi}_i = \sqrt{i/(i+1)} \left[\mathbf{r}_{i+1} - (1/i) \sum_{j=1}^{i} \mathbf{r}_j \right], \quad i = 1, \ldots, A-1. \tag{2.2}$$

Now, in the $(3A - 3)$-dimensional space spanned by these coordinates we define hyperspherical coordinates, consisting of a hyperradius

$$r^2 = \sum_{i=1}^{A-1} \boldsymbol{\xi}_i^2 \equiv \sum_{i=1}^{A} (\mathbf{r}_i - \mathbf{R})^2 \tag{2.3}$$

and a set Ω of $3A - 4$ hyperangles. The volume element is given by

$$d\tau_{3A} \equiv d^3 r_1 d^3 r_2 \cdots d^3 r_A = A^{-3/2} d^3 R d\tau_{3A-3}, \tag{2.4a}$$

$$d\tau_{3A-3} \equiv d^3 \xi_1 d^3 \xi_2 \cdots d^3 \xi_{A-1} = r^{3A-4} dr d\Omega, \tag{2.4b}$$

where $d\Omega$ is the surface element over the unit hypersphere.

The internal wavefunction is expanded on an HH basis as

$$\psi_{\text{int}}(1, \ldots, A) = \sum_{[L]} \psi_{[L]}(r) D_{[L]}(\Omega), \tag{2.5}$$

[a] Laboratoire associé au C.N.R.S.

 0022-2488/83/081992-05$02.50

where $D_{[L]}(\Omega)$ is a normalized HH of order L, including the full spin–isospin dependence, completely antisymmetric with respect to the exchange of any pair of nucleons. The symbol $[L]$ stands for a set of $3A-4$ quantum numbers, including the HH order L. The hyperradial functions are determined by the boundary conditions in the Schrödinger equation. They are solution of a system of coupled second-order differential equations, in which the coupling coefficients are given by the potential matrix.

A. The harmonic oscillator potentials

Consider now the harmonic oscillator (HO) potential

$$V=\tfrac{1}{2}m\omega^2\sum_{i=1}^{A}r_i^2=\frac{\hbar^2}{2mb^2}\sum_{i=1}^{A}\left(\frac{r_i}{b}\right)^2. \tag{2.6}$$

From the definition (2.3) of the hyperradius we immediately get another equivalent expression for the potential

$$V=\frac{\hbar^2}{2mb^2}\left[\left(\frac{r}{b}\right)^2+A\left(\frac{R}{b}\right)^2\right], \tag{2.7}$$

which also has an HO form for both internal and center-of-mass potentials. The potential (2.6) gives an IPM description of the system, where the eigenstates are described by a Slater determinant $D_{[L]}(\mathbf{r},b)$ constructed from the individual HO eigenfunctions

$$\phi^{(b)}_{nlm}(\mathbf{r})=\left[\frac{2n!}{b^3\Gamma(n+l+\frac{3}{2})}\right]^{1/2}\left(\frac{r}{b}\right)^l Y^m_l(\omega)$$
$$\times L_n^{l+1/2}\left(\left(\frac{r}{b}\right)^2\right)e^{-(r/b)^2/2}. \tag{2.8}$$

The potential (2.7) gives a collective description of the system, and the eigenstates are expressed in terms of functions

$$D_{[L]}(\Omega)\psi^{(b)}_{L,n'}(r)\phi^{(b)}_{nlm}(\mathscr{R}) \tag{2.9}$$

where $D_{[L]}(\Omega)$ is an HH and $\psi^{(b)}_{L,n'}(r)$ is a hyperradial HO function given by

$$\psi^{(b)}_{L,n}(r)=r^{-(3A-4)/2}\left[\frac{2n!}{b\Gamma(n+\nu+1)}\right]^{1/2}$$
$$\times e^{-(r/b)^2/2}\left(\frac{r}{b}\right)^{\nu+1/2}L_n^{\nu}\left(\left(\frac{r}{b}\right)^2\right) \tag{2.10}$$

with $\nu=L+(3A-5)/2$, and the notation $\mathscr{R}=\sqrt{A}\,\mathbf{R}$ has been used.

The same HO problem has been solved in the IPM and in the HHM. Therefore, the eigenfunctions of one model are linear combinations of the eigenfunctions of the other method, provided that the total energy is preserved. We thus obtain the relation

$$D_{[L]}(\mathbf{r},b)=\sum_{n'nlm}C(n',n,l,m,[L])\phi^{(b)}_{nlm}(\mathscr{R})$$
$$\times\psi^{(b)}_{L',n'}(r)D_{[L']}(\Omega), \tag{2.11}$$

with the following restriction in the sum:

$$L=L'+2n'+2n+l=\sum_{i=1}^{A}(2n_i+l_i), \tag{2.12}$$

where n_i and l_i refer to the individual HO quantum numbers in the Slater determinant (2.13). Of course, the coefficients C of the linear combination are to be determined.

Equation (2.11) is still valid when $D_{[L]}(\mathbf{r},b)$ is replaced by a sum of Slater determinants with the same L (i.e., the same energy) but different individual configurations. For instance, the coefficients of such a sum can be determined in order to produce a linear combination of Slater determinants with definite total angular momentum J and isospin T. Then the HH of highest order L in (2.11) (i.e., for $n=l=n'=0$) would exhibit the same definite quantum numbers J and T.

In order to simplify the expressions below, we introduce the following quantities:

$$p(n)=1 \quad \text{for } n \text{ odd}$$
$$=0 \quad \text{for } n \text{ even}, \tag{2.13}$$

$$F_{nlm}(\mathscr{R},b)=\phi^{(b)}_{nlm}(\mathscr{R})/\phi^{(b)}_{000}(\mathscr{R}) \tag{2.14}$$

and

$$G(r,b)=\psi^{(b)}_{Lm,1}(r)/\psi^{(b)}_{Lm,0}(r), \tag{2.15}$$

where $\phi^{(b)}_{nlm}(\mathscr{R})$ and $\psi^{(b)}_{Lm,n}(r)$ are given in (2.8) and (2.10).

B. Calculation of the coefficients up to L_m+2

For the sake of simplicity we limit ourselves to the case of closed $N=Z$ shell nuclei. Then the HO determinant of order L_m is a closed-shell core, denoted by $|b,L_m\rangle$, having quantum numbers $J=T=0$. Any HO determinant of order $L>L_m$ can be obtained through particle–hole excitations on the core; the specific orbitals as well as the number of particle–hole couples are both limited by the order of the determinant. The symbol $|bLJT\alpha\rangle$ will be used for a suitable linear combination of HO determinants of order L having well-defined quantum numbers J and T; the index α distinguishes between the various possible individual configurations in a determinant of order L. For instance, a determinant of order L_m+1 is obtained by creating a hole in a state (i) of the last occupied shell of the core and creating a particle in a state (m) of the next unoccupied shell. In that case, the notation $i^{-1}m$ will substitute the symbol α. Finally, the symbol $D_L^{(\alpha JT)}(\Omega)$ will stand for a completely antisymmetric HH of order L with quantum numbers JT associated with a specific configuration α. Obviously, the superscript (αJT) will be omitted in the case $L=L_m$ since we are dealing with closed shell nuclei.

Now, in order to expand a wavefunction $|bLJT\alpha\rangle$ in terms of HH having well-defined quantum numbers, they must be suitably coupled with the center-of-mass angular momentum. This coupling obviously gives some restrictions on the coefficients which will be explained. The expansion (2.11) is written now up to the order L_m+2:

$$|b,L_m\rangle=\phi^{(b)}_{000}(\mathscr{R})\psi^{(b)}_{L_m,0}(r)D_{L_m}(\Omega), \tag{2.16a}$$

$$|b,L_{m+1},JT\alpha\rangle=A_1(\alpha JT)\phi^{(b)}_{000}(\mathscr{R})\psi^{(b)}_{L_m+1,0}(r)D^{(\alpha JT)}_{L_m+1}(\Omega)$$
$$+\delta(J,1)\delta(T,0)A_2(\alpha JT)\phi^{(b)}_{01M}(\mathscr{R})\psi^{(b)}_{L_m,0}(r)D_{L_m}(\Omega), \tag{2.16b}$$

$$|b,L_m+2,JT\alpha\rangle = B_1(\alpha JT)\phi^{(b)}_{000}(\mathscr{R})\psi^{(b)}_{L_m+2,0}(r)D^{(\alpha JT)}_{L_m+2}(\Omega)$$
$$+\sum_{\beta J'M'} B_2(\alpha JT\beta J'M')\phi^{(b)}_{01\,M-M'}(\mathscr{R})$$
$$\times\psi^{(b)}_{L_m+1,0}(r)D^{(\beta J'M'T)}_{L_m+1}(\Omega)$$
$$+\delta(J,2)\delta(T,0)B_3(\alpha,JT)\phi^{(b)}_{02M}(\mathscr{R})\psi^{(b)}_{L_m,0}(r)D_{L_m}(\Omega),$$
$$+\delta(J,0)\delta(T,0)B_4(\alpha,JT)\phi^{(b)}_{100}(\mathscr{R})\psi^{(b)}_{L_m,0}(r)D_{L_m}(\Omega),$$
$$+\delta(J,0)\delta(T,0)B_5(\alpha,JT)\phi^{(b)}_{000}(\mathscr{R})\psi^{(b)}_{L_m,1}(r)D_{L_m}(\Omega). \tag{2.16c}$$

This set of equations constitute eventually a method for generating suitable HH in terms of HO determinants.

Let us consider now (2.16b). The coefficient A_2 is identically given by

$$A_2(\alpha,JT) = \langle\phi^{(b)}_{01M}\,\psi^{(b)}_{L_m,0}D_{L_m}|b,L_m+1,JT\alpha\rangle$$
$$= \langle bL_m|F^*_{01M}(\mathscr{R},b)|b,L_m+1,JT\alpha\rangle. \tag{2.17a}$$

It is a one-body matrix element between the core and a 1p–1h excitation. Using the conventions of Ref. 5 about time-reversal invariance, one finds

$$A_2(i^{-1}mJT) = -\left(\frac{2j_i+1}{3A}\right)^{1/2} p(l_i+l_m)(j_i1j_m;\tfrac{1}{2}\ 0\ \tfrac{1}{2})\cdot$$
$$\times\langle n_il_i|r|n_ml_m\rangle\delta(J,1)\delta(T,0), \tag{2.17b}$$

where $\langle nl|r|n'l'\rangle$ is an HO matrix element with parameter $b=1$. The notation $i^{-1}m$ is substituted for the index α according to the comments above. A normalized HH of order L_m+1 is obtained in (2.16b) taking

$$A^2_1(i^{-1}mJT) = 1 - A^2_2(i^{-1}mJT). \tag{2.18}$$

For the case L_m+2, one similarly gets

$$B_5(\alpha JT) = \langle bL_m|G(r,b)|b,L_m+2,JT\alpha\rangle, \tag{2.19a}$$
$$B_4(\alpha JT) = \langle bL_m|F_{100}(\mathscr{R},b)|b,L_m+2,JT\alpha\rangle, \tag{2.19b}$$
$$B_3(\alpha JT) = \langle bL_m|F^*_{02,M}(\mathscr{R},b)|b,L_m+2,JT\alpha\rangle, \tag{2.19c}$$
$$B_2(\alpha JT\beta J'M') = \frac{1}{A_1(\beta,J'M'T')}\langle b,L_m+1,J'M'T\beta|$$
$$\times F^*_{01\,M-M'}(\mathscr{R},b)|b,L_m+2,JT\alpha\rangle$$
$$-\frac{A_2(\beta,J'M'T)}{A_1(\beta,J'M'T')}\langle b,L_m|F^*_{01\,M-M'}$$
$$\times(\mathscr{R},b)F^*_{01M'}(\mathscr{R},b)|b,L_m+2,JT\alpha\rangle. \tag{2.19d}$$

The operators occurring in the matrix elements are one- and two-body operators. We restrict our calculations to the simplest case where we have to deal with the 1p–1h excitations. We obtain therefore the following results:

$$B_5(i^{-1}mJT) = -(1/A)\sqrt{2(2j_i+1)/(\nu_m+1)}$$
$$\times\delta(j_ij_m)\delta(J,0)\delta(T,0)$$
$$\times\Big[(A-1)\langle n_il_i|r^2|n_ml_m\rangle\delta(l_il_m)$$
$$+2\sum_{n_kl_k} p(l_i+l_k)\cdot p(l_k+l_m)\langle n_il_i|r|n_kl_k\rangle$$
$$\times\langle n_kl_k\rangle|r|n_ml_m\rangle\Big], \tag{2.20a}$$

$$B_4(i^{-1}mJT) = (2/A)\sqrt{(2j_i+1)/3}\,\delta(j_ij_m)\delta(J,0)\delta(T,0)$$
$$\times\Big[-\langle n_il_i|r^2|n_ml_m\rangle\delta(l_il_m)$$
$$+2\sum_{n_kl_k} p(l_i+l_k)\cdot p(l_k+l_m)$$
$$\times\langle n_il_i|r|n_kl_k\rangle\langle n_kl_k|r|n_ml_m\rangle\Big], \tag{2.20b}$$

$$B_3(i^{-1}mJT) = (8/15A)\sqrt{(2j_i+1)/2}\,p(l_i+l_m+1)$$
$$\times(j_i2j_m;\tfrac{1}{2}\ 0\ \tfrac{1}{2})\delta(J,2)\delta(T,0)$$
$$\times\Big[\langle n_il_i|r^2|n_ml_m\rangle + 2\sum_{n_kl_k} p(l_i+l_k)$$
$$\cdot p(l_k+l_m)\langle n_il_i|r|n_kl_k\rangle\langle n_kl_k|r|n_ml_m\rangle\Big], \tag{2.20c}$$

$$B_2(i^{-1}mJT,k^{-1}pJ'M')$$
$$=\sqrt{2(2J'+1)/A}\,[(1J'J;M'-M,M')/A_1(k^{-1}pJ'M')]$$
$$\times\Big[\sqrt{2j_p+1}(j_p1j_m;\tfrac{1}{2}\ 0\ \tfrac{1}{2})W(1j_pJj_i;j_mJ')$$
$$\times p(l_m+l_p)\langle n_pl_p|r|n_ml_m\rangle$$
$$+\sqrt{2j_i+1}(j_i1j_k;\tfrac{1}{2}\ 0\ \tfrac{1}{2})(-)^{1+J'-J}W(1j_kJj_m;j_iJ')$$
$$\times p(l_i+l_k)\langle n_il_i|r|n_kl_k\rangle\Big]$$
$$-[A_2(k^{-1}p;J'M')/A_1(k^{-1}p;J'M')]$$
$$\times[\sqrt{2}(112;M-M',M')B_3(i^{-1}m,JT)$$
$$+(-)^{1+M'}\sqrt{\tfrac{2}{3}}\delta(M,0)B_4(i^{-1}mJT)]. \tag{2.20d}$$

The sum over n_k and l_k refers to the quantum numbers of the core. The case of 2 p–2h can be treated using standard shell model techniques to obtain the 2 p–2h configurations as a linear combination of HO determinants with definite total angular momentum and isospin which allow us to calculate the expectation values given in the formulae (2.16).

III. HYPERSPHERICAL MATRIX ELEMENTS

The calculation of the matrix element

$$\langle D^{(\alpha JT)}_L|Q|D^{(\beta J'T')}_{L'}\rangle\rangle$$
$$=\int d\Omega\, D^{(\alpha JT)^+}_L(\Omega)Q(\xi)D^{(\beta J'T')}_{L'}(\Omega) \tag{3.1}$$

can be performed using the Surkov theorem,[6] which states that for any translationally invariant operator $O(\xi)$ we have the identity

$$\int d\Omega\, O = (\pi^{3/2}r^{3A-5}\pi i)^{-1}\int_C ds\, e^{sr^2}s^{3/2}$$
$$\times\int d\tau_{3A}\exp\Big(-s\sum_{i=1}^{A} r_1^2\Big)O(\xi), \tag{3.2}$$

where the path C is along the imaginary axis. The integration over the variable s proceeds from an integral representation of the delta function, which enables one to exclude the hyperradius in the integration over all the variables $\mathbf{r}_i$. In this respect, when the hyperradial dependence in $O(\xi)$ is factorized, it can be obviously extracted out of the integral over τ_{3A}. Indeed, the result of the integral (3.1) depends on the hyperradius only.

A. A procedure of calculation

The matrix element (3.1) can be calculated using Surkov's theorem with

$$O(\xi) = D_L^{(\alpha JT)*}(\Omega)Q(\xi)D_{L'}^{(\beta J'T')}(\Omega). \tag{3.3}$$

In order to emphasize the essential aspects of the procedure, let us particularize to with the case $L = L' = L_m + 1$. One can extract from (2.16b) the HH of order $L_m + 1$

$$\begin{aligned} D^{(\alpha JT)}_{L_m+1}(\Omega) = A_1^{-1}(\alpha JT)\left[\phi^{(b)}_{000}(\mathscr{R})\psi^{(b)}_{L_m+1,0}(r)\right]^{-1} \\ \times\left[|b,L_m+1,JT\alpha\rangle - A_2(\alpha JT)\phi^{(b)}_{01m}(\mathscr{R})\right. \\ \left.\times\psi^{(b)}_{L_m,0}(r)D_{L_m}(\Omega)\right], \end{aligned} \tag{3.4}$$

where from (2.8) and (2.10) we have

$$\begin{aligned} &\left[\phi^{(b)}_{000}(\mathscr{R})\psi^{(b)}_{L_m+1,0}(r)\right]^{-1} \\ &\quad = \left[\frac{\pi^{3/2}\Gamma(\nu_m+1)}{2}\right]^{1/2} b^{\nu_m+7/2}\frac{\exp\left(\frac{1}{2}\sum_{i=1}^{A}\frac{r_i^2}{b^2}\right)}{r^{L_m+1}}. \end{aligned} \tag{3.5}$$

The left-hand side of (3.4) does not obviously depend on the center of mass, the hyperradius, and the HO parameter in spite of their occurrence at the right-hand side. We must keep in mind that (3.4) is an identity valid for any value of $\mathbf{R}$, r, and b. Therefore, the term $\exp(-s\Sigma_{i=1}^{A} r_i^2)$ of the Surkov theorem can be eliminated by choosing $b^2 = s^{-1}$ as Baz' and Zhukov[3] did for the case $L = L' = L_m$. The matrix element (3.1) can be obtained in two steps: First, one calculates all the matrix elements between HO determinants of order L_m, $L_m + 1$ and, then, with $b^2 = s^{-1}$, we perform an additional integration over s.

However, the operator occurring between the HO determinants is the original Q multiplied by some center-of-mass functions, which are included in the linear combination (3.4). For instance, in the case $L = L' = L_m + 1$ the following term is present:

$$\begin{aligned} &A_2(\alpha JT)\langle\phi^{(b)}_{01M}\psi^{(b)}_{L_m,0}D_{L_m}|Q|b,L_m+1,J'T'\beta\rangle \\ &\quad = A_2(\alpha JT)\langle bL_m|F^{*}_{01M}(\mathscr{R},b)Q|b,L_m+1,J'T'\beta\rangle. \end{aligned} \tag{3.6}$$

This event is avoided when $L = L' = L_m$, since, using (2.16a) and the Surkov theorem, one gets

$$\begin{aligned} &\langle D_{L_m}|Q|D_{L_m}\rangle \\ &\quad = \frac{\Gamma(\nu_m+1)}{2\pi i r^{2\nu_m}}\int_C ds\frac{e^{sr^2}}{s^{\nu_m+1}}\langle s^{-1/2}L_m|Q|s^{-1/2}L_m\rangle, \end{aligned} \tag{3.7}$$

which is the result of Baz' and Zhukov. At first sight, it seems that for $L > L_m$ the calculation is much too cumbersome since we have to deal with an A-body operator between HO determinants. However, this difficulty can be overcome by noticing that the center of mass is factorized in both the volume element and the integrand, enabling one to perform the integration over the center of mass. For instance, the term (3.6) can be written as

$$\begin{aligned} &A_2(\alpha JT)A_1(\beta J'T')\langle\phi^{(b)}_{01m}|\phi^{(b)}_{000}\rangle\langle\psi^{(b)}_{L_m,0}D_{L_m}|Q|\psi^{(b)}_{L_m+1,0}D^{(\beta J'T')}_{L_m+1}\rangle \\ &\quad + A_2(\alpha JT)A_2(\beta J'T')\langle\phi^{(b)}_{01M}|\phi^{(b)}_{01M'}\rangle \\ &\quad\times\langle\psi^{(b)}_{L_m,0}D_{L_m}|Q|\psi^{(b)}_{L_m,0}D_{L_m}\rangle \\ &\quad = A_2(\alpha JT)A_2(\beta J'T')\delta(M,M')\langle bL_m|Q|bL_m\rangle. \end{aligned} \tag{3.8}$$

This last expression exhibits the useful structure required to use the Surkov theorem as in (3.7). For larger L the situation is very similar but, of course, more complex. Once again, the choice $b^2 = s^{-1}$ enables one to eliminate the exponential of the Surkov theorem, and to perform the integration over the factorized center-of-mass functions. Finally, the hyperradial functions, which do not occur in the case $L = L' = L_m + 1$, can be isolated out of the integral, according to the remark given at the beginning of Sec. III.

B. Hyperspherical matrix elements up to $L_m + 2$

This procedure can be applied to any order L, but of course, the complexity of the formulae increases with the order. The matrix elements up to the order $L_m + 2$ are given below without details:

$$\begin{aligned} &\langle D_{L_m}|Q|D_{L_m}\rangle \\ &\quad = \frac{\Gamma(\nu_m+1)}{2\pi i r^{2\nu_m}}\int_C ds\frac{e^{sr^2}}{s^{\nu_m+1}}\langle s^{-1/2},L_m|Q|s^{-1/2},L_m\rangle, \end{aligned} \tag{3.9}$$

$$\begin{aligned} \langle D_{L_m}|Q|D^{(\alpha JT)}_{L_m+1}\rangle &= \frac{\Gamma(\nu_m+1)\sqrt{\nu_m+1}}{2\pi i A_1(\alpha JT)r^{2\nu_m+1}} \\ &\times\int ds\frac{e^{sr^2}}{s^{\nu_m+3/2}} \\ &\times\langle s^{-1/2},L_m|Q|s^{-1/2},L_m+1,JT\alpha\rangle, \end{aligned} \tag{3.10a}$$

$$\begin{aligned} &\langle D^{(\alpha JT)}_{L_m+1}|Q|D^{(\beta,J'T')}_{L_m+1}\rangle \\ &\quad = \frac{\Gamma(\nu_m+3)}{2\pi i A_1(\alpha,JT)A_2(\beta,J'T')r^{2\nu_m+2}} \\ &\quad\times\int_C ds\frac{e^{sr^2}}{s^{\nu_m+2}} \\ &\quad\times[\langle s^{-1/2},L_m+1,JT,\alpha|Q|s^{-1/2},L_m+1,J'T',\beta\rangle \\ &\quad - A_2(\alpha,JT)A_2(\beta,J'T')\delta(M,M') \\ &\quad\times\langle s^{-1/2},L_m|Q|s^{-1/2},L_m\rangle], \end{aligned} \tag{3.10b}$$

$$\begin{aligned} &\langle D_{L_m}|Q|D^{(\alpha,JT)}_{L_m+2}\rangle \\ &\quad = \frac{\Gamma(\nu_m+1)\sqrt{(\nu_m+1)(\nu_m+2)}}{2\pi i B_1(\alpha,JT)r^{2\nu_m+2}} \\ &\quad\times\int_C ds\frac{e^{sr^2}}{s^{\nu_m+2}}[\langle s^{-1/2},L_m|Q|s^{-1/2},L_m+2,JT,\alpha\rangle \\ &\quad - B_5(\alpha,JT)G(r,s^{-1/2})\cdot\langle s^{-1/2},L_m|Q|s^{-1/2},L_m\rangle], \end{aligned} \tag{3.11a}$$

$$\begin{aligned} &\langle D^{(\alpha JT)}_{L_m+1}|Q|D^{(\beta J'T')}_{L_m+2}\rangle \\ &\quad = \frac{\Gamma(\nu_m+2)\sqrt{\nu_m+2}}{2\pi i A_1(\alpha,JT)B_1(\beta,J'T')r^{2\nu_m+3}} \\ &\quad\times\int_C ds\frac{e^{sr^2}}{s^{\nu_m+5/2}}\Big[\langle s^{1/2},L_m+1,JT,\alpha|Q|s^{-1/2}, \\ &\quad\times L_m+2,J'T',\beta\rangle - B_5(\beta,J'T')G(r,s^{-1/2}) \\ &\quad\cdot\langle s^{-1/2},L_m+1,JT,\alpha|Q|s^{-1/2},L_m\rangle \\ &\quad - A_2(\alpha,JT)\sum_{(\gamma J'')}\frac{B_2(\beta,J'T';\gamma,J'',M'-M)}{A_1(\gamma,J''T')} \\ &\quad\times\langle s^{-1/2},L_m|Q|s^{-1/2},L_m+1,J''T'_\gamma\rangle\Big], \end{aligned} \tag{3.11b}$$

$$\langle D^{(\alpha,JT)}_{L_m+2}|Q|D^{(\beta,J'T')}_{L_m+2}\rangle = \frac{\Gamma(\nu_m+3)}{2\pi i B_1(\alpha,JT)B_1(\beta,J'T')r^{2\nu_m+4}}\int_C ds\frac{e^{sr^2}}{s^{\nu_m+3}}\Big\{\langle s^{-1/2},L_m+2,JT,\alpha|Q|s^{-1/2},$$

$$\times L_m+2,J'T',\beta\rangle - \sum_{\substack{(\gamma J''M'')\\(\delta J'''M''')}}\delta(M-M'',M'-M''')\frac{B_2(\alpha,JT;\gamma,J''M'')B_2(\beta,J'T';\delta,J'''M''')}{A_1(\gamma,J''T)A_1(\delta,J'''T')}$$

$$\times\langle s^{-1/2},L_m+1,J''T,\gamma|Q|s^{-1/2},L_m+1,J'''T',\delta\rangle - G(r,s^{-1/2})[B_5(\alpha,JT)\langle s^{-1/2},L_m|Q|s^{-1/2},L_m+2,J'T',\beta\rangle$$

$$+B_5(\beta,J'T')\langle s^{-1/2},L_m+2,JT,\alpha|Q|s^{-1/2},L_m\rangle] - [B_3(\alpha,JT)B_3(\beta,J'T')\delta(M,M')+B_4(\alpha,JT)B_4(\beta,J'T')$$

$$-G^2(r,s^{-1/2})\cdot B_5(\alpha,JT)B_5(\beta,J'T) - \sum_{\substack{(\gamma J''M'')\\(\delta J''')}}\frac{B_2(\alpha,JT;\gamma,J''M'')B_2(\beta,J'T';\delta J'''M''')}{A_1(\gamma,J''T)A_1(\delta,J'''T')}$$

$$\times\delta(M-M',M'-M''')A_2(\gamma,J''T)A_2(\delta,J'''T')]\langle s^{-1/2},L_m|Q|s^{-1/2},L_m\rangle\Big\}. \tag{3.11c}$$

We have considered here a general translationally invariant operator Q. Of course, some simplifications are obtained when Q has specific properties. For instance, the matrix elements (3.10a) and (3.11b) vanish when Q is a parity-conserving operator. Also, when Q is a scalar (isoscalar) the angular momentum (isospin) mus be conserved.

IV. FINAL COMMENTS

In our calculation we used systematic connections between HH and HO determinants. This procedure enables one to utilize standard techniques of the HO shell model. It is clear that the procedure is particularly suitable for operators Q leading to an analytical integrations over $s = b^{-2}$. In the various L_m calculations previously performed zero-range density dependent potentials and central potentials consisting of superposition of Gaussians have been used. All of them lead to analytical expressions in b which are easily integrated over s. The situation is similar for orders higher than L_m because the mathematical problem is essentially the same.

We have only given explicit expressions for the hyperspherical matrix elements in the cases L_m, L_m+1, and L_m+2. It is obvious that the procedure can be extended to higher orders, but it becomes rapidly tedious. The difficulty proceeds from the elimination of both the center-of-mass motion and the hyperradial excitations included in an HO determinant, which must be taken out in order to obtain a pure HH. Especially, the elimination of the center-of-mass motion cannot be accomplished using a standard prescription, but must be done case by case, leading to more complicated matrix elements. Anyway, the problem can be handled for the lowest orders as we did for L_m+1 and L_m+2. These orders include most of the 1p–1h and 2p–2h lowest nuclear excited states.

Let us say, finally, a few words about the internal motion. Any determinantal wavefunction can be expressed as a linear combination of center-of-mass and internal motions (see, e.g., Refs. 7 and 8). In this respect the hyperspherical expansion (2.11) of an HO determinant is nothing more that a particular choice for the internal motion given in terms of hyperspherical coordinates. We have shown explicitly how the various center-of-mass states contribute to an HO determinant. It seems worthwhile to recall that in general a single HO determinant describing an excited state contains not only spurious center-of-mass motion but also hyperradial excitations. In Refs. 9 and 10 it has been shown that the hyperradial excitations are to be identified with the giant monopole resonances or breathing mode. These collective excitations described by the hyperradial functions must be eliminated in order to select pure internal excitations.

ACKNOWLEDGMENTS

We would like to thank Professor J. Bernabéu for some productive discussions as well as for a critical reading of the manuscript. We acknowledge the C.A.I.C.T. (Ministerio de Educación y Ciencias, Spain) for partial financial support.

[1] M. Fabre de la Ripelle and J. Navarro, Ann. Phys. (N.Y.) **123**, 185 (1979).
[2] D. Levi and M. Moshinsky, Nuovo Cimento A **20**, 107 (1974).
[3] A. L. Baz' and N.Y. Zhukov, Sov. J. Nucl. Phys. **11**, 435 (1970).
[4] J. A. Castilho-Alcaras, J. Math. Phys. **19**, 1714 (1978).
[5] I. S. Towner, *A Shell Model Decription of Light Nuclei* (Oxford U.P., Oxford, 1977).
[6] E. L. Surkov, Sov. J. Nucl. Phys. **5**, 644 (1966).
[7] J. P. Elliott and T. H. R. Skyrme, Proc. Roy Soc. (London) A **232**, 561 (1955).
[8] C. M. Vincent, Phys. Rev. C **8**, 929 (1973).
[9] M. Sotona and J. Zofka, Phys. Lett. B **57**, 27 (1975).
[10] J. Navarro, Phys. Lett. B **62**, 22 (1976).

A new approach to perturbation theory: star diagrams

J. C. Houard and M. Irac-Astaud
Laboratoire de Physique Théorique et Mathématique, Université Paris VII—Tour 33-43—1er étage, 2, place Jussieu, 75251 Paris Cedex 05, France

(Received 1 June 1982; accepted for publication 1 October 1982)

Perturbation theory for first-order nonlinear differential equations with source is developed in a new way, and associated with diagrams that we call star diagrams. In some cases the method allows one to express the n-point functions in explicit form.

PACS numbers: 02.30.Hq

1. INTRODUCTION

The motions of dynamical systems submitted to external excitations are governed by differential equations involving a forcing term (or source term).[1] When the source function is considered as an arbitrary datum, the corresponding solutions are functionals of this source. As is well known, in the case of linear equations each of these functionals is expressed with the help of an integral involving a Green function. For nonlinear equations, this notion is generalized into that of n-point functions defined as the kernels of the functional Taylor expansion of the solutions with respect to the source. These n-point functions are, in particular, currently used in field theory, where they are related to the vacuum expectation values of the fields.[2] We intend studying here the structure of these functionals, that is to say to explain the relations between the different orders of the Taylor expansion. In a primary approach, we limit ourselves to first-order equations.

More specifically, we are interested in the retarded solution of the equations of the following type:

$$\frac{dx}{dt} = \eta + P_t(x), \tag{E}$$

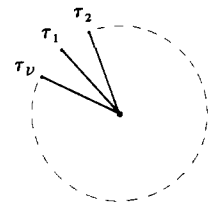
where P_t is a given polynomial, the coefficients of which depend on t:

$$P_t(x) = \sum_{\nu \geq 1} \alpha_\nu(t)\frac{x^\nu}{\nu!},$$

the arbitrary source function η vanishing for negative t. The retarded solution is a functional of η, which can be written in the form

$$x[t,\eta] = \sum_{n=1}^{\infty} \frac{1}{n!}\int d\tau_1 \cdots d\tau_n G_n(t;\tau_1,\ldots,\tau_n)\eta(\tau_1)\cdots\eta(\tau_n).$$

This expansion actually exists and converges for an interval $0 \leqslant t \leqslant T$ for some strictly positive T, if only η and α_ν, $\nu \geqslant 1$, are continuous functions in $[0, \infty[$. This can be easily established by using the expressions of the G_n's, which result from the usual iterative process applied to (E). Our concern is to search for an explicit expression of the n-point functions G_n. Such an expression was previously exhibited in the particular case where P reduces to a monomial $\alpha(t)(x^\nu/\nu!)$.[3] In this case, the n-point functions are all expressed as linear combinations of products only built with the help of the first of them, G_ν. This basic element will be termed a ν-star, and graphically represented by the schema



$$= \theta(t - \sup(\tau_1,\ldots,\tau_\nu))\int_{\sup(\tau_1,\ldots,\tau_\nu)}^{t} d\tau\,\alpha(\tau).$$

The n-point function is then represented by diagrams obtained by a juxtaposition of ν-stars linked by their extremities, its value $G_n(t;\tau_1,\ldots,\tau_n)$ being thus a polynomial in a primitive of α taken for various arguments. This result suggests looking for a generalization, in the case of any polynomial P, in which ν-stars for different values of ν would occur.

In what follows, we show that this is indeed the case, provided that the functions α_ν, associated with each type of star, are replaced by modified functions. These functions are functionals of the α's, which represent a dressing of the basic stars.

Section 2 introduces some functionals $X_\nu(t,\tau)$, the determination of which is equivalent to that of the solution $x(t)$. For comparison with our method, the usual Feynman expansion is recalled in Sec. 3. In Sec. 4, the equations obtained in Sec. 2 are used to develop a perturbation theory for the functionals X_ν, in such a manner that the dressed stars naturally appear. Equations for the dressed stars, and their expansions, are also derived. The result is a set of rules characterizing all the diagrams in our theory and giving their values. In Sec. 5, these rules are used to find the ones defining the solution $x(t)$. In particular, the numerical coefficient of each diagram contains a factor depending on the whole topology of the diagram. Further properties of this factor are developed in the Appendix. Finally, Sec. 6 is devoted to a possible physical interpretation of perturbation theory based on our expansion, as compared to the Feynman one, and to a discussion of the occurrence of some simplifications when the α_ν's are proportional.

2. BASIC RELATIONS

To the retarded solution $x(t)$ of Eq. (E), one associates the functionals $X_\nu(t,\tau)$ of η defined by

$$X_\nu(t,\tau) = \exp\left(\frac{1}{\nu!}\int_\tau^t \alpha_\nu(\tau')x^{\nu-1}(\tau')d\tau'\right). \tag{2.1}$$

The product $\theta(t-\tau)\Pi_{\nu\geq 1}X_\nu(t,\tau)$ is the retarded Green function of the differential operator

 0022-2488/83/081997-09$02.50

$$\frac{d}{dt} - \sum_{\nu\geq 1} \frac{\alpha_\nu(t)}{\nu!} x(t)^{\nu-1},$$

so that the solution $x(t)$ of (E) may be written as

$$x(t) = \int_0^t d\tau \left(\prod_{\nu\geq 1} X_\nu(t,\tau) \right) \eta(\tau). \tag{2.2}$$

In what follows the problem of determining $x(t)$ is replaced by the problem of determining the X_ν's. Equation (E) is in fact equivalent to the set constituted by (2.2) and the equations

$$\frac{d}{dt} X_\nu(t,\tau) = \frac{1}{\nu!} \alpha_\nu(t) x(t)^{\nu-1} X_\nu(t,\tau), \quad X_\nu(\tau,\tau) = 1. \tag{2.3}$$

The latter, or the equivalent expression (2.1), implies the composition relation

$$X_\nu(t_1,t_2) X_\nu(t_2,t_3) = X_\nu(t_1,t_3), \quad \forall\, t_1,t_2,t_3,\nu. \tag{2.4}$$

The functionals X_ν are related to the functional derivative of x with respect to η: differentiation of (E) gives, for $\tau > 0$,

$$\frac{d}{dt} \frac{\delta x(t)}{\delta \eta(\tau)} = \delta(t-\tau) + \sum_{\nu\geq 1} \frac{\alpha_\nu(t)}{(\nu-1)!} x(t)^{\nu-1} \frac{\delta x(t)}{\delta \eta(\tau)}, \tag{2.5}$$

$$\frac{\delta x(0)}{\delta \eta(\tau)} = 0$$

whose solution reads

$$\frac{\delta x(t)}{\delta \eta(\tau)} = \theta(t-\tau) \prod_{\nu\geq 1} (X_\nu(t,\tau))^\nu, \quad \tau > 0. \tag{2.6}$$

Equations (2.2), (2.4), and (2.6) are equivalent to the original equation (E) in the case where the polynomial P is homogeneous:

Proposition: When ν takes one value only, for any solution (x,X_ν), of Eqs. (2.2), (2.4), and (2.6), the function x is a solution of Eq. (E) for some function α.

Proof: From (2.4) it follows that $(d/dt) \ln X_\nu(t,\tau)$ does not depend on τ. This allows one to define α_ν (*a priori*, depending on η), for $t > 0$, by

$$\alpha_\nu(t,\eta) = \nu! x^{1-\nu}(t) \frac{d}{dt} \ln X_\nu(t,\tau), \quad t > 0. \tag{2.7}$$

The definition (2.2) of x then implies, since $X_\nu(t,t) = 1$,

$$\frac{d}{dt} x(t) = \eta(t) + \frac{1}{\nu!} \alpha_\nu(t,\eta) x^\nu(t), \quad x(0) = 0. \tag{2.8}$$

This is Eq. (E) if α does not depend on η, which we prove now. From (2.8) one gets, with the help of (2.6),

$$\frac{d}{dt} \frac{\delta x(t)}{\delta \eta(\tau)} = \delta(t-\tau) + \frac{1}{(\nu-1)!} \alpha_\nu(t,\eta) x(t)^{\nu-1} \theta(t-\tau) X_\nu(t,\tau) + \frac{1}{\nu!} x(t)^\nu \frac{\delta \alpha_\nu(t,\eta)}{\delta \eta(\tau)}. \tag{2.9}$$

Replacing $\delta x(t)/\delta\eta(\tau)$ by (2.6) and using (2.7) gives the first two terms on the right-hand side of (2.9), so that it remains

$$\frac{\delta \alpha_\nu(t,\eta)}{\delta \eta(\tau)} = 0. \quad \text{Q.E.D.} \tag{2.10}$$

In the case where P is not homogeneous, (2.10) is replaced by

$$\sum_{\nu\geq 1} \frac{1}{\nu!} x(t)^\nu \frac{\delta \alpha_\nu(t,\eta)}{\delta \eta(\tau)} = 0. \tag{2.11}$$

This does not suffice to conclude the independence on η of the α_ν's, which has to be explicitly postulated; that is, in terms of (x,X),

$$\frac{\delta}{\delta \eta(\tau)} \left(x^{1-\nu} \frac{d}{dt} \ln X_\nu \right) = 0, \quad \forall \nu. \tag{2.12}$$

The basic equations (2.2) and (2.3) will be taken as the starting point of perturbation theory. At first, one recovers the usual Feynman expansion.

3. FEYNMAN DIAGRAMS

For the product

$$K(t,\tau) = \prod_{\nu\geq 1} X_\nu(t,\tau),$$

Eq. (2.3) gives

$$\frac{d}{dt} K(t,\tau) = \left(\sum_{\nu\geq 1} \frac{1}{\nu!} \alpha_\nu(t) x(t)^{\nu-1} \right) K(t,\tau), \tag{3.1}$$

$$K(\tau,\tau) = 1$$

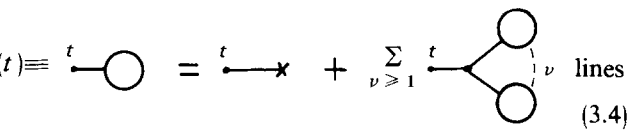
and, therefore,

$$K(t,\tau) = 1 + \int_\tau^t d\tau' \left(\sum_{\nu\geq 1} \frac{1}{\nu!} \alpha_\nu(\tau') x(\tau')^{\nu-1} \right) K(\tau',\tau). \tag{3.2}$$

By multiplying the two members of this equation by $\eta(\tau)$, integrating on τ between 0 and t, and permuting the integrations in the last term, one gets

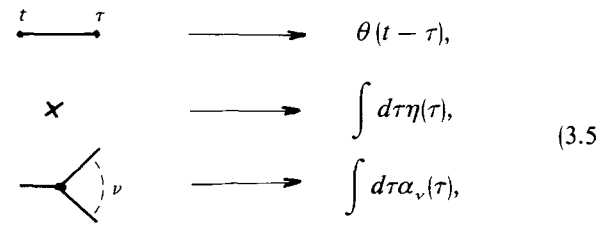
$$x(t) = \int_0^t \eta(\tau)\, d\tau + \int_0^t d\tau \sum_{\nu\geq 1} \frac{1}{\nu!} \alpha_\nu(\tau) x(\tau)^\nu. \tag{3.3}$$

This is nothing but the integral form of Eq. (E). The iteration method leads to the well-known Feynman expansion: diagrammatically (3.3) reads



(3.4)

with the correspondence rules



(3.5)

the factor $1/\nu!$ accounting for the symmetry of the diagram with ν lines. The diagrams appearing in the iteration of (3.4) are tree diagrams stretching to the right from the point t, at the vertices of which occur the interactions corresponding to the different terms of the polynomial P; the associated values in the expansion of (3.3) are the integrals defined by the rules (3.5) applied to each part of the diagrams. These integrals contribute to the nth order in η by an expression of the general form

$$\int d\tau_1 \cdots d\tau_n \eta(\tau_1) \cdots \eta(\tau_n) G(t,\tau_1,\ldots,\tau_n), \tag{3.6}$$

where G is the n-point function. This latter is then a sum of terms coming from various diagrams; according to (3.5) these terms are integrals on the variables attached to the interaction vertices. In the next section, we develop another version of perturbation theory which, in some cases, furnishes an expression for G completely explicitly, that is to say free from integrations.

4. STAR DIAGRAMS

Instead of solving the equation for the product $K = \Pi_\nu X_\nu$ one considers separately the equations for the X_ν's. The solution now developed needs, at first, the elimination of the linear term of P; this is easily done by replacing the function $x(t)$ by $x(t)\exp[-\int_0^t d\tau\, \alpha_1(\tau)]$, which satisfies an equation of the type (E), without the linear term, and with modified functions η and α_ν, $\nu \neq 1$. From now on one thus assumes that P does not contain a linear term. With the help of (2.4), one transforms (2.3) into

$$\frac{d}{d\tau} X_\nu(t,\tau) = -\frac{1}{\nu!}\alpha_\nu(\tau) x(\tau)^{\nu-1} X_\nu(t,\tau), \tag{4.1}$$

then

$$\frac{d}{d\tau} X_\nu(t,\tau)^{\nu-1} = -\frac{\nu-1}{\nu!}\alpha_\nu(\tau) x(\tau)^{\nu-1} (X_\nu(t,\tau))^{\nu-1} \tag{4.2}$$

and, introducing the expression (2.2) of x and the composition law (2.4), one gets

$$\frac{d}{d\tau} X_\nu(t,\tau)^{\nu-1} = -\frac{\nu-1}{\nu!}\alpha_\nu(\tau) \int_0^\tau d\tau_1 \cdots d\tau_{\nu-1}\, \eta(\tau_1)\cdots\eta(\tau_{\nu-1}) \prod_\mu (X_\mu(\tau,\tau_1)\cdots X_\mu(\tau,\tau_{\nu-1}))\,(X_\nu(t,\tau))^{\nu-1}$$

$$= -\frac{\nu-1}{\nu!}\alpha_\nu(\tau) \int_0^\tau d\tau_1 \cdots d\tau_{\nu-1} \eta(\tau_1)\cdots\eta(\tau_{\nu-1}) \prod_\mu (X_\mu(t,\tau_1)\cdots X_\mu(t,\tau_{\nu-1})) \prod_{\rho\neq\nu} (X_\rho(\tau,t))^{\nu-1}. \tag{4.3}$$

By integration, this gives

$$X_\nu(t,\tau)^{\nu-1} = 1 + \frac{\nu-1}{\nu!}\int_\tau^t d\tau' \alpha_\nu(\tau') \int_0^{\tau'} d\tau_1 \cdots d\tau_{\nu-1}$$

$$\times \eta(\tau_1)\cdots\eta(\tau_{\nu-1}) \prod_\mu (X_\mu(t,\tau_1)\cdots X_\mu(t,\tau_{\nu-1})) \prod_{\rho\neq\nu} (X_\rho(\tau',t))^{\nu-1}. \tag{4.4}$$

Multiplying by $\theta(t-\tau)$ and permuting the integrations, (4.4) becomes

$$\theta(t-\tau)X_\nu(t,\tau) = \theta(t-\tau)\left[1 + \frac{\nu-1}{\nu!}\int_0^t d\tau_1\cdots d\tau_{\nu-1}\,\eta(\tau_1)\cdots\eta(\tau_{\nu-1})\right.$$

$$\left.\times\left(\prod_\mu \prod_{i=1}^{\nu-1} \theta(t-\tau_i) X_\mu(t,\tau_i)\right) \int_{\sup(\tau_1,\ldots,\tau_{\nu-1},\tau)}^t d\tau' \alpha_\nu(\tau') \prod_{\rho\neq\nu} (\theta(t-\tau')X_\rho(\tau',t))^{\nu-1}\right]^{1/(\nu-1)}. \tag{4.5}$$

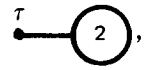
One now takes these equations as the starting point of a perturbation theory. To simplify the subsequent exposition we limit ourselves to a polynomial P of the form $P(x) = (\alpha/2!)x^2 + (\beta/3!)x^3$ and change X_2 and X_3 into X and Y. Equations (4.5) now read

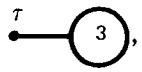
$$\theta(t-\tau)X(t,\tau) = \theta(t-\tau)\left[1 + \frac{1}{2}\int_0^t d\tau'\eta(\tau')[\theta(t-\tau')X(t,\tau')]\,[\theta(t-\tau')Y(t,\tau')]\int_{\sup(\tau,\tau')}^t d\tau''\alpha(\tau'')\theta(t-\tau'')Y(\tau'',t)\right], \tag{4.6}$$

$$\theta(t-\tau)Y(t,\tau) = \theta(t-\tau)\left[1 + \frac{1}{3}\int_0^t d\tau' d\tau''\eta(\tau')\eta(\tau'')[\theta(t-\tau')X(t,\tau')]\,[\theta(t-\tau')Y(t,\tau')]\right.$$

$$\left.\times[\theta(t-\tau'')X(t,\tau'')]\,[\theta(t-\tau'')Y(t,\tau'')]\int_{\sup(\tau,\tau',\tau'')}^t d\tau'''\beta(\tau''')\,[\theta(t-\tau''')X(\tau''',t)]^2\right]^{1/2}.$$

Equations (4.6) may be translated into diagrams with the help of the following representations, t being considered as fixed:

$\theta(t-\tau)X(t,\tau)$ ⟶  ,

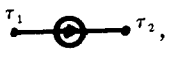
$\theta(t-\tau)Y(t,\tau)$ ⟶  ,

$\theta(t-\tau)X(t,\tau)Y(t,\tau)$ ⟶ 

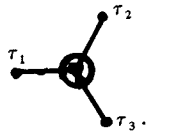
(4.7)

and

$\theta(t-\sup(\tau_1,\tau_2))$

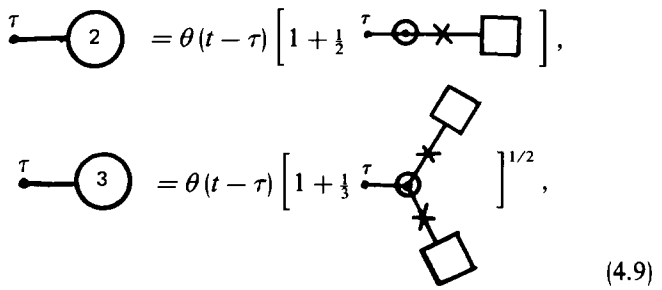
$\times \int_{\sup(\tau_1,\tau_2)}^{t} d\tau\,\alpha(\tau)Y(\tau,t)$ ⟶  ,

$\theta(t-\sup(\tau_1,\tau_2,\tau_3))$

$\times \int_{\sup(\tau_1,\tau_2,\tau_3)}^{t} d\tau\,\beta(\tau)(X(\tau,t))^2$ ⟶  .

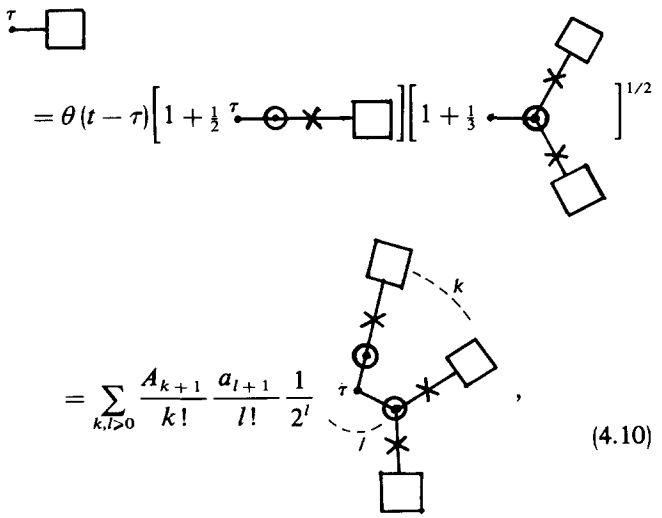
(4.8)

The last two diagrams will be respectively called *2-star* and *3-star*. Equations (4.6) now take the form



(4.9)

in which the symbol $\times$ has the same meaning as in (3.5) for Feynman diagrams. From these equations, one deduces

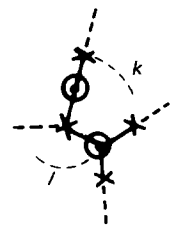


(4.10)

with

$$A_1 = 1, \quad A_2 = \tfrac{1}{2}, \quad A_k = 0,\ k \geqslant 3,$$
$$a_{l+1} = \frac{(-1)^{l+1}}{6^l(2l-1)}\frac{(2l)!}{l!}. \tag{4.11}$$

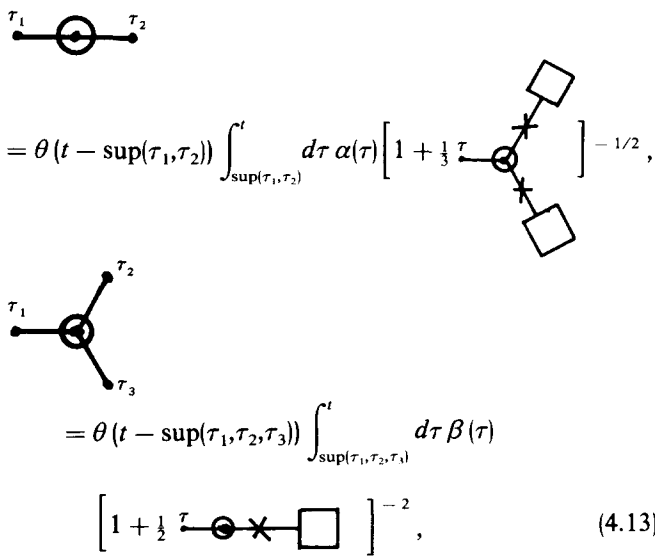
The iteration of (4.10) leads to tree diagrams radiating from the point labeled by τ, each of them being calculated by the rule: From any source η starts a number of 2-stars and 3-stars, say k and l, respectively, according to the drawing



(4.12)

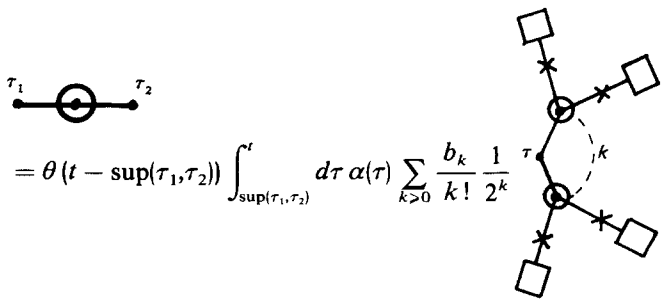
Such a vertex as the one labeled by τ standing furthest on the left gives a factor $A_{k+1}a_{l+1}$, the symmetry being accounted for by an extra factor $1/S$, where S is the order of the symmetry group of the whole diagram. According to this rule, the expansion of $x(t)$ is expressed by a sum of diagrams consisting of a juxtaposition of 2- and 3-stars, bound by the integrations on η. Until then, however, the stars have been considered as given functions. In fact, these functions still have to be submitted to the iteration process.

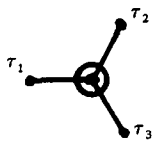
From (4.7) to (4.9) follows



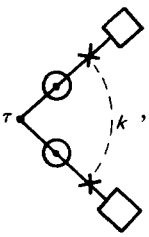
(4.13)

the expansion of which reads





$$= \theta(t-\sup(\tau_1,\tau_2,\tau_3)) \int_{\sup(\tau_1,\tau_2,\tau_3)}^{t} d\tau\, \beta(\tau) \sum_{k>0} \frac{B_k}{k!}$$



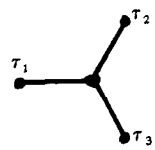
(4.14)

with

$$b_k = (-1)^k \frac{(2k)!}{6^k\, k!}, \qquad B_k = (-1)^k \frac{(k+1)!}{2^k}. \tag{4.15}$$

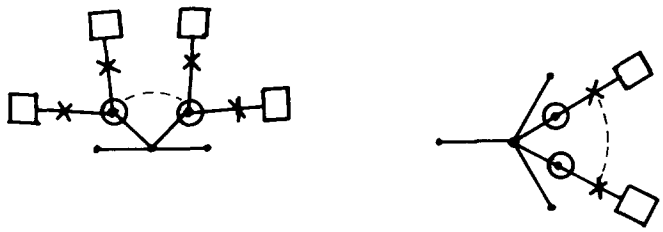
The first terms ($k = 0$) in the expansions (4.14) will be called *simple stars*, and represented by

$$= \theta(t - \sup(\tau_1,\tau_2)) \int_{\sup(\tau_1,\tau_2)}^{t} d\tau\, \alpha(\tau),$$



$$= \theta(t - \sup(\tau_1,\tau_2,\tau_3)) \int_{\sup(\tau_1,\tau_2,\tau_3)}^{t} d\tau\, \beta(\tau). \tag{4.16}$$

The next terms can be interpreted as a dressing of the center of these stars, corresponding to a modification of the functions α and β by a multiplicative term, according to the drawings



(4.17)

From (4.12) and (4.17) the following complete rule results: The perturbation expansion of the function $\theta(t-\tau)X(t,\tau)Y(t,\tau)$ is represented by the sum of the tree diagrams radiating from the point labeled by τ, so that

R_1: from each source η start arbitrary numbers k and l of 2-stars and 3-stars with the coefficient $A_{k+1}a_{l+1}$,

R_2: from each center α of 2-stars (respectively, β of 3-stars) starts an arbitrary number k of 3-stars (respectively, 2-stars) with a coefficient b_k (respectively, B_k),

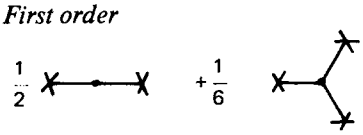
R_3: to the whole diagram is furthermore assigned the symmetry factor $1/S$.

From (2.2), the expansion for $x(t)$ is then obtained by integrating on τ between 0 and t, after multiplying by $\eta(\tau)$; this is graphically represented by simply adding a cross ($\times$) at the point τ. To illustrate these rules, let us give the lowest orders in α and β for $x(t)$:

Zeroth order

$$\times = \int_0^t d\tau \eta(\tau),$$

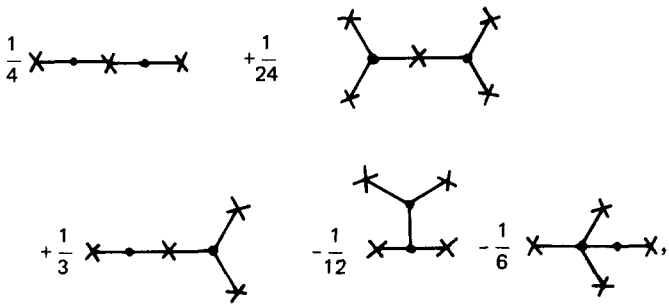
First order



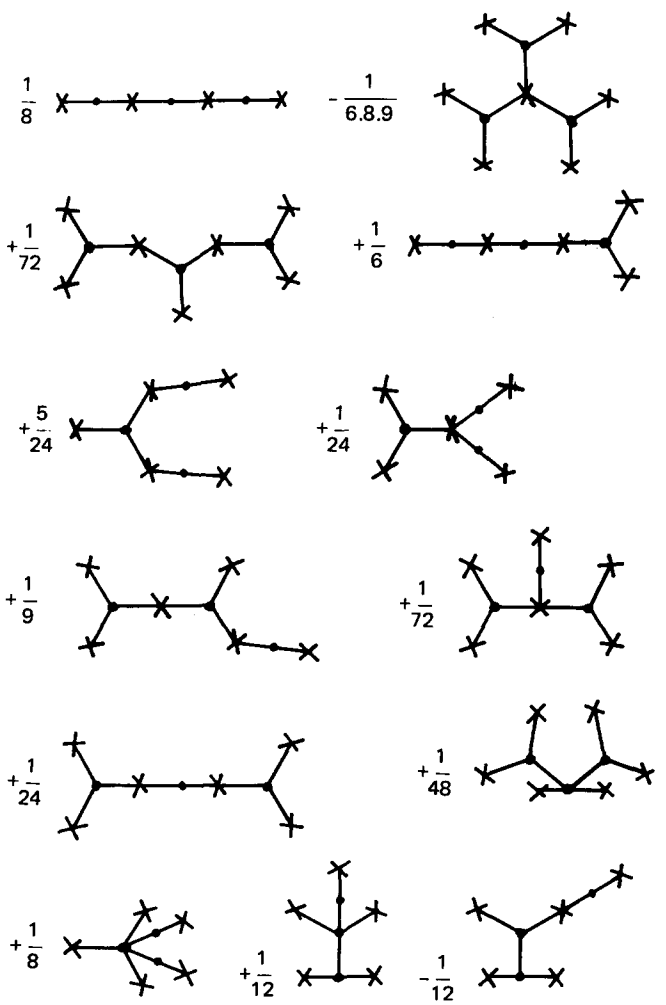
$$= \tfrac{1}{2} \int_0^t d\tau_1\, d\tau_2\, \eta(\tau_1)\eta(\tau_2) \int_{\sup(\tau_1,\tau_2)}^{t} d\tau\, \alpha(\tau)$$

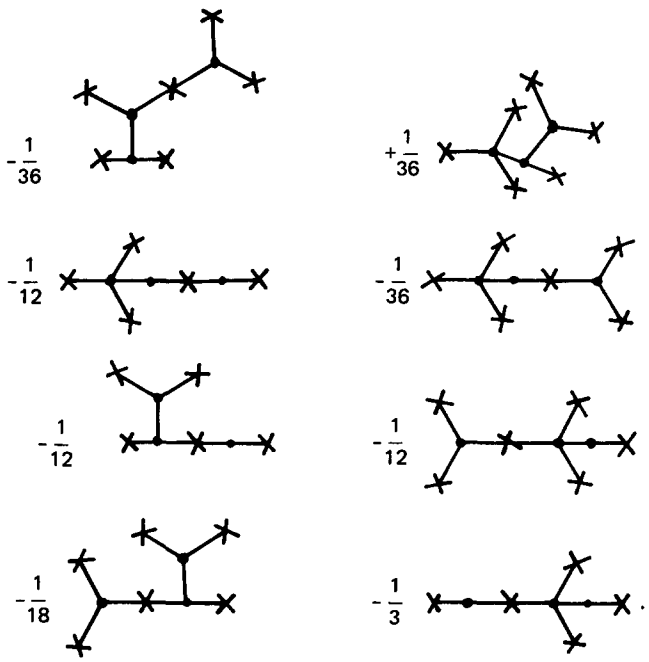
$$+ \tfrac{1}{6} \int_0^t d\tau_1\, d\tau_2\, d\tau_3\, \eta(\tau_1)\eta(\tau_2)\eta(\tau_3) \int_{\sup(\tau_1,\tau_2,\tau_3)}^{t} d\tau\, \beta(\tau),$$

Second order



Third order

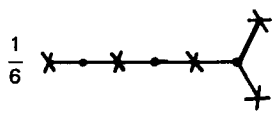




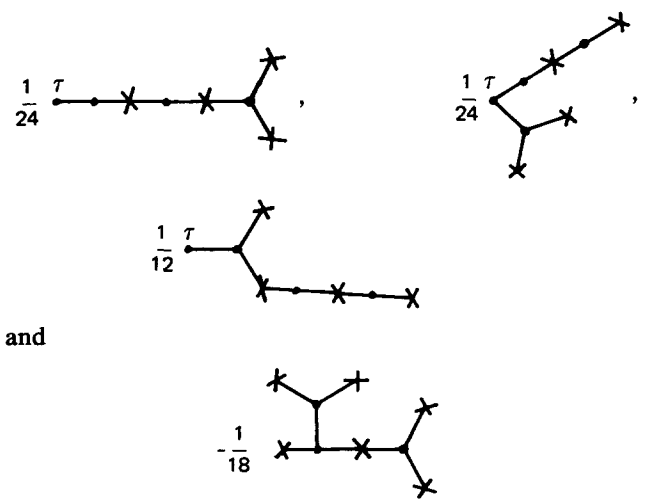
Among these terms, those containing only simple stars furnish explicit contributions to the n-point functions. In fact, for these terms, the integrations relative to the centers of the stars are factorized, and are directly expressed in terms of the primitives of α and β. This favorable circumstance occurs for any diagram when the polynomial P is reduced to a monomial, that is, when $\alpha = 0$ or $\beta = 0$. This case will be more completely discussed in the next section.

5. STAR STRUCTURE FOR THE SOLUTION $x(t)$

The rules given in the preceding section concern the diagrams describing the function $\theta(t-\tau)X(t,\tau)Y(t,\tau)$. It is to be noted that the same diagram for $x(t)$ can be obtained from different diagrams for θXY, the coefficients of which have to be added. For instance, the diagram

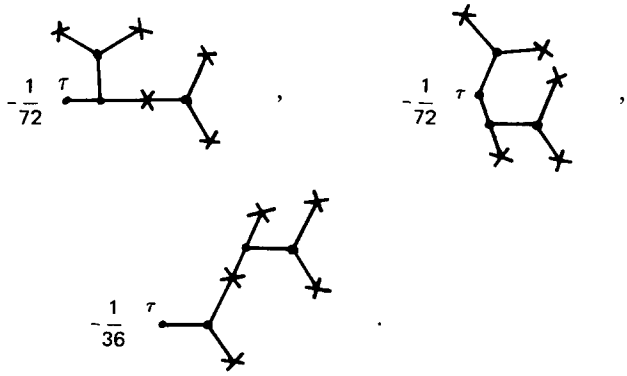


comes from the diagrams



and

from



Therefore, let D be a diagram for $x(t)$. In the diagrams of θXY contributing to D, according to (4.12), the point τ cannot belong to the dressing of a center, so that it suffices to consider the diagrams with dressed centers. Henceforth, D will be constituted by elements like (4.12). A diagram deduced from D by distinguishing a source η and labeling it by τ, thus corresponding to a diagram for θXY, will be called a *modified diagram*.

Let us write the coefficient of any diagram in the form $(1/S)F$, where S is the symmetry factor. For the modified diagrams, F is then defined by rule R_1, the centers being dressed.

Proposition: The factor F_D of D is equal to the sum of the factors of the modified diagrams, obtained by successively modifying *every* source η in D.

Proof: The proposition is obvious when D has no symmetry. Otherwise, one has

$$\frac{1}{S_D} F_D = \sum_\alpha \frac{1}{S_{D_\alpha}} F_{D_\alpha} , \tag{5.1}$$

where the sum runs over all the different modified diagrams D_α deduced from D. Let us call a *labeled diagram* any diagram the sources of which are labeled by the numbers 1,2,... . Let N_D and N_{D_α} be the numbers of *different* labeled diagrams, respectively, obtained from D and D_α. If n is the number of sources of D, one has

$$N_D = n!/S_D \quad \text{and} \quad N_{D_\alpha} = n!/S_{D_\alpha} \tag{5.2}$$

so that (5.1) becomes

$$N_D F_D = \sum_\alpha N_{D_\alpha} F_{D_\alpha} . \tag{5.3}$$

The right member is the sum of the factors of the labeled modified diagrams. Since these diagrams are obtained from the labeled diagrams associated with D by choosing in any way the place of the modification, and since the factors do not depend on the labeling, this sum is also equal to $N_D \Sigma_i F_i$, where i runs over *all* the sources of D and F_i denotes the factor of the diagram modified at the ith source. This gives

$$F_D = \sum_i F_i . \qquad \text{Q.E.D.} \tag{5.4}$$

Rules to calculate F_D are now deduced from (5.4) and R_1. One has to collect all the modified diagrams constructed from D. Let $i = 1,2,...,n$ be an index labeling the sources and k_i (respectively, l_i) the number of 2-stars (respectively, 3-

stars) tied at the source i. According to R_1, the factor F_i is the product $F_i = \Pi_j C_i^j$, the coefficients C_i^j assigned to the source j being given by

$$C_i^i = A_{k_i+1}\, a_{l_i+1}\,,$$

$$C_i^j = A_{k_j}\, a_{l_j+1} \quad \text{(respectively, } A_{k_j+1}\, a_{l_j}\text{)}$$

if the shortest path traced on D and joining j ends with a 2-star (respectively, 3-star). These expressions are condensed into

$$C_i^j = A_{k_j+p(i,j)}\, a_{l_j+q(i,j)}\,, \tag{5.5}$$

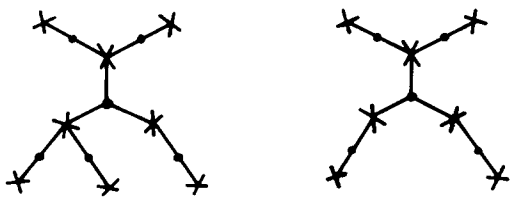
where $p(i,i) = q(i,i) = 1$ and $p(i,j) = 1$ (respectively, 0), $q(i,j) = 0$ (respectively, 1) if the path joining i to j ends with a 3-star (respectively 2-star). The recursion relations

$$\begin{aligned} A_{k+1} &= (1 - k/2)\, A_k\,, \\ a_{l+1} &= (1 - 2l/3)\, a_l \end{aligned} \tag{5.6}$$

implied by (4.11) then give

$$\begin{aligned} F_D &= \left(\prod_i A_{k_i}\, a_{l_i}\right) \mathbf{f}_D\,, \\ f_D &= \sum_i \prod_j \left[\left(1 - p(i,j)\frac{k_j}{2}\right)\left(1 - q(i,j)\frac{2l_j}{3}\right)\right]. \end{aligned} \tag{5.7}$$

The first factor in F_D depends only on the configuration of the stars around each source, while f_D depends on the whole topology of D. For instance, the diagrams



have, respectively, $f_D = 0$ and $f_D \neq 0$.

According to formula (5.7), the factor f_D seems to depend on all the details of the topology of D, for (i,j) runs over all the pairs of sources. In fact, as shown in the Appendix, f_D is determined by a rougher structure in which the only relevant sources are those binding both types of stars.

The star structure of $x(t)$ gets simpler when α or β vanishes, that is, when the polynomial P reduces to a monomial. If $\alpha = 0$, for instance, the 2-star vanishes according to (4.8). Therefore, any diagram for $x(t)$ only contains 3-stars, the centers of which are not dressed on account of R_2. Since only simple stars occur, the integrations involving β are factorized so that the n-point functions are algebraically expressed in terms of the primitive of β. Finally, the factor F_D is reduced to

$$F_D = \prod_i a_{l_i}\,. \tag{5.8}$$

It is thus simply obtained by assigning to each source a coefficient only depending on the number of stars that it binds. The structure of these diagrams looks like that of the usual Feynman diagrams in field theory, the coefficients a_l acting as interaction constants, the star replacing the propagator.

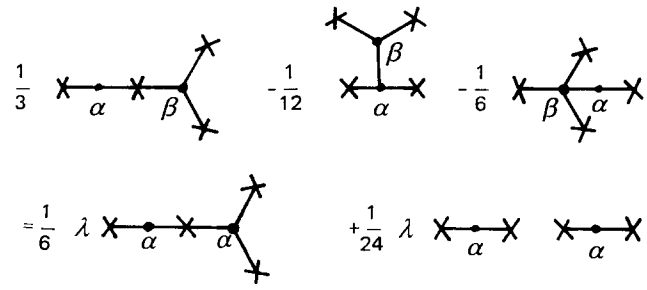
6. CONCLUSION

The Feynman expansion and the star expansion lead to different physical interpretations. In any term of the first one, some of the given sources $\eta(\tau_1),\eta(\tau_2),\ldots$ distribute amongst disjointed bundles; each of them has a number ν of elements equal to one of the exponents appearing in P and creates a new source $\alpha_\nu(\tau)\eta(\tau_{i_1})\cdots\eta(\tau_{i_\nu})$ at a subsequent time τ. This process repeats itself with the remaining η-sources and new ones, and so on, until the time t to give a contribution to $x(t)$. On the contrary, in terms of the star expansion, when $P = \alpha x^\nu/\nu!$, all the η-sources gather in a single stage into bundles having ν elements; these bundles are jointed so that each source can belong to different ones, and the resulting contribution to $x(t)$ is factorized. When P is an arbitrary polynomial, a similar description holds for each step of a more complicated process, in which each bundle eventually creates a new source (corresponding to a dressed center) acting in the next step.

In the general case where the polynomial P is not reduced to a monomial, the expression of the n-point functions furnished by the star expansion is not completely free from integrations (i.e., is not a function of the primitives of the α_ν's), because the integrations appearing in the dressed centers intertwine and therefore remain. However, these integrations can be carried out when the α_ν's are proportional to a single function. Let us prove by induction this property in the case previously considered when $\beta = \lambda\alpha$. This property holds for diagrams with no dressed center, for they contain simple stars only. If it is true for diagrams having less than n dressed centers, the same holds for $n + 1$: In fact, according to the recursion hypothesis, the expression for every center, the dressing of which contains less than n dressed centers, is a linear combination of terms of the type

$$A(\bar{\alpha}) \int_a^b d\tau\, \alpha(\tau)(\bar{\alpha}(\tau))^m = A(\bar{\alpha}) \left|\frac{(\bar{\alpha}(\tau))^{m+1}}{m+1}\right|_a^b,$$

where A is an algebraic function of $\bar{\alpha}$, the primitive of α, and where a and b depend on $t, \tau_1,\tau_2,\ldots$. For instance, the $\alpha\beta$ term in the solution $x(t)$ can be represented by



We remark that the last expression only involves simple stars, but contains disconnected diagrams. This property can be verified for the lowest orders, and we thus conjecture that it holds in general. Let us give a solvable example that supports this assertion: The retarded solution of the equation

$$\frac{dx}{dt} = \eta(t) + \alpha(t)\exp(\lambda x)$$

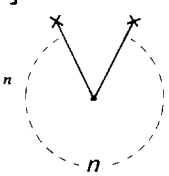
is easily obtained and reads

$$x(t) = \bar{\eta}(t) - \frac{1}{\lambda} \ln\left(1 - \lambda \int_0^t d\tau\, \alpha(\tau) e^{\lambda \bar{\eta}(\tau)}\right),$$

where $\bar{\eta}(t) = \int_0^t d\tau\, \eta(\tau)$; the expansion with respect to λ is

$$x(t) = \bar{\eta}(t) + \sum_{m=1}^{\infty} \frac{\lambda^{m-1}}{m} \times \left[\sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \int_0^t d\tau\, \alpha(\tau)(\bar{\eta}(\tau))^n\right]^m$$

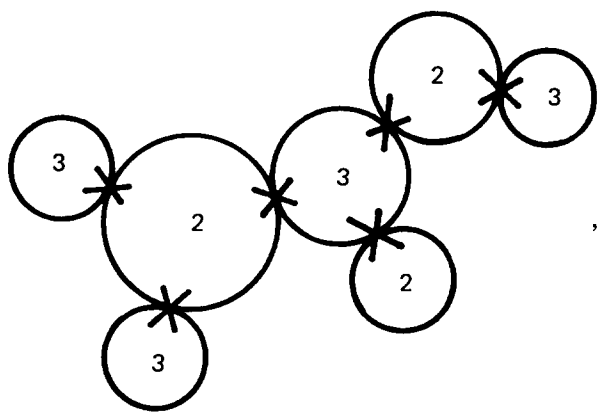
$$= \bar{\eta}(t) + \sum_{m=1}^{\infty} \frac{\lambda^{m-1}}{m} \left[\sum_{n=0}^{\infty} \lambda^n \right]^m .$$



The solution $x(t)$ is then represented by a sum of diagrams constituted by disconnected simple ν-stars, where ν take any integer value.

APPENDIX: STRUCTURE OF THE FACTOR f_D

Qne calls homogeneous tree (respectively, 2- or 3-tree) a tree diagram constructed with a unique type of star (respectively, 2- or 3-stars). Any diagram D for $x(t)$ may be considered as consisting of the juxtaposition of homogeneous trees joined by sources. Such a juxtaposition is described by a diagram of the type



where the bubble labeled by 2 or 3 represent 2- or 3-trees, the internal structure of which is left out.

Let us at first remark that, in formula (5.7) for f_D, if the source j different from i is internal to a 2-bubble (respectively, 3-bubble) one has $p(i,j) = l_j = 0$ [respectively, $q(i,j) = k_j = 0$], so that this source does not contribute to the product Π_j. The only contributions to Π_j thus come from the modified point ($j = i$) and the points common to a 2-bubble and a 3-bubble (frontier points). Let us label the 2-bubble by $\alpha,\beta,\cdots$ and the 3-bubble by $\alpha',\beta',\cdots$. When the bubble α touches the bubble α', one denotes by $(\alpha\alpha')$ the corresponding frontier point. Let us write

$$f_D = \sum_i f_i\,, \quad f_i = \prod_j \left(1 - p(i,j)\frac{k_j}{2}\right)\left(1 - q(i,j)\frac{2l_j}{3}\right). \quad \text{(A1)}$$

If i_1 and i_2 are internal to the bubble α (respectively, α'), the paths joining, respectively, i_1 or i_2 to any frontier point ($\beta\beta'$) end by stars of the same kind; due to the definitions of $p(i,j)$ and $q(i,j)$, the contributions of all the frontier points to f_{i_1} and f_{i_2} are equal, and thus only depend on α (respectively, α'). Let g_α (respectively, $g_{\alpha'}$) be this contribution. By taking into account the contribution of the point i itself, one thus obtains $f_i = (1 - k_i/2)\, g_\alpha$ [respectively, $f_i = (1 - 2l_i/3)\, g_{\alpha'}$]. If $i = (\alpha\alpha')$ is a frontier point, the contribution to $f_{(\alpha\alpha')}$ of all the frontier points different from $(\alpha\alpha')$ is equal to that occurring in g_α or $g_{\alpha'}$, the contribution of $i = (\alpha\alpha')$ here being

$$\left(1 - \frac{k_{(\alpha\alpha')}}{2}\right)\left(1 - \frac{2l_{(\alpha\alpha')}}{3}\right)$$

this gives the two relations

$$f_{(\alpha\alpha')} = \left(1 - \frac{k_{(\alpha\alpha')}}{2}\right) g_\alpha = \left(1 - \frac{2l_{(\alpha\alpha')}}{3}\right) g_{\alpha'}\,.$$

By adding all these contributions, one obtains

$$f_D = \sum_\alpha \left[\sum_{i\in\alpha}\left(1 - \frac{k_i}{2}\right)\right] g_\alpha + \sum_{\alpha'} \left[\sum_{i\in\alpha'}\left(1 - \frac{2l_i}{3}\right)\right] g_{\alpha'} - \sum_{(\alpha\alpha')} f_{(\alpha\alpha')}\,, \quad \text{(A2)}$$

where the last term exists because the frontier points occur twice in the preceding terms. Since for the homogeneous trees one has

$$\sum_{i\in\alpha}\left(1 - \frac{k_i}{2}\right) = 1 \quad \text{and} \quad \sum_{i\in\alpha'}\left(1 - \frac{2l_i}{3}\right) = 1\,,$$

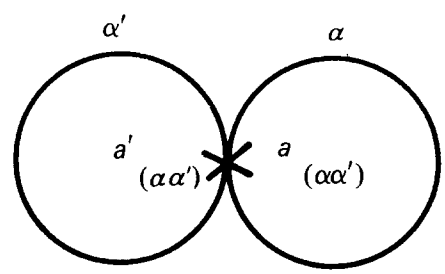
it remains the formula

$$f_D = \sum_\alpha g_\alpha + \sum_{\alpha'} g_{\alpha'} - \sum_{(\alpha\alpha')} f_{(\alpha\alpha')}\,. \quad \text{(A3)}$$

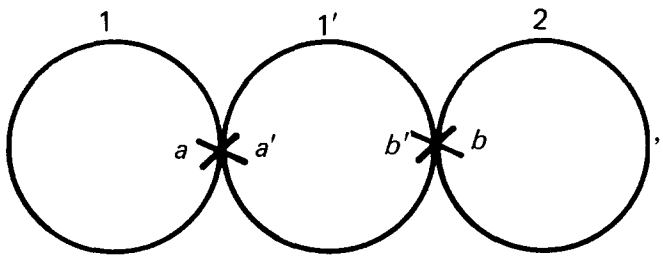
Contrary to (A1), this expression only depends on the bubble structure of the diagram and the configuration of stars at the frontier points. This structure essentially occurs in the definition of the g_α's and $g_{\alpha'}$'s. These terms may be calculated as follows: For any frontier point $(\alpha\alpha')$ let us put

$$a_{(\alpha\alpha')} = 1 - \frac{k_{(\alpha\alpha')}}{2} \quad \text{and} \quad a'_{(\alpha\alpha')} = 1 - \frac{2l_{(\alpha\alpha')}}{3}\,; \quad \text{(A4)}$$

one decides to affect each of these coefficients, respectively, to the bubbles α or α', on each side of the frontier point according to the schema



Then g_α is obtained by drawing all the paths starting from an arbitrary point internal to the bubble α and ending at any frontier point, each of these paths contributing to g_α by a factor a (respectively, a'), according as the frontier point is reached on the side of a' (respectively, a). For instance, to the diagram



in which $a, a', \cdots$ represent $a_{(1')}, a'_{(1l')}, \cdots$, correspond the values $g^1 = a'b$, $g_1' = ab$, $g_2 = ab'$ and consequently $f_{11}' = aa'b$, $f_{21}' = abb'$, and

$$\begin{aligned} f_D &= a'b + ab + ab' - aa'b - abb' \\ &= 1 - (1-a)(1-a') - (1-b)(1-b') \\ &\quad + (1-a)(1-a'-b')(1-b). \end{aligned} \tag{A5}$$

The last expression for f_D illustrates a general rule. Let us call a cut diagram a diagram D_c deduced from D by suppressing some frontier points together with the associated coefficients (A4); let us put $\epsilon(D_c) = (-1)^{n(D_c)}$, where $n(D_c)$ is the number of remaining frontier points in D_c; finally, let $\Sigma_\gamma(D_c)$ be the sum of the coefficients (A4) contained in the bubble $\gamma = \alpha$ or α' of D_c.

Proposition: The factor f_D is given by

$$f_D = \sum_{D_c} \epsilon(D_c) \prod_\gamma (1 - \Sigma_\gamma D_c)). \tag{A6}$$

Proof: Let us proceed by induction. One assumes that the proposition holds for the diagrams containing a number of bubbles lower than N. Let D be a diagram having N bubbles and D' obtained by adding one bubble to D. Let δ be this bubble, γ_0 the bubble of D in contact with δ, and a (respectively, b) the coefficient relative to the frontier point $(\gamma_0\delta)$ contained in γ_0 (respectively, δ). Due to (A3) and the recursion hypothesis one has, for D,

$$f_D = \sum_\gamma g_\gamma - \sum_{(\alpha\alpha')} f_{(\alpha\alpha')} = \sum_{D_c} \prod (D_c)(1 - \Sigma_{\gamma_0}(D_c)), \tag{A7}$$

where

$$\prod (D_c) = \epsilon(D_c) \prod_{\gamma \neq \gamma_0} (1 - \Sigma_\gamma(D_c)). \tag{A8}$$

Now, according to (A3) and the calculation rules of the g's and the f's, one has

$$\begin{aligned} f_{D'} &= \sum_\gamma bg_\gamma + ag_{\gamma_0} - \sum_{(\alpha\alpha')} bf_{(\alpha\alpha')} - abg_{\gamma_0} \\ &= b \sum_{D_c} \prod (D_c)(1 - \Sigma_{\gamma_0}(D_c)) + a(1-b) g_{\gamma_0}, \end{aligned} \tag{A9}$$

then

$$\begin{aligned} f_{D'} &= \sum_{D_c} \prod (D_c)(1 - \Sigma_{\gamma_0}(D_c)) \\ &\quad + (1-b)\left[ag_{\gamma_0} - \sum_{D_c} \prod (D_c)(1 - \Sigma_{\gamma_0}(D_c))\right]. \end{aligned} \tag{A10}$$

Let us prove the equality

$$g_{\gamma_0} = \sum_{D_c} \prod (D_c). \tag{A11}$$

To this end, let D_0 be the diagram identical to D, except for the coefficients internal to γ_0 that are taken equal to zero; for D_0, (A7) reads

$$f_{D_0} = g_{\gamma_0} = \sum_{D_c} \prod (D_c). \qquad \text{Q.E.D.} \tag{A12}$$

Thus (A10) becomes

$$\begin{aligned} f_{D'} &= \sum_{D_c} \prod (D_c)(1 - \Sigma_{\gamma_0}(D_c)) \\ &\quad - \sum_{D_c} \prod (D_c)(1-b)(1 - \Sigma_{\gamma_0}(D_c) - a). \end{aligned} \tag{A13}$$

This proves the proposition for D', the first (respectively, second) sum being the contribution to $f_{D'}$ of all the cut diagrams of D' for which the frontier point $(\gamma_0\delta)$ is suppressed (respectively, conserved).

[1]Les vibrations forcées dans les systèmes non linéaires. Colloque interne du CNRS $N°$ 148 (Marseille, 1964) (Editions du Centre National de la Recherche Scientifique, Paris, 1965).
[2]C. Itzykson and J. B. Zuber, *Quantum Field Theory* (McGraw-Hill, New York, 1980).
[3]J. C. Houard, Lett. Nuovo Cimento **33**, 519 (1982).

Symmetries of differential equations. IV

F. González-Gascón[a)]
Departamento de Física Teórica, Universidad de Salamanca, Instituto de Estructura de la Materia, CSIC, GIFT, Salamanca, Spain

A. González-López
Departamento de Métodos Matemáticos, Universidad de Complutense, Facultad de Ciencias Físicas, Complutense, Spain

(Received 27 January 1982; accepted for publication 11 March 1983)

By an application of the geometrical techniques of Lie, Cohen, and Dickson it is shown that a system of differential equations of the form $x_i^{(r_i)} = F_i$ (where $r_i > 1$ for every $i = 1,...,n$) cannot admit an infinite number of pointlike symmetry vectors. When $r_i = r$ for every $i = 1,...,n$, upper bounds have been computed for the maximum number of independent symmetry vectors that these systems can possess: The upper bounds are given by $2n^2 + nr + 2$ (when $r > 2$), and by $2n^2 + 4n + 2$ (when $r = 2$). The group of symmetries of $\bar{x}^{(r)} = \bar{0}$ ($r > 1$) has also been computed, and the result obtained shows that when $n > 1$ and $r > 2$ the number of independent symmetries of these equations does not attain the upper bound $2n^2 + nr + 2$, which is a common bound for all systems of differential equations of the form $\bar{x}^{(r)} = \bar{F}(t,\bar{x},...,\bar{x}^{(r-1)})$ when $r > 2$. On the other hand, when $r = 2$ the first upper bound obtained has been reduced to the value $n^2 + 4n + 3$; this number is equal to the number of independent symmetry vectors of the system $\ddot{\bar{x}} = \bar{0}$, and is also a common bound for all systems of the form $\ddot{\bar{x}} = \bar{F}(t,\bar{x},\dot{\bar{x}})$.

PACS numbers: 02.30.Hq, 02.30.Jr, 02.20. + b

I. INTRODUCTION

This paper should be considered as a continuation of a series of papers by the authors,[1] in this and other journals, on the fascinating subject of the symmetries of systems of differential equations. In these papers both the direct and the inverse problem concerning the symmetries have been studied, as well as certain connections between the symmetry vectors and the first integrals of systems of differential equations. Although some global results have been obtained, most of the results obtained are of a local character.

In the present paper we obtain, following the geometrical and local techniques contained in the classical treatises of Lie and Scheffers, Cohen, and Dickson,[2] upper bounds for the number of independent pointlike symmetry vectors of differential equations of the form

$$\bar{x}^{(r)} = \bar{F}(t,\bar{x},...,\bar{x}^{(r-1)}), \tag{i}$$

where $r > 1$ and $\bar{x}$ stands for $(x_1,...,x_n)$. The case $r = 1$ has not been studied, since it is well known—see, for instance, the first and fourth papers quoted in Ref. 1—that when $r = 1$ the number of independent symmetries is always infinite.

We obtain in Sec. III the upper bound $2n^2 + nr + 2$ ($r > 2$), as well as the number of independent symmetry vectors of the system $\bar{x}^{(r)} = \bar{0}$, which is given by $n^2 + nr + 3$, and the explicit expression of them. Since $2n^2 + nr + 2$ is greater than $n^2 + nr + 3$ when $n > 1$, the problem arises of knowing whether or not the upper bound $2n^2 + nr + 2$ is attained by a system of differential equations of this type, when $n > 1$.

Similarly, for a system of the form $\ddot{\bar{x}} = \bar{F}(t,\bar{x},\dot{\bar{x}})$, we obtain in Sec. IV the upper bound $2n^2 + 4n + 2$, which is reduced in Sec. V to $n^2 + 4n + 3$ by using a remarkable property of the projective group. This last upper bound is attained by the system $\ddot{\bar{x}} = \bar{0}$, whose symmetry group is the projective group of pointlike transformations of the space $\{(t,\bar{x})\}$.

When $n = 1$, i.e., when only a single differential equation is considered, the upper bounds obtained reduce to $r + 4$ (when $r > 2$) and 8 (when $r = 2$). These two results are classical and well known, and the proof we give of them in Sec. II tries only to be a bit more careful than the classical proofs, at the same time preparing the reader for a clearer understanding of the more complicated case of a normal system of differential equations of the form

$$x_i^{(r_i)} = F_i, \quad r_i > 1, \quad \forall i = 1,...,n. \tag{ii}$$

As is shown in Sec. VI, a system of this type possesses only a *finite* number $N(n;r_1,...,r_n)$ of independent symmetry vectors, and this number grows without limit when either n or some of the r_i's tend to infinity. The conclusion is that a system of differential equations of the type $x_i^{(r_i)} = F_i$, with $r_i > 1$ for every i, does not admit a Lie group (in the generalized sense of a group of transformations with an *infinite* number of essential parameters) as its symmetry group.

The reader should consult the classical treatises cited in Refs. 2 and 5 for most of the definitions and the notation used here, as well as the first three papers of this series cited in Ref. 1.

II. MAXIMUM NUMBER OF INDEPENDENT SYMMETRY VECTORS OF A DIFFERENTIAL EQUATION OF ORDER $r > 1$

In order that the reader can follow us without difficulty in the more complicated case of a normal system of differential equations, it is convenient to treat first the relatively simple case of a single differential equation of the form

$$x^{(r)} = F(t,x,\dot{x},...,x^{(r-1)}). \tag{1}$$

We remind the reader that when $r = 1$ Eq. (1) always possesses an infinite number of independent symmetry vec-

[a)] Postal address for reprints: F. G. Gascon, Serrano 119, Madrid 6, Spain.

 0022-2488/83/082006-16$02.50

tors.[1] On the contrary, when $r>1$ Eq. (1) does *not* admit, in general, pointlike families of symmetries of the form

$$t' = t + \epsilon\cdot a(t,x),$$
$$x' = x + \epsilon\cdot b(t,x). \tag{2}$$

In particular, when $r=2$, one can even classify[1] all the differential equations of the form

$$\ddot{x} = F(t,x), \tag{3}$$

admitting at least one symmetry vector of type (2) different from zero.

Moreover, concerning the pointlike symmetry vectors, it is a classical result that when $r>1$, Eq. (1) admits no more than eight symmetry vectors (if $r=2$) and no more than $(r+4)$ if $r>2$. The proof of this result, or at least the fundamental ideas behind it, can be found in the classical treatises of Lie, Cohen and Dickson.[2] For the sake of completeness, we present here a proof of this classical result, which tries to be a bit more careful than the one presented by the above-mentioned authors, and at the same time prepares the reader for the more complicated case of a normal system of differential equations of the following type:

$$x_i^{(r_i)} = F_i, \quad r_i > 1, \quad \forall i = 1,\ldots,n, \tag{4}$$

where the smooth functions F_i appearing in (4) depend, of course, on the variables $t;x_1,\ldots,x_1^{(r_1-1)};x_n,\ldots,x_n^{(r_n-1)}$.

We begin by studying the case $r>2$:

(a) Consider the unique solution $\phi(t;\lambda)$ of (1) corresponding to the initial conditions $(t_0,x_0,\ldots,x_0^{(r-2)};\lambda)$, and let $P_1=(t_1,x_1=\phi(t_1))$, with t_1 sufficiently close to t_0, and

$$\phi(t) = \phi(t;x_0^{(r-1)}) \tag{5}$$

for an arbitrary, but fixed, $x_0^{(r-1)}$. We shall now show that for certain neighborhoods U_1 of P_1 and I_1 of $x_0^{(r-1)}$ there exists a unique smooth (i.e., C^∞) function $\theta_1: U_1 \to I_1$ satisfying

(i) $\theta_1(P_1) = x_0^{(r-1)}$;

(ii) If $P=(t,x)\in U_1$ and $x^{(r-1)}\in I_1$, then

$$\phi(t;x^{(r-1)}) = x \quad \text{iff } x^{(r-1)} = \theta_1(P).$$

That is, through every point of U_1 there passes a *unique* integral curve of (1) whose $(r-1)$th derivative lies on I_1 having a contact of order $(r-2)$ at $P_0=(t_0,x_0)$ with the integral curve γ_0 of (1) corresponding to the initial conditions $(t_0,x_0,\ldots,x_0^{(r-1)})$.

The proof follows from the fact that, regarded as functions of t and of the initial conditions $t_0,x_0,\ldots,x_0^{(r-1)}$, the solutions of (1) are C^∞ functions, provided only that the function F appearing in (1) is, as we shall assume throughout this paper, a C^∞ function of its variables. Therefore, $\phi(t;\lambda)$ will be also smooth in t and λ, and since the triplet $(t_1,x_1,x_0^{(r-1)})$ satisfies the equation

$$x = \phi(t;\lambda), \tag{6}$$

in order to complete our proof, it suffices to show that for t_1 sufficiently close to t_0 the "transversality condition"

$$\left.\frac{\partial\phi}{\partial\lambda}\right|_{(t_1,x_0^{(r-1)})} \neq 0 \tag{7}$$

holds; indeed, if this were the case, the implicit function theorem[3] applied to (6) in a neighborhood of the point $(t_1,x_1,x_0^{(r-1)})$ would yield λ as a smooth function θ_1 of the variables t and x.

Now, one can obviously write

$$\phi(t;\lambda) = x_0 + \dot{x}_0(t-t_0) + \cdots + x_0^{(r-2)}(t-t_0)^{r-2}/(r-2)! + \lambda(t-t_0)^{r-1}/(r-1)! + (t-t_0)^r R(t,\lambda), \tag{8}$$

$R(t,\lambda)$ being a C^∞ function of t and λ near $(t_0,x_0^{(r-1)})$.[4]

Therefore, one can also write

$$\frac{\partial\phi}{\partial\lambda} = \frac{(t-t_0)^{r-1}}{(r-1)!} + (t-t_0)^r\frac{\partial R}{\partial\lambda} \tag{9}$$

and, accordingly,

$$\left.\frac{\partial\phi}{\partial\lambda}\right|_{(t_1,x_0^{(r-1)})} = (t_1-t_0)^{r-1}\left[\frac{1}{(r-1)!} + (t_1-t_0)R_1\right] \tag{10}$$

where $R_1 = \left.\dfrac{\partial R}{\partial\lambda}\right|_{(t_1,x_0^{(r-1)})}$.

This last expression guarantees that (7) holds provided only that one chooses $t_1\neq t_0$ satisfying

$$|(t_1-t_0)\cdot R_1| < 1/(r-1)!, \tag{11}$$

which is possible since R is continuous (C^∞ in fact).

Summarizing, the implicit function theorem applied to (6) yields the unique smooth function θ_1 satisfying conditions (i) and (ii) above.

(b) Let now $\phi_1(t;\lambda)$ be the maximal solution of (1) corresponding to the initial conditions $(t_1,x_1,\ldots,x_1^{(r-2)};\lambda)$, where

$$x_1^{(k} = \phi^{(k}(t_1) \tag{12}$$

and $\phi(t)$ is the function defined by (5). Choosing now a third point P_2 on $\gamma_0\cap U_1$ sufficiently close to P_1, and repeating the construction sketched in (a) with P_0 and P_1 replaced respectively by P_1 and P_2, we obtain a second function $\theta_2: U_2\to I_2$ satisfying:

(a) $\theta_2(P_2) = x_1^{(r-1}\in I_2$;

(b) If $P=(t,x)\in U_2$ and $x^{(r-1}\in I_2$, then

$$\phi_1(t;x^{(r-1)}) = x \quad \text{iff} \quad x^{(r-1} = \theta_2(P).$$

Since $U = U_1\cap U_2\neq\emptyset$ and $U\subset U_1$, the mapping $\theta: U\to I_1\times I_2$ defined by

$$P\to\theta(P) = (\theta_1(P),\theta_2(P)) \tag{13}$$

is such that, given any two integral curves of (1), $\gamma_1=(t,f_1(t))$ and $\gamma_2=(t,f_2(t))$, having a contact of order $(r-2)$ with γ_0, respectively, at P_0 and P_1 and satisfying

$$f_1^{(r-1}(t_0)\in I_1, \quad f_2^{(r-1}(t_1)\in I_2, \tag{14}$$

then γ_1 and γ_2 will pass through a point $P\in U$ if and only if

$$(f_1^{(r-1}(t_0), f_2^{(r-1}(t_1)) = \theta(P). \tag{15}$$

(c) Let now U_0 be an open subset of $(U_1\cap U_2)-\gamma_0$: If $P\in U_0$, then P will be an isolated point of $\gamma_1\cap\gamma_2$ [where γ_1 and γ_2 are, of course, the curves defined in (b) passing through P].

In fact, if this were not the case one could immediately write

$$f_1^{(k}(t_p) = f_2^{(k}(t_p), \quad k=0,1,2,\cdots, \tag{16}$$

where $P=(t_P,x_P)$. Clearly, we can restrict ourselves to the case $t_0<t_P<t_1$ $\forall P\in U_0$. We then define $f(t)$ as follows:

$$f(t)=\begin{cases} f_1(t) & \text{when } t\leqslant t_P, \\ f_2(t) & \text{when } t\geqslant t_P. \end{cases} \tag{17}$$

Then condition (16) guarantees that $f(t)$ is a C^∞ function in some interval $J\supset(t_0,t_1)$ such that

$$f^{(r-1)}(t_0)=f_1^{(r-1)}(t_0)\in I_1. \tag{18}$$

Therefore, the curve $\gamma=\{(t,f(t))\}$ is an integral curve of (1) having a contact of order $(r-2)$ with γ_0 at P_0 and passing through $P_1\in U_1$: Hence (18) and the properties of the function θ_1 [see (i) and (ii) above] imply that

$$f^{(r-1)}(t_0)=\theta_1(P_1)=x_0^{(r-1)}=\phi^{(r-1)}(t_0). \tag{19}$$

It follows, by uniticity, that $f(t)=\phi(t)$ and, in particular, $P\in\gamma_0$, contrary to the definition of U_0.

Therefore, one can safely assume that, for every $P\in U_0$, γ_1 and γ_2 meet transversally at P, that is,

$$f_1^{(s)}(t_P)\neq f_2^{(s)}(t_P), \quad \text{for some } s,\ 0<s<r. \tag{20}$$

The results obtained in Sec. II(a)–(c) imply the existence of an open neighborhood U_0 near P_0 having the following property: Through every point P of U_0 it is possible to draw two integral curves of (1), γ_1 and γ_2, such that P is isolated in $\gamma_1\cap\gamma_2$ and in addition γ_1 and γ_2 have a contact of order $(r-2)$ with γ_0, respectively, at P_0 and P_1.

(d) Assume now the **S** is a pointlike symmetry vector of (1) such that any integral curve of (1) having a contact of order $(r-2)$ with γ_0 either at P_0 or P_1 is invariant under the local one-parameter group of transformations generated by **S**. That is, the graph $\{(t,f(t))\}$ corresponding to any solution $f(t)$ having this property will be left invariant by any member g of the local one-parameter group G generated by **S**.

Under these circumstances, γ_1 and γ_2 will be invariant under G and, accordingly, the same thing will happen with $\gamma_1\cap\gamma_2$. Now, since P is isolated in $\gamma_1\cap\gamma_2$, P will be left invariant under the action of any $g\in G$ sufficiently close to the identity transformation, by continuity. This proves that **S** vanishes at P: since P was an arbitrary point of U_0, we conclude that **S** vanishes on U_0.

(e) Let us now compute the number of conditions sufficient in order that any integral curve of (1) having a contact of order $(r-2)$ with γ_0 at P_0 or P_1 be, as a subset of R^2, invariant under the local one-parameter group G generated by **S** (in short, under **S**).

If **S** is given by

$$\mathbf{S}=\varphi(t,x)\frac{\partial}{\partial t}+\psi(t,x)\frac{\partial}{\partial x}, \tag{21}$$

then $\mathbf{S}^{(r-1)}$, the extension of **S** to the variables $t,x,\dot{x},\ldots,x^{(r-1)}$, will be given by

$$\mathbf{S}^{(r-1)}=\mathbf{S}+\sum_{i=1}^{r-1}\psi^i\frac{\partial}{\partial x^{(i)}}, \tag{22}$$

where

$$\psi^j=\frac{d\psi^{j-1}}{dt}-x^{(j}\frac{d\varphi}{dt}, \qquad \psi^0=\psi \quad \text{by definition} \tag{23}$$

and, of course,

$$\frac{d}{dt}=\frac{\partial}{\partial t}+\dot{x}\frac{\partial}{\partial x}+\cdots+x^{(k}\frac{\partial}{\partial x^{(k-1}}+\cdots. \tag{24}$$

First of all, we notice that a sufficient condition in order that an integral curve of (1), $\gamma=\{(t,x(t))\}$, be invariant under **S** is that $\mathbf{S}^{(r-1)}$ vanish on its initial conditions $(t_0,x(t_0),\ldots,x^{(r-1}(t_0))$, since **S** is by hypothesis a symmetry vector of (1). Therefore, in order that **S** leave invariant *any* integral curve of (1) having a contact of order $(r-2)$ with γ_0 at P_0 or P_1 it will be sufficient that

$$\mathbf{S}^{(r-1)}\big|_{(t_0,x_0,\ldots,x_0^{(r-2},x^{(r-1)})}=0 \tag{25a}$$

and

$$\mathbf{S}^{(r-1)}\big|_{(t_1,x_1,\ldots,x_1^{(r-2},x^{(r-1)})}=0 \tag{25b}$$

hold for *every* value of $x^{(r-1}$.

Conditions (25a) are clearly equivalent to the following set of $(r+1)$ equalities:

$$\begin{aligned} &\varphi(t_0,x_0)=0,\\ &\psi(t_0,x_0)=0,\\ &\psi^1(t_0,x_0,\dot{x}_0)=0,\\ &\cdots\\ &\psi^{r-2}(t_0,x_0,\ldots,x_0^{(r-2})=0,\\ &\psi^{r-1}(t_0,x_0,\ldots,x_0^{(r-2},x^{(r-2})=0, \quad \forall x^{(r-1}\in\mathbb{R}, \end{aligned} \tag{26}$$

where the functions ψ^i were defined by (23). Since, for $i>1$, the functions ψ^i are easily seen to have the following affine structure,

$$\psi^i=A_i(t,x,\ldots,x^{(i-1})x^{(i}+B_i(t,x,\ldots,x^{(i-1}), \tag{27}$$

conditions (26) are equivalent to the following set of $(r+2)$ equations:

$$\begin{aligned} &\varphi(t_0,x_0)=\psi(t_0,x_0)=\psi^i(t_0,x_0,\ldots,x_0^{(i})=0,\\ &i=1,\ldots,r-2,\\ &A_{r-1}(t_0,x_0,\ldots,x_0^{(r-2})=B_{r-1}(t_0,x_0,\ldots,x_0^{(r-2})=0 \end{aligned} \tag{28}$$

[notice that $r>2$ by hypothesis, and therefore $r-1>1$ implies that ψ^{r-1} has indeed the affine structure (27) with $i=r-1$].

Conditions (25a), and hence (28), imply (as has been remarked above) that any integral curve of (1) having a contact of order $(r-2)$ with γ_0 at P_0 is invariant under **S**. In particular, if (28) holds, then γ_0 itself is invariant under **S**, and therefore **S** has to be parallel to the tangent vector to γ_0 on every point of γ_0, that is,

$$\begin{aligned} &\mathbf{S}_P=a(t)\left[\frac{\partial}{\partial t}+\dot{\phi}(t)\frac{\partial}{\partial x}\right]\\ &\forall P=(t,\phi(t))\in\gamma_0 \end{aligned} \tag{29}$$

for some C^∞ function $a(t)$; by setting equal the coefficients of $\partial/\partial t$ in both members of (29), we conclude that $a(t)=\varphi(P)$ and therefore

$$\mathbf{S}_P=\varphi(P)\left[\frac{\partial}{\partial t}+\dot{\phi}(t)\frac{\partial}{\partial x}\right] \quad \forall P=(t,\phi(t))\in\gamma_0. \tag{30}$$

Therefore, in order that P_1 be invariant under **S** a single condition suffices, namely,

$$\varphi(t_1,x_1)=0. \tag{31}$$

When this last condition holds, the invariance of γ_0 under $\mathbf{S}$ implies that any linear element of order k at P_1, $(t_1,x_1,\ldots,x_1^{(k)})$, is invariant under $\mathbf{S}^{(k)}$, for every value of k. Consequently, $\mathbf{S}^{(k)}$ vanishes at the point $(t_1,x_1,\ldots,x_1^{(k)})$ for every value of k, in particular for $k=1,\ldots,r-1$; hence we have

$$\psi^i(t_1,x_1,\ldots,x_1^{(i)})=0, \quad i=1,\ldots,r-2,$$
$$A_{r-1}(t_1,x_1,\ldots,x_1^{(r-2)})\cdot x_1^{(r-1)}+B_{r-1}(t_1,\ldots,x_1^{(r-2)})=0 \tag{32}$$

as a consequence of (28) and (31). Therefore, in order that (25b) be also satisfied, only one additional condition is sufficient (and not two, as it would seem), namely,

$$A_{r-1}(t_1,x_1,\ldots,x_1^{(r-2)})=0. \tag{33}$$

Indeed, using the last Eq. (32), we get

$$A_{r-1}x^{(r-1)}+B_{r-1}\overset{(32)}{=}A_{r-2}(x^{(r-1)}-x_1^{(r-1)})=0 \tag{34}$$

for every value of $x^{(r-1)}$ if and only if (33) holds [of course, A_{r-1} and B_{r-1} have to be evaluated at $(t_1,x_1,\ldots,x_1^{(r-2)})$ in (34)].

Therefore, the $(r+4)$ conditions (28), (31), and (33) are sufficient in order that any integral curve of (1) having a contact of order $(r-2)$ with γ_0 at P_0 or P_1 be invariant under the symmetry vector of (1), $\mathbf{S}$ given by (21).

(f) Let us show finally that, when $r>2$, Eq. (1) does not admit more than $r+4$ linearly independent symmetry vectors.

Indeed, suppose that $\mathbf{S}_1,\ldots,\mathbf{S}_{r+5}$ are $r+5$ symmetry vectors of (1). Since the conditions in order that a vector field be a symmetry vector of (1) constitute a system of *linear* partial differential equations, any linear combination

$$\mathbf{X}=\sum_{i=1}^{r+5}c_i\mathbf{S}_i \tag{35}$$

of $\mathbf{S}_1,\ldots,\mathbf{S}_{r+5}$ will also be a symmetry vector of (1).

On the other hand, conditions (28), (31), and (33) are easily seen to be linear in the components of the vector field $\mathbf{S}$, by the linearity of the functions ψ^i in these components. Therefore, imposing that $\mathbf{X}$ satisfy conditions (28), (31), and (33), we obtain a linear and homogeneous system of $r+4$ algebraic equations in the unknowns $c_1,\ldots,c_{r+5}$, whose coefficients are real numbers depending on the vector fields $\mathbf{S}_1,\ldots,\mathbf{S}_{r+5}$ and on the fixed values of $(t_0,x_0,\ldots,x_0^{(r-2)})$ and $(t_1,x_1,\ldots,x_1^{(r-2)})$. Since the number of equations in this system exceeds the number of unknowns, it has a nontrivial solution $c_1^0,\ldots,c_{r+5}^0$, and, consequently, the vector field

$$\mathbf{X}_0=\sum_{i=1}^{r-5}c_i^0\mathbf{S}_i \tag{36}$$

will satisfy conditions (28), (31), and (33). Hence $\mathbf{X}_0$ must vanish on U_0, the open set defined in II(c), and, consequently,

$$\sum_{j=1}^{r+5}c_j^0\mathbf{S}_j=\mathbf{0} \tag{37}$$

on U_0, implying that $\mathbf{S}_1,\ldots,\mathbf{S}_{r-5}$ are linearly dependent on U_0, contrary to our initial assumption. This completes the proof that for $r>2$ there are at most $(r+4)$ independent symmetry vectors of (1).

(g) The case $r=2$ must be considered separately, since for $i=1$ the affine structure of ψ^i, given by

$$\psi^i=A_i(t,x,\ldots,x^{(i-1)})x^{(i)}+B_i(t,x,\ldots,x^{(i-1)}) \tag{38}$$

is no longer valid, and therefore the previous reasonings fail. Indeed, we are going to see that the maximum number of independent symmetry vectors of (1) is equal to *eight* when $r=2$.

In order to prove this statement, we start from the expression of $\mathbf{S}^1$, the first extension of $\mathbf{S}$:

$$\mathbf{S}^1=\mathbf{S}+[\psi_{,t}+(\psi_{,x}-\varphi_{,t})\dot{x}-\varphi_{,x}\dot{x}^2]\frac{\partial}{\partial\dot{x}}. \tag{39}$$

The line element $(t_0,x_0,\dot{x})$ will be invariant under $\mathbf{S}^1\forall\dot{x}$ provided that the following *five* conditions are satisfied:

$$\varphi|_{P_0}=\psi|_{P_0}=0,$$
$$\psi_{,t}|_{P_0}=(\psi_{,x}-\varphi_{,t})|_{P_0}=\varphi_{,x}|_{P_0}=0, \tag{40}$$

where $P_0=(t_0,x_0)$ as before. Denoting again by $\phi(t)$ the solution of the differential equation

$$\ddot{x}=F(t,x,\dot{x}), \tag{41}$$

corresponding to the initial conditions $(t_0,x_0,\dot{x}_0)$, only *one* condition is now sufficient in order that a second point $P_1=(t_1,x_1)$ chosen on the integral curve of (41) associated to the solution $\phi(t)$ be invariant under the symmetry vector $\mathbf{S}$ of (41), namely,

$$\varphi(P_1)=0 \tag{42}$$

exactly as in Sec. II(e).

When (40) and (42) are satisfied, both P_1 and the integral curve γ_0 of (41) associated with the solution $\phi(t)$ are invariant under S, and, consequently, the line element $(t_1,\phi(t_1),\dot{\phi}(t_1))$ will be also invariant under S^1. The following relation is therefore automatically satisfied:

$$\psi_{,t}|_{P_1}=-(\psi_{,x}-\varphi_{,t})|_{P_1}\dot{\phi}(t_1)+\varphi_{,x}|_{P_1}\dot{\phi}^2(t_1), \tag{43}$$

leading to

$$\mathbf{S}^1|_{(t_1,x_1,\dot{x})}=\mathbf{S}|_{P_1}+(\psi_{,x}-\varphi_{,t})|_{P_1}(\dot{x}-\dot{x}_1)+\varphi_{,x}|_{P_1}(\dot{x}_1^2-\dot{x}^2), \tag{44}$$

where $\dot{x}_1=\dot{\phi}(t_1)$.

If one now imposes on $\mathbf{S}^1$ the *two* additional conditions

$$(\psi_{,x}-\varphi_{,t})|_{P_1}=0, \quad \varphi_{,x}|_{P_1}=0, \tag{45}$$

then any line element of the form $(t_1,x_1,\dot{x})$ will be left invariant by $\mathbf{S}^1\forall\dot{x}$.

Consequently, the *eight* conditions (40), (42), and (45) replace the $(r+4)$ conditions obtained when $r>2$, and, therefore, by the reasoning following in Sec. II(f), we conclude that Eq. (41) has at most *eight* independent symmetry vectors.

(h) We shall see in this section that the upper bounds on the number of independent symmetry vectors of Eq. (1) obtained above cannot be improved. Indeed, it is a standard result[5] that for $r=2$ the equation

$$\ddot{x}=0 \tag{46}$$

has exactly *eight* independent symmetry vectors; on the other hand, we are going to prove now that the equation

$$x^{(r} = 0 \tag{47}$$

has exactly $r + 4$ independent symmetry vectors when $r > 2$. Thus the upper bounds obtained above are actually attained by (47) for every $r \geqslant 2$ and therefore cannot be improved.

Let us prove that (47) has exactly $r + 4$ independent symmetry vectors when $r > 2$.

Indeed, calling $\mathbf{S}^n$ the nth extension of $\mathbf{S}$, we have

$$\mathbf{S}^n = \varphi(t,x)\frac{\partial}{\partial t} + \psi(t,x)\frac{\partial}{\partial x} + \psi^1(t,x,\dot{x})\frac{\partial}{\partial \dot{x}} + \cdots + \psi_n(t,x,\ldots,x^{(n})\frac{\partial}{\partial x^{(n}}. \tag{48}$$

It is easy to verify that the following identity holds:

$$\psi^i = \frac{d^i\psi}{dt^i} - \sum_{k=1}^{i}\binom{i}{k}x^{(i-k+1}\frac{d^k\varphi}{dt^k}. \tag{49}$$

The condition to be satisfied in order that Eq. (1) admit $\mathbf{S}$ as a symmetry vector can be written in compact form as follows:

$$\mathbf{S}^r(x^{(r} - F) = 0 \quad \text{if } x^{(r} - F(t,x,\ldots,x^{(r-1}) = 0, \tag{50}$$

i.e., the subset of the space $\{(t,x,\ldots,x^{(r})\}$ defined by

$$x^{(r} - F(t,x,\ldots,x^{(r-1}) = 0 \tag{51}$$

must be invariant under the rth extension of $\mathbf{S}$.

For the particular case of Eq. (47), condition (50) reads

$$\mathbf{S}^r(x^{(r}) = 0 \quad \text{if } x^{(r} = 0, \tag{52}$$

that is,

$$\psi^r(t,x,\ldots,x^{(r-1},0) = 0. \tag{53}$$

Taking into account the structure of ψ^i, given by (49), Eq. (53) reduces to

$$\left[\frac{d^r\psi}{dt^r} - \sum_{k=1}^{r}\binom{r}{k}x^{(r-k+1}\frac{d^k\varphi}{dt^k}\right]_{x^{(r}=0} = 0. \tag{54}$$

Let us see now that (54) has indeed $(r + 4)$ independent solutions $(\varphi(t,x),\psi(t,x))$.

In order to show this, consider first the solutions of (54) with $\varphi = 0$ given by

$$\psi(t,x) = c_1x + c_2 + c_3t + \cdots + c_{r+1}t^{r-1}, \qquad \varphi(t,x) = 0. \tag{55}$$

These particular solutions of (54) provide a set of $(r + 1)$ independent symmetry vectors of (47).

Next, since (54) is free from $\dot{\varphi}$ [the coefficient of $\dot{\varphi}$ in (54) being $x^{(r}$, which must be set equal to zero], another solution of (54) is obviously given by

$$\dot{\varphi} = a, \quad a \in R, \quad \psi = 0, \tag{56}$$

that is,

$$\varphi = at + b, \quad \psi = 0. \tag{57}$$

We have therefore $(r + 3)$ independent solutions of (54), given by (55) and (57). The additional independent solution of (54) is easily found taking into account the identity

$$\frac{d^p(tx)}{dt^p} = tx^{(p} + \binom{p}{1}x^{(p-1}, \quad p \in N, \tag{58}$$

whence we get

$$\left.\frac{d^r(tx)}{dt^r}\right|_{x^{(r}=0} = rx^{(r-1}. \tag{59}$$

Therefore, if we look for a symmetry vector having the structure

$$\mathbf{S} = tx\frac{\partial}{\partial x} + \varphi(t,x)\frac{\partial}{\partial t}, \tag{60}$$

the following relation should be satisfied by $\varphi(t,x)$:

$$rx^{(r-1} - \binom{r}{2}x^{(r-1}\ddot{\varphi} - \cdots - \binom{r}{r}\dot{x}\frac{d^r\varphi}{dt^r} = 0. \tag{61}$$

A particular solution of (61) is obviously

$$\varphi = t^2/(r-1). \tag{62}$$

Multiplying $\varphi(t,x) = t^2/(r-1)$ and $\psi(t,x) = tx$ by the factor $(r-1)$, we arrive at the following solution of (54):

$$\varphi = t^2, \quad \psi = (r-1)tx, \tag{63}$$

which is clearly independent of the other $(r + 3)$ solutions of (54) previously found, given by (55) and (57).

Therefore, (54) has at least $r + 4$ independent solutions (55), (57), and (63), and hence (47) has *at least* $r + 4$ independent symmetry vectors: since for $r > 2$ it has *at most* $r + 4$ independent symmetry vectors, as we proved in Sec. II(f), it follows that (47) has *exactly* $r + 4$ independent symmetry vectors when $r > 2$.

The reader should notice that these $(r + 4)$ symmetry vectors do behave, under the Lie–Jacobi bracket, as the generators of a Lie group. That is, one can write

$$[\mathbf{S}_i,\mathbf{S}_j] = \sum_{k=1}^{r+4}c_{ij}^k\mathbf{S}_k, \quad i,j = 1,\ldots,r+4. \tag{64}$$

This property follows from the fact that if $\mathbf{S}_i$ and $\mathbf{S}_j$ are two symmetries of (1), then the same thing will happen also with their Lie–Jacobi bracket $[\mathbf{S}_i,\mathbf{S}_j]$.

Indeed, the condition that $\mathbf{S}_k$ be a symmetry vector of (1) can be written as follows[6]:

$$[\mathbf{S}_k^{r-1},\mathbf{X}] = f_k(t,x,\ldots,x^{(r-1})\mathbf{X}, \tag{65}$$

$\mathbf{X}$ being the vector field canonically associated with Eq. (1):

$$\mathbf{X} = \frac{\partial}{\partial t} + \dot{x}\frac{\partial}{\partial x} + \cdots + F(t,x,\ldots,x^{(r-1})\frac{\partial}{\partial x^{(r-1}}. \tag{66}$$

On the other hand, we have the following identity[7]:

$$[\mathbf{A},\mathbf{B}]^P = [\mathbf{A}^P,\mathbf{B}^P], \quad p \in N, \tag{67}$$

where $\mathbf{A}$ and $\mathbf{B}$ are arbitrary vector fields.

Therefore, since $\mathbf{S}_i$ and $\mathbf{S}_j$ are by hypothesis symmetries of (1), we have

$$\begin{aligned}[[\mathbf{S}_i,\mathbf{S}_j]^{r-1},\mathbf{X}] &\overset{(67)}{=} [[\mathbf{S}_i^{r-1},\mathbf{S}_j^{r-1}],\mathbf{X}] \\ &= (\text{Jacobi's identity}) - [[\mathbf{S}_j^{r-1},\mathbf{X}],\mathbf{S}_i^{r-1}] \\ &\quad - [[\mathbf{X},\mathbf{S}_i^{r-1}],\mathbf{S}_j^{r-1}] \\ &\overset{(65)}{=} -[f_j\mathbf{X},\mathbf{S}_i^{r-1}] + [f_i\mathbf{X},\mathbf{S}_j^{r-1}] \\ &\overset{(65)}{=} f_j(f_i\mathbf{X}) + (\mathbf{S}_i^{r-1}f_j)\mathbf{X} \\ &\quad - f_i(f_j\mathbf{X}) - (\mathbf{S}_j^{r-1}f_i)\mathbf{X} \\ &= g_{ij}\mathbf{X},\end{aligned} \tag{68}$$

with

$$g_{ij} = (\mathbf{S}_i^{r-1} f_j) - (\mathbf{S}_j^{r-1} f_i). \tag{69}$$

Hence the Lie–Jacobi bracket $[\mathbf{S}_i, \mathbf{S}_j]$ satisfies (65) and is therefore a symmetry vector of (1).

It easily follows that $[\mathbf{S}_i, \mathbf{S}_j]$ must be a linear combination of $\mathbf{S}_1, \ldots, \mathbf{S}_{r+4}$, since if this were not the case (47) would have $r+5$ independent symmetry vectors: $\mathbf{S}_1, \ldots, \mathbf{S}_{r+4}$ and $[\mathbf{S}_i, \mathbf{S}_j]$, contrary to what has been already proved in Sec. II(f) (since $r > 2$). Obviously, the same conclusion holds for Eq. (46).

III. MAXIMUM NUMBER OF INDEPENDENT SYMMETRY VECTORS OF THE SYSTEM $\mathbf{x}^{(r)} = \mathbf{0}$ $(r > 2)$

We show in this section that a system of differential equations of the form

$$\mathbf{x}^{(r)} = \mathbf{F}(t, \mathbf{x}, \ldots, \mathbf{x}^{(r-1)}),$$
$$\mathbf{x} = (x_1, \ldots, x_n) \in R^n, \quad \mathbf{F} \in C^\infty, \quad \text{and} \quad r > 2, \tag{70}$$

does not admit more than $2n^2 + nr + 2$ independent symmetry vectors. It would be nice to produce an example of a system of differential equations of the form (70) with $n > 1$ possessing this maximum number of independent symmetry vectors. Unfortunately, the system

$$\mathbf{x}^{(r)} = \mathbf{0}, \quad r > 2, \tag{71}$$

has only $n^2 + nr + 3$ independent symmetry vectors, which is equal to the previously quoted upper bound $2n^2 + nr + 2$ only when $n = 1$. Therefore, the open problem remains of either showing that the system $\mathbf{x}^{(r)} = \mathbf{0}$ has more independent symmetry vectors than any system of type (70)—in which case the number $2n^2 + nr + 2$ should be substituted by the number $n^2 + nr + 3$ as an upper bound on the number of independent symmetry vectors of (70)—or of producing a concrete example of a differential system of type (70) with the maximum number s of independent symmetry vectors $(n^2 + nr + 3 < s \leqslant 2n^2 + nr + 2)$.

(a) Let γ_0 be the integral curve of (70) corresponding to the initial conditions

$$(t_0, \mathbf{x}_0, \dot{\mathbf{x}}_0, \ldots, \mathbf{x}_0^{(r-1)}) \tag{72}$$

and $P_1 = (t_1, \mathbf{x}_1)$ be a point on γ_0 sufficiently close to $P_0 = (t_0, \mathbf{x}_0)$. By a reasoning completely similar to that followed in Secs. II(a), (b), (c), one can prove that there exists an open neighborhood $U \subset R \times R^n$ near P_0 such that through every point P of U it is possible to draw two integral curves of (70), γ_1 and γ_2, with the following two properties:

(i) γ_1 and γ_2 have a contact of order $(r-2)$ with γ_0, respectively, at P_0 and P_1.

(ii) P is isolated in $\gamma_1 \cap \gamma_2$.

(b) Assume now that the vector $\mathbf{S}$ defined by

$$\mathbf{S} = \varphi(t, \mathbf{x}) \frac{\partial}{\partial t} + \sum_{i=1}^{n} \psi_i(t, \mathbf{x}) \frac{\partial}{\partial x_i} \tag{73}$$

is a symmetry vector of Eqs. (70). If we were able to construct $\mathbf{S}$ in such a way that any integral curve of (70) having a contact of order $(r-2)$ with γ_0 either at P_0 or P_1 be invariant under the local one-parameter group G generated by $\mathbf{S}$, then in particular the two paths γ_1 and γ_2 considered above would be invariant under G, and, consequently,

$$g(\gamma_1 \cap \gamma_2) \subset \gamma_1 \cap \gamma_2 \quad \forall g \in G. \tag{74}$$

But, by construction, P is isolated in $\gamma_1 \cap \gamma_2$; therefore, we can write

$$g(P) = P \tag{75}$$

for any $g \in G$ sufficiently close to the identity transformation. Hence $\mathbf{S}$ must vanish at P, and, since P was an arbitrary point of U, we conclude that $\mathbf{S}$ is identically zero on U.

(c) Let us now show that in order that any integral curve of (70) having a contact of order $(r-2)$ with γ_0 at P_0 or P_1 be invariant under $\mathbf{S}$, $2n^2 + nr + 2$ linear conditions on $\mathbf{S}$ suffice.

First, we must impose that the linear element of order $r-1$

$$(t_0, \mathbf{x}_0, \ldots, \mathbf{x}_0^{(r-2)}, \mathbf{x}^{(r-1)}) \tag{76}$$

be invariant under $\mathbf{S}^{r-1}$ for every value of $\mathbf{x}^{(r-1)}$, that is,

$$\mathbf{S}^{r-1}\big|_{(t_0, \mathbf{x}_0, \ldots, \mathbf{x}_0^{(r-2)}, \mathbf{x}^{(r-1)})} = \mathbf{0} \quad \forall \mathbf{x}^{(r-1)} \in R^n. \tag{77}$$

Condition (77) can be written in detail as follows:

$$\begin{aligned}
&\varphi(t_0, \mathbf{x}_0) = 0, \\
&\psi(t_0, \mathbf{x}_0) = 0, \\
&\psi^k(t_0, \mathbf{x}_0, \ldots, \mathbf{x}_0^{(k)}) = \mathbf{0}, \quad k = 1, \ldots, r-2, \\
&\psi^{r-1}(t_0, \mathbf{x}_0, \ldots, \mathbf{x}_0^{(r-2)}, \mathbf{x}^{(r-1)}) = \mathbf{0}, \quad \forall \mathbf{x}^{(r-1)} \in R^n,
\end{aligned} \tag{78}$$

where, of course, $\psi^k = (\psi_1^k, \ldots, \psi_n^k)$. Taking into account the identity [analogous to (49)]

$$\psi_i^{r-1} = \frac{d^{r-1}\psi_i}{dt^{r-1}} - \sum_{k=1}^{r-1} \binom{r-1}{k} x_i^{(r-k)} \frac{d^k \varphi}{dt^k}, \quad i = 1, \ldots, n \tag{79}$$

and the structure of $d^k f / dt^k$, given by

$$\frac{d^k f}{dt^k} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} x_i^{(k)} + B(t, \mathbf{x}, \ldots, \mathbf{x}^{(k-1)}), \tag{80}$$
$$f\colon (t, \mathbf{x}) \mapsto f(t, \mathbf{x}) \in \mathbb{R},$$

we conclude that ψ_i^{r-1} has the following affine structure:

$$\psi_i^{r-1} = \sum_{j=1}^{n} A_{ij}(t, \mathbf{x}, \dot{\mathbf{x}}) x_j^{(r-1)} + B_i(t, \mathbf{x}, \ldots, \mathbf{x}^{(r-2)}), \tag{81}$$
$$A_{ij} = \frac{\partial \psi_i}{\partial x_j} - \dot{\mathbf{x}}_i \frac{\partial \varphi}{\partial x_j} - (r-1) \frac{d\varphi}{dt} \delta_{ij},$$

provided that $r - 1 > 1$, i.e., $r > 2$.

Therefore, taking (81) into account, (78) is equivalent to the following set of $n^2 + nr + 1$ linear conditions on the components of $\mathbf{S}$:

$$\begin{aligned}
&\varphi(t_0, \mathbf{x}_0) = \psi_i(t_0, \mathbf{x}_0) = 0, \\
&\psi_i^k(t_0, \mathbf{x}_0, \ldots, \mathbf{x}_0^{(k)}) = 0, \\
&A_{ij}(t_0, \mathbf{x}_0, \dot{\mathbf{x}}_0) = 0, \\
&B_i(t_0, \mathbf{x}_0, \ldots, \mathbf{x}_0^{(r-2)}) = 0, \\
&i, j = 1, \ldots, n, \quad k = 1, \ldots, r-2.
\end{aligned} \tag{82}$$

Next, in order to assure that the linear element at P_1

$$(t_1,\mathbf{x}_1,\ldots,\mathbf{x}_1^{(r-2)},\mathbf{x}^{(r-1)}),\quad \mathbf{x}_1^{(k}=\boldsymbol{\phi}^{(k}(t_1), \tag{83}$$

is invariant under $\mathbf{S}^{(r-1}$ for every value of $\mathbf{x}^{(r-1}$ [where $\boldsymbol{\phi}(t)$ is of course the solution of (70) corresponding to the initial conditions $(t_0,\mathbf{x}_0,\ldots,\mathbf{x}_0^{(r-1})$] we must impose that

$$\mathbf{S}^{r-1}|_{(t_1,\mathbf{x}_1,\ldots,\mathbf{x}_1^{(r-2},\mathbf{x}^{(r-1})}=\mathbf{0}\quad \forall\mathbf{x}^{(r-1}\in\mathbb{R}^n. \tag{84}$$

Now, if $\mathbf{S}$ satisfies conditions (82), then the integral curve γ_0 will be invariant under $\mathbf{S}$, since $\mathbf{S}$ is by hypothesis a symmetry vector of (70). Therefore, $\mathbf{S}$ must be parallel to the tangent vector to γ_0 on every point of γ_0, that is,

$$\mathbf{S}_P=\varphi(P)\left[\frac{\partial}{\partial t}+\sum_{i=1}^{n}\phi_i(t)\frac{\partial}{\partial x_i}\right],\quad \forall P=(t,\boldsymbol{\phi}(t))\in\gamma_0. \tag{85}$$

Consequently, P_1 will remain invariant under $\mathbf{S}$ if

$$\varphi(t_1,\mathbf{x}_1)=0. \tag{86}$$

This last condition automatically implies that the linear element at P_1

$$(t_1,\mathbf{x}_1,\ldots,\mathbf{x}_1^{(k)}) \tag{87}$$

is invariant under $\mathbf{S}^k$ for every value of k; therefore, for $k=r-1$ we have, taking into account (81);

$$B_i(t_1,\mathbf{x}_1,\ldots,\mathbf{x}_1^{(r-2)})=-\sum_{j=1}^{n}A_{ij}(t_1,\mathbf{x}_1,\dot{\mathbf{x}}_1)x_1^{(r-1},\quad i=1,\ldots,n. \tag{88}$$

Consequently, the linear element (83) will be invariant under $\mathbf{S}^{r-1}$ if

$$\begin{aligned}&\sum_{j=1}^{n}A_{ij}(t_1,\mathbf{x}_1,\dot{\mathbf{x}}_1)(x_j^{(r-1}-x_{1j}^{(r-1})=0,\quad i=1,\ldots,n,\\&\mathbf{x}^{(r-1}=(x_1^{(r-1},\ldots,x_n^{(r-1}),\\&\mathbf{x}_1^{(r-1}=(x_{11}^{(r-1},\ldots,x_{1n}^{(r-1}).\end{aligned} \tag{89}$$

Since (89) must hold for every value of $\mathbf{x}^{(r-1}$, we must finally impose that

$$A_{ij}(t_1,\mathbf{x}_1,\dot{\mathbf{x}}_1)=0,\quad i,j=1,\ldots,n. \tag{90}$$

The $2n^2+nr+2$ equations (82), (86), and (90) guarantee that any integral curve of (70) having a contact of order $(r-2)$ with γ_0 at P_0 or P_1 be invariant under the symmetry vector of (70) $\mathbf{S}$. The linearity of these equations in the components of $\mathbf{S}$ is a direct consequence of the linearity of $\mathbf{S}^{(r-1}$.

(d) We shall now compute the maximum number of independent symmetry vectors of the system

$$\mathbf{x}^{(r}=0,\quad \mathbf{x}=(x_1,\ldots,x_n),\quad r>2. \tag{91}$$

By the reasoning given in Sec. II(h) they will automatically close as a Lie algebra under the Lie-Jacobi bracket.

The necessary and sufficient conditions in order that the vector field

$$\mathbf{S}=\varphi(t,\mathbf{x})\frac{\partial}{\partial t}+\sum_{i=1}^{n}\psi_i(t,\mathbf{x})\frac{\partial}{\partial x_i}$$

be a symmetry vector of (91) can be written as follows:

$$\psi^r|_{\mathbf{x}^{(r}=0}=\mathbf{0} \tag{92}$$

or, taking into a account (79),

$$\left.\frac{d^r\psi_i}{dt^r}\right|_{\mathbf{x}^{(r}=0}-\sum_{k=2}^{r}\binom{r}{k}x_i^{(r-k+1}\left.\frac{d^k\varphi}{dt^k}\right|_{\mathbf{x}^{(r}=0}=0.\quad i=1,\ldots,n. \tag{93}$$

At this point it is important to have in mind the structure of $d^sf(t,\mathbf{x})/dt^s$, which can be shown to be

$$\begin{aligned}\frac{d^sf}{dt^s}=&\sum_{|r|+q=s}c^s_{r_1\cdots r_p}\\&\times\sum_{i_1,\ldots,i_p=1}^{n}\frac{\partial^{p+q_f}}{\partial t^q\,\partial x_{i_1}\cdots\partial x_{i_p}}x_{i_1}^{(r_1}\cdots x_{i_p}^{(r_p},\\&|r|=r_1+\cdots+r_p,\quad r_1<r_2<\cdots<r_p,\quad c^s_{r_1\cdots r_p}\in N.\end{aligned} \tag{94}$$

An immediate consequence of (94) is that a term of the form $x_i^{(r-1}\ddot{x}_j$ cannot appear in (93) from the development of $d^r\psi_i/dt^{r-1}$, since $(r+1)+2=r+1>r$. A term of this type can only arise, therefore, from the expressions

$$-\binom{r}{2}x_i^{(r-1}\frac{d^2\varphi}{dt^2},\quad -\binom{r}{r-1}\frac{d^{r-1}\varphi}{dt^{r-1}} \tag{95}$$

also appearing in (93). These two terms are different when $r-1\neq 2$, i.e., when $r>3$, and therefore for $r>3$ the coefficient of the term $x_i^{(r-1}\ddot{x}_j$ is either

$$\begin{aligned}&-\binom{r}{2}\varphi_{,j},\quad \text{when } i\neq j,\\ &\text{or}\\&-\left[\binom{r}{2}+r\right]\varphi_{,j},\quad \text{when } i=j,\end{aligned} \tag{96}$$

whereas, for $r=3$, $x_i^{(r-1}\ddot{x}_j$ reduces to $\ddot{x}_i\ddot{x}_j$, whose coefficient is simply

$$-3\varphi_{,j},\quad i,j=1,\ldots,n. \tag{97}$$

Since (93) must be an *identity in* $\dot{\mathbf{x}},\ddot{\mathbf{x}},\ldots,\mathbf{x}^{(r-1}$, and φ,ψ_i do not depend on these variables, the coefficient of the term $x_i^{(r-1}\ddot{x}_j$ must equal zero; taking into account (96) (for $r>3$) and (97) (for $r=3$), we conclude that

$$\varphi_{,j}=0,\quad j=1,\ldots,n. \tag{98}$$

Accordingly, for every symmetry vector of (91) we have

$$\varphi(t,x)=f(t). \tag{99}$$

Note that the above reasoning obviously fails when $r=2$, since then the term $x_i^{(r-1}\ddot{x}_j$ reduces to $\dot{x}_i\ddot{x}_j$, which is absent from (93) by the restriction $\ddot{\mathbf{x}}=0$.

Substituting (99) into (93), we obtain

$$\left.\frac{d^r\psi_i}{dt^r}\right|_{\mathbf{x}^{(r}=0}-\sum_{k=2}^{r}\binom{r}{k}x_i^{(r-k+1}f^{(k}(t)=0,\quad i=1,\ldots,n. \tag{100}$$

Remembering (94) again, we realize that the term $x_j^{(r-1}\dot{x}_k$ appears in (100) only through $d^r\psi_i/dt^r|_{\mathbf{x}^{(r}=0}$ and its coefficient is (up to the positive integer $c^r_{1,r-1}$) $\psi_{i,jk}$. Therefore, we must have

$$\frac{\partial^2\psi_i}{\partial x_j\,\partial x_k}=0,\quad i,j,k=1,\ldots,n. \tag{101}$$

Similarly, considering the coefficients of the terms $x_j^{(r-1}$ with $j\neq 1$, which again only appear in (100) through $d^r\psi_i/dt^r|_{\mathbf{x}^{(r}=0}$, we obtain

$$\frac{\partial^2\psi_i}{\partial t\,\partial x_j}=0,\quad i,j=1,\ldots,n,\ i\neq j. \tag{102}$$

When $i=j$, considering the coefficient of the term x_i^{r-1} in (100), we get

$$c_{r-1}^r\frac{\partial^2\psi_i}{\partial t\,\partial x_i}-\binom{r}{2}\ddot{f}(t)=0,\quad i=1,\ldots,n. \tag{103}$$

Since c_{r-1}^r is a positive integer, we can rewrite (103) as follows:

$$\frac{\partial^2\psi_i}{\partial t\,\partial x_i}=K\cdot f(t),\quad i=1,\ldots,n\quad\left[K=\binom{r}{2}\quad c_{r-1}^r>0\right]. \tag{104}$$

Considering now the coefficient of the term independent of $\dot{\mathbf{x}},\ddot{\mathbf{x}},\ldots,\mathbf{x}^{(r-1)}$ in (100), we are led to

$$\frac{\partial^r\psi_i}{\partial t^r}=0,\quad i=1,\ldots,n. \tag{105}$$

From Eq. (101) we readily obtain

$$\psi_i=\sum_{j=1}^n a_{ij}(t)x_j+b_i(t), \tag{106}$$

and, taking (102) and (104) into account, we immediately arrive at

$$\psi_i=\sum_{j=1}^n a_{ij}x_j+K\dot{f}(t)x_i+b_i(t),\qquad a_{ij}\in R\ \ \forall i,j=1,\ldots,n, \tag{107}$$

and, substituting (107) into (105), we finally get

$$Kf^{(r+1)}(t)x_i+b_i^{(r)}(t)=0. \tag{108}$$

Therefore, we must have

$$f(t)=P_r(t),\quad b_i(t)=Q_{r-1}^i(t),\quad i=1,\ldots,n, \tag{109}$$

P_r and Q_{r-1}^i being polynomials of maximum degree r and $(r-1)$, respectively. From (107) and (109) we get the following structure of ψ_i,

$$\psi_i=\sum_{j=1}^n a_{ij}x_j+Kx_i\dot{P}_r(t)+Q_{r-1}^i(t)\quad(a_{ij}\in R), \tag{110}$$

and, substituting it back into (100), we arrive at

$$K\frac{d^r}{dt^r}\left[x_i\dot{P}_r(t)\right]\Big|_{\mathbf{x}^{(r)}=\mathbf{0}}-\sum_{k=2}^r\binom{r}{k}x_i^{(r-k+1)}P_r^{(k)}(t)=0. \tag{111}$$

Applying Leibnitz's theorem to the first term of (111), we obtain

$$K\sum_{k=1}^{r-1}\binom{r}{k}x_i^{(r-k)}P_r^{(k+1)}(t)=\sum_{k=1}^{r-1}\binom{r}{k+1}x_i^{(r-k)}P_r^{(k+1)}(t). \tag{112}$$

Since we are considering now the case $r>2$, we can compare the coefficients of $x_i^{(r-1)}$ and $x_i^{(r-2)}$ in both members of (112), obtaining

$$K\cdot r\cdot P_r(t)=\binom{r}{2}\cdot P_r(t),\qquad K\cdot\binom{r}{2}\cdot P_r(t)=\binom{r}{3}\cdot P_r(t). \tag{113}$$

It is easy to prove by induction that $c_{r-1}^r=r$; hence $K=(r-1)/2$ [see (104)]. The first equation in (113) reduces to an identity and the second one leads to

$$\dddot{P}_r(t)=0,\quad \forall t\in R, \tag{114}$$

$$\text{i.e.,}\quad P_r(t)=a+bt+ct^2\quad(a,b,c\in R).$$

Conversely, if (114) holds, then (112) is automatically satisfied. Therefore, the "general solution" of (93) is obtained by substituting $P_r(t)=a+bt+ct^2$ into (110), and, consequently, the general solution of (92) is

$$\varphi=a+bt+ct^2,$$

$$\psi_i=\sum_{j=1}^n A_{ij}x_j+c(r-1)tx_i+Q_{r-1}^i(t), \tag{115}$$

$$i=1,\ldots,n,\quad a,b,c,A_{ij}\in R,$$

where we have set $A_{ij}=a_{ij}+\frac{1}{2}b(r-1)\delta_{ij}$ (δ_{ij} being, of course, the Kronecker delta).

From (115) we immediately obtain the following set of n^2+nr+3 independent symmetry vectors of (91):

$$x_i\frac{\partial}{\partial x_j},\quad i,j=1,\ldots,n,$$

$$t^p\frac{\partial}{\partial x_i},\quad p=0,1,\ldots,r-1,\ i=1,\ldots,n, \tag{116}$$

$$\frac{\partial}{\partial t},\quad t\frac{\partial}{\partial t},\quad t^2\frac{\partial}{\partial t}+(r-1)t\cdot\sum_{k=1}^n x_k\frac{\partial}{\partial x_k}.$$

This establishes the point we wanted to make: when $n>1$ and $r>2$, the system of differential equations $\mathbf{x}^{(r)}=\mathbf{0}$ does not provide us (as happened for $n=1$) with a maximum number of independent symmetry vectors equal to the upper bound $2n^2+nr+2$ obtained in III(a)–(c). Therefore, it remains an open problem to find systems of differential equations—if any—whose maximum number of independent symmetry vectors is greater than the number n^2+nr+3.

Finally, note that, when $n=1$, the symmetry vectors (116) reduce to the symmetry vectors of $x^{(r)}=0$ computed in Sec. II(h), as it should be.

IV. MAXIMUM NUMBER OF INDEPENDENT SYMMETRY VECTORS OF THE SYSTEM $\ddot{\mathbf{x}}=\mathbf{F}$

We show in this section that a system of differential equations of the form

$$\ddot{\mathbf{x}}=\mathbf{F}(t,\mathbf{x},\dot{\mathbf{x}}),\quad \mathbf{x}=(x_1,\ldots,x_n) \tag{117}$$

cannot possess more than $2(n+1)^2$ independent symmetry vectors. We also compute, by a direct procedure, all the symmetry vectors of the system $\ddot{\mathbf{x}}=\mathbf{0}$, obtaining only n^2+4n+3 independent vectors. Since this number is less than the upper bound $2(n+1)^2$ mentioned above, the open question arises of whether or not there exist differential systems admitting more than n^2+4n+3 independent symmetry vectors.

In Sec. V we show that this is not the case: In other words, the maximum number of independent symmetry vectors admitted by any system of the form (117) is never greater than n^2+4n+3, the number of independent symmetry vectors of the system $\ddot{\mathbf{x}}=\mathbf{0}$.

(a) Let

$$S = \varphi \frac{\partial}{\partial t} + \sum_{i=1}^{n} \psi_i \frac{\partial}{\partial x_i} \tag{118}$$

be a pointlike symmetry vector of (117); then it is easy to verify that the structure of $\mathbf{S}^1$ (the first extension of $\mathbf{S}$ to the variables $t,\mathbf{x},\dot{\mathbf{x}}$) is the following:

$$\mathbf{S}^1 = \mathbf{S} + \sum_{i=1}^{n} \psi_i^1 \cdot \frac{\partial}{\partial \dot{x}_i},$$

$$\psi_i^1 = \sum_{j=1}^{n} \psi_{i,j}\dot{x}_j + \psi_{i,t} - \dot{x}_i\left(\sum_{j=1}^{n} \varphi_{,j}\dot{x}_j + \varphi_{,t}\right). \tag{119}$$

Therefore, it is clear that the linear element $(t_0,\mathbf{x}_0,\dot{\mathbf{x}}) = (P_0,\dot{\mathbf{x}})$ will be left invariant by $\mathbf{S}^1$ if for every value of $\dot{\mathbf{x}}$ the following set of $n^2 + 3n + 1$ linear equations in the components of $\mathbf{S}$ holds:

$$\left.\begin{aligned} &\varphi(P_0) = 0, \quad \psi(P_0) = 0 \\ &\psi_{i,t}(P_0) = \varphi_{,j}(P_0) = 0 \\ &(\psi_{i,j} - \varphi_{,t}\delta_{ij})(P_0) = 0 \end{aligned}\right\}, \quad i,j = 1,\ldots,n. \tag{120}$$

Similarly, a second point $P_1 = (t_1,\mathbf{x}_1)$ lying on the integral curve γ_0 of (117) corresponding to the initial conditions $(t_0,\mathbf{x}_0,\dot{\mathbf{x}}_0)$ will be left invariant by $\mathbf{S}$ provided only that

$$\varphi(P_1) = 0 \tag{121}$$

since, exactly as in Secs. II and III, (121) and the fact that $\mathbf{S}$ is a symmetry vector of (117) and P_1 lies on an integral curve of (117) imply that $\psi(P_1) = \mathbf{0}$ as well.

Finally, from all that has been said in Secs. II and III, it should be clear by now that, in order that any linear element at P_1, $(P_1,\dot{\mathbf{x}})$, be invariant under $\mathbf{S}^1$, the following $n^2 + n$ linear conditions in φ and ψ suffice:

$$\left.\begin{aligned} &\varphi_{,j}(P_1) = 0 \\ &(\psi_{i,j} - \varphi_{,t}\delta_{ij})(P_1) = 0 \end{aligned}\right\}, \quad i,j = 1,\ldots,n, \tag{122}$$

since when (120), (121) and (122) hold $\psi_{i,t}(P_1)$ automatically vanishes, due to the fact that the linear element $(P_1,\dot{\mathbf{x}}_1)$ tangent to γ_0 is then invariant under $\mathbf{S}^1$.

Accordingly, the $2(n+1)^2$ conditions (120), (121), and (122) are sufficient in order that any linear element at P_0 or P_1 be invariant under $\mathbf{S}^1$; since these conditions are linear in the components of $\mathbf{S}$, the same construction followed in Secs. II and III can be repeated now, with the result that Eq. (117) does not admit more than $2(n+1)^2$ independent symmetry vectors.

(b) We now compute all the pointlike symmetry vectors of the system

$$\ddot{\mathbf{x}} = \mathbf{0}, \quad \mathbf{x} = (x_1,\ldots,x_n) \tag{123}$$

in order to establish whether or not the dimension of the vector space generated by these symmetries equals the upper bound $2(n+1)^2$ obtained above.

Since the necessary and sufficient conditions in order that (118) be a symmetry vector of (123) are

$$\psi_i^2|_{\dot{\mathbf{x}}=\mathbf{0}} = 0, \quad i = 1,\ldots,n, \tag{124}$$

computing $\psi_i^2|_{\dot{\mathbf{x}}=\mathbf{0}}$ and setting equal to zero the coefficients of $1, \dot{x}_i$, and $\dot{x}_i\dot{x}_j$, we arrive at the following system of partial differential equations in φ, ψ:

$$\left.\begin{aligned} &\varphi_{,jk} = 0 \\ &\psi_{i,tt} = 0 \\ &\psi_{i,jk} = \varphi_{,kt}\delta_{ij}(1+\delta_{jk}) \\ &\psi_{i,jt} = \tfrac{1}{2}\varphi_{,tt}\delta_{ij} \end{aligned}\right\}, \quad i,j,k = 1,\ldots,n, \tag{125a,b,c,d}$$

From (125a) and (125b) we get

$$\varphi = \sum_{j=1}^{n} C_j(t)x_j + D(t), \qquad \psi_i = A_i(x)t + B_i(x). \tag{126}$$

Substituting (126) into (125c) and (125d), we obtain

$$A_i(x) = a_i(x_i), \qquad B_i(x) = b_i(x_i) + \sum_{\substack{j=1 \\ (j\neq 1)}}^{n} b_{ij}(x_i)x_j. \tag{127}$$

Substituting (127) back into (126), we obtain, after some easy calculations, the general solution of (125):

$$\varphi = \sum_{j=1}^{n} (c_j t + c_j')x_j + at^2 + dt + d',$$

$$\psi_i = (ax_i + a_i)t + \sum_{j=1}^{n} c_j x_i x_j + \sum_{j=1}^{n} b_{ij}x_j + b_i. \tag{128}$$

From (128) we obtain the following set of $n^2 + 4n + 3$ independent generators of the vector space of the symmetries of (123):

$$\left.\begin{aligned} &\frac{\partial}{\partial t}, \quad t\frac{\partial}{\partial t}, \quad x_i\frac{\partial}{\partial t} \\ &\frac{\partial}{\partial x_i}, \quad t\frac{\partial}{\partial x_i}, \quad x_j\frac{\partial}{\partial x_i} \\ &tx_i\frac{\partial}{\partial t} + x_i\sum_{k=1}^{n} x_k\frac{\partial}{\partial x_k} \\ &t^2\frac{\partial}{\partial t} + t\sum_{k=1}^{n} x_k\frac{\partial}{\partial x_k} \end{aligned}\right\}, \quad i,j = 1,\ldots,n. \tag{129}$$

By the reasoning followed in Sec. II, the set of vectors (129) closes as a Lie algebra under the Lie–Jacobi bracket.

It is not difficult to verify that the set of symmetry vectors given by (129) is a set of generators for the projective pseudogroup of the space $\{(t,\mathbf{x})\} = R^{n+1}$, whose finite expression is given by

$$x_i' = \frac{\sum_{j=1}^{n+1} a_{ij}x_j + a_{i,n+2}}{\sum_{j=1}^{n+1} b_j x_j + b_{n+2}}, \qquad x_{n+1} = t, \quad i = 1,\ldots,n+1. \tag{130}$$

The projective pseudogroup does precisely possess $(n+2)^2 - 1 = n^2 + 4n + 3$ essential parameters, and, therefore, $n^2 + 4n + 3$ independent generators (see the Appendix).

V. REDUCTION OF THE MAXIMUM NUMBER OF INDEPENDENT SYMMETRIES OF THE SYSTEM $\ddot{\mathbf{x}} = \mathbf{F}$

We show in this section that the system (117) does not admit more than $n^2 + 4n + 3$ independent symmetry vectors, thereby achieving an improvement of the maximum number of independent symmetry vectors of (117), $2n^2 + 4n + 2$, derived in Sec. IV. The new upper bound obtained in this section cannot be further improved, since in Sec. IV, it has been shown that the system $\ddot{\mathbf{x}} = \mathbf{0}$ has precisely $n^2 + 4n + 3$ independent symmetry vectors.

The proof given here uses the following remarkable property of the projective pseudogroup of R^{n+1}:

If a projective transformation T of R^{n+1} leaves $n + 3$ points of R^{n+1} fixed, and these points are in "*generic position*," then T is the identity transformation.[8] (We say that $n + 3$ points of R^{n+1} are in *generic position* if for *every* selection of $n + 2$ of them the $n + 1$ vectors obtained choosing one of these $n + 2$ points as the origin and the rest as end points are linearly independent.)

(a) Let $P_1,...,P_{n+3}$ be $n + 3$ points of R^{n+1} such that

$$P_i = (t_i, \mathbf{x}_i), \quad i \neq j \Rightarrow t_i \neq t_j. \tag{131}$$

Let us assume for the moment that these points can be chosen in such a way that to any couple of them (P_i, P_j) with $i \neq j$ there corresponds an integral curve $\gamma_{ij} = \{(t, \boldsymbol{\phi}_{ij}(t)) \times | t \in R\}$ of (117) passing through P_i and P_j: We shall prove in Sec. V (e) that this assumption can indeed be satisfied.

Assuming then that we have chosen the points $P_1,...,P_{n+3}$ in a such a way that this last assumption holds true, by a straightforward generalization of the argument given in Sec. II(a) one can prove the following result:

If the points $P_1,...,P_{n+3}$ are sufficiently close to each other, then for every pair (i, j) with $i \neq j$ there exist open neighborhoods U_{ij} and P_j and V_{ij} of $\dot{\mathbf{x}}_{ij} = \dot{\boldsymbol{\phi}}_{ij}(t_i)$ such that through every point P of U_{ij} there passes exactly one integral curve of (117) containing P_i, with velocity (= derivative with respect to time t) at t_i lying in V_{ij}.

Suppose now that the vector field $\mathbf{S}$ given by (118) is a symmetry of (117) leaving all the points $P_1,...,P_{n+3}$ invariant. It is clear that in order to achieve it the following $(n + 1)(n + 3)$ conditions are sufficient:

$$\varphi(P_i) = 0, \quad \psi_j(P_i) = 0, \quad i = 1,...,n+3, j = 1,...,n. \tag{132}$$

Equations (132) automatically imply that the integral curves γ_{ij} $(i \neq j)$ are subsets of R^{n+1} invariant under $\mathbf{S}$, and therefore that the $n + 2$ linear elements at P_i

$$(P_i, \dot{\mathbf{x}}_{ij}), \quad i \neq j, \tag{133}$$

are also invariant under $\mathbf{S}^1$, for every $i = 1,...,n + 3$.

Indeed, if g is a transformation belonging to the local one-parameter group generated by S and sufficiently close to the identity, then we have by continuity

$$\dot{\mathbf{x}}'_{ij} = \frac{d}{ds}\bigg|_{s=t_i} (g\boldsymbol{\phi}_{ij})(s) \in V_{ij} \tag{134}$$

$$[\, g\gamma_{ij} = \{(s, (g\boldsymbol{\phi}_{ij})(s)) | s \in R\}\,]$$

since $\dot{\mathbf{x}}_{ij} \in V_{ij}$ by construction. But this necessarily implies that $g\gamma_{ij} = \gamma_{ij}$, since both γ_{ij} and its transform $g\gamma_{ij}$ pass through P_i with velocity at P_i lying in V_{ij}, and the equality of γ_{ij} and its transform implies obviously that $\dot{\mathbf{x}}_{ij}$ equals $\dot{\mathbf{x}}'_{ij}$, its transform under S^1, as claimed.

(b) Consider now a finite transformation

$$t' = g(t, \mathbf{x}), \quad \mathbf{x}' = \mathbf{f}(t, \mathbf{x}) \tag{135}$$

such that P_i is invariant under (135), for fixed $i \in \{1,...,n + 3\}$. As is well known, the transformation induced by (135) on the derivatives $\dot{\mathbf{x}}$ at P_i is given by

$$\dot{\mathbf{x}}' = \frac{\sum_{j=1}^{n} \mathbf{f}_{,j}(P_i)\dot{\mathbf{x}}_j + \mathbf{f}_{,t}(P_i)}{\sum_{j=1}^{n} g_{,j}(P_i)\dot{\mathbf{x}}_j + g_{,t}(P_i)}, \tag{136}$$

i.e., any curve $\{(t, a(t)) | t \in R\}$ passing through P_i such that $\dot{\mathbf{a}}(t_i) = \dot{\mathbf{x}}$ will be transformed under (135) into another curve $\{(s, \mathbf{b}(s)) | s \in \mathbb{R}\}$ through P_i, with $\dot{\mathbf{b}}(t_i) = \dot{\mathbf{x}}'$.

Since P_i is fixed, (136) implies that the velocities at P_i transform under a projective transformation, whose parameters depend, of course, on the point P_i that is being kept fixed. Denoting now by

$$t' = g(t, \mathbf{x}; \alpha), \quad \mathbf{x}' = \mathbf{f}(t, \mathbf{x}; \alpha), \tag{137}$$

the local one-parameter group of transformations generated by the symmetry $\mathbf{S}$ satisfying conditions (132), then $\mathbf{S}^1$ acts on the velocities at P_i as a one-parameter subgroup G_i of the projective pseudogroup of $R^n = \{\dot{\mathbf{x}}\}$. Furthermore, every transformation $g \in G_i$ leaves invariant the $n + 2$ linear elements at P_i given by (133), as we have just seen: therefore, if we are able to choose the velocities $\dot{\mathbf{x}}_{ij}$ $(i \neq j, i$ fixed) in generic position (by an appropriate selection of the points $P_1,...,P_{n+3}$), then, by the property of the projective pseudogroup quoted at the beginning of this section, we can conclude that G_i reduces to the identity transformation and, therefore, that $\mathbf{S}^1$ leaves *every* linear element at P_i invariant.

(c) Suppose now that we are able to find a set of $n + 3$ points of R^{n+1} $\{P_1,...,P_{n+3}\}$ satisfying (131), and the following additional requirement: The *two* sets of $(n + 2)$ vectors of R^n given by

$$\{\dot{\mathbf{x}}_{ij} | j = 2,3,...,n+3\}, \quad \{\dot{\mathbf{x}}_{2k} | k = 1,3,4,...,n+3\} \tag{138}$$

are in generic position in R^n. According to III(a), we can find an open neighborhood U in R^{n+1} such that through every point P of U there pass two integral curves of (117), γ_1 and γ_2, joining P, respectively, with P_1 and P_2 in such a way that P is isolated in $\gamma_1 \cap \gamma_2$. Since (131), (132), and (138) imply that every linear element at P_1 or P_2 is invariant under $\mathbf{S}$ and $\mathbf{S}$ is by hypothesis a symmetry vector of (117), it follows that γ_1 and γ_2 are both invariant under $\mathbf{S}$; therefore, P has to be invariant under $\mathbf{S}$, since P is isolated in $\gamma_1 \cap \gamma_2$. Hence every point of U is invariant under $\mathbf{S}$, implying that $\mathbf{S} = \mathbf{0}$ on U.

Since conditions (132) are clearly linear in the components of $\mathbf{S}$, we can again apply the argument of Sec. II(f) to conclude that (117) does not admit more than $n^2 + 4n + 3$ independent symmetry vectors.

The only point meriting a separate treatment in order that our proof be complete is the following: We have to show that it is indeed possible to find a set of $n + 3$ points of R^{n+1} satisfying conditions (131) and (138), such that every pair of points of this set can be joined by an integral curve of (117). In order to prove this statement, the lemma that follows is of

great practical value since, as we shall explain below, it reduces the problem of finding the set of points $P_1,...,P_{n+3}$ with the properties mentioned above to an easier one.

(d) *Lemma*: Let $P_0 = (t_0,\mathbf{x}_0)$ be a point of R^{1+n} and call $\boldsymbol{\phi}(t,\boldsymbol{\xi})$ the unique solution of (117) corresponding to the initial condition $(P_0,\boldsymbol{\xi})$. Consider the straight line of R^{1+n} parallel to $(1,\mathbf{v})$ and passing through P_0, whose equation is

$$t = t_0 + s, \quad \mathbf{x} = \mathbf{x}_0 + s\mathbf{v} \quad \forall s \in R. \tag{139}$$

Then one can find $\epsilon > 0$ such that for every s such that $0 < |s| < \epsilon$ there is an integral curve of (117) passing through P_0 and $(t_0 + s,\mathbf{x}_0 + s\mathbf{v})$, whose derivative at t_0, $\mathbf{h}(s)$, satisfies

$$\lim_{s\to 0} \mathbf{h}(s) = \mathbf{v}. \tag{140}$$

Proof: The function $\mathbf{f}(s,\boldsymbol{\xi})$ defined by

$$\mathbf{f}(s,\boldsymbol{\xi}) = \boldsymbol{\phi}(t_0 + s,\boldsymbol{\xi}) - \mathbf{x}_0 - s\boldsymbol{\xi} \tag{141}$$

is a C^∞ function [since the function $\mathbf{F}$ appearing in (117) is assumed in what follows to be of class C^∞][9] and satisfies

$$\mathbf{f}(0,\boldsymbol{\xi}) = \mathbf{0} \quad \forall \boldsymbol{\xi} \in R^n. \tag{142}$$

Therefore, we can write

$$\mathbf{f}(s,\boldsymbol{\xi}) = \int_0^1 \frac{\partial}{\partial\theta}\, \mathbf{f}(\theta s,\boldsymbol{\xi})\, d\theta, \tag{143}$$

and, since

$$\frac{\partial}{\partial\theta}\, \mathbf{f}(\theta s,\boldsymbol{\xi}) = s\cdot \mathbf{f}_{,s}(\theta s,\boldsymbol{\xi}), \tag{144}$$

$\mathbf{f}(s,\boldsymbol{\xi})$ can be factorized as follows:

$$\mathbf{f}(s,\boldsymbol{\xi}) = s\cdot \mathbf{g}(s,\boldsymbol{\xi}),$$

$$\left(g(s,\boldsymbol{\xi}) = \int_0^1 f_{,s}(\theta s,\boldsymbol{\xi})\, d\theta\right), \tag{145}$$

where $\mathbf{g}(s,\boldsymbol{\xi})$ is C^∞ since $\mathbf{f}$ is C^∞.

Therefore, we have

$$\boldsymbol{\phi}(t_0 + s,\boldsymbol{\xi}) = s\cdot \mathbf{g}(s,\boldsymbol{\xi}) + x_0 + s\boldsymbol{\xi}, \quad \mathbf{g} \in C^\infty \tag{146}$$

and the intersection of the integral curve $\{(t,\boldsymbol{\phi},(t,\boldsymbol{\xi}))\,|\,t \in R\}$ with the straight line (139) leads to the equation

$$s\cdot \mathbf{g}(s,\boldsymbol{\xi}) + \mathbf{x}_0 + s\boldsymbol{\xi} = \mathbf{x}_0 + s\mathbf{v} \tag{147a}$$

or, since $s \neq 0$,

$$\mathbf{v} = \boldsymbol{\xi} + \mathbf{g}(s,\boldsymbol{\xi}). \tag{147b}$$

Equation (144) implicitly defines $\boldsymbol{\xi}$ as a C^∞ function of s, $\boldsymbol{\xi} = \mathbf{h}(s)$. Indeed, define a function $\boldsymbol{\psi}$, $(s,\boldsymbol{\xi})$ as follows:

$$\boldsymbol{\psi}(s,\boldsymbol{\xi}) = \boldsymbol{\xi} + \mathbf{g}(s,\boldsymbol{\xi}) - \mathbf{v}. \tag{148}$$

Then we have

$$\boldsymbol{\psi}(0,\mathbf{v}) = \mathbf{g}(0,\mathbf{v}) = \mathbf{0} \tag{149}$$

[since $\mathbf{g}(0,\boldsymbol{\xi}) = \int_0^1 \mathbf{f}_{,s}(0,\boldsymbol{\xi})\, d\theta = \mathbf{0}\ \forall \boldsymbol{\xi} \in \mathbb{R}$ on account of the definition (141)], and

$$(D_{\boldsymbol{\xi}}\, \boldsymbol{\psi})(0,\mathbf{v}) = I + (D_{\boldsymbol{\xi}}\, \mathbf{g})(0,\mathbf{v}) = I \tag{150}$$

(I = identity matrix of dimension n)

[taking into account that $\mathbf{g}(0,\boldsymbol{\xi}) = \mathbf{0}$ for every $\boldsymbol{\xi}$, as we have just shown].

Equations (149) and (150) allow us to apply the implicit function theorem to the function $\boldsymbol{\psi}(s,\boldsymbol{\xi})$ at the point $(0,\mathbf{v})$, thus obtaining $\boldsymbol{\xi}$ as a function of s, $\boldsymbol{\xi} = \mathbf{h}(s)$, in a sufficiently small neighborhood $|s| < \epsilon$ of $s = 0$. The function $\mathbf{h}(s)$ satisfies

$$\mathbf{h}(0) = \mathbf{v}, \tag{151a}$$

$$\boldsymbol{\psi}(s,\mathbf{h}(s)) = \mathbf{0} \quad \text{if } |s| < \epsilon. \tag{151b}$$

It follows that the integral curve of (117) corresponding to the initial condition $(P_0,\mathbf{h}(s))$ passes through P_0 and through the point $(t_0 + s,\mathbf{x}_0 + s\mathbf{v})$ [by (146)–(148) and (151)]; in addition, we have

$$\lim_{s\to 0} \mathbf{h}(s) = \mathbf{h}(0) = \mathbf{v} \tag{152}$$

on account of (151a) since $\mathbf{h}(s)$ is a continuous function (as a matter of fact, $\mathbf{h}$ is C^∞, as follows from the fact that $\mathbf{g}$ is C^∞ and the implicit function theorem). This completes the proof of the lemma.

(e) *Consequences of the lemma*: Let $\{P_1,...,P_{n+3}\}$ be a set of $n + 3$ points of R^{1+n}, $P_i = (t_i,\mathbf{x}_i)$, satisfying the following conditions:

$$i \neq j, \quad t_i \neq t_j; \tag{153a}$$

the two sets of points of R^n

$$\left\{\frac{\mathbf{x}_i - \mathbf{x}_1}{t_i - t_1}\,\middle|\, i = 2,3,...,n+3\right\},$$

$$\left\{\frac{\mathbf{x}_j - \mathbf{x}_2}{t_j - t_2}\,\middle|\, j = 1,3,...,n+3\right\} \tag{153b}$$

are in generic position in R^n.

We shall indicate at the end of this section how to construct sets of $n + 3$ points of R^{1+n} satisfying conditions (153).

Consider now the transformation $H_a : R^{1+n} \to R^{1+n}$ defined as follows:

$$H_a(P) = P_1 + a(P - P_1) \equiv P^a, \quad a \in R,\ a > 0. \tag{154}$$

If $\{P_1,...,P_{n+3}\}$ satisfy conditions (153), the same will happen with $\{P_1^a,...,P_{n+3}^a\}$, since we have

$$\frac{\mathbf{x}_i^a - \mathbf{x}_k^a}{t_i^a - t_k^a} = \frac{a(\mathbf{x}_i - \mathbf{x}_k)}{a(t_i - t_k)} = \frac{\mathbf{x}_i - \mathbf{x}_k}{t_i - t_k}. \tag{152}$$

When $a \to 0$, $P_i^a \to P_1^a$ for every $i = 1,...,n + 3$, but the directions $(t_i - t_j,\mathbf{x}_i - \mathbf{x}_j)$ defined by every pair of points P_i,P_j with $i \neq j$ remain invariant under H_a.

Therefore, by repeated application of the lemma proved above, it follows that, for sufficiently small a, for every pair of points P_i,P_j with $i \neq j$ there is an integral curve of (117) joining P_i with P_j and satisfying

$$\lim_{a\to 0} \dot{\boldsymbol{\phi}}_{ij}^a(t_i) = \frac{\mathbf{x}_j - \mathbf{x}_i}{t_j - t_i}, \tag{156}$$

$$i,j = 1,...,n+3, \quad i \neq j,$$

where $\boldsymbol{\phi}_{ij}^a(t)$ is the solution of (117) whose associated integral curve passes through P_i and P_j.

Furthermore, it is easy to verify that if $m + 2$ points of R^m are in generic position, any sufficiently small perturbation applied to them will lead again to a set of $m + 2$ points in generic position; this is essentially due to the fact that genericity is defined in terms of linear independence of certain sets of vectors, and linear independence is preserved by suffi-

ciently small perturbations. It follows [by (153b)] that the two sets of vectors of R^n defined by

$$\{\dot{\phi}^a_{1i} | i = 2,...,n+3\},$$
$$\{\dot{\phi}^a_{2j} | j = 1,3,...,n+3\} \tag{157}$$

are in generic position in R^n, if we choose a sufficiently small.

The conclusion is, therefore, that if $P_1,...,P_{n+3}$ satisfy condition (153), then one can find $a \in \mathbb{R}$ such that the new set of points $P^a_1,...,P^a_{n+3}$ satisfy conditions (131) and (138). The only point that remains to be proved is, therefore, that it is indeed possible to find $P_1,...,P_{n+3}$ such that conditions (153) are satisfied.

To this end, notice that if the following points of R^n

$$\{0,\mathbf{v}_1,...,\mathbf{v}_n\} \tag{158}$$

are in generic position, it immediately follows that the following set of $n+3$ points of R^{1+n},

$$\{(0,\mathbf{0}),(r_0,\mathbf{0}),(r_1,\mathbf{v}_1),...,(r_{n+1},\mathbf{v}_{n+1})\},$$
$$r_i \neq 0 \quad \forall i = 0,1,...,n+1, \quad r_i \neq r_j \quad \forall i \neq j, \tag{159}$$

satisfies conditions (153), provided only that the numbers

$$r_0, \quad r_i - 1, \quad i = 1,...,n+1, \tag{160}$$

are chosen sufficiently small.

Indeed, choosing $P_1 = (0,\mathbf{0})$ and $P_2 = (r_0,\mathbf{0})$, the two sets of vectors

$$\left\{\frac{\mathbf{0}}{r_0}, \frac{\mathbf{v}_i}{r_i}, \quad i = 1,...,n+1\right\},$$
$$\left\{\frac{\mathbf{0}}{-r_0}, \frac{\mathbf{v}_i}{r_i - r_0}, \quad i = 1,...,n+1\right\} \tag{161}$$

are both in generic position in R^n, since they are obtained by applying an arbitrarily small perturbation to the set of vectors (158), which are by hypothesis generic in R^n.

VI. MAXIMUM NUMBER OF INDEPENDENT SYMMETRIES OF THE SYSTEM $x_i^{(r_i)} = F_i(t; x_1,..., x_1^{(r_1-1)};...;x_n,...,x_n^{(r_n-1)}), r_i > 1$

The results obtained in Secs. II–V indicate that systems of differential equations of the form

$$\mathbf{x}^{(r)} = \mathbf{F}(t,\mathbf{x},...,\mathbf{x}^{(r-1)}),$$
$$\mathbf{x} = (x_1,...,x_n), \quad r > 1, \tag{162}$$

possess a finite number of independent symmetry vectors and that the system $\mathbf{x}^{(r)} = 0$ possesses a number of independent symmetries that tends to infinity when either r or n tend to infinity, thus showing that the upper bound for the maximum number of independent symmetry vectors of (162) tends to infinity when either r or n tend to infinity.

We shall see in this section that these results hold as well for the more general class of systems of the form

$$x_i^{(r_i)} = F_i(t;\mathbf{y}),$$
$$\mathbf{y} = (x_1,...,x_1^{(r_1-1)};...;x_n,...,x_n^{(r_n-1)}),$$
$$i = 1,...,n, \quad 1 < r_i. \tag{163}$$

The restriction $r_i > 1$ for every i is essential for the validity of these results, since it is not difficult to give examples of systems of the form (163) with $r_{i_0} = 1$ for some i_0 possessing an infinite number of independent symmetries. This is what happens, for example, with "split" systems of the form

$$\dot{x}_1 = F_1(t,x_1),$$
$$x_i^{(r_i)} = F_i(t;x_2,...,x_2^{(r_2-1)};...;x_n,...,x_n^{(r_n-1)}),$$
$$r_i > 1 \quad \text{for every } i = 2,...,n, \tag{164}$$

admitting an infinite number of independent symmetries of the form

$$t' = t, \quad x_1' = x_1 + \epsilon\cdot\psi(t,x_1),$$
$$x_i' = x_i \quad \text{for every } i = 2,...,n, \tag{165}$$

where $\psi(t,x_1)$ is such that $\psi(t,x_1)\partial/\partial x_1$ is a symmetry vector of the equation

$$\dot{x}_1 = F_1(t,x_1) \tag{166}$$

[since it is well known that every first-order equation like (166) admits an infinite number of independent symmetries[10]].

A less trivial example of a differential system of the form (163) with $r_{i_0} = 1$ for some i_0 admitting an infinite number of independent symmetries is the following:

$$\dot{x} = F(t,x), \quad \ddot{y} = G(t,x). \tag{167}$$

Indeed, the necessary and sufficient condition in order that $\mathbf{S}(t,x,y) = \eta(t,x)\,\partial/\partial y$ be a symmetry vector of (167) turns out to be the following linear partial differential equation in η:

$$\eta_{tt} = -2\eta_{tx}F - \eta_{xx}F^2 - \eta_x\dot{F} \quad (\dot{F} = F_t + F_xF). \tag{168}$$

Equation (168) is Kowalewskian in the variable t, and therefore[11] possesses an infinite number of local solutions, depending on two arbitrary functions $f(x)$ and $g(x)$; for instance;

$$f(x) = \eta(0,x), \quad g(x) = \eta_t(0,x). \tag{169}$$

Therefore, the system (167) possesses an infinite number of independent symmetries, as claimed.

(a) We begin now the proof of the assertions made at the beginning of this section.

As in previous sections [II(a), (b), (c); III(a)] it is not difficult to show that, given the initial value $(t_0,\mathbf{y}_0)$, where

$$\mathbf{y}_0 = (x_{01},...,x_{01}^{(r_1-1)};...;x_{0n},...,x_{0n}^{(r_n-1)}), \tag{170}$$

and denoting by $\phi(t)$ the solution of (163) corresponding to this initial condition, for $P_1 = (t_1,\phi(t_1)) = (t_1,\mathbf{x}_1)$ sufficiently close to $P_0 = (t_0,x_{01},...,x_{0n}) = (t_0,\mathbf{x}_0)$ one can find an open neighborhood U in R^{1+n} such that through every point P of U there pass two integral curves of (163), $\gamma_1 = \{(t,\phi_1(t)) | t \in I_1 \subset R\}$ and $\gamma_2 = \{(t_1,\phi_2(t)) | t \in I_2 \subset R\}$, satisfying

$$\phi_{1i}^{(k)}(t_0) = x_{0i}^{(k)}, \quad \phi_{2i}^{(k)}(t_1) = x_{1i}^{(k)} \tag{171a}$$

for every $k = 1,...,r_i - 2$ and $i = 1,...,n$,

P is isolated in $\gamma_1 \cap \gamma_2$, (171b)

where we have set $x_{1i}^{(k)} = \phi_i^{(k)}(t_1)$.

(b) Let

$$\mathbf{S} = \varphi(\mathrm{t},\mathbf{x})\frac{\partial}{\partial t} + \sum_{i=1}^{n} \psi_i(t,\mathbf{x})\frac{\partial}{\partial x_i} \tag{172}$$

be a symmetry vector of (163). If for every $P \in U$ the two integral curves of (163), γ_1 and γ_2, defined above are invariant under $\mathbf{S}$, then P will be invariant under $\mathbf{S}$, since by (171b) P is isolated in $\gamma_1 \cap \gamma_2$, and, consequently, $\mathbf{S}$ will vanish at P for every $P \in U$, i.e., $\mathbf{S}$ will vanish identically on U. Therefore, by the arguments given in Sec. II(c), (e), to compute an upper bound for the maximum number of independent symmetry vectors of (163) it suffices to find the number of linear equations in the components of $\mathbf{S}$ that guarantee the invariance of the following linear elements of order $(r_n - 1)$ under $S^{r_n - 1}$:

$$\begin{aligned}\mathbf{z}_0 = (&t_0, x_{01},\ldots,x_{01}^{(r_1-2)},\xi_1,F_{01},\ldots,F_{01}^{(r_n-r_1)};\ldots;\\ &x_{0n},\ldots,x_{0n}^{(r_n-2)},\xi_n),\\ \mathbf{z}_1 = (&t_1, x_{11},\ldots,x_{11}^{(r_1-2)},\xi_1,F_{11},\ldots,F_{11}^{(r_n-r_1)};\ldots;\\ &x_{1n},\ldots,x_{1n}^{(r_n-2)},\xi_n)\end{aligned} \tag{173}$$

for *every* $\xi = (\xi_1,\ldots,\xi_n)$, $\quad r_n = \max_i r_i$,

where we have set

$$F_{\sigma i}^{(k} = \frac{d^k}{dt^k} F_i \Big|_{(t,x_{\sigma 1}^{(r_1-2)},\ldots,x_{\sigma 1}^{(r_1-2)},\xi_1;\ldots;x_{\sigma n},\ldots,x_{\sigma n}^{(r_n-2)},\xi_n)},$$

$$\sigma = 0,1,$$

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \sum_{i=1}^{n}\left[\dot{x}_i\frac{\partial}{\partial x_i} + \cdots + x_i^{(r_i-1}\frac{\partial}{\partial x_i^{(r_i-2}} + F_i\frac{\partial}{\partial x_i^{(r_i-1}}\right]. \tag{174}$$

[Equations (174) simply state that the derivatives of $\mathbf{F}$ appearing in (173) are to be computed along the integral curves of (163), γ_1—for $F_{0i}^{(k}$—and γ_2—for $F_{1i}^{(k}$.]

The invariance of the linear elements (173) for *every value* of ξ is in turn equivalent to the following set of linear equations in the functions φ and ψ_i:

$$\begin{aligned}&\forall\xi, \quad \varphi(P_0) = \psi_i(P_0) = \psi_i^k(\mathbf{z}_0) = 0,\\ &k = 1,\ldots,r_n - 1, \quad i = 1,\ldots,n,\end{aligned} \tag{175}$$

for $\mathbf{z}_0$, and

$$\begin{aligned}&\forall\xi, \quad \varphi(P_1) = \psi_i(P_1) = \psi_i^k(\mathbf{z}_1) = 0\\ &k = 1,\ldots,r_n - 1, \quad i = 1,\ldots,n\end{aligned} \tag{176}$$

for $\mathbf{z}_1$. At this point it is important to note that $\psi_i^k(\mathbf{z}_\sigma)$ depends on ξ not only explicitly, but also implicitly, through $F_{\sigma j}^{(p}$ $(p = 1,\ldots,r_n - r_j; j = 1,\ldots,n)$.

More precisely, taking into account the structure of ψ_i^k, given by (79) and (94), we see that $\psi_i^k(\mathbf{z}_\sigma)$ depends on the variables ξ, $F_{\sigma j}^{(p}$ *polynomially*; $\psi_i^k(\mathbf{z}_\sigma)$ is a polynomial in the variables ξ, $F_{\sigma j}^{(p}$ whose coefficients are linear combinations of the partial derivatives of the functions φ and ψ evaluated at P_σ. The *constant* values of $(t_\sigma; x_{\sigma 1},\ldots,x_{\sigma 1}^{(r_1-2)};\ldots;x_{\sigma n},\ldots,x_{\sigma n}^{(r_n-2)})$ appear in $\psi_i^k(\mathbf{z}_\sigma)$ implicitly through the variables $F_{\sigma j}^{(p}$ and explicitly as coefficients of the partial derivatives of the functions φ and ψ.

As a consequence, it immediately follows that the fulfillment of (175)–(176) is guaranteed by a *finite* number of *linear* conditions on the functions φ and ψ, namely the vanishing of $\varphi(P_\sigma)$, $\psi(P_\sigma)$ and of all the coefficients appearing in $\psi_i^k(\mathbf{z}_\sigma)$ regarded as a polynomial in the variables ξ and $F_{\sigma j}^{(p}$ $(p = 1,\ldots,r_n - r_j; j = 1,\ldots,n)$, for every value of $k = 1,\ldots,r_n - 1$, $i = 1,\ldots,n$ and for $\sigma = 0,1$. Indeed, these conditions involve only the constant values of $(t_\sigma, \mathbf{x}_\sigma; \dot{x}_{\sigma 1},\ldots,x_{\sigma 1}^{(r_1-2)};\ldots;\dot{x}_{\sigma n},\ldots,x_{\sigma n}^{(r_n-2)})$, and their linearity in φ, ψ is a direct consequence of the linearity of (175), (176) in φ and ψ.

Clearly, not all of the above conditions are independent: For instance, following the reasoning of Sec. III(c), it would be easy to verify that the vanishing of $\psi(P_1)$ and of the term independent of the variables ξ, $F_{1j}^{(p}$ in $\psi_i^k(\mathbf{z}_1)$ are a consequence of all the other conditions, and therefore this condition could be omitted. But the point here is that, at any rate, the number of the conditions obtained above is *finite*; therefore, the argument given in Sec. II(f) shows that the number of independent symmetry vectors of (163) is also finite, since it cannot exceed the number of these conditions.

(b) We shall now see that the least upper bound on the number of independent symmetries of (163) tends to infinity when either n or some of the r_i tend to infinity. Indeed, consider the system

$$x_i^{(r_i} = 0, \quad i = 1,\ldots,n. \tag{177}$$

The necessary and sufficient condition in order that (172) be a symmetry vector of (177) can be expressed as follows:

$$\psi_i^{r_i}\big|_{x_k^{(r_k}=0,\ k=1,\ldots,n} = 0, \quad i = 1,\ldots,n. \tag{178}$$

Taking into account the structure of $\psi_i^{r_i}$, given by Eq. (79), we observe that (177) admits the particular solutions

$$\varphi = 0, \quad \frac{d^{r_i}\psi_i}{dt^{r_i}}\Big|_{x_k^{(r_k}=0,\ k=1,\ldots,n} = 0. \tag{179}$$

A particular solution of (179) is the following one, dependent on $r_1 + r_2 + \cdots + r_n$ arbitrary constants:

$$\begin{aligned}&\varphi = 0, \quad \psi_i = a_i^0 + a_i^1 t + \cdots + a_i^r i^{-1} t^{r\,i-1},\\ &i = 1,\ldots,n.\end{aligned} \tag{180}$$

From (180) we obtain the following set of $r_1 + r_2 + \cdots + r_n$ independent symmetry vectors of (177):

$$\frac{\partial}{\partial x_i};\quad t\frac{\partial}{\partial x_i};\quad \cdots;\quad t^{r\,i-1}\frac{\partial}{\partial x_i}. \tag{181}$$

Since the number $r_1 + r_2 + \cdots + r_n$ evidently tends to infinity when either n or some of the r_i tend to infinity, it follows that the same thing will happen with the least upper bound on the number of independent symmetries of (163), since the least upper bound by definition is greater than or equal to the number of independent symmetries of (177), which in turn exceeds the number $r_1 + r_2 + \cdots + r_n$, as we have just shown.

VII. FINAL REMARKS

It has been shown that a system of differential equations of the type (163) can only admit an ordinary local Lie group

(i.e., a local Lie group with a *finite* number of essential parameters) of pointlike symmetries. This result precludes the possibility that a system of differential equations of this kind admit a Lie group of symmetries with an infinite number of parameters (as the formal group of locally invertible transformations of the manifold $\{(t,\mathbf{x})\} = R^{1+n}$, for instance). As is well known, this result is no longer valid when dynamical symmetries are considered (see, e.g., the paper by the authors cited in Ref. 1).

It has also been shown that a system of differential equations of the kind (70), with $r > 2$, does not admit more than $N(r,n)$ independent symmetries, where the number $N(r,n)$ satisfies the following inequalities:

$$n^2 + nr + 3 \leqslant N(r,n) \leqslant 2n^2 + nr + 2. \tag{182}$$

In addition, the system $\mathbf{x}^{(r)} = \mathbf{0}$ has exactly $n^2 + nr + 3$ independent symmetries: Therefore, it would be nice to show that, when $n > 1$, this number cannot be surpassed by the number of independent symmetries of any system of the kind (70), or, if this were not the case, to exhibit a system of this kind having more than $n^2 + nr + 3$ independent symmetries. Also open is the problem of obtaining computational algorithms for constructing systems of the kind (70) with any preassigned number of symmetries s [not exceeding the maximum number of independent symmetries allowed to every equation of the kind (70), for given n and r].

When $n = 1$, the least upper bound to the number of independent symmetries of of an equation of the kind (1) when $r > 2$ is given by the number $r + 4$, this number being equal to the number of independent symmetries of the equation $x^{(r)} = 0$ when $r > 2$.

If $r = 2$, the least upper bound to the number of independent symmetries of (70) is given by $n^2 + 4n + 3$, the number of independent symmetries of the system $\ddot{\mathbf{x}} = \mathbf{0}$. Therefore, in this case no new feature distinguishes the two cases $n > 1$ and $n = 1$, since in both cases the maximum number of symmetries is attained by the system (or equation) $\ddot{\mathbf{x}} = \mathbf{0}$ ($\ddot{x} = 0$).

It is also interesting to notice that the least upper bound to the number of independent symmetries of a system of the kind (163) tends to infinity when either n or some of the r_i tend to infinity; this result is not completely unexpected, in view of the fact that the general solution of (163) depends on $r_1 + r_2 + \cdots + r_n$ parameters.

Another interesting consequence of the previous results is that, when r is kept fixed—say $r = 2$, which is the case of Newtonian mechanics—and a certain group G of transformations of the manifold $\{(t,x_1,\ldots,x_n)\}$ depending on s parameters is given, then no equation of the form $\ddot{\mathbf{x}} = F(t,\mathbf{x},\dot{\mathbf{x}})$ can possess as many symmetries as G if $s > n^2 + 4n + 3$. But considering the action of G^N, the group of transformations of the manifold $R^{1+n} \times \overset{N}{\cdots} \times R^{1+n}$ induced by G, the possibility remains open that, for N sufficiently high, the group G^N, which also possesses s essential parameters, is a symmetry group (of generally nonpointlike transformations) of some system of the form

$$\ddot{\mathbf{x}}_i = \mathbf{F}_i(t,\mathbf{x}_1,\ldots,\mathbf{x}_N,\dot{\mathbf{x}}_1,\ldots,\dot{\mathbf{x}}_N)$$
$$i = 1,\ldots,N, \quad N > 1 \quad [x_i = (x_1,\ldots,x_{ni})]. \tag{183}$$

If this were the case for any G, we could then assert that any group of pointlike transformations of the manifold $\{(t,\mathbf{x})\} = R^{1+n}$ could be considered, when extended in the natural way to systems of more than one Newtonian particle, as a symmetry group of a system of this kind. The problem would be, of course, to find the number N appropriate for a given group G and, more importantly, the functions $\mathbf{F}_i$ appearing in (183).

Further work on these open problems is going on and will appear in forthcoming papers of this series.

ACKNOWLEDGMENTS

It is a pleasure to express our gratitude to Dr. C. Ruiz and Dr. M. Amores for useful discussions with them and for providing some bibliography. It is also a pleasure to acknowledge the constant encouragement given by M. C. Hidalgo-Brinquis.

APPENDIX

For completeness reasons, we give here some definitions concerning the projective group and a direct proof showing that this group is a symmetry group of the system $\ddot{\mathbf{x}} = \mathbf{0}$.

(i) Real $(m-1)$-dimensional projective space RP^{m-1} is usually defined as the quotient set

$$RP^{m-1} = (R^m - \{0\})/\sim \tag{A1}$$

where $\sim$ denotes the following equivalence relation:

$$\mathbf{y} \sim \mathbf{x} \Leftrightarrow \mathbf{y} = c\mathbf{x}, \qquad \mathbf{x},\mathbf{y} \in R^m - \{0\}, \quad c \in R - \{0\}. \tag{A2}$$

Therefore, the elements of RP^{m-1} are straight lines passing through the origin, with the origin removed. It is a standard result[12] that RP^{m-1} is a differentiable manifold, with the differentiable structure induced by the charts (U_i, φ_i) defined by

$$U_i = \{[\mathbf{x}] \in RP^{m-1} \mid x_i \neq 0\},$$
$$\varphi_i([\mathbf{x}]) = (x_1/x_i,\ldots,x_{i-1}/x_i,x_{i+1}/x_i,\ldots,x_m/x_i), \qquad i = 1,\ldots,m, \tag{A3}$$

where $[\mathbf{x}]$ denotes the equivalence class of $\mathbf{x} \in R^m - \{0\}$.

Geometrically, $(x_1/x_i,\ldots,x_{i-1}/x_i,1,x_{i+1}/x_i,\ldots,x_m/x_i)$ are nothing but the coordinates of the point of R^m defined by the intersection of the straight line $[\mathbf{x}]$ with the hyperplane $x_i = 1$.

Every linear nonsingular transformation $L: R^m \to R^m$ canonically induces a so-called projective transformation $\hat{L}$: $RP^{m-1} \to RP^{m-1}$ as follows:

$$\hat{L}([\mathbf{x}]) = [L\mathbf{x}] \quad \forall [\mathbf{x}] \in RP^{m-1}. \tag{A4}$$

$\hat{L}$ is well defined, since L is by hypothesis nonsingular and therefore $\mathbf{x} \in (R^m - \{0\}) \Leftrightarrow L\mathbf{x} \in (R^m - \{0\})$.

Geometrically, $\hat{L}([\mathbf{x}])$ is nothing but the straight line (with the origin removed) obtained by transforming the straight line $[\mathbf{x}]$ under L.

Let us see now what is the expression of a projective transformation in terms of the coordinates of one of the charts (A3), for instance, the chart (U_m, φ_m). If we denote by

l_{ij} $(i,j=1,...,m)$ the matrix elements of L relative to the canonical basis of R^m, then we have

$$U_m \cap \hat{L}^{-1}(U_m) = V_m, \tag{A5}$$

where V_m is the open subset of U_m (and hence of RP^{m-1}, since U_m is itself open in RP^{m-1}) defined by

$$V_m = \{[\mathbf{x}] \in RP^{m-1} \mid \mathbf{x} \notin \Pi\} \cap U_m, \tag{A6}$$

Π being the hyperplane of R^m whose equation is

$$\Pi: \quad \sum_{i=1}^{m} l_{mi} x_i = 0. \tag{A7}$$

If we denote by

$$u_i = x_i / x_m, \quad i = 1,...,m-1, \tag{A8}$$

the coordinates of $[\mathbf{x}] \in V_m$ relative to the chart (U_m, φ_m), then the coordinates of $\hat{L}([\mathbf{x}]) \in U_m$ relative to the same chart will be given by

$$u_i' = \frac{(L\mathbf{x})_i}{(L\mathbf{x})_m} = \frac{\sum_{j=1}^{m} l_{ij} x_j}{\sum_{j=1}^{m} l_{mj} x_j} = \frac{\sum_{j=1}^{m} l_{ij} u_j + l_{im}}{\sum_{j=1}^{m} l_{mj} u_j + l_{mm}}, \quad i = 1,...,m-1. \tag{A9}$$

Note that $\hat{L}$ depends on $m^2 - 1$ essential parameters since, for any $c \neq 0$, L and cL induce the same projective transformation $\hat{L}$.

From the identities

$$\hat{L}_1 \hat{L}_2 = (L_1 L_2)\hat{}, \quad (\hat{L})^{-1} = (L^{-1})\hat{} \tag{A10}$$

it follows that the set of all projective transformations forms a group, called the *projective group*: the dimension of the projective group of RP^m is, according to what has been said above, equal to $(m+1)^2 - 1 = m^2 + 2m$.

(ii) Let us show now that the system of differential equations

$$\ddot{\mathbf{x}} = \mathbf{0}, \quad \mathbf{x} = (x_1,...,x_n) \tag{A11}$$

is symmetrical under the local transformations (sufficiently close to the identity) defined by

$$x_i' = \frac{\sum_{j=1}^{n+1} l_{ij} x_j + l_{i,n+2}}{\sum_{j=1}^{n+1} l_{n+2,j} x_j + l_{n+2,n+2}} \quad \text{with } x_{n+1} = t, \tag{A12}$$

where it is understood that the point $(x_1,...,x_n,t)$ belongs to a certain open subset W of R^{n+1} such that the denominator appearing in (A12) does not vanish on W.

We can regard $(x_1,...,x_n,t)$ as the coordinates relative to the chart (U_{n+2}, φ_{n+2}) of the point $[\mathbf{y}] \in RP^{n+1}$ defined as follows:

$$[\mathbf{y}] = [(x_1,...,x_n,t,1)]. \tag{A13}$$

Similarly, we consider (A12) as the expression in the chart (U_{n+2}, φ_{n+2}) of the projective transformation $\hat{L}$ induced by the linear transformation $L: R^{n+2} \to R^{n+2}$ whose matrix elements (relative to the canonical basis of R^{n+2}) are the numbers l_{ij} $(i,j=1,...,n+2)$ appearing in (A12).

The general solution of the system (A11) is the following:

$$x_i(t) = a_i t + b_i, \quad i = 1,...,n, \; a_i, b_i \in R, \tag{A14}$$

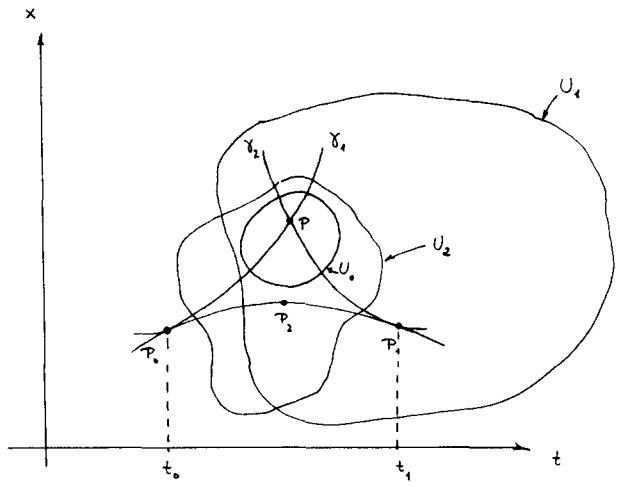


FIG. 1

which can be regarded as the implicit equation of the straight line of RP^{n+1} whose parametric equations are

$$x_i = a_i \lambda + b_i, \quad x_{n+1} = \lambda, \quad i = 1,...,n, \; \lambda \in \mathbb{R}. \tag{A15}$$

To the "straight line" (A15) there corresponds the following subset of R^{n+2}:

$$y_i = \mu(a_i \lambda + b_i), \quad y_{n+1} = \mu\lambda, \quad y_{n+2} = \mu, \quad i = 1,...,n, \quad \lambda \in R, \mu \in R - \{0\}. \tag{A16}$$

Geometrically, (A16) is obtained from (A15) as follows: for each point of the form $(a_1\lambda + b_1,...,a_n\lambda + b_n, t, 1)$ in the hyperplane $y_{n+2} = 1$ of R^{n+2}, we draw the straight line joining this point to the origin of R^{n+2}; the union of all the straight lines thus obtained with the origin removed is precisely the subset of R^{n+2} defined by (A16).

It is not difficult to verify that (A16) can be alternatively obtained from the two-dimensional subspace Π_2 or R^{n+2} defined by

$$y_i = \nu a_i + \mu b_i, \quad i = 1,...,n, \quad y_{n+1} = \nu, \quad y_{n+2} = \mu, \quad \nu, \mu \in R, \tag{A17}$$

by simply removing all the points of the straight line $r \subset \Pi_2$ given by

$$y_i = \eta a_i, \quad i = 1,...,n, \quad y_{n+1} = \eta, \quad y_{n+2} = 0, \quad \eta \in R \tag{A18}$$

(see Fig. 2).

Since L is a linear, nonsingular transformation, it transforms $\Pi_2 - r$ into $\Pi_2' - r'$, where $\Pi_2' = L(\Pi_2)$ is a two-dimensional subspace of R^{n+2} and $r' = L(r)$ is a straight line contained in Π_2'. Furthermore, since (A12) can be chosen arbitrarily close to the identity (whose parameters are given by $l_{ij} = c\delta_{ij}$, for every $c \in R - \{0\}$), it follows that $\Pi_2' - r'$ intersects the hyperplane $y_{n+2} = 1$, since this hyperplane intersects the set $\Pi_2 - r$.

The intersection of $\Pi_2' - r'$ with the hyperplane $y_{n+2} = 1$ is a straight line, whose equation we write in the form

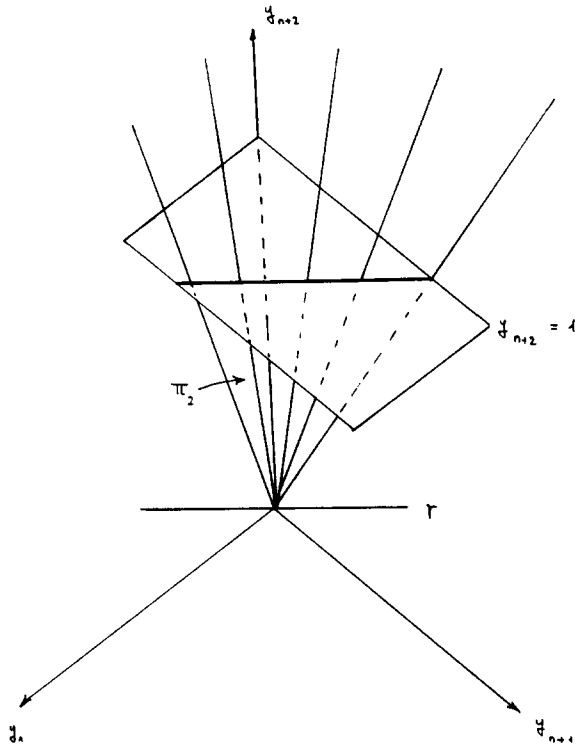


FIG. 2.

$$y_i = \alpha_i \mu + \beta_i, \quad y_{n+2} = 1,$$
$$i = 1,...,n+1, \quad \mu \in R, \tag{A19}$$

where α_i and β_i are fixed real numbers depending on a_j and b_j $(j = 1,...,n)$ and on the matrix elements l_{ij} $(i,j = 1,...,n+2)$ of L.

It follows that the equations in the chart (U_{n+2},φ_{n+2}) of the subset of RP^{n+1} obtained by applying $\hat{L}$ to the subset of RP^{n+1} whose equation—in the same chart—is (A15) are the following:

$$x_i' = \alpha_i \mu + \beta_i, \quad i = 1,...,n+1. \tag{A20}$$

Since $\hat{L}$ is arbitrarily close to the identity, and clearly $\alpha_{n+1} = 1$ when $\hat{L}$ is equal to the identity, it follows that $\alpha_{n+1} \neq 0$; therefore, we can use the $(n+1)$th equation of (A20) to solve for μ as a function of t':

$$\mu = (t' - \beta_{n+1})/\alpha_{n+1}. \tag{A21}$$

Substituting back into (A20), we see that (A20) is equivalent to the following set of equations:

$$x_i' = A_i t' + B_i, \quad i = 1,...,n, \tag{A22}$$

where

$$A_i = \frac{\alpha_i}{\alpha_{n+1}}, \quad B_i = \frac{\alpha_{n+1}\beta_i - \alpha_i \beta_{n+1}}{\alpha_{n+1}}. \tag{A23}$$

Since we have shown that by applying (A12) to an arbitrary solution (A14) of (A11) we obtain another solution of (A11), given by (A22), it follows that (A12) is a symmetry of the system (A11), as we had claimed.

ADDENDUM

We shall show here that the function $R(t,\lambda)$ defined by (8) is a C^∞ function.

Indeed, we have

$$(t - t_0)^r R(t,\lambda) = \frac{1}{(r-1)!}\int_{t_0}^{t} \phi^{(r}(s,\lambda)(t-s)^{(r-1} \, ds \tag{i}$$

by Cauchy's integral form for the remainder of the Taylor expansion of $\phi(t,\lambda)$ around $t = t_0$ (for fixed λ).[13]

Performing the change of variable,

$$s = t_0 + u(t - t_0), \quad u \in [0,1], \tag{ii}$$

we immediately obtain

$$\int_{t_0}^{t} \phi^{(r}(s,\lambda)(t-s)^{(r-1} \, ds = (t - t_0)^r \int_0^1 \phi^{(r}(t_0 + u(t-t_0),\lambda)(1-u)^{(r-1} \, du. \tag{iii}$$

Comparing (i) with (iii), we get

$$R(t,\lambda) = \frac{1}{(r-1)!}\int_0^1 \phi^{(r}(t_0 + u(t-t_0),\lambda) \times (1-u)^{(r-1} \, du, \tag{iv}$$

which is clearly of class C^k in the variables (t,λ) provided that $\phi(t,\lambda)$ is of class C^{r+k}.

Since by hypothesis $\phi(t,\lambda)$ is of class C^∞, it follows that $R(t,\lambda)$ is also of class C^∞, as claimed.

[1] F. G. Gascón, J. Math. Phys. **18**, 1763 (1977); **21**, 2046 (1980); Hadronic J. **3**, 1457 (1980). F. G. Gascón and A. G. López, Lett. Nuovo Cimento **32**, 353 (1981). For a global formulation of certain relations arising between symmetries and first integrals see the following papers: F. G. Gascón, Lett. Nuovo Cimento **20**, 54 (1977); F. G. Gascón, F. Moreno, and E. Rodríguez, Lett. Nuovo Cimento **21**, 595 (1978); F. G. Gascón, Lett. Nuovo Cimento **26**, 385 (1979); F. G. Gascón and E. Rodríguez, Lett. Nuovo Cimento **27**, 363 (1980); Hadronic J. **3**, 1059 (1980); F. A. Gascón, Phys. Lett. A **76**, 205 (1980); Lett. Nuovo Cimento **29**, 73 (1980) and the references therein. See also the following papers: F. G. Gascón, F. Moreno, and E. Rodríguez, Hadronic J. **3**, 1491 (1980); F. G. Gascón and E. Rodríguez, Phys. Lett. A **80**, 133 (1980); F. G. Gascón, Phys. Lett. A **87**, 385 (1982); Lett. Nuovo Cimento **33**, 97 (1982); Lett. Nuovo Cimento **34**, 35 (1982). See also the paper "The inverse problem concerning the Poincaré symmetry for second order differential equations" by F. G. Gascón and A. G. López (Hadronic J., accepted for publication).

[2] S. Lie and G. Scheffers, *Vorlesungen über Continuierliche Gruppen* (Chelsea, Bronx, NY, 1971). A. Cohen, *An Introduction to the Lie Theory of One-Parameter Groups* (Heath, Boston, 1911). L. E. Dickson, Ann. Math. **25**, 287 (1924).

[3] W. Fleming, *Functions of Several Variables* (Springer-Verlag, New York, 1977), p. 148.

[4] See the Addendum, following the Appendix.

[5] J. E. Campbell, *Introductory Treatise on Lie's Theory of Finite Continuous Transformation Groups* (Chelsea, Bronx, NY, 1966), pp. 23–27.

[6] See the first two papers quoted in Ref. 1.

[7] See Cohen, Ref. 2, p. 211.

[8] K. Gruenberger and A. Weis, *Linear Geometry*, University Series Undergraduate Mathematics (Van Nostrand, Princeton, NJ, 1967).

[9] V. Arnold: *Equations Differentielles Ordinaires* (Ed. MIR, Moscow, 1974), p. 222.

[10] See the first four papers quoted in Ref. 1.

[11] I. G. Petrovskii, *Partial Differential Equations* (Iliffe, London, 1967), p. 19.

[12] S. Sternberg, *Lectures on Differential Geometry* (Prentice-Hall, Englewood Cliffs, NJ, 1964). S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962). M. Spivak, *Calculus on Manifolds* (Benjamin, New York, 1965).

[13] G. Chilov, *Analyse Mathematique (Fonctions d'une Variable, Tome I)* (Ed. MIR, Moscow, 1974), p. 290.

Completely integrable relativistic Hamiltonian systems and separation of variables in Hermitian hyperbolic spaces

C. P. Boyer
IIMAS, Universidad Nacional Autonoma de México, Ap. Postal 20-726, México 20, D. F., Mexico

E. G. Kalnins
Mathematics Department, University of Waikato, Hamilton, New Zealand

P. Winternitz[a)]
Centre de Recherche de Mathématiques Appliquées, Université de Montréal, Montréal, Québec, Canada H3C 3J7

(Received 29 June 1982; accepted for publication 7 October 1982)

The Hamilton–Jacobi and Laplace–Beltrami equations on the Hermitian hyperbolic space HH(2) are shown to allow the separation of variables in precisely 12 classes of coordinate systems. The isometry group of this two-complex-dimensional Riemannian space, SU(2,1), has four mutually nonconjugate maximal abelian subgroups. These subgroups are used to construct the separable coordinates explicitly. All of these subgroups are two-dimensional, and this leads to the fact that in each separable coordinate system two of the four variables are ignorable ones. The symmetry reduction of the free HH(2) Hamiltonian by a maximal abelian subgroup of SU(2,1) reduces this Hamiltonian to one defined on an O(2,1) hyperboloid and involving a nontrivial singular potential. Separation of variables on HH(2) and more generally on HH(n) thus provides a new method of generating nontrivial completely integrable relativistic Hamiltonian systems.

PACS numbers: 02.30.Jr, 02.20. + b, 02.40. + m

I. INTRODUCTION

The purpose of this article is to discuss the separation of variables in the four (real) -dimensional Hermitian hyperbolic space HH(2) for the following two equations:

(i) The Hamilton–Jacobi equation (HJ)

$$\sum_{i,j} g^{ij} \frac{\partial S}{\partial x^i} \frac{\partial S}{\partial x^j} = E; \tag{1.1}$$

(ii) the Laplace–Beltrami equation (LB)

$$\Delta\psi = \sum_{i,j} \frac{1}{\sqrt{g}} \frac{\partial}{\partial x^i} \sqrt{g}\, g^{ij} \frac{\partial}{\partial x^j} = \lambda\psi\,. \tag{1.2}$$

In a previous paper[1] (further to be referred to as I) we have considered the separation of variables in complex projective spaces $\mathbb{C}P(n)$. The isometry group of $\mathbb{C}P(n)$ is the compact group SU($n+1$), and its Cartan subgroup was used to generate n ignorable variables and to reduce the problem of variable separation on $\mathbb{C}P(n)$ to the separation of variables on the real sphere S^n. We refer to this paper for a discussion of the motivation and for some historical background.

Here let us just mention the relation between separation of variables in the HJ equation and complete integrability of the corresponding Hamiltonian system. Indeed, separability for the HJ equation is defined to mean that a solution S of (1.1) exists satisfying

$$S = \sum_i S_i(x^i,\lambda_1,\ldots,\lambda_n), \quad \det \frac{\partial^2 S}{\partial x^i \partial \lambda_j} \neq 0, \tag{1.3}$$

where λ_i are n constants: the separation constants. We associate n second-order operators in involution with each separable coordinate system in an n-dimensional space (one of them is the Hamiltonian); the constants λ_i are the eigenvalues of these operators. The existence of these operators assures that the system is integrable.

For studies of the separation of variables in Hamilton–Jacobi equations on Riemannian and pseudo-Riemannian manifolds, see also Refs. 2–5.

The additive separation of variables (1.3) in the HJ equation corresponds to multiplicative separation in the LB equation (1.2):

$$\psi = \prod_i \psi_i(x^i,\lambda_1,\ldots,\lambda_n). \tag{1.4}$$

Indeed, for Einstein spaces every coordinate system that separates the HJ equation will also separate the LB equation[2–4] (the converse is always true). Separation of variables in LB equations makes it possible to use powerful methods of group theory to study broad classes of special functions.[5–9]

II. THE SPACE HH(n) AND ITS ISOTROPY GROUP SU(n,1)

We introduce the Hermitian hyperbolic (or complex hyperbolic) space HH(n) following Kobayashi and Nomizu[10] and Helgason.[11] Let $(e_0,e_1,\ldots,e_n)$ be a standard basis in $\mathbb{C}^{n+1}$ and consider the Hermitian form

$$F(x,y) = -\bar{x}_0 y_0 + \sum_{k=1}^{n} \bar{x}_k y_k, \tag{2.1}$$

where the overbar denotes complex conjugation. This form is invariant under the action of the group U(n,1):

$$g \in U(n,1), \quad F(gx,gy) = F(x,y), \quad x,y \in \mathbb{C}^{n+1}, \tag{2.2}$$

which acts transitively on the real hypersurface M in $\mathbb{C}^{n+1}$ defined by

$$F(y,y) = -1. \tag{2.3}$$

[a)] Research supported in part by the Natural Science and Engineering Research Council of Canada and by the Fonds FCAC pour l'aide et le soutin à la recherche du Gouvernement du Québec."

 0022-2488/83/082022-13$02.50

The group $U(1) = \{e^{i\theta}\}$ acts freely on this manifold by $y \to e^{i\theta} y$. The space of orbits with suitable complex manifold structure and Kaehler metric is identified as $HH(n)$. The corresponding natural projection $\pi: M \to HH(n)$ defines a principal bundle with $U(1)$ as structure group. The $U(n,1)$ action commutes with that of $U(1)$, and it hence projects to an action on the base $HH(n)$. The isotropy subgroup of $U(n,1)$ at the point $p_0 = \pi(e_0)$ is $U(1) \times U(n)$, and we obtain the diffeomorphism

$$U(n,1)/[U(n) \times U(1)] \sim HH(n). \tag{2.4}$$

The group $SU(n,1)$ acts almost effectively on this space.

In addition to the homogeneous coordinates $\{y_0, y_1, \ldots, y_n\}$, let us introduce affine coordinates on $HH(n)$:

$$\pi(y_0, y_1, \ldots, y_n) = (z_1, \ldots, z_n), \quad z_k = y_k / y_0, \quad k = 1, \ldots, n. \tag{2.5}$$

The space $HH(n)$ can then be identified with an open unit ball in $\mathbb{C}^n$

$$z \in \mathbb{C}^n, \quad \sum_{k=1}^{n} \bar{z}_k z_k < 1. \tag{2.6}$$

The real part of the Hermitian form (2.1) determines in a natural manner a metric on $HH(n)$, which is the noncompact version of the well-known Fubini-Study metric[10]:

$$ds^2 = -\frac{4}{c} \times \frac{\left(1 - \sum \bar{z}_k z_k\right)\left(\sum d\bar{z}_k dz_k\right) + \left(\sum \bar{z}_k dz_k\right)\left(\sum z_k d\bar{z}_k\right)}{\left(1 - \sum \bar{z}_k z_k\right)^2}, \tag{2.7}$$

where $c < 0$ is the (constant) holomorphic sectional curvature.

We now limit ourselves to the case under consideration, namely $n = 2$.

The Hamiltonian associated with the metric (2.7) for $n = 2$ ($c = -4$) is

$$H = 4(1 - |z_1|^2 - |z_2|^2)[(|z_1|^2 - 1) p_1 \bar{p}_1 + (|z_2|^2 - 1) \times p_2 \bar{p}_2 + z_1 \bar{z}_2 p_1 \bar{p}_2 + \bar{z}_1 z_2 \bar{p}_1 p_2]. \tag{2.8}$$

The Lie algebra $u(2,1)$ in the representation acting on the homogeneous coordinates (y_0, y_1, y_2) is realized by 3×3 complex matrices X satisfying

$$X^+ J + JX = 0, \quad J = \begin{pmatrix} -1 & & \\ & 1 & \\ & & 1 \end{pmatrix} \tag{2.9}$$

(the superscript $+$ denotes Hermitian conjugation).

Two convenient bases are given by the matrices X_i, or alternatively Y_i, $i = 0,1,\ldots,8$:

$$X_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} = Y_7, \quad X_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & i \\ 0 & i & 0 \end{pmatrix} = Y_8,$$

$$X_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & i & 0 \\ 0 & 0 & -i \end{pmatrix} = \frac{Y_1 + Y_6}{2},$$

$$X_4 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = Y_5, \quad X_5 = \begin{pmatrix} 0 & i & 0 \\ -i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = Y_6 - Y_4,$$

$$X_6 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} = Y_7 - Y_2,$$

$$X_7 = \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix} = Y_8 - Y_3,$$

$$X_8 = \frac{i}{\sqrt{3}} \begin{pmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} = \frac{Y_1 - 3Y_6}{2\sqrt{3}},$$

$$X_0 = \begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix} = Y_0 \tag{2.10}$$

[the Y_i basis is particularly appropriate for considering solvable subalgebras of $su(2,1)$].

With these conventions the second order Casimir operator of $su(2,1)$ can be written as

$$C_2 = X_1^2 + X_2^2 + X_3^2 - X_4^2 - X_5^2 - X_6^2 - X_7^2 + X_8^2 = \tfrac{1}{3} Y_1^2 - Y_2^2 - Y_3^2 - Y_4^2 - Y_5^2 + \{Y_2, Y_7\} + \{Y_3, Y_8\} + \{Y_4, Y_6\}, \tag{2.11}$$

where $\{\ ,\ \}$ denotes the anticommutator.

A Killing vector L on the cotangent bundle with local coordinates $(z_i, \bar{z}_i, p_i, \bar{p}_i, i = 1,2)$ is a linear polynomial in p_i, $\bar{p}_i$:

$$L = \sum_i c_i(z_1, z_2, \bar{z}_1, \bar{z}_2)\, p_i + \text{c.c.} \tag{2.12}$$

(where c.c. indicates the complex conjugate quantity), such that

$$[H, L]_P = 0, \tag{2.13}$$

i.e., the Poisson bracket of H with L is zero. The Killing vectors for $HH(2)$ provide a realization of the algebra $su(2,1)$. Using the basis X_i ($i = 1,\ldots,8$) of (2.10) for the infinitesimal operators, we calculate the corresponding Killing vectors in affine and homogeneous coordinates to be, respectively,

$$X_1 = -z_2 p_{z_1} + z_1 p_{z_2} + \text{c.c.} = -y_2 p_{y_1} + y_1 p_{y_2} + \text{c.c.},$$

$$X_2 = -i(z_2 p_{z_1} + z_1 p_{z_2}) + \text{c.c.} = i(y_2 p_{y_1} + y_1 p_{y_2}) + \text{c.c.},$$

$$X_3 = i(-z_1 p_{z_1} + z_2 p_{z_2}) + \text{c.c.} = i(y_1 p_{y_1} - y_2 p_{y_2}) + \text{c.c.},$$

$$X_4 = (z_1^2 - 1) p_{z_1} + z_1 z_2 p_{z_2} + \text{c.c.} = y_1 p_{y_0} + y_0 p_{y_1} + \text{c.c.},$$

$$X_5 = i[(z_1^2 + 1) p_{z_1} + z_1 z_2 p_{z_2}] + \text{c.c.} = i(-y_1 p_{y_0} + y_0 p_{y_1}) + \text{c.c.},$$

$$X_6 = z_1 z_2 p_{z_1} + (z_2^2 - 1) p_{z_2} + \text{c.c.} = y_2 p_{y_0} + y_0 p_{y_2} + \text{c.c.},$$

$$X_7 = i[z_1 z_2 p_{z_1} + (z_2^2 + 1) p_{z_2}] + \text{c.c.}$$
$$= i(-y_2 p_{y_0} + y_0 p_{y_2}) + \text{c.c.},$$
$$X_8 = i\sqrt{3}(z_1 p_{z_1} + z_2 p_{z_2}) + \text{c.c.}$$
$$= (i/\sqrt{3})(2y_0 p_{y_0} - y_1 p_{y_1} - y_2 p_{y_2}) + \text{c.c.} \tag{2.14}$$

Throughout we shall make use of the moment map; whenever convenient we use the operators $\partial/\partial z_i$ or $\partial/\partial y_\mu$ instead of the functions P_{z_i} or P_{y_μ} and commutator brackets instead of Poisson brackets.

III. SUBGROUPS OF SU(2,1) AND COMPLETE SETS OF COMMUTING SECOND-ORDER OPERATORS

According to the operator approach to the separation of variables,[6–9] each separable system on HH(2) will be characterized by four second-order operators $\{H, T_1, T_2, T_3\}$ that are in involution with respect to the appropriate Lie bracket (one of them being the Hamiltonian H, or correspondingly the Laplace operator Δ). The first task is to classify the triplets of operators $\{T_1, T_2, T_3\}$ into equivalence classes under the action of the group SU(2,1), leaving H invariant.

The task in the present case of HH(2) is greatly simplified by two circumstances:

(1) It has recently been shown[12] that for HH(2) all second-order Killing tensors, i.e., operators

$$T = \sum_{i,k=1}^{2} \{c_{ik}(z_1, \bar{z}_1, z_2, \bar{z}_2) p_1 p_k + d_{ik}(z_1, \bar{z}_1, z_2, \bar{z}_2) p_i \bar{p}_k + \text{c.c.}\}, \tag{3.1}$$

satisfying

$$[T, H] = 0 \tag{3.2}$$

lie in the enveloping algebra of su(2,1). Each of the operators T_i can hence be written in the form

$$T_i = \sum_{a,b=1}^{8} A^i_{ab} X_a X_b, \quad A^i_{ab} = A^i_{ba} \in \mathbb{R}. \tag{3.3}$$

(2) We have shown in I, Theorem 4, that every separable coordinate system in $\mathbb{C}P(2)$ and HH(2) has precisely two ignorable variables. We recall that an ignorable variable in a certain coordinate system is a variable that does not figure in the metric tensor g_{ik} expressed in this system.[4] An ignorable variable ϕ is obtained by setting a Killing vector, say L_1, equal to the momentum p_ϕ canonically conjugate to ϕ. The square of this Killing vector is then a second-order Killing tensor

$$T_1 = L_1^2 = p_\phi^2. \tag{3.4}$$

This can be done[13] if the corresponding Killing tensor T_1 is the square of a Killing vector, i.e., in our case the square of an element of su(2,1). Since two variables must be ignorable in each separable coordinate system, it follows that two of the operators T_i, say T_1 and T_2, must be squares of elements of su(2,1):

$$T_1 = L_1^2 = \left(\sum_{\alpha=1}^{8} a_\alpha X_\alpha\right)^2,$$
$$T_2 = L_2^2 = \left(\sum_{\alpha=1}^{8} b_\alpha X_\alpha\right)^2. \tag{3.5}$$

Since T_1 and T_2 commute, the operators L_1 and L_2 must generate an abelian subalgebra of su(2,1). All subalgebras of su[2,1) are known,[14] and work is in progress on the classification of the maximal abelian subalgebras (MASA's) of all classical Lie algebras.[15,16] In particular, su(2,1) has four different MASA's [each representing a conjugacy class of MASA's under the action of SU(2,1)]. Each of them is two-dimensional.

The procedure of finding all triplets of operators $\{T_1, T_2, T_3\}$ related to separable coordinates on HH(2) thus reduces to the following:

(i) Take T_1 and T_2 as in (3.5), where L_1 and L_2 run through the four different MASA's of su(2,1).

(ii) For each MASA L_1, L_2, find the most general operator $Q = T_3 \in S^2(\text{su}(2,1))$ [second-order symmetric tensor in the enveloping algebra of su(2,1)] commuting with L_1 and L_2. The operator T_3 has the form (3.3).

(iii) Simplify each T_3 by linear combinations, with L_1^2, L_2^2, $L_1 L_2$ and C_2 (2.11) and classify the operators T_3 into conjugacy classes under the action of the normalizer of $\{L_1, L_2\}$ in SU(2,1) (the normalizer is the group of transformations leaving the algebra $\{L_1, L_2\}$ invariant).

A particularly important and simple class of coordinates are called "subgroup type coordinates,"[5,7,8,13] and they occur when T_3 is the Casimir operator of a subgroup of SU(2,1).

In Fig. 1 we show all subalgebras of su(2,1) that are relevant for our purposes (for a complete classification see Ref. 14). The basis elements $\{X_\alpha\}$ and $\{Y_\alpha\}$ are defined in (2.10), we use the two bases interchangeably. The lowest row in Fig. 1 is occupied by the four MASA's: $\{X_3, X_8\}$ and $\{Y_1, Y_6 - Y_4\}$ are the compact and noncompact Cartan subalgebras, respectively, $\{Y_1, Y_4\}$ contains a nilpotent element

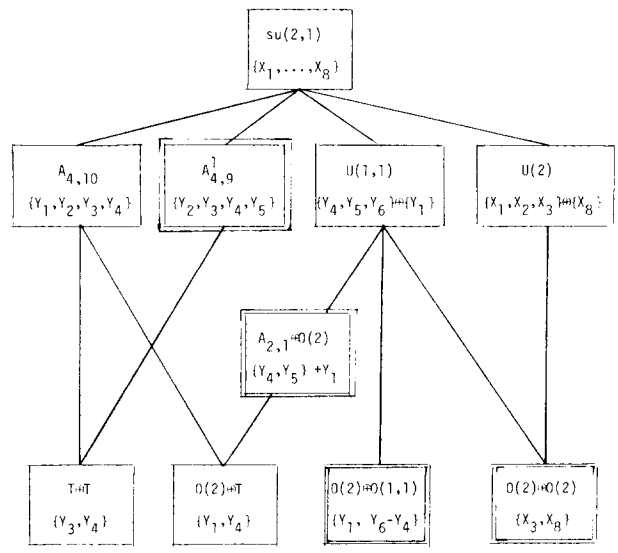


FIG. 1. Maximal abelian subalgebras of su(2,1) and some subalgebras containing them. The basis elements X_i and Y_i are defined in (2.10). The four MASA's constitute the lowest row, and double boxes indicate their normalizers; $A_{4,10}$, $A^1_{4,9}$, and $A_{2,1}$ are solvable algebras, and T denotes a translation type subalgebra.

Y_4 (Y_4 is represented by a nilpotent matrix in any finite-dimensional representation). All elements of $\{Y_2, Y_4\}$ are nilpotent, i.e., this is a maximal abelian nilpotent subalgebra (MANS).[15,16] The letter T in the boxes denotes the presence of such nilpotent elements ("translations" on a light cone). The double boxes indicate normalizers of the MASA's. By definition, Cartan subalgebras are self-normalizing. A classification of all real Lie algebras of dimension $d \leqslant 5$ exists[17]; the notation $A_{4,10}$, $A^{1}_{4,9}$, $A_{2,1}$ refers to that article. The algebras $A_{4,10}$, U(1,1), and u(2) are the only subalgebras of su(2,1) [up to conjugacy under SU(2,1)] containing at least one MASA and having a second-order Casimir operator.

These algebras and their Casimir operators play an important role below; so let us discuss them in more detail.

(1) The su(2) subalgebra of u(2) is $\{X_1, X_2, X_3\}$ and its Casimir operator is

$$I(\mathrm{su}(2)) = X_1^2 + X_2^2 + X_3^2. \tag{3.6}$$

(2) Two mutually conjugate su(1,1) subalgebras and their Casimir operators are

$$\{X_4, X_5, \tfrac{1}{2}(X_3 - \sqrt{3}\, X_8)\} \sim \{Y_4, Y_5, Y_6\}, \tag{3.7}$$

$$I_1(\mathrm{su}(1,1)) = X_4^2 + X_5^2 - \tfrac{1}{4}(X_3 - \sqrt{3}X_8)^2$$

and

$$\{X_6, X_7, \tfrac{1}{2}(X_3 + \sqrt{3}X_8)\},$$

$$I_2(\mathrm{su}(1,1)) = X_6^2 + X_7^2 - \tfrac{1}{4}(X_3 + \sqrt{3}X_8)^2. \tag{3.8}$$

(3) The solvable algebra $A_{4,10}$:

$$\{Y_1, Y_2, Y_3, Y_4\} \sim \{X_3 + (1/\sqrt{3})X_8, \; -X_5 + \tfrac{1}{2}(X_3 - \sqrt{3}X_8), X_1 - X_6, X_2 - X_7\}.$$

Its invariant is[14]

$$I_{4,10} = 4Y_1Y_4 + 3(Y_2^2 + Y_3^2). \tag{3.9}$$

Notice that one realization of $A_{4,10}$ is related to the one-dimensional harmonic oscillator. If we put

$$Y_1 \equiv \frac{3}{2}\left(\frac{\partial^2}{\partial x^2} + x^2\right), \quad Y_2 = x, \quad Y_3 = \frac{\partial}{\partial x}, \quad Y_4 = \tfrac{1}{2},$$

then the commutation relations for Y_i are satisfied, and we have $I_{4,10} = 4Y_1$.

Let us now return to the classification of triplets of operators outlined above.

A. The compact Cartan subalgebra

We have

$$T_1 = X_3^2, \quad T_2 = X_8^2,$$

and $[T_1, T_3] = 0$, $[T_2, T_3] = 0$ implies

$$Q_{\mathrm{I}} = T_3 = aI(\mathrm{su}(2)) + bI_1(\mathrm{su}(1,1)) + cI_2(\mathrm{su}(1,1)). \tag{3.10}$$

The Cartan subalgebras are self-normalizing; hence the only freedom left is to subtract some multiple of C_2. The following possibilities occur:

(1) $b = c \neq -a$: $Q_1 = I(\mathrm{su}(2))$,
(2) $c = -a \neq b$: $Q_2 = I_1(\mathrm{su}(1,1))$,
(3) $b \neq c \neq -a \neq b$: $Q_{3,4} = I_1(\mathrm{su}(1,1)) + \mu I_2(\mathrm{su}(1,1))$,
Q_3: $0 < \mu < 1$, Q_4: $-1 \leqslant \mu < 0$,

(the case $|\mu| > 1$ can be rotated into one of the cases with $|\mu| \leqslant 1$).

B. The noncompact Cartan subalgebra

$$T_1 = \tfrac{9}{4}[X_3 + (1/\sqrt{3})X_8]^2, \quad T_2 = X_5^2,$$

$$Q_{\mathrm{II}} = T_3 = aI_1(\mathrm{su}(1,1)) + b(X_1X_6 + X_6X_1 + X_2X_7 + X_7X_2). \tag{3.11}$$

Two possibilities should be distinguished:

Q_5: $b = 0$, $a = 1$,
Q_6: $b = 1$, $a \geqslant 0$

(the relative sign of b and a can be changed by a rotation through the angle π, hence the restriction $a \geqslant 0$ in Q_6).

C. The MASA $\{Y_1, Y_4\}$

$$T_1 = Y_1^2 = \tfrac{9}{4}[X_3 + (1/\sqrt{3})X_8]^2,$$

$$T_2 = Y_4^2 = (-X_5 + \tfrac{1}{2}X_3 - \tfrac{1}{2}\sqrt{3}\, X_8)^2,$$

$$Q_{\mathrm{III}} = T_3 = aI_{4,10} + bI_1(\mathrm{su}(1,1)). \tag{3.12}$$

Four possibilities occur:

Q_7: $a = 0$, $b = 1$,
Q_8: $a = 1$, $b = 0$,
Q_9: $a = b = 1$,
Q_{10}: $a = -b = 1$.

Indeed, if $ab \neq 0$, we make use of the external part of the normalizer of $\{Y_1, Y_4\}$, namely the operator Y_5 to scale a with respect to b: For $ab > 0$ we can scale so that we get $a = b$; for $ab < 0$ so that we get $a = -b$.

D. The maximal abelian nilpotent subalgebra

$$T_1 = Y_3^2 = (X_2 - X_7)^2,$$

$$T_2 = (Y_4)^2 = (-X_5 + \tfrac{1}{2}X_3 - \tfrac{1}{2}\sqrt{3}\, X_8)^2,$$

$$Q_{\mathrm{IV}} = T_3 = aI_{4,10} + b[Y_1Y_3 + Y_3Y_1 - 3(Y_2Y_5 + Y_5Y_2) - 6(Y_4Y_8 + Y_8Y_4)]. \tag{3.13}$$

Two cases should be distinguished:

Q_{11}: $a = 1$, $b = 0$,
Q_{12}: $a = 0$, $b = 1$.

Indeed, if $a \neq 0$, we set $a = 1$ and use the external part of the normalizer of $\{Y_3, Y_4\}$ to transform $b \to 0$ [this is achieved by a transformation of the type $Q' = \exp(\alpha Y_2) Q \exp(-\alpha Y_2)$].

We have thus obtained 12 orbits of operators $\{T_1, T_2, T_3\}$. Among them six are of the subgroup type, i.e, such that Q is the Casimir operator of some subgroup of SU(2,1). These are the sets involving Q_1, Q_2, Q_5, Q_7, Q_8, and Q_{11}.

In the following section we shall establish a one-to-one correspondence between the above-classified triplets of operators in involution and 12 types of separable coordinates on HH(2).

IV. SEPARABLE COORDINATES ON HH (2)

A. Introduction of ignorable coordinates and reduction to separation on an O(2,1) hyperboloid

Our purpose now is to find all separable coordinates in HH(2), i.e., to transform from the affine coordinates $\{z_1,z_2,\bar{z}_1,\bar{z}_2\}$ to four real variables $\{A,B,x,y\}$ such that x and y are ignorable and that Eqs. (1.1) and (1.2) separate in the new variables. This transformation can be performed in two different manners, starting with the affine coordinates z_i $(i=1,2)$ or the homogeneous coordinates y_μ $(\mu=0,1,2)$, respectively. In each case the procedure is repeated four times, separately for each MASA of su(2,1).

Using affine coordinates, we proceed as follows:

(1) Choose a basis $\{L_1,L_2\}$ for the considered MASA, express L_1 and L_2 in terms of z_i as in (2.14) and put

$$L_1=P_x,\quad L_2=P_y. \tag{4.1}$$

Solve equations (4.1): This provides the explicit dependence of z_1 and z_2 on the ignorable variables. The dependence on the essential variables A, B is as yet unknown and is contained in the integration "constants" of (4.1).

(2) To obtain the dependence on A, B make use of a procedure outlined in Ref. 4, for arbitrary four-dimensional Riemannian spaces. Since HH(2) is a positive-definite metric space and since each separable system must involve precisely two ignorable variables, only case "C" of Ref. 4 occurs. Hence a pseudogroup P of coordinate transformations (described in I and Ref. 4) must exist, transforming the Fubini–Study metric (2.7) into a form in which the metric tensor satisfies:

$$\begin{aligned}
&g^{AA}=g^{BB}=\frac{1}{k_1(A)+k_2(B)},\quad g^{AB}=0,\\
&g^{xx}=\frac{e_1(A)+e_2(B)}{k_1(A)+k_2(B)},\quad g^{yy}=\frac{f_1(A)+f_2(B)}{k_1(A)+k_2(B)},\\
&g^{xy}=\frac{h_1(A)+h_2(B)}{k_1(A)+k_2(B)},\\
&g^{Ax}=g^{Ay}=g^{Bx}=g^{By}=0,
\end{aligned} \tag{4.2}$$

where k_i,e_i,f_i, and h_i are functions of the indicated variables satisfying

$$\frac{\partial^2}{\partial A\partial B}\ln\left[\frac{(k_1+k_2)^2}{(e_1+e_2)(f_1+f_2)-(h_1+h_2)^2}\right]=0 \tag{4.3}$$

[i.e., $R_{AB}=0$, where R_{ij} is the Ricci tensor]. Solve Eqs. (4.2) and (4.3) to obtain the dependence of $\{z_1,z_2\}$ on A and B.

Following this procedure, we find that the MASA $\{X_3,X_8\}$ leads to four different types of coordinates, $\{Y_1,Y_6-Y_4\}$ to two types, $\{Y_1,Y_4\}$ to four types, and finally $\{Y_3,Y_4\}$ to two. The computations are quite long and involved, but the results are relatively simple and coincide with those obtained using a different, more geometrical and group-theoretical method, described below.

The second procedure is an adaptation of the general method of the reduction of phase space in classical mechanics by ignorable variables.[18] The procedure is related to that used by Marsden and Weinstein[19] and Kazhdan, Kostant, and Sternberg[20] to obtain completely integrable Hamiltonian systems. In I we applied this procedure to reduce by the maximal torus, i.e., the Cartan subgroup of SU$(n+1)$. We thus reduced the problem of separating variables on CP(n) to that of separating on the sphere S_{n+1}. The free Hamiltonian on CP(n) was reduced to a singular Hamiltonian on S_{n+1} with a specific inverse square type potential. We shall see that the situation is very similar for HH(2) and that the reduction can be performed by any of the maximal abelian subgroups (not just the maximal torus).

Instead of MASA's of su(2,1), we shall use MASA's of u(2,1), i.e., to the basis L_1,L_2 of each MASA we add a further operator

$$X_0=y_0p_{y_0}+y_1p_{y_1}+y_2p_{y_2}+\text{c.c.} \tag{4.4}$$

When acting on functions $f(y_0,y_1,y_2)$ that project properly onto HH(2), i.e., homogeneous functions satisfying

$$f(y_0,y_1,y_2)=f(y_1/y_0,y_2/y_0) \tag{4.5}$$

we have

$$\left(y_0\frac{\partial}{\partial y_0}+y_1\frac{\partial}{\partial y_1}+y_2\frac{\partial}{\partial y_2}\right)f=0 \tag{4.6}$$

and for the corresponding constant of motion on HH(2) we have

$$X_0=y_0p_{y_0}+y_1p_{y_1}+y_2p_{y_2}=0. \tag{4.7}$$

The procedure is:

(1) Choose a basis $\{L_1,L_2\}$ for the considered MASA, express L_1, L_2, and X_0 in terms of y as in (2.14) and (4.4) and put

$$L_1=p_x,\quad L_2=p_y,\quad X_0=p_\rho. \tag{4.8}$$

Solve equations (4.8) to obtain the explicit dependence of y on the ignorable variables x, y, and ρ [upon projection from C(3) to HH(2) ρ will cancel out].

The variables y_μ depend on three more real variables, say s_0, s_1, and s_2, which are contained in the integration constants of Eqs. (4.8). These must be introduced in such a manner that s_μ, x, y, and ρ parametrize all of C(3), that x and y project into ignorable variables on HH(2), and that the variables s_μ are compatible with the projection, i.e.,

$$\begin{aligned}
|y|^2\equiv|y_0^2|-|y_1|^2-|y_2|^2&=s_0^2-s_1^2-s_2^2\\
&\equiv s^2=\text{const.}
\end{aligned} \tag{4.9}$$

In order to obtain the space HH(2), we put $s^2=1$; other homogeneous spaces with SU(2,1) actions are obtained by putting $s^2=-1$ or $s^2=0$.

(2) Express the su(2,1) infinitesimal operators X_i $(i=1,\ldots,8)$ the Hamiltonian H and the Killing tensor $T_3=Q$ in terms of the variables $\{x,y,s_0,s_1,s_2\}$ (setting $p_\rho=0$, or correspondingly dropping a term containing $\partial/\partial\rho$). The essential variables s_μ are constrained by the condition (4.9). The corresponding momenta p_{s_μ} figure in the infinitesimal operators X_i only via the expressions

$$\begin{aligned}
&I_{12}=s_1p_{s_2}-s_2p_{s_1},\quad I_{01}=s_0p_{s_1}+s_1p_{s_0},\\
&I_{02}=s_0p_{s_2}+s_2p_{s_0}.
\end{aligned} \tag{4.10}$$

The quantities $I_{\mu\nu}$ $(\mu,\nu=0,1,2)$ generate an o(2,1) algebra under the corresponding Lie bracket. This o(2,1) is in general not a subalgebra of su(2,1); however, if we restrict ourselves to the manifold (4.9) by setting the ignorable variables equal

to zero, then we obtain

$$X_1 = I_{12}, \quad X_4 = I_{01}, \quad X_6 = I_{02}, \tag{4.11}$$

i.e., the O(2,1) group acting on the variables s coincides with the real O(2,1) subgroup of SU(2,1). In the new variables the Hamiltonian H and the Killing tensor Q are expressed as

$$H = I_{12}^2 - I_{01}^2 - I_{02}^2 + f_1(s_\mu)p_x^2 + f_2(s_\mu)p_y^2 + f_3(s_\mu)p_x p_y, \tag{4.12}$$

$$Q = \sum_{\substack{\mu\nu \\ \mu'\nu'}} A_{\mu\nu,\mu'\nu'} I_{\mu\nu} I_{\mu'\nu'} + h_1(s_\mu)p_x^2 + h_2(s_\mu)p_y^2 + h_3(s_\mu)p_x p_y, \tag{4.13}$$

where f_i and h_i are functions of the essential variables s_μ and $A_{\mu\nu,\mu'\nu'} = A_{\mu'\nu',\mu\nu}$ is a symmetric constant matrix. The problem of separating variables for the free Hamiltonian on HH(2) has thus been reduced to that of separating variables in the Hamiltonian (4.12). This is an O(2,1) Hamiltonian, which is, however, not a free one: It includes a "potential" term depending on the O(2,1) variables s_μ. We recall that the momenta p_x and p_y corresponding to the ignorable variables should be set equal to constants

$$p_x = c_1, \quad p_y = c_2. \tag{4.14}$$

Notice that we have

$$I_{12}^2 - I_{01}^2 - I_{02}^2 = (p_{s_0}^2 - p_{s_1}^2 - p_{s_2}^2), \tag{4.15}$$

where we have used the fact that

$$\sum_{\mu=0}^{2} s_\mu p_{s_\mu} = 0. \tag{4.16}$$

(3) Introduce separable coordinates on the hyperboloid (4.9), compatible with the form of the operator Q and the potential in (4.12).

Let us now implement the first two steps of this procedure for each of the four MASA's of su(2,1).

1. The compact Cartan subalgebra $\{X_3, X_8\}$

We first introduce the ignorable variables (ρ,α_1,α_2), putting

$$\tfrac{1}{2}[X_3 - (1/\sqrt{3})X_8] = p_{\alpha_1}, \quad -\tfrac{1}{2}[X_3 + (1/\sqrt{3})X_8] = p_{\alpha_2}, \quad X_0 = p_\rho. \tag{4.17}$$

Using (2.14), we obtain a system of equations that is easily solved to express the homogeneous coordinates as

$$y_0 = s_0 e^{i(3\rho - \alpha_1 - \alpha_2)/3}, \quad y_1 = s_1 e^{i(3\rho + 2\alpha_1 - \alpha_2)/3}, \quad y_2 = s_2 e^{i(3\rho - \alpha_1 + 2\alpha_2)/3}. \tag{4.18}$$

The infinitesimal operators are expressed in these coordinates in the Appendix. Putting $\alpha_1 = \alpha_2 = 0$, we obtain (4.11); X_2, X_3, X_5, X_7, X_8 then involve only the essential variables and the momenta conjugate to the ignorable ones. Expressions (4.12) and (4.13) for the Hamiltonian H and Killing tensor $Q_{\rm I}$ (3.10) reduce to

$$H = -I_{12}^2 + I_{01}^2 + I_{02}^2 + \left[\frac{1}{s_1^2}p_{\alpha_1}^2 + \frac{1}{s_2^2}p_{\alpha_2}^2 - \frac{1}{s_0^2}(p_{\alpha_1} + p_{\alpha_2})^2\right], \tag{4.19}$$

$$Q_{\rm I} = a\left[I_{12}^2 + \left(1 + \frac{s_2^2}{s_1^2}\right)p_{\alpha_1}^2 + \left(1 + \frac{s_1^2}{s_2^2}\right)p_{\alpha_2}^2\right] + b\left[I_{01}^2 + \left(-1 + \frac{s_0^2}{s_1^2}\right)p_{\alpha_1}^2 + \left(-1 + \frac{s_1^2}{s_0^2}\right)(p_{\alpha_1} + p_{\alpha_2})^2\right] + c\left[I_{02}^2 + \left(-1 + \frac{s_0^2}{s_2^2}\right)p_{\alpha_2}^2 + \left(-1 + \frac{s_2^2}{s_0^2}\right) \times (p_{\alpha_1} + p_{\alpha_2})^2\right]. \tag{4.20}$$

Setting $p_{\alpha_i} = 0$ we obtain a free O(2,1) Hamiltonian and a Killing tensor of a specific type: it involves the squares $I_{\mu\nu}^2$ only. Separation of variables on an O(2,1) hyperboloid H_2 is discussed below.[5,7,21] Nine distinct separable coordinate systems exist on H_2 but only four of them have Killing tensors of the type $Q_{\rm I}$. Precisely these four occur in our HH(2) problem.

Setting $p_{\alpha_i} = c_i \neq 0$, we reduce (4.19) to an O(2,1) Hamiltonian with an inverse square type singular potential, and $Q_{\rm I}$ reduces to the corresponding integral of motion. We have thus generated a nontrivial relativistic completely integrable Hamiltonian system. Similar systems with singular inverse square potentials have been studied in a nonrelativistic context.[22–24]

2. The noncompact Cartan subalgebra $\{X_3 + (1/\sqrt{3})X_8, X_5\}$

Introduce the ignorable variables (ρ,α,u) by putting

$$-\tfrac{1}{2}[X_3 + (1/\sqrt{3})X_8] = p_\alpha, \quad X_5 = p_u, \quad X_0 = p_\rho. \tag{4.21}$$

Expressing X_i in terms of the homogeneous coordinates y_μ, we obtain a system of partial differential equations that can be solved to yield

$$y_0 = e^{i(3\rho - \alpha)/3}(is_0 chu + s_1 shu), \quad 0 \leqslant \rho < 2\pi, \quad 0 \leqslant \alpha < 2\pi,$$
$$y_1 = e^{i(3\rho - \alpha)/3}(is_1 chu - s_0 shu), \quad 0 \leqslant u < \infty, \tag{4.22}$$
$$y_2 = e^{i(3\rho + 2\alpha)/3} is_2.$$

The infinitesimal operators are given in the Appendix. Putting $\alpha = u = 0$, we again obtain (4.11). The Hamiltonian and Killing tensor Q [(3.11)] in this case are

$$H = -I_{12}^2 + I_{01}^2 + I_{02}^2 + \left\{-\frac{s_0^2 - s_1^2}{(s_0^2 + s_1^2)^2}p_u^2 + \left[\frac{s_0^2 - s_1^2}{(s_0^2 + s_1^2)^2} - \frac{1}{s_2^2}\right]p_\alpha^2 + \frac{4s_0 s_1}{(s_0^2 + s_1^2)^2}p_u p_\alpha\right\}, \tag{4.23}$$

$$Q_{\rm II} = a\left[I_{01}^2 + \frac{(s_0^2 - s_1^2)^2}{(s_0^2 + s_1^2)^2}(p_u^2 - p_\alpha^2) - 4\frac{s_0 s_1 (s_0^2 - s_1^2)}{(s_0^2 + s_1^2)^2}p_u p_\alpha\right] + b\left[\{I_{12}, I_{02}\} + 2\frac{s_0 s_1 s_2^2}{(s_0^2 + s_1^2)^2}p_u^2 + 2\frac{s_0 s_1[(s_0^2 + s_1^2)^2 - s_2^4]}{(s_0^2 + s_1^2)^2 s_2^2}p_\alpha^2 + 2\frac{(s_0^2 + s_1^2) + s_2^2(s_0^2 - s_1^2)}{(s_0^2 + s_1^2)^2}p_u p_\alpha\right]. \tag{4.24}$$

Setting $p_u = p_\alpha = 0$, we again obtain a free O(2,1) Hamiltonian and a specific O(2,1) Killing tensor (leading to only two of the nine separable systems on H_2). For $p_u = c_1$ and $p_\alpha = c_2$, we obtain a new nontrivial completely integrable Hamiltonian system with a singular potential.

3. The orthogonally decomposable MASA $\{Y_1, Y_4\}$

To introduce the ignorable variables (ρ,α,t), we put

$$-\tfrac{1}{3}Y_1 = p_\alpha, \quad -Y_4 = p_t, \quad X_0 = p_\rho \tag{4.25}$$

and obtain

$$\begin{aligned} y_0 &= e^{i(3\rho-\alpha)/3}\,[s_0 + i(s_0 - s_1)t\,], \quad -\infty < t < \infty, \\ y_1 &= e^{i(3\rho-\alpha)/3}\,[s_1 + i(s_0 - s_1)t\,], \\ &\quad 0 \leqslant \rho < 2\pi, \ 0 \leqslant \alpha < 2\pi, \\ y_2 &= e^{i(3\rho+2\alpha)/3} s_2\,. \end{aligned} \tag{4.26}$$

The infinitesimal operators are given in the Appendix. The Hamiltonian and Killing tensor (3.12) are

$$\begin{aligned} H = &-I_{12}^2 + I_{01}^2 + I_{02}^2 \\ &+ \left[\frac{1}{s_2^2}p_\alpha^2 + \frac{s_0+s_1}{(s_0-s_1)^3}p_t^2 + \frac{2}{(s_0-s_1)^2}p_\alpha p_t\right], \end{aligned} \tag{4.27}$$

$$\begin{aligned} Q_{\rm III} = 3a\Bigg[&(I_{02}-I_{12})^2 + \frac{(s_0-s_1)^2}{s_2^2}p_\alpha^2 \\ &+ \frac{s_2^2}{(s_0-s_1)^2}p_t^2 + 2p_\alpha p_t\Bigg] \\ &+ b\left[I_{01}^2 + \frac{(s_0+s_1)^2}{(s_0-s_1)^2}p_t^2 + 2\frac{s_0+s_1}{s_0-s_1}p_t p_\alpha\right]. \end{aligned} \tag{4.28}$$

For $\alpha = t = 0$ we again have pure O(2,1) quantities. The specific form of $Q_{\rm III}$ leads to four of the nine separable O(2,1) systems. For $p_\alpha = c_1$ and $p_t = c_2$, we obtain yet another O(2,1) Hamiltonian with a new nontrivial singular interaction.

4. The maximal abelian nilpotent subalgebra $\{Y_3, Y_4\}$

To introduce the ignorable variables (ρ,t,u), we put

$$Y_3 = p_t, \quad Y_4 = -p_u, \quad X_0 = p_\rho \tag{4.29}$$

and obtain

$$\begin{aligned} y_0 &= e^{i\rho}\,[(s_0 - s_1)(u - \tfrac{1}{2}it^2) + s_2 t - is_0], \quad -\infty < u < \infty, \\ y_1 &= e^{i\rho}\,[(s_0 - s_1)(u - \tfrac{1}{2}it^2) + s_2 t - is_1], \quad -\infty < t < \infty, \\ y_2 &= e^{i\rho}\,[-is_2 - (s_0 - s_1)t\,], \quad 0 \leqslant \rho < 2\pi. \end{aligned} \tag{4.30}$$

The infinitesimal operators are in the Appendix; the Hamiltonian and Killing tensor (3.13) are

$$\begin{aligned} H = &-I_{12}^2 + I_{01}^2 + I_{02}^2 + \left[\frac{1}{(s_0-s_1)^2}p_t^2\right. \\ &\left. - \frac{4s_2}{(s_0-s_1)^3}p_u p_t + \frac{3s_2^2 - s_0^2 - s_1^2}{(s_0-s_1)^4}p_u^2\right], \end{aligned} \tag{4.31}$$

$$\begin{aligned} Q_{\rm IV} = 3a\Bigg[&(I_{12}-I_{02})^2 + \left(p_t - \frac{2s_2}{s_0-s_1}p_u\right)^2 \\ &+ 3b\left[\{I_{02}-I_{12}, I_{01}\} + \frac{2s_2}{s_0-s_1}p_t^2\right. \\ &+ 4s_2\frac{s_0^2 - s_1^2 + s_2^2}{(s_0-s_1)^3}p_u^2 \\ &\left. + 2\frac{2s_1^2 - 3s_2^2 - 2s_0 s_1}{(s_0-s_1)^2}p_u p_t\right]. \end{aligned} \tag{4.32}$$

For $p_t = p_u = 0$ the operator $Q_{\rm IV}$ reduces to an O(2,1) operator related to variable separation in two of the nine separable systems on H_2. For $p_t = c_1$ and $p_u = c_2$ we again obtain a nontrivial interaction term in (4.31).

B. Separation of variables on an O(2,1) hyperboloid

Let us now consider the separation of variables in the free Hamilton–Jacobi equation or free Laplace–Beltrami equation on the O(2,1) homogeneous space

$$s^2 = s_0^2 - s_1^2 - s_2^2 = K^2 \quad (K^2 = \pm 1 \text{ or } 0). \tag{4.33}$$

Nine separable coordinate systems have been shown to exist[21] and to be in one-to-one correspondence with orbits of second-order operators in the enveloping algebra of O(2,1).[5,7] Since the results are not readily available and were not presented in a convenient form for our purposes, we summarize them here.

Let $I_{\mu\nu}$ be the O(2,1) operators (4.10), satisfying

$$[I_{01},I_{02}] = -I_{12}, \quad [I_{12},I_{01}] = I_{02}, \quad [I_{12},I_{02}] = -I_{01}. \tag{4.34}$$

A general second-order operator in the O(2,1) enveloping algebra can be written as

$$R = (I_{12}I_{01}I_{02})X\begin{pmatrix} I_{12} \\ I_{01} \\ I_{02}\end{pmatrix}, \quad X = X^T \in \mathbb{R}^{3\times 3}. \tag{4.35}$$

Under an O(2,1) transformation, R is transformed into R' given by (4.35) with X replaced by X':

$$X' = G^T X G, \quad GJG^T = J, \tag{4.36}$$

where J is a nonsingular 3×3 real symmetric matrix with signature $(-\,+\,+)$. We rewrite (4.36) as

$$\tilde{X}' = G^{-1}\tilde{X}G, \quad \tilde{X} = JX, \quad \tilde{X}^T J = J\tilde{X}. \tag{4.37}$$

Thus, X is symmetric under the involution that defines o(2,1). Such symmetric matrices have recently been classified for all classical Lie algebras.[25] For O(2,1) the results are quite simple, namely any pair of matrices $(\tilde{X},J)$ satisfying (4.37) is SL(3,$\mathbb{R}$) conjugate to one of the following:

(I) $\tilde{X}_{\rm I}$ orthogonally decomposable with three real eigenvalues:

$$\tilde{X}_{\rm I} = \begin{pmatrix} -c & & \\ & a & \\ & & b\end{pmatrix}, \quad J = \begin{pmatrix} -1 & & \\ & 1 & \\ & & 1\end{pmatrix}, \quad a,b,c \in \mathbb{R}. \tag{4.38}$$

(II) $\tilde{X}_{\rm II}$ orthogonally decomposable with one real eigenvalue and one pair of complex conjugate eigenvalues:

$$\tilde{X}_{\rm II}=\begin{pmatrix} a & b & 0\\ -b & a & 0\\ 0 & 0 & c\end{pmatrix},\quad J=\begin{pmatrix}0&1&0\\1&0&0\\0&0&1\end{pmatrix},\quad a,b,c\in\mathbb{R},\ b>0\,. \tag{4.39}$$

(III) $\tilde{X}_{\rm III}$ orthogonally decomposable with two real eigenvalues:

$$\tilde{X}_{\rm III}=\begin{pmatrix} a & 0 & 0\\ 1 & a & 0\\ 0 & 0 & b\end{pmatrix},\quad J=\begin{pmatrix}0&1&0\\1&0&0\\0&0&1\end{pmatrix},\quad a,b\in\mathbb{R}\,. \tag{4.40}$$

(IV) $\tilde{X}_{\rm IV}$ indecomposable (one real eigenvalue):

$$\tilde{X}_{\rm IV}=\begin{pmatrix} a & 0 & 0\\ 1 & a & 0\\ 0 & 1 & a\end{pmatrix},\quad J=\begin{pmatrix} & & 1\\ & 1 & \\ 1 & & \end{pmatrix},\quad a\in\mathbb{R}\,. \tag{4.41}$$

Returning to a basis in which J is as (4.38) and simplifying by linear combinations with the O(2,1) Casimir operator

$$\Delta=I_{01}^2+I_{02}^2-I_{12}^2, \tag{4.42}$$

we obtain four classes of quadratic operators R:

$$\begin{aligned} R_{\rm I}&=\lambda_3 I_{12}^2+\lambda_1 I_{01}^2+\lambda_2 I_{02}^2,\quad \lambda_i\in\mathbb{R},\\ R_{\rm II}&=\lambda I_{01}^2+\{I_{02},I_{12}\},\\ R_{\rm III}&=\lambda I_{01}^2+\mu(I_{02}-I_{12})^2,\quad \mu\neq 0,\ \lambda,\mu\in\mathbb{R},\\ R_{\rm IV}&=\{I_{02}-I_{12},I_{01}\}\end{aligned} \tag{4.43}$$

(the brackets $\{\ ,\ \}$ denote an anticommutator). The operator $R_{\rm I}$ can be further simplified by combinations with Δ; in $R_{\rm II}$ we can assume $\lambda\geqslant 0$; in $R_{\rm III}$ we can scale μ with respect to λ by means of the O(2,1) transformation $\exp\alpha I_{01}$ and hence only distinguish three cases: $\lambda=0,\mu=1;\lambda=\mu=1;$ $\lambda=-\mu=1$.

Finally we obtain nine classes of operators R_a $(a=1,\ldots,9)$ and the corresponding coordinate systems for which the O(2,1) Hamilton-Jacobi and Laplace–Beltrami equations separate. The separable coordinates, Hamiltonians H and integrals of motion R_a, for the two-sheeted hyperboloid, i.e., $K^2=1$ are as follows.

1. Spherical: $R_{\rm I}$ with $\lambda_1=\lambda_2\neq\lambda_3$

$$\begin{aligned} &s_0=\cosh A,\quad s_1=\sinh A\cos B,\quad s_2=\sinh A\sin B,\\ &0\leqslant A<\infty,\ 0\leqslant B<2\pi,\\ &H=p_A^2+\frac{1}{\sinh^2 A}p_B^2,\quad R_1=I_{12}^2=p_B^2.\end{aligned} \tag{4.44}$$

2. Hyperbolic: $R_{\rm I}$ with $-\lambda_3=\lambda_2\neq\lambda_1$

$$\begin{aligned} &s_0=\cosh A\cosh B,\quad s_1=\cosh A\sinh B,\\ &s_2=\sinh A,\quad A,B\in\mathbb{R},\\ &H=p_A^2+\frac{1}{\cosh^2 A}p_B^2,\quad R_2=I_{01}^2=p_B^2.\end{aligned} \tag{4.45}$$

3. Elliptic I: $R_{\rm I}$ with $-\lambda_3\neq\lambda_1\neq\lambda_2\neq-\lambda_3$, $(\lambda_1+\lambda_3)/(\lambda_2+\lambda_3)>0$

$$\begin{aligned} &s_0^2=\nu\rho/a,\quad s_1^2=(\nu-1)(\rho-1)/(a-1),\\ &s_2^2=(\nu-a)(a-\rho)/(a-1)a,\quad 1\leqslant\rho\leqslant a\leqslant\nu<\infty,\ 1<a\\ &H=[4/(\nu-\rho)]\,[\nu(\nu-1)(\nu-a)\,p_\nu^2\\ &\qquad+\rho(\rho-1)(a-\rho)p_\rho^2\,],\\ &R_3=aI_{01}^2+I_{02}^2\\ &\quad=[4\nu\rho/(\nu-\rho)]\,[(\nu-1)(\nu-a)\,p_\nu^2\\ &\qquad+(\rho-1)(a-\rho)\,p_\rho^2\,].\end{aligned} \tag{4.46}$$

4. Elliptic II: $R_{\rm I}$ with $-\lambda_3\neq\lambda_1\neq\lambda_2\neq-\lambda_3$, $(\lambda_1+\lambda_3)/(\lambda_2+\lambda_3)<0$

$$\begin{aligned} &s_0^2=(\nu-1)(1-\rho)/(a-1),\quad s_1^2=-\nu\rho/a,\\ &s_2^2=(\nu-a)(a-\rho)/(a-1)a,\\ &\rho\leqslant 0,\ 1<a\leqslant\nu,\ 0<a-1\leqslant 1,\\ &H=[4/(\nu-\rho)]\,[\nu(\nu-1)(\nu-a)\,p_\nu^2\\ &\qquad+\rho(\rho-1)(a-\rho)\,p_\rho^2\,],\\ &R_4=(a-1)I_{01}^2-I_{02}^2\\ &\quad=-[4(1-\rho)(\nu-1)/(\nu-\rho)]\,[\nu(\nu-a)\,p_\nu^2\\ &\qquad+(a-\rho)\,\rho p_\rho^2\,].\end{aligned} \tag{4.47}$$

5. Complex elliptic: $R_{\rm II}$

$$\begin{aligned} &\tfrac{1}{2}(s_0+is_1)^2=(\nu-a)(\rho-a)/a(a-a^*),\quad s_2^2=-\nu\rho/|a|^2,\\ &\nu<0<\rho,\ a=\alpha+i\beta,\ \beta>0,\ \alpha,\beta\in\mathbb{R},\\ &H=[4/(\rho-\nu)]\,[\rho(\rho-a)(\rho-a^*)\,p_\rho^2\\ &\qquad-\nu(\nu-a)(\nu-a^*)\,p_\nu^2\,],\\ &R_5=aI_{01}^2-\beta\{I_{12},I_{02}\}\\ &\quad=[4\rho\nu/(\rho-\nu)]\,[(\rho-a)(\rho-a^*)\,p_\rho^2\\ &\qquad-(\nu-a)(\nu-a^*)\,p_\nu^2\,].\end{aligned} \tag{4.48}$$

6. Horospheric: $R_{\rm III}$ with $\lambda=0$

$$\begin{aligned} &s_0=\cosh A+\tfrac{1}{2}r^2e^{-A},\quad s_1=\sinh A+\tfrac{1}{2}r^2e^{-A},\\ &s_2=re^{-A},\quad -\infty<A<\infty,\quad -\infty<r<\infty,\\ &H=p_A^2+e^{2A}p_r^2,\quad R_6=(I_{02}-I_{12})^2=p_r^2.\end{aligned} \tag{4.49}$$

7. Elliptic parabolic: $R_{\rm III}$ with $\lambda\mu>0$

$$\begin{aligned} &s_0^2=\tfrac{1}{4}(\nu+\rho)^2/\nu\rho,\quad s_1^2=\tfrac{1}{4}(\nu+\rho-2\nu\rho)^2/\nu\rho,\\ &s_2^2=(1-\nu)(\rho-1),\quad 0<\nu<1<\rho\,,\\ &H=[4/(\rho-\nu)]\,[\rho^2(\rho-1)\,p_\rho^2+\nu^2(1-\nu)p_\nu^2\,],\\ &R_7=I_{01}^2+(I_{02}-I_{12})^2\\ &\quad=[4\rho\nu/(\rho-\nu)]\,[\rho(\rho-1)\,p_\rho^2+\nu(1-\nu)p_\nu^2\,].\end{aligned} \tag{4.50}$$

8. Hyperbolic parabolic: R_{III} with $\lambda\mu < 0$

$$s_0^2 = (\nu+\rho-2\nu\rho)^2/(-4\nu\rho), \quad s_1^2 = (\nu+\rho)^2/(-4\nu\rho),$$
$$s_2^2 = (1-\nu)(\rho-1), \quad \nu<0<1<\rho, \tag{4.51}$$
$$H = [4/(\rho-\nu)]\,[\rho^2(\rho-1)p_\rho^2 + \nu^2(1-\nu)p_\nu^2],$$
$$R_8 = I_{01}^2 - (I_{02}-I_{12})^2$$
$$= [4\rho\nu/(\rho-\nu)]\,[\rho(\rho-1)p_\rho^2 + \nu(1-\nu)p_\nu^2].$$

9. Semicircular parabolic: R_{IV}

$$s_0^2 = 1/[-16(\rho\nu)^3]\,[(\rho-\nu)^2+\rho^2\nu^2]^2,$$
$$s_1^2 = 1/[-16(\rho\nu)^3]\,[(\rho-\nu)^2-\rho^2\nu^2]^2,$$
$$s_2^2 = (\rho+\nu)^2/(-4\rho\nu), \quad \nu<0<\rho, \tag{4.52}$$
$$H = [4/(\rho-\nu)](\rho^3 p_\rho^2 - \nu^3 p_\nu^2), \quad R_9 = \{I_{02}-I_{12}, I_{01}\}$$
$$= 2[\nu\rho/(\nu-\rho)](\rho^2 p_\rho^2 - \nu^2 p_\nu^2).$$

Three of these coordinate systems are of the "subgroup type," namely spherical, hyperbolic, and horospheric, corresponding to the group reductions

$$O(2,1)\supset O(2), \quad O(2,1)\supset O(1,1), \text{ and } O(2,1)\supset T,$$

respectively (T being the group of translations generated by $I_{02}-I_{12}$).

All coordinate systems are written so as to parametrize the upper sheet of a one-sheeted hyperboloid. It is not difficult to modify the coordinates so as to parametrize the one sheeted hyperboloid ($s^2 = -1$).

C. Separable coordinates on HH(2) and the Hamiltonian systems

In Sec. III we have classified triplets of operators $\{T_1,T_2,T_3\}$ into 12 orbits under SU(2,1). In Sec. IVA we have introduced ignorable variables on HH(2). Each different MASA of SU(2,1) leads to specific coordinates in which the Hamiltonian H and integral of motion $Q = T_3$ reduce to an O(2,1) form corresponding to an O(2,1) Hamiltonian system with a nontrivial interaction. In Sec. IVB we reviewed separation on the O(2,1) hyperboloid $s^2 = 1$. Combining all these results together, we obtain the following theorem.

Theorem 1: (1) There exist precisely 12 systems of coordinates on HH (2) in which the Hamiltonian–Jacobi and Laplace–Beltrami equations separate.

(2) Each separable system has two ignorable and two nonignorable variables. The nonignorable variables are introduced so as to separate variables on the O(2,1) hyperboloid $s^2 = s_0^2 - s_1^2 - s_2^2 = 1$.

(3) The separable coordinate systems in HH (2) are in one-to-one correspondence with orbits of triplets of second-order operators $\{T_1,T_2,T_3\}$ in the enveloping algebra of su(2,1). The operators T_i are in involution, two of them, $T_1 = L_1^2$ and $T_2 = L_2^2$, are squares of the generators L_1,L_2 of a MASA of su(2,1), the third $Q = T_3$ is a general operator of the form (3.3). The operator Q takes one of the forms $Q_1,\ldots,Q_{12}$ listed in Sec. III.

(4) The compact Cartan subalgebra $\{X_3,X_8\}$ for which Q has the form Q_1 of (4.20) leads to four types of coordinate systems, namely, (4.18) with (s_0,s_1,s_2), expressed in spherical (Q_1), hyperbolic (Q_2), elliptic I (Q_3), or elliptic II (Q_4) coordinates on the O(2,1) hyperboloid H_2.

(5) The noncompact Cartan subalgebra $\{X_3 + (1/\sqrt{3})X_8, X_5\}$ for which Q has the form Q_{II} of (4.24) leads to two types of coordinate systems, namely, (4.22) with (s_0,s_1,s_2) expressed in hyperbolic (Q_5) or complex elliptic (Q_6) coordinates on H_2.

(6) The decomposable non-Cartan subalgebra $\{Y_1,Y_4\}$ for which Q has the form Q_{III} of (4.28) leads to four separable coordinate systems, namely, (4.26) with (s_0,s_1,s_2) expressed in hyperbolic (Q_7), horospheric (Q_8), elliptic parabolic (Q_9), or elliptic hyperbolic (Q_{10}) coordinates on H_2.

(7) The MANS $\{Y_3,Y_4\}$ for which Q has the form Q_{IV} of (4.32) leads to two separable systems, namely, (4.30) with (s_0,s_1,s_2) expressed in horospheric (Q_{11}) or semicircular parabolic coordinates (Q_{12}) on H_2.

Finally, let us list the 12 separable coordinate systems and in the process show that the "potentials" in the O(2,1) Hamiltonians are indeed compatible with separation in each of the 12 cases. We shall use the affine coordinates (2.5).

1. The compact Cartan subalgebra $\{X_3,X_8\}$

$$\tfrac{1}{2}[X_3 - (1/\sqrt{3})X_8] = p_{\alpha_1} = c_1,$$
$$-\tfrac{1}{2}[X_3 + (1/\sqrt{3})X_8] = p_{\alpha_2} = c_2.$$

a. Spherical coordinates:

$$z_1 = \tanh A\cos Be^{i\alpha_1}, \quad z_2 = \tanh A\sin Be^{i\alpha_2}, \tag{4.53}$$
$$Q_1 = p_B^2 + (1/\cos^2 B)\,p_{\alpha_1}^2 + (1/\sin^2 B)\,p_{\alpha_2}^2 = c_3,$$
$$H = p_A^2 + (1/\sinh^2 A)\,Q_1 - (1/\cosh^2 A)(p_{\alpha_1}+p_{\alpha_2})^2 = E.$$

b. Hyperbolic coordinates:

$$z_1 = \tanh Be^{i\alpha_1}, \quad z_2 = (\tanh A/\cosh B)e^{i\alpha_2}, \tag{4.54}$$
$$Q_2 = p_B^2 + (1/\sinh^2 B)\,p_{\alpha_1}^2 - (1/\cosh^2 B)(p_{\alpha_1}+p_{\alpha_2})^2 = c_3,$$
$$H = p_A^2 + (1/\cosh^2 A)\,Q_2 + (1/\sinh^2 A)\,p_\alpha^2 = E.$$

c. Elliptic I coordinates:

$$z_1^2 = [a(\nu-1)(\rho-1)/(a-1)\nu\rho]\, e^{2i\alpha_1},$$
$$z_2^2 = [(\nu-a)(a-\rho)/(a-1)\nu\rho]\, e^{2i\alpha_2}, \tag{4.55}$$
$$Q_3 = [1/(\nu-\rho)]\{4\rho\nu(\nu-1)(\nu-a)p_\nu^2 + 4\nu\rho(\rho-1)(a-\rho)p_\rho^2$$
$$+ [(a-\rho)\nu/(\rho-1) + (\nu-a)\rho/(\nu-1)]p_{\alpha_1}^2 + a[(\rho-1)\nu/(a-\rho) + (\nu-1)\rho/(\nu-a)]p_{\alpha_2}^2$$
$$- a(\nu/\rho - \rho/\nu)(p_{\alpha_1}+p_{\alpha_2})^2\} = c_3,$$
$$H = [1/(\nu-\rho)]\,[4\nu(\nu-1)(\nu-a)p_\nu^2 + 4\rho(\rho-1)(a-\rho)p_\rho^2$$
$$+ (a-1)[1/(\rho-1) - 1/(\nu-1)]p_{\alpha_1}^2 + a(a-1)[1/(a-\rho) + 1/(\nu-a)]p_{\alpha_2}^2$$
$$- a(1/\rho - 1/\nu)(p_{\alpha_1}+p_{\alpha_2})^2] = E.$$

d. Elliptic II coordinates:

$$z_1^2 = -(a-1)\nu\rho/a(\nu-1)(1-\rho), \quad z_2^2 = (\nu-a)(a-\rho)/a(\nu-1)(1-\rho), \tag{4.56}$$
$$\rho\leqslant 0, \quad 1<a\leqslant\nu, \quad 0<a-1\leqslant 1,$$
$$H = [1/(\nu-\rho)]\,\{4\nu(\nu-1)(\nu-a)p_\nu^2 + 4\rho(\rho-1)(a-\rho)p_\rho^2 + a(-1/\rho + 1/\nu)p_{\alpha_1}^2$$
$$+ a(a-1)[1/(a-\rho) + 1/(\nu-a)]p_{\alpha_2}^2 - (a-1)[1/(1-\rho) + 1/(\nu-1)](p_{\alpha_1}+p_{\alpha_2})^2 = E,$$
$$Q_4 = [4(1-\rho)(\nu-1)/(\nu-\rho)]\,[\nu(\nu-a)p_\nu^2 + \rho(a-\rho)p_\rho^2] + [(\nu\rho - a\rho - a\nu + a)/\nu\rho]p_{\alpha_1}^2$$
$$+ (a-1)(\rho\nu - a)p_{\alpha_2}^2 + [(a-1)(2-\nu-\rho)/(\nu-1)(1-\rho)](p_{\alpha_1}+p_{\alpha_2})^2 = c_3.$$

2. The noncompact Cartan subalgebra $\{X_3 + (1/\sqrt{3})X_8, X_5\}$

$$-\tfrac{1}{2}[X_3 + (1/\sqrt{3})X_8] = p_\alpha = c_1, \quad X_5 = p_u = c_2.$$

e. Hyperbolic coordinates:

$$z_1 = \frac{i\sinh B\cosh u - \cosh B\sinh u}{i\cosh B\cosh u + \sinh B\sinh u}, \quad z_2 = ie^{i\alpha}\frac{\tanh A}{i\cosh B\cosh u + \sinh B\sinh u}, \tag{4.57}$$
$$Q_5 = p_B^2 + (1/\cosh^2 2B)(p_u^2 - p_\alpha^2) - [2\sinh 2B/\cosh^2 2B]\,p_u p_\alpha,$$
$$H = p_A^2 + (1/\cosh^2 A)\,Q_5 + (1/\sinh^2 A)\,p_\alpha^2.$$

f. Complex elliptic coordinates: The coordinates are

$$z_1 = \frac{is_1\cosh u - s_0\sinh u}{is_0\cosh u + s_1\sinh u}, \quad z_2 = ie^{i\alpha}\frac{s_2}{is_0\cosh u + s_1\sinh u} \tag{4.58}$$

with s_0, s_1, and s_2 as in (4.48)

$$Q_6 = [1/(\rho-\nu)]\,\Big\{4\rho\nu(\rho-a)(\rho-a^*)p_\rho^2 - 4\rho\nu(\nu-a)(\nu-a^*)p_\nu^2$$
$$+ \tfrac{1}{4}|a-a^*|^2\rho\nu\,[1/(\rho-a)(\rho-a^*) - 1/(\nu-a)(\nu-a^*)](-p_u^2 + p_\alpha^2)$$
$$+ (|a|^2/\rho\nu)(\rho^2-\nu^2)p_\alpha^2$$
$$+ \tfrac{1}{2}i(a-a^*)\{[(a^*+a)\nu\rho - 2|a|^2\rho]/(\nu-a)(\nu-a^*) - [(a^*+a)\nu\rho - 2|a|^2\nu]/(\rho-a)(\rho-a^*)\}p_u p_\alpha\Big\},$$
$$H = [1/(\rho-\nu)]\,\Big\{4\rho(\rho-a)(\rho-a^*)p_\rho^2 - 4\nu(\nu-a)(\nu-a^*)p_\nu^2$$
$$+ \tfrac{1}{4}|a-a^*|^2\,[\rho/(\rho-a)(\rho-a^*) - \nu/(\nu-a)(\nu-a^*)](-p_u^2 + p_\alpha^2)$$
$$+ (|a|^2/\rho\nu)(\rho-\nu)p_\alpha^2$$
$$+ \tfrac{1}{2}i(a-a^*)\{[(a^*+a)\nu + 2]|a|^2/(\nu-a)(\nu-a^*) - [(a^*+a)\rho - 2|a|^2]/(\rho-a)(\rho-a^*)\}p_u p_\alpha\Big\}.$$

3. The orthogonally decomposable MASA $\{Y_1, Y_4\}$

$$-\tfrac{1}{3}Y_1 = -\tfrac{1}{2}[X_3 + (1/\sqrt{3})X_8] = p_\alpha = c_1,$$
$$-Y_2 = X_5 - \tfrac{1}{2}(X_3 - \sqrt{3}X_8) = p_t = c_2.$$

g. Hyperbolic coordinates [$a = 0$ in (4.28)]:

$$z_1 = (\sinh B + ite^{-B})/(\cosh B + ite^{-B}),$$
$$z_2 = \tanh Ae^{i\alpha}/(\cosh B + ite^{-B}), \tag{4.59}$$
$$Q_7 = p_B^2 + (e^{2B}p_t + p_\alpha)^2 - p_\alpha^2 = c_3,$$
$$H = p_A^2 + (1/\cosh^2 A)\,Q_7 + (1/\sinh^2 A)p_\alpha^2 = E.$$

h. Horospheric coordinates [$b = 0$ *in* (4.28)]:

$$z_1 = (-1 + e^{2A} + B^2 + 2it)/(1 + e^{2A} + B^2 + 2it),$$
$$z_2 = Be^{i\alpha}/(1 + e^{2A} + B^2 + 2it), \tag{4.60}$$
$$Q_8 = p_B^2 + [(1/B)p_\alpha + Bp_t]^2,$$
$$H = p_A^2 + e^{2A}Q_8 + e^{4A}p_t^2.$$

i. Elliptic parabolic coordinates [$3a = b$ *in* (4.28)]:

$$z_1 = (\nu + \rho - 2\nu\rho + 2i\nu\rho t)/(\nu + \rho + 2i\nu\rho t),$$
$$z_2^2 = 4\nu\rho(1-\nu)(\rho-1)e^{2i\alpha}/(\nu + \rho + 2i\nu\rho t)^2, \tag{4.61}$$
$$Q_9 = [1/(\rho-\nu)]\{4\rho\nu[\rho(\rho-1)p_\rho^2 + \nu(1-\nu)p_\nu^2] + [\rho(1-\nu)/\nu^2 + \nu(\rho-1)/\rho^2]p_t^2 + \nu\rho[1/(1-\nu) + 1/(\rho-1)]p_\alpha^2 + 2(\rho/\nu - \nu/\rho)p_t p_\alpha\} = c_3,$$
$$H = [1/(\rho-\nu)]\{4\rho^2(\rho-1)p_\rho^2 + 4\nu^2(1-\nu)p_\nu^2 + [1/(1-\nu) + 1/(\rho-1)]p_\alpha^2 + [(1-\nu)/\nu^2 + (\rho-1)/\rho^2]p_t^2 + 2(1/\nu - 1/\rho)p_\alpha p_t\} = E.$$

j. Hyperbolic parabolic coordinates [$3a = -b$ *in* (4.28)]:

$$z_1 = (\nu + \rho - 2i\nu\rho t)/(\nu + \rho - 2\nu\rho - 2i\nu\rho t), \quad z_2^2 = [-4\nu\rho(1-\nu)(\rho-1)e^{2i\alpha}]/(\nu + \rho - 2\nu\rho - 2i\nu\rho t)^2, \tag{4.62}$$
$$Q_{10} = [1/(\rho-\nu)]\{4\rho\nu[\rho(\rho-1)p_\rho^2 + \nu(1-\nu)p_\nu^2] + [\rho(1-\nu)/\nu^2 + \nu(\rho-1)/\rho^2]p_t^2 + \nu\rho[1/(1-\nu) + 1/(\rho-1)]p_\alpha^2 - 2(\rho/\nu - \nu/\rho)p_\alpha p_t\},$$
$$H = [1/(\rho-\nu)]\{4\rho^2(\rho-1)p_\rho^2 + 4\nu^2(1-\nu)p_\nu^2 + [1/(1-\nu) + 1/(\rho-1)]p_\alpha^2 + [(1-\nu)/\nu^2 + (\rho-1)/\rho^2]p_t^2 - 2(1/\nu - 1/\rho)p_\alpha p_t\}.$$

4. The maximal abelian nilpotent subalgebra $\{Y_3, Y_4\}$

$$Y_3 = X_2 - X_7 = p_t = c_1, \quad -Y_4 = X_5 - \tfrac{1}{2}X_3 + \tfrac{1}{2}\sqrt{3}X_8 = p_u = c_2.$$

k. Horospheric coordinates [$b = 0$ *in* (4.32)]:

$$z_1 = [2(u + Bt) - i(e^{2A} + B^2 + t^2 - 1)]/[2(u + Bt) - i(e^{2A} + B^2 + t^2 + 1)],$$
$$z_2 = -2(t + iB)/[2(u + Bt) - i(e^{2A} + B^2 + t^2 + 1)], \tag{4.63}$$
$$Q_{11} = p_B^2 + (p_t - 2Bp_u)^2 = c_3,$$
$$H = p_A^2 + e^{2A}Q_{11} + e^{4A}p_u^2 = E.$$

l. Semicircular parabolic coordinates [$a = 0$ *in* (4.32)]:

$$z_1 = \frac{2\rho^2\nu^2 u - 2\rho\nu(\rho+\nu)t - i[(\rho-\nu)^2 + \rho^2\nu^2(t^2-1)]}{2\rho^2\nu^2 u - 2\rho\nu(\rho+\nu)t - i[(\rho-\nu)^2 + \rho^2\nu^2(t^2+1)]}, \tag{4.64}$$
$$z_2 = \frac{2\rho\nu t - 2i(\rho+\nu)}{2\rho^2\nu^2 u - 2\rho\nu(\rho+\nu)t - i[(\rho-\nu)^2 + \rho^2\nu^2(t^2+1)]},$$
$$Q_{12} = [2/(\rho-\nu)]\{\nu\rho^3 p_\rho^2 - \nu^3\rho p_\nu^2 + (\nu/\rho - \rho/\nu)p_t^2 + 4(\nu/\rho^3 - \rho/\nu^3)p_u^2 + (4\nu/\rho^2 - 4\rho/\nu^2 + \rho - \nu)p_u p_t\},$$
$$H = [4/(\rho-\nu)]\{\rho^3 p_\rho^2 - \nu^3 p_\nu^2 + (1/\rho - 1/\nu)p_t^2 + 4(1/\rho^3 - 1/\nu^3)p_u^2 + 4(1/\rho^2 - 1/\nu^2)p_u p_t\}.$$

To summarize: The nonsubgroup type coordinates on H_2, namely elliptic I and II, complex elliptic, elliptic parabolic, hyperbolic parabolic, and semicircular parabolic each occur precisely once. The subgroup type coordinates on H_2 occur as follows: spherical coordinates once [since the compact subalgebra u(2) contains only one MASA], hyperbolic coordinates three times [u(1,1) contains three MASA's] and horospheric coordinates twice [$A_{4,10}$ contains two MASA's (see Fig. 1)].

V. CONCLUSION

The results of this article should be viewed in the context of three different but related research programs. One is a systematic study of the group theoretical, algebraic, and geometrical aspects of the separation of variables in linear and nonlinear partial differential equations. From this point of view we should stress that the hermitian hyperbolic space HH(2) is a noncompact manifold of nonconstant curvature (it does, however, have constant holomorphic sectional curvature). The fact that it has a large isometry group, namely SU(2,1), made it possible to apply essentially the same techniques as for spaces with constant curvature. We have shown that all 12 separable coordinate systems on HH(2) have their origin in the properties of the algebra su(2,1), its subalgebras, and its enveloping algebra.

The second context is that of the classification of subgroups of Lie groups, in particular, maximal Abelian sub-

groups of classical Lie groups, and its application to the study of differential equations. Indeed, the classification of all MASA's of su(2,1) into four conjugacy classes was the basis of our calculations providing the explicit forms of the 12 separable coordinate systems. In passing, we comment that other applications of this classification are being pursued. In addition to the separation of variables, these include the derivation of superposition principles for certain systems of nonlinear differential equations[26–28] and the symmetry reduction of certain nonlinear partial differential equations to ordinary ones.[29]

Finally, the reduction of the problem of separating variables for the free Hamiltonian on HH(2) to that of a Hamiltonian with a nontrivial interaction, defined on a lower-dimensional manifold, namely the O(2,1) hyperboloid H_2, is an example of a more general method of introducing interactions, in particular completely integrable Hamiltonian systems, by symmetry reduction on group manifolds or homogeneous spaces.

All three above aspects are being actively pursued. In particular, we are currently generalizing the results of this paper to the space HH(n) making use of the MASA's of SU(n,1). The completely integrable Hamiltonian systems obtained in this article are being investigated (explicit solutions, properties of trajectories, special functions occurring as solutions of the Laplace–Beltrami equations, etc.). The related problem of separating variables in Hamiltonians on HH(2) with specific potentials that reduce by symmetry to more general completely integrable relativistic Hamiltonian systems than the ones treated in this article is also under consideration.

ACKNOWLEDGMENTS

We are indebted to J. Harnad, W. Miller, Jr., and H. Zassenhaus for very helpful discussions.

APPENDIX: THE su(2,1) INFINITESIMAL OPERATORS IN TERMS OF IGNORABLE VARIABLES AND O(2,1) VARIABLES

1. Compact Cartan subalgebra variables

$$X_1 = \cos(\alpha_1 - \alpha_2)I_{12} + \sin(\alpha_1 - \alpha_2)[(s_2/s_1)p_{\alpha_1} + (s_1/s_2)p_{\alpha_2}],$$

$$X_2 = -\sin(\alpha_1 - \alpha_2)I_{12} + \cos(\alpha_1 - \alpha_2)[(s_2/s_1)p_{\alpha_1} + (s_1/s_2)p_{\alpha_2}],$$

$$X_3 = p_{\alpha_1} - p_{\alpha_2},$$

$$X_4 = \cos\alpha_1 I_{01} - \sin\alpha_1[(s_1/s_0 + s_0/s_1)p_{\alpha_1} + (s_1/s_0)p_{\alpha_2}],$$

$$X_5 = \sin\alpha_1 I_{01} + \cos\alpha_1[(s_1/s_0 + s_0/s_1)p_{\alpha_1} + (s_1/s_0)p_{\alpha_2}],$$

$$X_6 = \cos\alpha_2 I_{02} - \sin\alpha_2[(s_2/s_0)p_{\alpha_1} + (s_2/s_0 + s_0/s_2)p_{\alpha_2}],$$

$$X_7 = \sin\alpha_2 I_{02} + \cos\alpha_2[(s_2/s_0)p_{\alpha_1} + (s_2/s_0 + s_0/s_2)p_{\alpha_2}],$$

$$X_8 = -\sqrt{3}(p_{\alpha_1} + p_{\alpha_2}).$$

2. Noncompact Cartan subalgebra variables

$$X_1 = \cosh u\cos\alpha I_{12} + \sinh u\sin\alpha I_{02} - [s_2/(s_0^2 + s_1^2)](s_0\cosh u\sin\alpha - s_1\sinh u\cos\alpha)p_u - [1/s_2(s_0^2 + s_1^2)][-s_0(s_0^2 + s_1^2 + s_2^2)\sinh u\cos\alpha + s_1(s_0^2 + s_1^2 - s_2^2)\cosh u\sin\alpha]p_\alpha,$$

$$X_2 = \cosh u\sin\alpha I_{12} + \sinh u\cos\alpha I_{02} + [s_2/(s_0^2 + s_1^2)](s_0\cosh u + \cos\alpha + s_1\sinh u\sin\alpha)p_u + [1/s_2(s_0^2 + s_1^2)][s_0(s_0^2 + s_1^2 + s_2^2)\sinh u\sin\alpha + s_1(s_0^2 + s_1^2 - s_2^2)\cosh u\cos\alpha]p_\alpha,$$

$$X_3 = \tfrac{1}{2}(-\sinh 2u I_{01} + [2s_0s_1/(s_0^2 + s_1^2)]\cosh 2u p_u + \{[(s_0^2 - s_1^2)/(s_0^2 + s_1^2)]\cosh 2u - 3\}p_\alpha),$$

$$X_4 = \cosh 2u I_{01} - [2s_0s_1/(s_0^2 + s_1^2)]\sinh 2u p_u - [(s_0^2 - s_1^2)/(s_0^2 + s_1^2)]\sinh 2u p_\alpha,$$

$$X_5 = p_u,$$

$$X_6 = -\sinh u\sin\alpha I_{12} + \cosh u\cos I_{02} - [s_2/(s_0^2 + s_1^2)](s_0\sinh u\cos\alpha + s_1\cosh u\sin\alpha)p_u - [1/s_2(s_0^2 + s_1^2)][s_0(s_0^2 + s_1^2 + s_2^2)\cosh u\sin\alpha + s_1(s_0^2 + s_1^2 - s_2^2)\sinh u\cos\alpha]p_\alpha,$$

$$X_7 = \sinh u\cos\alpha I_{12} + \cosh u\sin\alpha I_{02} - [s_2/(s_0^2 + s_1^2)](s_0\sinh u\sin\alpha - s_1\cosh u\cos\alpha)p_u - [1/s_2(s_0^2 + s_1^2)][-s_0(s_0^2 + s_1^2 + s_2^2)\cosh u\cos\alpha + s_1(s_0^2 + s_1^2 - s_2^2)\sinh u\sin\alpha]p_\alpha,$$

$$X_8 = \tfrac{1}{2}\sqrt{3}(\sinh 2u I_{01} - [2s_0s_1/(s_0^2 + s_1^2)]\cosh 2u p_u - \{[(s_0^2 - s_1^2)/(s_0^2 + s_1^2)]\cosh 2u + 1\}p_\alpha).$$

3. Variables corresponding to orthogonally decomposable non-Cartan subalgebra

$$Y_1 = -3p_\alpha,$$

$$Y_2 = -\cos\alpha(I_{02} - I_{12}) + \sin\alpha\{[(s_0 - s_1)/s_2]p_\alpha - [s_2/(s_0 - s_1)]p_t\},$$

$$Y_3 = -\sin\alpha(I_{02} - I_{12}) - \cos\alpha\{[(s_0 - s_1)/s_2]p_\alpha - [s_2/(s_0 - s_1)]p_t\},$$

$$Y_4 = -p_t, \quad Y_5 = I_{01} + 2tp_t,$$

$$Y_6 = -2tI_{01} + [(s_0 + s_1)/(s_0 - s_1)]p_\alpha + \{[2s_0s_1 - 2t^2(s_0 - s_1)^2]/(s_0 - s_1)^2\}p_t,$$

$$Y_7 = \cos\alpha I_{12} + t\sin\alpha(I_{02} - I_{12}) + \Big\{\{[-s_1(s_0 - s_1) - s_2^2]\sin\alpha + (s_0 - s_1)^2 t\cos\alpha\}/s_2(s_0 - s_1)\Big\}p_\alpha - \{s_2[s_0\sin\alpha + t(s_0 - s_1)\cos\alpha]/(s_0 - s_1)^2\}p_t,$$

$$Y_8 = \sin\alpha I_{12} - t\cos\alpha(I_{02} - I_{12}) + (\{[s_1(s_0 - s_1) + s_2^2]\cos\alpha + (s_0 - s_1)^2 t\sin\alpha\}/s_2(s_0 - s_1))p_\alpha + \{s_2[s_0\cos\alpha - t(s_0 - s_1)\sin\alpha]/(s_0 - s_1)^2\}p_t.$$

4. Variables corresponding to the maximal abelian nilpotent subalgebra

$$Y_1 = 3\{t(I_{12} - I_{02}) + [s_2/(s_0 - s_1)]p_t + (t^2 - s_2^2/(s_0 - s_1)^2)p_u\},$$

$$Y_2 = I_{12} - I_{02} + 2tp_u, \quad Y_3 = p_t,$$

$$Y_4 = -p_u, \quad Y_5 = I_{01} + tp_t + 2up_u,$$

$$Y_6 = -2uI_{01} + t^3(I_{02} - I_{12}) + t(I_{02} + I_{12}) - \left[\frac{3s_2t^2}{s_0 - s_1} + 2tu + \frac{(s_0 + s_1)s_2}{(s_0 - s_1)^2}\right]p_t + \left[-\tfrac{1}{2}t^4 + \frac{3t^2s_2^2}{(s_0 - s_1)^2} - 2u^2 + \frac{s_2^2(s_0 + s_1)}{(s_0 - s_1)^3} + \frac{2s_0s_1}{(s_0 - s_1)^2}\right]p_u,$$

$$Y_7 = I_{12} + \frac{3t^2}{2}(I_{02} - I_{12}) - \left(u + \frac{3s_2t}{s_0 - s_1}\right)p_t - t\left[t^2 - 1 - \frac{3s_2^2}{(s_0 - s_1)^2}\right]p_u,$$

$$Y_8 = tI_{01} + u(I_{12} - I_{02}) + \left[\frac{t^2}{2} + \frac{s_1^2 - s_2^2 - s_0s_1}{(s_0 - s_1)^2}\right]p_t + \left[2tu + \frac{s_2(s_0^2 - s_1^2 + s_2^2)}{(s_0 - s_1)^3}\right]p_u.$$

[1]C. P. Boyer, E. G. Kalnins, and P. Winternitz, "Separation of Variables for the Hamilton Jacobi Equation on Complex Projective Spaces," Preprint CRMA-1064, Montréal, 1981.

[2]S. Benenti and M. Francoviglia, "The Theory of Separability of the Hamilton-Jacobi Equation and Its Application to General Relativity," in *General Relativity and Gravitation*, edited by A. Held (Plenum, New York, 1980), Vol. 1.

[3]E. G. Kalnins and W. Miller, Jr., SIAM J. Math. Anal. **12**, 617 (1981).

[4]C. P. Boyer, E. G. Kalnins, and W. Miller, Jr., Comm. Math. Phys. **59**, 285 (1978).

[5]C. P. Boyer, E. G. Kalnins, and W. Miller, Jr., J. Math. Phys. **19**, 200 (1978).

[6]P. Winternitz and I. Friš, Yad. Fiz. **1**, 889 (1965) [Sov. J. Nucl. Phys. **1**, 636 (1965)].

[7]P. Winternitz, I. Lukač, and Ya. A. Smorodinskii, Yad. Fiz. **7**, 192 (1968) [Sov. J. Nucl. Phys. **7**, 139 (1968)].

[8]W. Miller, Jr., *Symmetry and Separation of Variables* (Addison-Wesley, Reading, Mass., 1977).

[9]J. Patera and P. Winternitz, J. Math. Phys. **14**, 1130 (1973); J. Chem. Phys. **65**, 2725 (1976).

[10]S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry*, (Interscience, New York, 1969), Vol. II.

[11]S. Helgason, *Diffferential Geometry, Lie Groups and Symmetric Spaces* (Academic, New York, 1978).

[12]P. Delong, Ph. D. thesis, University of Minnesota, 1982.

[13]W. Miller, Jr., J. Patera, and P. Winternitz, J. Math. Phys. **22**, 251 (1981).

[14]J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **15**, 1378 (1974).

[15]J. Patera, P. Winternitz, and H. Zassenhaus, Math. Rep. Acad. Sci. (Can.) **2**, 231,237 (1980).

[16]J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **24**, 1973 (1983).

[17]J. Patera, R. T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **17**, 986 (1976).

[18]R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Benjamin, Reading, Mass., 1978).

[19]J. Marsden and A. Weinstein, Rep. Math. Phys. **5**, 121 (1974).

[20]D. Kazhdan, B. Kostant, and S. Sternberg, Comm. Pure Appl. Math. **31**, 481 (1978).

[21]M. P. Olevsky, Mat. Sb. **27**(69), 379 (1950).

[22]P. Winternitz, I. Friš, M. Uhlíř, and Ya. A. Smorodinskiĭ, Yad. Fiz. **4**, 625 (1966) [Sov. J. Nucl. Phys. **4**, 444 (1967)].

[23]P. Kramer, M. Moshinsky, and T. Seligman, "Complex Extensions of Canonical Transformation and Quantum Mechanics," in *Group Theory and Applications*, edited by E. Loebl (Academic, New York, 1975), Vol. 3.

[24]M. Moshinsky, J. Patera, and P. Winternitz, J. Math. Phys. **16**, 82 (1975).

[25]D. Ž. Djokokić, J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **24**, 1363 (1983).

[26]R. L. Anderson, Lett. Math. Phys. **4**, 1 (1980).

[27]R. L. Anderson, J. Harnad, and P. Winternitz, Lett. Math. Phys. **5**, 143 (1981); Physica D **4**, 164 (1982); J. Math. Phys. **23**, 1062 (1982).

[28]P. Winternitz, Physica A **114**, 105 (1982).

[29]A. M. Grundland, J. Harnad, and P. Winternitz, Kinam **4**, 333 (1982).

On a new hierarchy of Hamiltonian soliton equations

M. Boiti, C. Laddomada, F. Pempinelli, and G. Z. Tu[a)]
Dipartimento di Fisica dell'Università, Lecce, Italy and Istituto Nazionale di Fisica Nucleare, Sezione di Bari, Italy

(Received 10 December 1982; accepted for publication 10 February 1983)

A method is suggested for studying the Hamiltonian structure of the nonlinear partial differential equations that can be solved by the use of the spectral transform (soliton equations). The method is applied to a new hierarchy of $N+1$ coupled partial differential equations related to a Schrödinger-like spectral problem. It is shown that these soliton equations are integrable Hamiltonian equations with commuting flows. For $N=1$ and $N=2$ a Miura-like transformation is computed and the corresponding modified equations are explicitly given.

PACS numbers: 02.30.Jr, 02.30.Tb

1. INTRODUCTION

Let us consider evolution equations of the form

$$q_t = s(q), \quad q \in M, \tag{1.1}$$

where the field function $q(x,t)$ is in a linear space M and $s(q) \equiv s(q, q_x, q_{xx}, \cdots)$ is a suitable C^∞ vector field.

One usually calls soliton equations the special evolution equations of the form (1.1), corresponding to isospectral deformations of the linear spectral problem

$$\Psi_x = U(q;\lambda)\Psi, \tag{1.2}$$

where the linear operator U is rational in the spectral parameter λ with singularities at fixed points with fixed multiplicities (see Ref. 1 and references quoted therein).

An infinitesimal isospectral deformation of U

$$U(q(x,t);\lambda) \to U(q(x,t+dt);\lambda) \tag{1.3}$$

can be generated by a gauge transformation of Ψ

$$\Psi \to [1 + dt\, V(q;\lambda)]\Psi. \tag{1.4}$$

Then the operators U and V satisfy the Lax representation

$$U_t - V_x + [U, V] = 0. \tag{1.5}$$

The Lax representation (1.5) can be equivalently obtained (AKNS method[2]) as a consistency condition by requiring that Ψ satisfy the so-called auxiliary spectral equation

$$\Psi_t = V\Psi. \tag{1.6}$$

If one requires that V have fixed rational singularities in λ, Eq. (1.5) can be understood as a system of λ-independent equations on residues of U and V at their fixed singularities in λ.

The solution of (1.5), when it exists, furnishes the explicit form of V and the soliton equation.

The obtained soliton equation can be solved by the use of the spectral transform[2,3] related to the principal spectral problem (1.2).

In general, with different special choices of V, one gets a hierarchy of soliton equations of the form

$$q_t = JL^n p(q) \quad (n = 0,1,\cdots), \tag{1.7}$$

where J is a symplectic operator with respect to a prefixed bilinear form $\langle , \rangle$ and L is an integrodifferential operator.

If $L^n p(q)$ is a potential operator, i.e.,

$$L^n p(q) = \frac{\delta \mathscr{H}_n}{\delta q}, \tag{1.8}$$

by introducing the following Poisson bracket for any two functionals F and G

$$\{F, G\} = \left\langle \frac{\delta F}{\delta q}, J\frac{\delta G}{\delta q} \right\rangle, \tag{1.9}$$

the hierarchy (1.7) is endowed with an Hamiltonian structure.[4–6]

Moreover, if the operators L and J satisfy the "coupling condition"

$$JL = L^+ J, \tag{1.10}$$

where L^+ is the adjoint operator of L with respect to the bilinear form $\langle , \rangle$, it is easy to verify that this coupling condition, together with the definition (1.9) for the Poisson bracket, implies that all the Hamiltonians $\mathscr{H}_n$ in (1.8) are conserved quantities in involution.[4–6]

Consequently, the hierarchy (1.7) consists of integrable Hamiltonian equations with commuting flows.

The crucial statement that one needs to arrive at this result is that $L^n p(q)$ is a potential operator for any n.

This problem was first solved in the framework of the so-called bi-Hamiltonian systems.[4,6]

Subsequently, it has been shown[5] that the two following conditions are sufficient for proving this statement:

(i) L^+ is a Nijenhuis operator or a hereditary symmetry according to the two equivalent definitions of Magri[5] and Fuchssteiner.[6]

(ii) The first two elements $p(q)$ and $Lp(q)$ in the sequence $L^n p(q)$ are potential operators.

However, in order to search for possibly different geometrical structures generating hierarchies of integrable Hamiltonian soliton equations, it is important to find a procedure for proving directly that $L^n p(q)$ is a gradient function for any n.

In fact, one needs only to take into account that the infinite set $\{\mathscr{H}_n\}$ of conserved quantities can be directly derived from the attached linear spectral problem. This procedure has been applied successfully in Ref. 7 to the soliton

[a)] On leave of absence from Computing Center of the Chinese Academy of Sciences, Beijing, China.

 0022-2488/83/082035-07$02.50

equations related to some spectral problems.

In this paper, we consider the spectral problem with $N+1$ independent potentials q_{N-l},

$$-y_{xx}+\sum_{l=0}^{N} q_{N-l}\lambda^{-l}y=\lambda y\,, \tag{1.11}$$

we derive the related soliton equations and, by means of the previously outlined procedure, we prove that they are integrable Hamiltonian equations with commuting flows.

We verify, *a posteriori*, that this case can be described according to the Magri picture.[5]

For $N=1$ and $N=2$, we show that the found soliton equations are quasi-Lagrangian equations[8] in the sense that they can be reduced by a canonical transformation[8] to "modified" soliton equations which are the Euler–Lagrange equations of a variational problem. The modified hierarchy obtained for $N=1$ coincides with that deduced by Itô by using different methods.[9]

2. A NEW HIERARCHY OF SOLITON EQUATIONS

Let us consider the one-dimensional stationary Schrödinger equation

$$-y_{xx}+Qy=\lambda y\,, \tag{2.1}$$

where the potential Q is supposed to be dependent on the spectral parameter λ in the following way:

$$Q=\sum_{l=0}^{N} q_{N-l}(x,t)\lambda^{-l} \tag{2.2}$$

and the potentials $q_l(x,t)$ decrease rapidly as $|x|\to\infty$.

Equation (2.1) can be cast into the form (1.2) by putting

$$\Psi=\begin{pmatrix} y \\ y_x \end{pmatrix} \tag{2.3}$$

and

$$U=\sigma_+ +(Q-\lambda)\sigma_-\,, \tag{2.4}$$

where $2\sigma_+=\sigma_1+i\sigma_2$, $2\sigma_-=\sigma_1-i\sigma_2$. The σ_i's $(i=1,2,3)$ are the Pauli 2×2 matrices.

The t-dependence is fixed by requiring that Ψ satisfy the spectral equation

$$\Psi_t=V\Psi\,, \tag{2.5}$$

where

$$V=a\sigma_3+b\sigma_-+c\sigma_+ \tag{2.6}$$

with a, b, and c functions of the q's and λ.

The compatibility condition for the principal spectral equation (1.2) and the auxiliary spectral equation (2.5) furnishes the Lax representation

$$U_t-V_x+[U,V]=0 \tag{2.7}$$

or, equivalently, the set of equations

$$a=-\tfrac{1}{2}\mathscr{D}c\,, \tag{2.8}$$

$$b=-\tfrac{1}{2}\mathscr{D}^2c+(Q-\lambda)c\,, \tag{2.9}$$

$$Q_t=-\tfrac{1}{2}\mathscr{D}^3c+2\mathscr{D}\mathscr{R}(Q-\lambda)c\,, \tag{2.10}$$

where we denote the x differentiation operator by

$$\mathscr{D}\equiv\frac{\partial}{\partial x} \tag{2.11}$$

and by $\mathscr{R}$ the operator

$$\mathscr{R}(Q)P\equiv QP-\tfrac{1}{2}\mathscr{I}(Q_xP)\,. \tag{2.12}$$

The operator $\mathscr{I}$ is the inverse of $\mathscr{D}$:

$$\mathscr{I}(P)\equiv\frac{1}{2}\left(\int_{-\infty}^{x}-\int_{x}^{+\infty}\right)dx'\,P\,. \tag{2.13}$$

We choose $c(q;\lambda)$ to be of the form

$$c(q;\lambda)=\sum_{j=0}^{n}\lambda^{n-j}c_j(q)\,. \tag{2.14}$$

By introducing for convenience $c_{-N-1}, c_{-N},\ldots,c_{-1}$ and by equating to zero the coefficients of the powers of λ in (2.10), we get the following recursion formula for the c_j's:

$$c_{-N-1}=c_{-N}=\cdots=c_{-1}=0\,, \tag{2.15}$$

$$\mathscr{D}c_{j+1}=-\frac{1}{4}\mathscr{D}^3c_j+\sum_{l=0}^{N}\mathscr{D}\mathscr{R}(q_{N-l})c_{j-l}$$

$$(j=-1,0,\ldots,n-1)\,, \tag{2.16}$$

and the $N+1$ coupled evolution equations

$$q_{N-j,t}=2\sum_{l=j}^{N}\mathscr{D}\mathscr{R}(q_{N-l})c_{n+j-l}\quad(j=0,1,\ldots,N)\,. \tag{2.17}$$

The recursion formula (2.16) can be once integrated to

$$c_{j+1}=-\frac{1}{4}\mathscr{D}^2c_j+\sum_{l=0}^{N}\mathscr{R}(q_{N-l})c_{j-l}+\gamma_{j+1}(t)\,, \tag{2.18}$$

where $\gamma_{j+1}(t)$ is an arbitrary function of t.

By introducing for convenience the vectors

$$C_j=\begin{pmatrix} c_{j-N} \\ c_{j-N+1} \\ \vdots \\ c_j \end{pmatrix}, \tag{2.19}$$

$$q=\begin{pmatrix} q_0 \\ q_1 \\ \vdots \\ q_N \end{pmatrix}, \tag{2.20}$$

and the matrix operator

$$L=\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ \mathscr{R}(q_0) & \mathscr{R}(q_1) & \mathscr{R}(q_2) & \cdots & \mathscr{R}(q_{N-1}) & -\tfrac{1}{4}\mathscr{D}^2+\mathscr{R}(q_N) \end{pmatrix}, \tag{2.21}$$

$$J=\mathscr{D}\begin{pmatrix} 0 & 0 & \cdots & 0 & \mathscr{R}(q_0) \\ 0 & 0 & \cdots & \mathscr{R}(q_0) & \mathscr{R}(q_1) \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & \mathscr{R}(q_0) & \cdots & \mathscr{R}(q_{N-2}) & \mathscr{R}(q_{N-1}) \\ \mathscr{R}(q_0) & \mathscr{R}(q_1) & \cdots & \mathscr{R}(q_{N-1}) & -\frac{1}{4}\mathscr{D}^2+\mathscr{R}(q_N) \end{pmatrix}, \tag{2.22}$$

the recursion formula (2.18) and the evolution equations (2.17) can be rewritten

$$C_{-1}=0\,, \tag{2.23}$$

$$C_{j+1}=LC_j+\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \gamma_{j+1} \end{pmatrix} \quad (j=-1,0,1,\ldots,n-1)\,, \tag{2.24}$$

and

$$q_t=2JC_n\,. \tag{2.25}$$

From (2.23) and (2.24), one easily deduces the explicit form of the C_j's and, therefore, from (2.25) the soliton equation we are looking for. Equations (2.8) and (2.9) furnish the explicit form of the auxiliary operator V.

With a convenient choice of the constants of integration γ_j, i.e., $\gamma_0=1$ and $\gamma_j=0$ for $j\geqslant 1$, one obtains the following hierarchy of soliton equations:

$$q_t=JL^n\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 2 \end{pmatrix} \quad (n=0,1,\cdots)\,. \tag{2.26}$$

We write explicitly the first few of them.

For $n=0$, one gets the trivial equation

$$q_t=q_x\,. \tag{2.27}$$

For $n=1$, one gets

$$\begin{pmatrix} q_0 \\ q_1 \\ \vdots \\ q_{N-1} \\ q_N \end{pmatrix}_t=\begin{pmatrix} q_0q_{N,x}+\frac{1}{2}q_Nq_{0,x} \\ q_{0,x}+q_1q_{N,x}+\frac{1}{2}q_Nq_{1,x} \\ \cdots\cdots\cdots\cdots \\ q_{N-2,x}+q_{N-1}q_{N,x}+\frac{1}{2}q_Nq_{N-1,x} \\ q_{N-1,x}-\frac{1}{4}q_{N,xxx}+\frac{3}{2}q_Nq_{N,x} \end{pmatrix}. \tag{2.28}$$

The more general soliton equation related to the spectral problem (2.4) with the special choice of V that we made in (2.14) is derived by taking a linear combination of Eqs. (2.26) with coefficients which are arbitrary functions of time.

3. REVIEW OF SOME BASIC NOTIONS

The basic notions referred in this section and used throughout the paper can be found in Refs. 4–7 and 10–17.

The $q(x,t)$ defined in (2.20) is considered as a point in the linear space M of the vector-valued field functions regarded as functions of the space coordinate x only. The space M will be called the configuration space associated with the soliton equations (2.26).

$q(x,t)$ is supposed to be defined on the whole real x-axis and to satisfy the homogeneous boundary condition $q(x,t)\to 0$ (rapidly) as $|x|\to\infty$.

To each point q in the configuration space, one associates a tangent space T_q whose elements $\alpha(x,t)=(\alpha_0(x,t),\alpha_1(x,t),\ldots,\alpha_N(x,t))^T$, called controvariant fields, define the infinitesimal displacements of the point $q(x,t)$,

$$\delta q=\epsilon\alpha\,. \tag{3.1}$$

The vector-valued field functions $\alpha(x,t)$ satisfy the same homogeneous boundary conditions as q.

The space of field functions $\beta(x,t)=(\beta_0(x,t),\ldots,\beta_N(x,t))$, obeying the same boundary conditions, which is put in duality with T_q via the bilinear form

$$\langle\beta,\alpha\rangle=\int_{-\infty}^{+\infty}\beta(x,t)\alpha(x,t)\,dx \equiv\sum_{j=0}^{N}\int_{-\infty}^{+\infty}\beta_j(x,t)\alpha_j(x,t)\,dx\,, \tag{3.2}$$

is called cotangent space T_q^* and its elements are named covariant fields.

The Gateaux derivative of an operator G defined in M with values in M or T_q or T_q^* is defined by

$$G'(q)[\alpha]=\frac{d}{d\epsilon}G(q+\epsilon\alpha)\Big|_{\epsilon=0}\,. \tag{3.3}$$

It is also called the directional derivative of G at the point q in the α direction.

Taking into account that the notation $G(q)$, in general, synthesizes a nonlinear differential operator of the form

$$G(q)\equiv G(q,q_x,q_{xx},\cdots)\,, \tag{3.4}$$

the Gateaux derivative is explicitly given by the formula

$$G'(q)[\alpha]=\mathscr{V}(G(q))\alpha\,, \tag{3.5}$$

where the operator $\mathscr{V}$ is defined by

$$\mathscr{V}(G(q))=\sum_{j\geq 0}(\partial_jG)\mathscr{D}^j \tag{3.6}$$

with

$$\partial_j\equiv\frac{\partial}{\partial q_{,j}} \tag{3.7}$$

and

$$q_{,j}\equiv\mathscr{D}^jq\,. \tag{3.8}$$

We next introduce

$$\mathscr{Y}(q)\equiv\sum_{j\geq 0}(-\mathscr{D})^j\partial_j\,, \tag{3.9}$$

the Euler operator of the calculus of variations. For a function $G(q)$, $\mathscr{Y}(q)G(q)$ is the functional derivative of $\int G(q)dx$

$$\frac{\delta}{\delta q}\int_{-\infty}^{+\infty} G(q)dx = \mathscr{Y}(q)G(q) \tag{3.10}$$

aside from boundary terms.

If one considers a point transformation

$$\bar{q} = \bar{q}(q) \tag{3.11}$$

between the configuration space M and a new configuration space $\bar{M}$, the controvariant fields α and the covariant fields β transform according to the formulas

$$\bar{\alpha} = \mathscr{V}(\bar{q}(q))\alpha\,, \tag{3.12}$$

$$\beta = \mathscr{V}^{+}(\bar{q}(q))\,\bar{\beta}\,, \tag{3.13}$$

where $\mathscr{V}^{+}$ is the adjoint operator of $\mathscr{V}$ with respect to the bilinear form $\langle\ ,\ \rangle$.

The Euler operator $\mathscr{Y}$ under the point transformation $\bar{q} = \bar{q}(q)$ transforms as

$$\mathscr{Y}(q) = \mathscr{V}^{+}(\bar{q}(q))\,\mathscr{Y}(\bar{q})\,, \tag{3.14}$$

and therefore $\delta/\delta q$ is a covariant operator.

A covariant operator $f: M \to T_q^*$ that can be expressed as a variational derivative of a scalar operator $F: M \to \mathbf{R}$

$$f = \frac{\delta F}{\delta q} \tag{3.15}$$

is called a potential operator.

It is convenient to introduce the commutator of two controvariant operators f, g: $M \to T_q$:

$$[f,g] = \mathscr{V}(f)g - \mathscr{V}(g)f. \tag{3.16}$$

Under this operation, the controvariant operators constitute a Lie algebra.

An operator $J(q)$: $T_q^* \to T_q$, $q \in M$, is called symplectic if it is skew symmetric,

$$\langle \beta, J\alpha\rangle = -\langle \alpha, J\beta\rangle\,, \tag{3.17}$$

and if the bracket

$$\{\alpha,\beta,\gamma\} = \langle \beta, J'(q)[J(q)\alpha]\gamma\rangle \tag{3.18}$$

satisfies the Jacobi identity for every $\alpha, \beta, \gamma \in T_q$.

The symplectic operator $J(q)$ enables us to introduce the following Poisson bracket for two scalar operators $F, G: M \to \mathbf{R}$:

$$\{F, G\} = \left\langle \frac{\delta F}{\delta q}, J\frac{\delta G}{\delta q}\right\rangle, \tag{3.19}$$

which is antisymmetric and satisfies the Jacobi identity.

Any equation that can be written in the form

$$q_t = J\frac{\delta \mathscr{H}}{\delta q} \tag{3.20}$$

is called a Hamiltonian system.

The flows of two Hamiltonian systems with the same symplectic structure and with Hamiltonians $\mathscr{H}$ and $\mathscr{K}$ are said to commute iff

$$\left[J\frac{\delta \mathscr{H}}{\delta q}, J\frac{\delta \mathscr{K}}{\delta q}\right] = 0\,. \tag{3.21}$$

The controvariant operators $J(\delta\mathscr{H}/\delta q)$ and $J(\delta\mathscr{K}/\delta q)$ can be considered as the infinitesimal generators of a symmetry transformation, respectively, for the Hamiltonian system $\mathscr{K}$ and $\mathscr{H}$.

One can prove, following e.g., Magri,[4] that

$$\left[J\frac{\delta F}{\delta q}, J\frac{\delta G}{\delta q}\right] = J\frac{\delta}{\delta q}\left\langle \frac{\delta F}{\delta q}, J\frac{\delta G}{\delta q}\right\rangle \tag{3.22}$$

and, consequently, if J is invertible, deduce that the two Lie algebras defined by [ ] and { } on the Hamiltonian operators $J(\delta F/\delta q)$ and on the potential operators $\delta F/\delta q$, respectively, are isomorphic. Specifically, two Hamiltonian flows commute iff their corresponding Hamiltonian are in involution.

A linear operator $N(q)$ mapping the tangent space T_q into itself is called a Nijenhuis operator or an hereditary symmetry iff

$$N'(q)[N(q)\alpha]\beta - N(q)N'(q)[\alpha]\beta \tag{3.23}$$

is symmetric with respect to α and β.

A symplectic operator J and a Nijenhuis operator N are said to satisfy the first "coupling condition" iff

$$NJ = JN^{+}\,. \tag{3.24}$$

An operator $N: T_q \to T_q$ is a strong symmetry (or a recursion operator in the terminology of Ref. 13) of an evolution equation $q_t = s(q)$ if it is invariant along the trajectory of the vector field $s(q)$. This is the case iff N satisfies the operator equation

$$N'[s] - s'N + Ns' = 0\,. \tag{3.25}$$

A hereditary symmetry N and a symplectic operator J are said to satisfy the additional "coupling condition" iff

$$\begin{aligned}\langle \alpha, N'[\phi]J\beta\rangle - \langle \alpha, N'[J\beta]\phi\rangle + \langle \beta, N'[J\alpha]\phi\rangle \\ + \langle \beta, NJ'[\phi]\alpha\rangle - \langle \beta, J'[N\phi]\alpha\rangle = 0\,.\end{aligned} \tag{3.26}$$

The proofs that the operator J defined in the previous section is a symplectic operator and that L^{+} is a Nijenhuis operator are given in the Appendices A and B.

The first coupling condition (3.24) can be easily verified, while the proof that the second coupling condition (3.26) is satisfied is sketched in the Appendix C.

According to the Magri general result (1st Theorem of Magri), the nonlinear evolution equations in the hierarchy

$$q_t = JL^{n}p(q) \quad (n = 0,1,\ldots)\,, \tag{3.27}$$

with J a symplectic operator and $N = L^{+}$ a hereditary symmetry satisfying the first coupling condition (3.24), are integrable Hamiltonian equations with commuting flows if $p(q)$ and $Lp(q)$ are potential operators.

If, in addition (2nd Theorem of Magri), J and $N = L^{+}$ satisfy the second coupling condition (3.26), N is a strong symmetry for all the equations in the hierarchy (3.27).

However, in the following section, by taking advantage of the existence of a spectral problem related to the found hierarchy (2.26), we show that the 1st Theorem of Magri can be proved without using the hereditary symmetry property of L^{+}.

Moreover, we are able to derive by a recursion relation the explicit form of the Hamiltonians of the soliton equations and, consequently, to show that, in spite of the integrodiffer-

ential character of the operators L and J, all the soliton equations are local equations.

4. THE HAMILTONIAN STRUCTURE

The soliton equations in the hierarchy (2.26) have an infinite set of polynomial conserved quantities $\{H_n\}$.

The explicit form of the $\{H_n\}$'s can be obtained by solving the Riccati equation

$$h_x + h^2 - (Q - \lambda) = 0 \tag{4.1}$$

that is derived from the spectral equation (2.1) by means of the transformation

$$h = y_x / y . \tag{4.2}$$

From (2.3), (2.5), and (2.6) it results that

$$h_t = b - 2ah - ch^2 , \tag{4.3}$$

and from (2.8), (2.9), and (4.1) that the density h satisfies the conservation law

$$h_t = (-\tfrac{1}{2} c + ch)_x . \tag{4.4}$$

By assuming the formal expansion in λ for h,

$$h = \sum_{j=0}^{\infty} h_j \lambda^{-(1/2)j} + i\lambda^{1/2} , \tag{4.5}$$

we get from (4.1) the following recursion formula for the coefficients h_j:

$$h_0 = 0 ,$$

$$h_{j+1} = \frac{1}{2} i \left(h_{j,x} + \sum_{l=0}^{j} h_l h_{j-l} - \sum_{l=0}^{N} q_{N-l} \delta_{2l,j} \right)$$

$$(j = 0,1,\cdots) , \tag{4.6}$$

which determines uniquely all the polynomial conserved densities h_j and, consequently, the corresponding conserved quantities

$$H_j = \int_{-\infty}^{+\infty} h_j \, dx . \tag{4.7}$$

Moreover, the Riccati equation (4.1) can be used to derive a differential equation for

$$K_l = \frac{\delta H}{\delta q_{N-l}} , \tag{4.8}$$

the functional derivative of $H = \int_{-\infty}^{+\infty} h \, dx$ with respect to q_{N-l}.

From (4.1) we derive the explicit dependence of q_{N-l} on h and its derivatives:

$$q_{N-l}(h) = \lambda^{l} (h_x + h^2 + \lambda - \sum_{\substack{r=0 \\ r \neq l}}^{N} q_{N-r} \lambda^{-r}) . \tag{4.9}$$

It can be used to compute the right-hand side of the identity

$$1 = \frac{\delta H}{\delta h} = \mathscr{Y}^{+}(q_{N-l}(h)) \frac{\delta H}{\delta q_{N-l}} \tag{4.10}$$

obtained taking into account the covariant transformation properties of the gradient operator $\delta / \delta q_{N-l}$ under a point transformation $q_{N-l} = q_{N-l}(h)$.

It results that K_l satisfies the differential equation

$$K_{l,x} = 2hK_l - \lambda^{-l} . \tag{4.11}$$

By differentiating (4.11) two times with respect to x and by using repeatedly the Riccati equation (4.1), one obtains

$$K_{l,xxx} = 4(Q - \lambda) K_{l,x} + 2Q_x K_l . \tag{4.12}$$

If we define the formal series

$$C = \sum_{j=0}^{\infty} C_j \lambda^{-j} \tag{4.13}$$

and we choose $\gamma_0 = 1$ and $\gamma_j = 0$ for $j \geqslant 1$, then the recursion formula (2.24) for the C_j's can be cast into the form

$$C_{xxx} = 4(Q - \lambda) C_x + 2Q_x C . \tag{4.14}$$

The two formal series K and C satisfy the same linear differential equation.

Since

$$H_0 = 0 \tag{4.15}$$

and

$$\frac{\delta H_1}{\delta q} = -\tfrac{1}{2} i C_0 , \tag{4.16}$$

it follows that

$$\lambda^{1/2} \frac{\delta H}{\delta q} = -\tfrac{1}{2} i C . \tag{4.17}$$

By identifying the coefficients of the powers of λ, we obtain that the conserved quantities H_{2n} $(n = 0,1,\cdots)$ are trivially zero and the C_n's are potential operators

$$C_n = L^n C_0 = 2i \frac{\delta H_{2n+1}}{\delta q} . \tag{4.18}$$

Therefore, the soliton equations in the hierarchy (2.26) are Hamiltonian systems, i.e.,

$$q_t = J \frac{\delta \mathscr{H}_{2n+1}}{\delta q} \quad (n = 0,1,\cdots) , \tag{4.19}$$

where the Hamiltonians $\mathscr{H}_{2n+1}$ are directly related to the conserved quantities H_{2n+1},

$$\mathscr{H}_{2n+1} = 4iH_{2n+1} . \tag{4.20}$$

The identity (4.18) can be used to prove that the nontrivial conserved quantities H_{2n+1} are in involution.

The Poisson bracket of a pair of conserved quantities is given by the formula

$$\{H_{2n+1}, H_{2m+1}\} = -\tfrac{1}{4} \langle L^n C_0, JL^m C_0 \rangle . \tag{4.21}$$

The coupling condition

$$JL = L^{+} J \tag{4.22}$$

that can be easily verified induces the following recursion formula on the Poisson bracket (4.21):

$$\{H_{2n+1}, H_{2m+1}\} = \{H_{2n+3}, H_{2m-1}\} , \tag{4.23}$$

and, by iteration $(n < m)$, one finds

$$\{H_{2n+1}, H_{2m+1}\} = \{H_{2m+1}, H_{2n+1}\} \tag{4.24}$$

and then

$$\{H_{2n+1}, H_{2m+1}\} = 0 \tag{4.25}$$

on account of the skew symmetry of the Poisson bracket.

5. QUASI-LAGRANGIAN CHARACTER OF THE SOLITON EQUATIONS FOR $N = 1$ AND $N = 2$

Let us, first, consider the case $N = 1$.

We factorize the symplectic operator J in the following

way:

$$J = \hat{J}L \tag{5.1}$$

with

$$\hat{J} = \begin{pmatrix} \mathscr{D}\mathscr{R}(q_0) & 0 \\ 0 & \mathscr{D} \end{pmatrix}. \tag{5.2}$$

Since the operator $\hat{J}$ is symplectic, the hierarchy (2.26) can be rewritten in another form:

$$q_t = \hat{J}L^n \begin{pmatrix} 0 \\ 2 \end{pmatrix} = \hat{J}\frac{\delta \mathscr{H}_{2n+1}}{\delta q} \quad (n = 1,2,\cdots). \tag{5.3}$$

If we consider the point transformation $q = q(u)$ defined by

$$q_0 = \tfrac{1}{4}u_0^2, \tag{5.4}$$

$$q_1 = u_1, \tag{5.5}$$

the operator $\hat{J}$ can be transformed into the x differential operator $\mathscr{D}$ according to the formula

$$\hat{J} = \mathscr{V}'(q(u))\mathscr{D}\mathscr{V}'^{+}(q(u)). \tag{5.6}$$

Taking into account the controvariant transformation properties of q_t and the covariant transformation properties of the Euler operator $\mathscr{Y}$, we get that the hierarchy (5.3) under the point transformation (5.4) and (5.5) transforms into the hierarchy

$$u_t = \mathscr{D}\frac{\delta \mathscr{H}_{2n+1}}{\delta u} \quad (n = 1,2,\ldots). \tag{5.7}$$

Since this hierarchy can be derived from a variational principle, according to the definition given in Ref. 8, we say that the equations in the hierarchy (5.3) are quasi-Lagrangian systems.

According to the general results on hereditary symmetries,[6] the new hierarchy of Hamiltonian equations (5.7) takes the form

$$\begin{pmatrix} u_0 \\ u_1 \end{pmatrix}_t = \mathscr{D}\tilde{L}^n \begin{pmatrix} u_0 \\ u_1 \end{pmatrix},$$

where

$$\tilde{L} = \mathscr{V}'^{+}(q)L\,[\mathscr{V}'(q)]^{+-1} = \frac{1}{2}\begin{pmatrix} 0 & u_0 \\ u_0 - \mathscr{I}u_{0x} & -\tfrac{1}{2}\mathscr{D}^2 + 2u_1 - \mathscr{I}u_{1x} \end{pmatrix}$$

is the new hereditary symmetry.

We write explicitly the first soliton equation in (5.7):

$$u_{0,t} = \tfrac{1}{2}(u_0 u_1)_x, \tag{5.8}$$

$$u_{1,t} = -\tfrac{1}{4}(u_{1,xx} - 3u_1^2 - u_0^2)_x. \tag{5.9}$$

Let us now consider the $N = 2$ case.

The symplectic operator J is factorized as in the previous case with

$$\hat{J} = \mathscr{D}\begin{pmatrix} 0 & \mathscr{R}(q_0) & 0 \\ \mathscr{R}(q_0) & \mathscr{R}(q_1) & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{5.10}$$

The point transformation $q = q(u)$ defined by

$$q_0 = \tfrac{1}{4}(u_0 + iu_1)^2, \tag{5.11}$$

$$q_1 = \tfrac{1}{4}(u_0^2 + u_1^2), \tag{5.12}$$

$$q_2 = u_2, \tag{5.13}$$

transforms, as previously, the symplectic operator $\hat{J}$ into the $\mathscr{D}$ operator.

Therefore, also in the $N = 2$ case, we can say that the found soliton equations can be transformed via a Miura-like transformation into modified versions which are Lagrangian systems.

The corresponding hierarchy is

$$u_t = \mathscr{D}\tilde{L}^n \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}, \quad u = \begin{pmatrix} u_0 \\ u_1 \\ u_2 \end{pmatrix},$$

where

$$\tilde{L} = \mathscr{V}'^{+}(q)L\,[\mathscr{V}'(q)]^{+-1} = \begin{pmatrix} 1 & i & \tfrac{1}{2}u_0 \\ i & -1 & \tfrac{1}{2}u_1 \\ \tfrac{1}{2}(u_0 - \mathscr{I}u_{0x}) & \tfrac{1}{2}(u_1 - \mathscr{I}u_{1x}) & -\tfrac{1}{4}\mathscr{D}^2 + u_2 - \tfrac{1}{2}\mathscr{I}u_{2x} \end{pmatrix}$$

is the new hereditary symmetry.

We write explicitly the first modified soliton equation in the hierarchy

$$u_{0,t} = (u_0 + iu_1 + \tfrac{1}{2}u_0 u_2)_x, \tag{5.14}$$

$$u_{1,t} = (iu_0 - u_1 + \tfrac{1}{2}u_1 u_2)_x, \tag{5.15}$$

$$u_{2,t} = \tfrac{1}{4}(u_0^2 + u_1^2 - u_{2,xx} + 3u_2^2)_x. \tag{5.16}$$

We conjecture that this result can be extended to any value of N but we have not succeeded in computing the needed point transformation.

APPENDIX A

We prove that the operator J [Eq. (2.22)] satisfies the following Jacobi identity:

$$\langle \alpha, J'[J\beta]\gamma\rangle + (\text{cycle}) = 0,$$

where α, β, and γ are vectors with $N+1$ components which decrease rapidly as $x \to \pm\infty$, and (cycle) means the cyclic permutation over α, β, and γ.

From (2.22) and Definition (3.3) of the Gateaux derivative, we find that the integrand of $\langle \alpha, J'[J\beta]\gamma\rangle + (\text{cycle})$ is

$$\sum_{m=0}^{N}\sum_{j=0}^{m}\sum_{k=0}^{j}\alpha_m(q_k\beta_{k+N-j,x}\gamma_{N+j-m,x} + \tfrac{1}{2}q_{kx}\beta_{k+N-j}\gamma_{N+j-m,x} + \tfrac{1}{2}q_k\beta_{k+N-j,xx}\gamma_{N+j-m} + \tfrac{1}{4}q_{kxx}\beta_{k+N-j}\gamma_{N+j-m} + \tfrac{3}{4}q_{kx}\beta_{k+N-1,x}\gamma_{N+j-m}) - \tfrac{1}{4}\alpha_N(\beta_{Nxxx}\gamma_{Nx} + \tfrac{1}{2}\beta_{Nxxxx}\gamma_N) + (\text{cycle}).$$

By changing the order of summation, it is not difficult to see that the above expression can be transformed into

$$\frac{1}{8}\frac{d}{dx}\left[\sum_{k=0}^{N}(6q_{kx}\phi_k + 4q_k\phi_{kx}) + ((\alpha_N\beta_{Nx} - \alpha_{Nx}\beta_N)\gamma_{Nxx} - \alpha_N\beta_N\gamma_{Nxxx} + (\text{cycle}))\right.$$

where

$$\phi_k = \sum_{m=k}^{N}\sum_{j=k}^{m}\alpha_m\beta_{k+N-j}\gamma_{N+j-m}.$$

Accordingly, the integrand of $\langle \alpha, J'[J\beta]\gamma\rangle + (\text{cycle})$ is a total x-derivative and the desired identity results.

APPENDIX B

We prove that $N \equiv L^+$ is a hereditary symmetry.

Let $(\omega)_i$ represent the ith component of a vector ω, and by $f \simeq g$ we mean that $f - g$ is symmetric with respect to α and β. Then, from (2.21) and (3.3), we see that

$$\begin{aligned}(N'[N\alpha]\beta)_i &= \alpha_{i-1}\beta_N + (q_i\alpha_N\beta_N) + \tfrac{1}{2}q_{ix}\beta_N\mathscr{I}\alpha_N - \tfrac{1}{4}\alpha_{Nxx}\beta_N\delta_{iN} \\ &\quad + \tfrac{1}{2}(\mathscr{I}\beta_N)[\alpha_{i-1,x} + (q_{ix}\alpha_N + q_i\alpha_{Nx}) + \tfrac{1}{2}q_{ixx}(\mathscr{I}\alpha_N) + \tfrac{1}{2}q_{ix}\alpha_N - \tfrac{1}{4}\alpha_{Nxxx}\delta_{iN}] \\ &\simeq \alpha_{i-1}\beta_N - \tfrac{1}{4}\alpha_{Nxx}\beta_N\delta_{iN} + \tfrac{1}{2}(\mathscr{I}\beta_N)[\alpha_{i-1,x} + q_i\alpha_{Nx} + \tfrac{1}{2}q_{ix}\alpha_N - \tfrac{1}{4}\alpha_{Nxxx}\delta_{iN}]\,.\end{aligned}$$

In the same way, we find

$$\begin{aligned}(NN'[\alpha]\beta)_i &\simeq \alpha_{i-1}\beta_N + \tfrac{1}{2}\alpha_{i-1,x}(\mathscr{I}\beta_N) + \tfrac{1}{2}q_i\alpha_{Nx}(\mathscr{I}\beta_N) \\ &\quad + \tfrac{1}{4}q_{ix}\mathscr{I}(\alpha_{Nx}\mathscr{I}\beta_N) - \tfrac{1}{8}\delta_{iN}(\alpha_{Nxxx}\mathscr{I}\beta_N + 2\alpha_{Nxx}\beta_N)\,.\end{aligned}$$

Since

$$\mathscr{I}(\alpha_{Nx}\mathscr{I}\beta_N) = \alpha_N\mathscr{I}\beta_N - \mathscr{I}\alpha_N\beta_N \simeq \alpha_N\mathscr{I}\beta_N\,,$$

we see that

$$(NN'[\alpha]\beta)_i \simeq (N'[N\alpha]\beta)_i\,,$$

which completes the proof.

APPENDIX C

It is convenient to write the second coupling condition between J and $N = L^+$ as follows:

$$\langle \alpha, N'[\phi]J\beta\rangle - \langle \alpha, J'[\phi]L\beta\rangle + \langle \alpha, J'[N\phi]\beta\rangle = \langle \alpha, N'[J\beta]\phi\rangle - \langle \beta, N'[J\alpha]\phi\rangle\,.$$

The terms with ϕ_m $(m = 0,1,\ldots, N-1)$ appear only in the lhs and, thanks to the identity $\mathscr{D}\mathscr{R}(q) = \mathscr{R}^+(q)\mathscr{D}$, they cancel each other. It results that

$$\begin{aligned}&\langle \alpha, N'[\phi]J\beta\rangle - \langle \alpha, J'[\phi]L\beta\rangle + \langle \alpha, J'[N\phi]\beta\rangle \\ &= \int_{-\infty}^{+\infty} dx\left\{\sum_{m=0}^{N}\alpha_m\sum_{k=0}^{m}\mathscr{R}^+(\mathscr{R}^+(q_k)\phi_N)\beta_{N-m+k,x} - \tfrac{1}{4}\alpha_N\phi_{N,xx}\beta_{N,x} - \tfrac{1}{8}\alpha_N\phi_{N,xxx}\beta_N\right\}.\end{aligned}$$

The terms in the rhs can be cast into the form

$$\begin{aligned}\langle \alpha, N'[J\beta]\phi\rangle - \langle \beta, N'[J\alpha]\phi\rangle &= \int_{-\infty}^{+\infty} dx\left\{\sum_{m=0}^{N}\alpha_m\sum_{k=0}^{m}\mathscr{R}^+(\mathscr{R}^+(q_k)\beta_{N-m+k,x})\phi_N\right. \\ &\quad \left. - \tfrac{1}{4}\alpha_N\mathscr{R}^+(\beta_{N,xxx})\phi_N - \sum_{m=0}^{N}\sum_{k=0}^{m}\beta_{N-m+k}\mathscr{R}^+(\mathscr{R}^+(q_k)\alpha_{m,x})\phi_N + \tfrac{1}{4}\beta_N\mathscr{R}^+(\alpha_{N,xxx})\phi_N\right\}.\end{aligned}$$

By direct calculations, one can verify that the terms containing α_m $(m = 0,1,\ldots, N)$ and α_N in the lhs equate, separately, the similar terms in the rhs.

We conclude that the second coupling condition is satisfied.

[1] M. Boiti and G. Z. Tu, Nuovo Cimento B **71**, 253 (1982).
[2] M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Stud. Appl. Math. **53**, 249 (1974).
[3] F. Calogero and A. Degasperis, Nuovo Cimento B **32**, 20 (1976); **39**, 1 (1977).
[4] F. Magri, J. Math. Phys. **18**, 1405 (1977).
[5] F. Magri, "A geometrical approach to the nonlinear solvable equations" in *Lecture Notes in Physics No. 120*, edited by M. Boiti, F. Pempinelli, and G. Soliani (Springer-Verlag, Berlin, 1980).
[6] A. S. Fokas and B. Fuchssteiner, Lett. Nuovo Cimento **28**, 299 (1980); A. S. Fokas and B. Fuchssteiner, Physica D **4**, 47 (1981); A. S. Fokas and R. L. Anderson, J. Math. Phys. **23**, 1066 (1982).
[7] G. Z. Tu, Nuovo Cimento B **73**, 15 (1983); "A simple approach to Hamiltonian structures of soliton equations. II," Sci. Exploration (to appear).
[8] L. Abellanas and A. Galindo, Lett. Math. Phys. **5**, 287 (1981); L. M. Alonso and F. G. Guerrero, J. Math. Phys. **22**, 2497 (1981).
[9] M. Ito, Phys. Lett. A **91**, 335 (1982).
[10] J. M. Gel'fand and I. Ya Dorfman, Functional Anal. Appl. **13**, 248 (1979); **14**, 71 (1980).
[11] B. Fuchssteiner, Nonlinear Anal. **3**, 849 (1979).
[12] G. Z. Tu, J. Phys. A: Math. Gen. **15**, 277 (1982).
[13] P. J. Olver, J. Math. Phys. **18**, 1212 (1977); P. J. Olver, Math. Proc. Cambridge Philos. Soc. **88**, 71 (1980).
[14] H. H. Chen, Y. C. Lee, and C. S. Liu, Phys. Scripta **20**, 480 (1979).
[15] M. D. Arthur and K. M. Case, J. Math. Phys. **23**, 1771 (1982).
[16] W. Wadati, Stud. Appl. Math. **59**, 153 (1978).
[17] M. Bruschi and O. Ragnisco, "The Hamiltonian structure of the nonabelian Toda hierarchy," preprint (Roma, 1982).

Connection between the existence of bisolitons for quadratic nonlinearities and a factorization of the associate linear operator

H. Cornille[a] and A. Gervois[a]
Division de la Physique, Service de Physique Théorique, Cen-Saclay, 91191 Gif-sur-Yvette Cedex, France

(Received 16 April 1982; accepted for publication 15 October 1982)

In a first part we investigate the exponential type bisolitons of the class of equations $(\epsilon + L_q)K = (\epsilon + \Sigma_{i+j=1}^{i+j=q} a_{ij}\,\partial_{x^i t^j}^{i+j})K = (\lambda + \mu\partial_x)K^2$ ($\epsilon \neq 0$, $\lambda \neq 0$ or $\epsilon = \lambda = 0$), which are rational solutions and we assume that their denominators have no coupling between the solitons. We have found that their existence requires a factorization property of either the operator $\epsilon + L_q$ for $\epsilon \neq 0$, $\lambda \neq 0$, $q \leqslant 3$ or L_q for $\epsilon = \lambda = 0$, $q \leqslant 3$ with the exception of the Burgers equation. We find three kinds of nontrivial bisolitons: either those associated with the mixed nonlinearity $(\lambda + \mu\partial_x)K^2$ alone, or those common to the mixed nonlinearity and KK_x or those belonging to KK_x alone. In the second part we look at the two last types of bisolitons for $q > 3$ and give a constructive method leading both to factorized linear operators and to the explicit determination of the bisolitons. We consider mainly $l_{q-2}(t)\,(a\partial_t + b\partial_x + c\partial_{xx})G = (\partial_x G)^2$, $K = G_x$, with the linear operator factorizing the linear part of the Burgers equation, $l_{q-2}(t)$ is a differential operator in variable t only. The general G solution is a linear combination of $\log\Delta$ and Δ^{-i}, $i = 1,\ldots,q-2$ with $\Delta = 1 + \Sigma_1^2 \exp(t + \gamma_i x)$. We determine different classes of solutions as well as the associated $l_{q-2}(t)$ operators.

PACS numbers: 02.30.Jr, 02.30.Tb

I. INTRODUCTION

A. General considerations

In the last 15 years one of the most important success in the study of nonlinear equations is the discovery that for a class of equations, the so-called "completely integrable equations," the solutions can be obtained from linear integral equations. Unfortunately, in general the physically interesting[1] equations do not belong to that class. For nonintegrable equations, maybe a method could be the extension of the direct Hirota[2] method originally developed for the integrable case. It is not clear that this extension can be generally done, outside some ad hoc model equations[3] not too far from the integrable case. The truth could be that, in fact, the nonintegrable nonlinear equation must be divided in different classes and that for each class one has to find appropriate and specific methods. It is such a particular class that we propose to study here. A way to classify nonlinear equations may be to investigate their bisolitons which certainly are nontrivial solutions.

Let us call exponential type solitons the solutions rational in the variable $\exp(\gamma x + \rho t)$ and exponential type bisolitons the rational solutions with the variables $\omega_i = \exp(\gamma_i x + \rho_i t)$. A further physical restriction is certainly that these solutions vanish when $\omega \to \infty$ or $\omega_i \to \infty$, but we do not retain this condition *a priori*.

We have recently[4] obtained general features for the existence of bisolitons associated with the quadratic nonlinear equations $(\epsilon + L_q)K = (\epsilon + \Sigma\, a_{ij}\,\partial_{x^i t^j}^{i+j})K = K^2$, $\epsilon \neq 0$, which correponds to a model Boltzmann[5,6] equation for $q = 2$. On the one hand, the nontrivial bisolitons have denominators of the type $(1 + \omega_1 + \omega_2)^q$ without the coupling const. $\times\ \omega_1\omega_2$ between the two solitons; on the other hand, the linear operators $\epsilon + L_q$ for either K or $\epsilon + K$ are necessarily factorized. This property was entirely proved for $q \leqslant 3$, and conjectured as always true for $q > 3$, and we explicitly exhibit classes of bisolitons having these properties. It is interesting to know whether these properties, factorization of the linear operator linked to denominator powers of $1 + \omega + \omega_2$, are particular to K^2 or are true for a larger class of nonlinearities. If the answer is positive, we will have defined a class of nonlinear nonintegrable equations with common features.

In this paper we investigate other quadratic nonlinearities which are KK_x and a nonlinearity which mixes K^2 and KK_x, and, as we shall see, the factorization property is still the main fact of the study. A preliminary brief report,[5] without proofs, was previously done. We study the class of nonlinear equations in $1 + 1$ variables:

$$(\epsilon + L_q)K = \left(\epsilon + \sum_{i+j=1}^{i+j=q} a_{ij}\,\partial_{x^i t^j}^{i+j}\right)K = (\lambda + \mu\partial_x)K^2 \tag{1.1}$$

with either $\epsilon \neq 0$, $\lambda \neq 0$ for the mixed nonlinearity or $\epsilon = \lambda = 0$ for KK_x alone. *Further we assume that the possible bisolitons have denominators which are powers of* $1 + \omega_1 + \omega_2$, and this is a restriction because we know for KK_x that the KdV equations have not this kind of denominator. In the first part of the paper we systematically investigate all possible bisolitons of Eq. (1.1) for $q \leqslant 3$, the result being that *if we except the Burgers equation, all other bisolitons found are associated with factorized linear operators.* In the second part of the paper, analyzing these results, we are able to develop a general formalism for $q > 3$, with factorized linear operators, giving us the possibility of explicitly determining the bisolitons. *The key point is that these linear operators factorize the linear part of the Burgers equation.*

[a] Chercheur au C. N. R. S.

 0022-2488/83/082042-14$02.50

Let us notice that in Eq. (1.1) if we substitute $K = \epsilon/\lambda + \tilde{K}$, we find for $\tilde{K}$ another equation of the same type, $(-\epsilon + \tilde{L}_q)\tilde{K} = (-\epsilon + L_q - 2\mu\epsilon\partial_x)\tilde{K} = (\lambda + \mu\partial_x)\tilde{K}^2$. Similarly, if $\epsilon = \lambda = 0$, $K = C + \tilde{K}$ leads to $\tilde{L}_q \tilde{K} = (L_q - 2C\mu\partial_x)\tilde{K} = \mu\partial_x \tilde{K}^2$. We do not distinguish between these two cases; for instance, the factorization property must be understood either for K or $\tilde{K}$. Similarly, for each soliton or bisoliton of K there exists a partner $\tilde{K}$. We note also that the nonlinearity in (1.1) is invariant under the transform $(\partial_x, \partial_t) \to (\partial_x, \partial_t + \text{const.}\,\partial_x)$, and all our results must be understood modulo these transforms.

B. Solitons and bisolitons

For a nonlinearity $(\lambda + \mu\partial_x)K^2$ there exist different types of solutions. First, those which are common to K^2: from a solution $(\epsilon + L_{q-1})K = K^2$, we trivially obtain $(\epsilon\lambda + L_q)K = (\lambda + \mu\partial_x)(\epsilon + L_{q-1})K = (\lambda + \mu\partial_x)K^2$, and in the following we do not consider this case. Secondly, those solutions which are specific to the mixed case and do not exist when $\lambda \to 0$. Thirdly, those common to the mixed nonlinearity and KK_x, which still survive when $\lambda \to 0$, $\epsilon \to 0$. In the last class are those specific to KK_x alone, and they are the most interesting because KK_x is the physically interesting nonlinearity. In order to avoid any misunderstanding, let us be precise that when we speak of KK_x alone, we mean practically always $\epsilon = 0$. This is due to the fact that if in (1.1) we put $\lambda = 0$ and $\epsilon \neq 0$, we have found that no solitons exist. This can also be seen directly from the explicit expression of the bisoliton in the mixed case: In fact, ϵ then disappears and we have to take the limit $\lambda \to 0$.

In Sec. II, in order to emphasize a closed connection between these quadratic nonlinearities K^2 with $\epsilon \neq 0$, KK_x with $\epsilon = 0$, and $(\lambda + \mu\partial x)K^2$, we compare their exponential type solitons $K = K(\omega), \omega = \exp\xi, \xi = \gamma x + \rho t$. For KK_x the solitons equations, integrated with respect to ξ lead to differential equations with K^2 nonlinearity. It is why the same finite number of algebraic forms exist for both KK_x and K^2. For $(\lambda + \mu\partial x)K^2$ in supplement to these solitons there appear two new kinds of solutions which are specific to that nonlinearity: first, a class of solitons with algebraic forms depending continuously on a parameter μ/λ and, secondly, a class which we call "bizarre solitons" because they have unusual asymptotic behavior—they are going to infinity when $\omega \to \infty$. These "bizarre solitons" are in finite number, and their existence is due to the presence of both nonlinearities K^2 and KK_x.

In Sec. III we investigate the possible bisolitons of Eq. (1.1) for $q \leqslant 3$, assuming that their denominators are powers of $1 + \omega_1 + \omega_2$. The method used in order to find the bisolitons has been fully[4-6] explained previously for the K^2 nonlinearity. It essentially consists of writing K as a polynomial in one variable, let us say ω_2

$$K = \sum_{l}^{l_{\max}} \omega_2^l F_l(Z), \quad Z = \omega_1 + \omega_2, \tag{1.2a}$$

and, substituting into the nonlinear partial differential equation (n.l.p.d.e.), we find that the coefficients $\{F_l(Z)\}$ satisfy systems of nonlinear differential equations, with one variable Z only. More explicitly, first the linear application $K \to (\epsilon + L_q)K$ leads to another polynomial

$$(\epsilon + L_q)K = \sum_{n}^{l_{\max}+q} \omega_2^m \overline{F}_m(Z), \tag{1.2b}$$

where $\overline{F}_m(Z)$ is a linear combination of $F_0, F_1, \ldots, F_{l\max}$ and the derivatives of various order with respect to Z. Secondly, the non-linear part of (1.1) can be rewritten

$$\begin{aligned}(\lambda + \mu\partial_x)K^2 &= \Sigma\, \omega_2^m \overline{\overline{F}}_m(Z),\\ \overline{\overline{F}}_m(Z) &= (\lambda + \mu m\gamma_2 + \mu\gamma_1 Z\partial_z)\sum_{p=0}^{m} F_p F_{m-p}\\ &\quad + \mu(\gamma_2 - \gamma_1)\partial_z \sum_{p=0}^{m-1} F_p F_{m-p-1}.\end{aligned} \tag{1.2c}$$

It follows that Eq. (1.1) is equivalent to $\overline{F}_m(Z) \equiv \overline{\overline{F}}_m(Z)$. For the set $\{F_l\}$ this represents $l_{\max} + q + 1$ coupled *nonlinear* differential equations in Z, becoming linear for the special m values such that $\overline{\overline{F}}_m \equiv 0$. In the pure K^2 case, the number of $\overline{F}_m = 0$, linear $\{F_l\}$ constraints, was sufficient to provide a factorization of $(\epsilon + L_q)$. It is not true in the full $(\lambda + \mu\partial_x)K^2$ case, and the factorization appears only after the complete resolution of the n.l.p.d.e. In Sec. 3, inessential technical modifications are performed. Mainly, in order to symmetrize the nonlinearity, we define $\omega_i = \exp x_i$, $i = 1, 2$, and the formalism is applied to (1.1) in the new variables x_1, x_2. Here we do not report the details and give only the results back in the (x,t) variables for $q = 2,3$.

(i) We begin with the bisolitons specific to the mixed nonlinearity without partners for KK_x. For $q = 2$ we have found two "bizarre bisolitons" and six for $q = 3$ with always factorized linear operators (see Table I). We still call them "bizarre" because they are going to infinity when both $\omega_1 \to \infty$ and $\omega_2 \to \infty$.

(ii) We go on with the bisolitons common to the mixed nonlinearity ($\lambda \neq 0$) and to KK_x ($\lambda = 0$). We have not found such an object for $q = 2$ but only for $q = 3$, which still has a factorized linear part. We write down the result in a potential formulation $K = (\lambda + \mu\partial_x)G$:

$$\begin{aligned}&(1 + \partial_t)\left[\lambda^2 + a\partial_t + (\lambda\mu + b)\partial_x + \tfrac{1}{6}\mu^2\partial_{xx}\right]G\\ &\quad = \left[(\lambda + \mu\partial_x)G\right]^2, \quad G = \Delta^{-1},\end{aligned} \tag{1.3}$$

$$a\Delta_t + b\Delta_x + \tfrac{1}{6}\mu^2\Delta_{xx} = 0, \tag{1.4}$$

$$\Delta = 1 + \omega_1 + \omega_2, \quad \omega_i = \exp(t + \gamma_i x),$$

$$a + b\gamma_i + \tfrac{1}{6}\mu^2\gamma_i^2 = 0.$$

(iii) We finish with the bisolitons specific to KK_x (or G_x^2). We have found only one: It has an associated factorized linear part, and we write down the result for the potential $G_x = K$:

$$\begin{aligned}&(1 + \partial_t/6)(a\partial_t + b\partial_x + c\partial_{xx})G = C(\partial_x G)^2,\\ &G = -\log\Delta + \Delta^{-1}\end{aligned} \tag{1.5}$$

still with the relations (1.4), $\mu^2/6$ being replaced by C.

We have found *only one example without factorized linear part. It is the Burgers equation* for $q = 2$, which we write down in the two formulations K and $G_x = K$:

$$(a\partial_t + b\partial_x - \mu\partial_{xx})K = \mu\partial_x K^2, \quad K = \Delta_x/\Delta, \tag{1.6a}$$

TABLE I. List of the bizarre bisolitons for $q = 2, 3$.

$\tilde{L}_q K = (-2 + \partial_x)K^2 \quad \Delta = 1 + \omega_1 + \omega_2 \quad \omega_1 = e^{x+t} \quad \omega_2 = e^{x-t}$	
$q = 2$	
$K_1 = \omega_2^2/6\Delta$	$\tilde{L}_2 = (1 - (\partial_x - \partial_t)/4)(1 - (\partial_x - \partial_t)/6)$
$K_2 = \omega_1\omega_2/\Delta$	$\tilde{L}_2 = (1 - (\partial_x + \partial_t)/2)(1 - (\partial_x - \partial_t)/2)$
$q = 3$	
$K_1 = \omega_2^3/10\Delta^2$	$\tilde{L}_3 = (1 - (\partial_x - \partial_t)/6)(1 - (\partial_x - \partial_t)/8)(1 - (\partial_x - \partial_t)/10)$
$K_2 = (\omega_2^2/23\Delta^2)(6 + 5\omega_1 + 5\omega_2)$	$\tilde{L}_3 = (1 + \frac{1}{23}\partial_x)(1 - (\partial_x - \partial_t)/4)(1 - (\partial_x - \partial_t)/6)$
$K_3 = \omega_2^2\omega_1/\Delta^2$	$\tilde{L}_3 = (1 - (\partial_x + \partial_t)/2)(1 - (\partial_x - \partial_t)/4)(1 - (\partial_x - \partial_t)/6)$
$K_4 = (\omega_2^2/6\Delta^2)(1 + \omega_2 + 2\omega_1)$	$\tilde{L}_3 = (1 - (\partial_x + \partial_t)/12)(1 - (\partial_x - \partial_t)/4)(1 - (\partial_x - \partial_t)/6)$
$K_5 = (3\omega_2^2/50\Delta^2)(1 + 2\omega_2 + \omega_1)$	$\tilde{L}_3 = (1 - (\partial_x - \partial_t)/20)(1 - (\partial_x - \partial_t)/10)(1 - (\partial_x - \partial_t)/4)$
$K_6 = \frac{8}{23}(\omega_1\omega_2/\Delta^2)(6 + 5\omega_1 + 5\omega_2)$	$\tilde{L}_3 = (1 + \frac{1}{23}\partial_x)(1 - (\partial_x + \partial_t)/2)(1 - (\partial_x - \partial_t)/2)$

$$(a\partial_t + b\partial_x - \mu\partial_{xx})G = \mu(\partial_x G)^2, \quad G = \log\Delta, \tag{1.6b}$$

with

$$\Delta = 1 + \Sigma\,\omega_i, \quad (a\Delta_t + b\Delta_x - \mu\Delta_{xx}) = 0,$$
$$a\rho_i + b\gamma_i - \mu\gamma_i^2 = 0\,, \tag{1.6c}$$

and we remark that (1.6c) contrary to (1.4) can be satisfied by an arbitrary number of ω_i terms. We can have multisolitons and not only bisolitons.

C. The Burgers equation as a germ for a class of Eq. (1.1) with factorized linear operators

Summarizing the above results concerning Eq. (1.1) and $q \leqslant 3$, it follows that, with the exception of the Burgers case, all other equations with bisolitons correspond to factorized linear operators. In the second part of the paper, we provide a general method for $q > 3$, leading to bisolitons associated with factorized linear operators and which include (1.3), (1.5), and (1.6) for $q = 2$ or 3. We are interested either in bisolitons common to the mixed nonlinearity and KK_x or specific to KK_x or $(G_x)^2$ alone. The main fact is that for $\lambda = 0$ the built linear operators factorize the linear part of the Burgers equations.

Let us define $G(\Delta)$, $\Delta = 1 + \Sigma\,\omega_i$, satisfying

$$l_{q-2}(t)G = \left(\sum_0^{q-2} b_i\partial_{t^i}\right) G(\Delta) \equiv H(\Delta),$$
$$\partial^2_{\Delta\Delta}\, H = \nu(G_\Delta)^2\,, \tag{1.7}$$
$$a\Delta_t + b\Delta_x + c\Delta_{xx} = 0\,,$$

and introduce a second-order differential operator

$$\tilde{l}_2 = \lambda^2 + a\partial_t + (b + \lambda\mu)\partial_x + c\partial_{xx},$$
$$b = 0 \text{ or } \neq 0,\ a \neq 0,\ c \neq 0, \tag{1.8}$$

which for $\lambda = 0$ reduces to the linear Burgers operator.

We want to build a class of nonlinear equations with factorized linear operator $L_q = l_{q-2}\,\tilde{l}_2$ and including the results (1.3), (1.5), and (1.6).

(i) We assume $H \equiv G, \lambda = 0, \nu = -1, C = -\mu$. From (1.7), and (1.8) we find $l_{q-2} \equiv 1$, $G = \log\Delta$ and we recover the Burgers equations (1.6).

(ii) We assume $H = G^2$; then $\nu c = \mu^2$, $\omega_i = \exp(t + \gamma_i x)$, $i = 1,2$. From (1.7), (1.8) we find

$$l_{q-2}\,G = G^2, \quad (G^2)_{\Delta\Delta} = \nu(G_\Delta)^2,$$
$$a + b\gamma_i + c\gamma_i^2 = 0\,, \tag{1.9}$$
$$l_{q-2}\,\tilde{l}_2 G = [(\lambda + \mu\partial_x)G]^2\,.$$

The general solution of (1.9) is easily obtained, $G = \Delta^{-q+2}$, $\nu = 2(2q-3)/(q-2)$, and we obtain, for the potentials G, a set of nonlinear equations valid in both cases $\lambda \neq 0$, $\lambda = 0$:

$$\prod_0^{q-3}\left(1 + \frac{\partial_t}{q - 2 + p}\right) \times\left(\lambda^2 + a\partial_t + (b + \lambda\mu)\partial_x + \frac{\mu^2(q-2)}{2(2q-3)}\partial_{xx}\right) G = [(\lambda + \mu\partial_x)G]^2, \quad G = \Delta^{2-q}\,, \tag{1.10}$$

which reduces to (1.3) for $q = 3$.

(iii) We assume $H \neq G$, $H \neq G^2$, which necessarily leads to $\lambda = 0$ and the nonlinearity KK_x alone or equivalently to the potential G and nonlinearity $(G_x)^2$. The study is done in Sec. IV. Further, we assume $\omega_i = \exp(t + \gamma^x)$ or $\Delta_t = \Delta - 1$, $a + b\gamma_i + c\gamma_i^2 = 0$, and the problem is reduced to the resolution of the equation

$$\frac{\partial^2}{\partial\Delta^2}\left(\sum_0^{q-2} b_i\,\partial_{t^i}\right) G(\Delta) = \nu\left(\frac{\partial G}{\partial\Delta}\right)^2, \quad b_0 = 1\,. \tag{1.11}$$

(1.11) is, in fact, a nonlinear quadratic differential equation depending on only one variable Δ. For the nonlinearity K^2 we have previously encountered similar equations and in Sec. IV we provide two methods in order to solve it. For each q value there exists a finite number of solutions, and, consequently, with $l_{q-2}\,\tilde{l}_2$ we can associate a finite number of n.l.p.d.e. with nonlinearities $(G_x)^2$ or KK_x. This number increases quickly with q: one for $q = 2$, two for $q = 3$, six for $q = 4$, twenty for $q = 5, \cdots$.

Consider, for instance, $q = 3$ and $\nu = 1$; then from (1.11) we find $G = -\log\Delta + \Delta^{-1}$ and we recover the nonlinear equation (1.5).

For each solution of Eq. (1.11) we have to find the set of b_i values or the operator l_{q-2}, and this is explained in Sec. IV. The general solutions of (1.11) can be written

TABLE II. Families of potential functions for KK_x or $(G_x)^2$ nonlinearity for any q.

$L_qG_q \equiv l_{q-2}\tilde{l}_2G_q = \nu_q(\partial_x G_q)^2$

$\tilde{l}_2 = a\partial_t + b\partial_x + c\partial_{xx}^2$, $\quad \Delta = 1 + \omega_1 + \omega_2$, $\quad \omega_i = e^{t+\gamma_i x}$, $\quad a + b\gamma_i + c\gamma_i^2 = 0$, $\quad i = 1, 2$

$G_{1,q} = \dfrac{1}{\Delta^{q-2}}$, $\quad l_{q-2} \equiv \prod_{l=0}^{q-3}\left(1 + \dfrac{\partial_t}{q-2+l}\right)$, $\quad \nu_q = 2(2q-3)/(q-2)$, $\quad q \geqslant 3$

$G_{2,q} = \dfrac{1}{\Delta^{q-2}} + \dfrac{p}{q-3}\cdot\dfrac{1}{\Delta^{q-3}}$, $\quad l_{q-2} \equiv \left(1 + \dfrac{(q-2)\partial_t}{3(q-2)(q-2+p) + 2(p-q+2)}\right)\prod_{l=0}^{q-4}\left(1 + \dfrac{\partial_t}{q-3+l}\right)$

$q \geqslant 4$, $\quad \nu_q = \dfrac{4(2q-5)(2q-3)}{3(q-2)(q-2+p) + 2(p-q+2)}$

$p^2(q-2) + p(2q^2 - 9q + 8) + (q-2)^2(q-3) = 0$

$G_{3,q} = \dfrac{1}{\Delta^{q-2}} + \dfrac{p}{q-3}\dfrac{1}{\Delta^{q-3}}$

$l_{q-2} = \prod_{l=0}^{p-1}\left(1 + \dfrac{\partial_t}{q-3+l}\right)\prod_{j=0}^{q-p-5}\left(1 + \dfrac{\partial_t}{p+q-2+l}\right)(1 + \tau_+\partial_t)(1 + \tau_-\partial_t)$

$q \geqslant 5$, $\quad p = 1, 2, \ldots, q-4$, $\quad \nu_q = \dfrac{2(q+p-3)(2q-5)(2q-3)}{(q-2)(q-3)(2q-5) + p(2q-3)(2q+p-6)}$

$[2(q-2)(q-3)(2q-5) + 2p(2q-3)(2q+p-6)](\tau_\pm)^2 - [(4q-11)(q-2) + 2p(2q-3)]\tau_\pm + (q-2) = 0$

$G_{4,q} = -\log\Delta + \sum_{k=1}^{q-2}\dfrac{1}{k\Delta^k}$, $\quad l_{q-2} = 1 + \sum_1^{q-2} b_l\partial_t{}^l$, $\quad q \geqslant 2$, $\nu_q = c$

$\bar{b}_0 = b_1$, $\quad \bar{b}_l = \sum_{j=l}^{q-3}\mathscr{C}_j^l b_{j+1}$ $\quad (l \geqslant 1)$, $\quad \sum_{l=m}^{q-3}\bar{b}_l\dfrac{(q-2+l)!}{(q-2)!}C_l^m(-1)^{l+m} = \dfrac{q-2-m}{(q-1+m)(q+m)}$

$\mathscr{C}_j^l$ stirling numbers

$$G = -\gamma\log\Delta + \sum_1^{q-2}\frac{\alpha_i}{\Delta^i} \tag{1.12}$$

and in Sec. IV we determine different classes of such solutions. The most simple one, $G = \Delta^{-(q-2)}$, was written down in (1.10) and corresponds to $H = G^2$. There is a very simple solution which generalizes the Burgers' ones (with the restriction that only two different ω_i are allowed), $G = -\log\Delta + \Sigma_{i=1}^{q-1}(1/i\Delta^i)$, and the associated l_{q-2} operator is determined in Sec. IV. Seeking solutions which are linear combinations of $\Delta^{-(q-3)}$ and $\Delta^{-(q-2)}$, we have found two general classes $G = p/(q-3)\Delta^{q-3} + 1/\Delta^{q-2}$: one for $q \geqslant 4$ for p complex and the other for $q \geqslant 5$ with p integer (see Table II). There exist also general classes of solutions which include three Δ^{-i}s. We have found in Sec. IV two such classes:

$$G = \frac{1}{(q-2)\Delta^{q-4}}\left(\frac{1}{\Delta^2} + \frac{1}{\Delta} + \frac{1}{(q-4)}\right) \quad \text{for } q \geqslant 6,$$

$$G = \frac{1}{(q-2)\Delta^{q-4}}\left(\frac{1}{\Delta^2} + \frac{2}{(q-3)\Delta} + \frac{2}{(q-3)(q-4)}\right)$$

TABLE III. Families of potential functions for KK_x or $(G_x)^2$ nonlinearity for any q when G has three Δ terms.

$L_qG_q = l_{q-2}\tilde{l}_2G_q = \nu_q(\partial_x G_q)^2$

$\tilde{l}_2 = a\partial_t + b\partial_x + c\partial_{xx}^2$, $\quad \Delta = 1 + \omega_1 + \omega_2$, $\quad \omega_i = e^{t+\gamma_i x}$, $\quad i = 1, 2$, $\quad a + b\gamma_i + c\gamma_i^2 = 0$

$G_{5,q} = \dfrac{1}{(q-2)}\left[\dfrac{1}{\Delta^{q-2}} + \dfrac{2}{(q-3)\Delta^{q-3}} + \dfrac{2}{(q-3)(q-4)\Delta^{q-4}}\right]$

$l_{q-2} = \left(1 + \dfrac{\partial_t}{q-4}\right)\prod_{l=0}^{q-8}\left(1 + \dfrac{\partial_t}{q-1+l}\right)\left(1 + \sum_{j=1}^{j=4}\sigma_j\partial_{t^j}\right)$

$q \geqslant 7$, $\quad \nu_q = 2(2q-7)(2q-6)(2q-5)(2q-4)(2q-3)\sigma_4$

$\sigma_3 - [8q - 26 + 4(2q-3)/(q-2)]\sigma_4 = 0$

$\sigma_2 - (6q-21)\sigma_3 + [24q^2 - 168q + 295 - 8(2q-5)(2q-3)/(q-2)(q-4)]\sigma_4 = 0$

$\sigma_1 - (4q-15)\sigma_2 + [12q^2 - 90q + 169]\sigma_3 - [(4q-15)(8q^2 - 60q + 113) + (2q-5)(2q-3)/(q-2)(q-3)]\sigma_4 = 0$

$1 - (2q-8)\sigma_1 + (2q-8)^2\sigma_2 - (2q-8)^3\sigma_3 + [(2q-8)^4 - 16(2q-5)(2q-3)/(q-2)(q-3)]\sigma_4 = 0$

$G_{6,q} = \dfrac{1}{(q-2)}\left[\dfrac{1}{\Delta^{q-2}} + \dfrac{1}{\Delta^{q-3}} + \dfrac{1}{(q-4)\Delta^{q-4}}\right]$

$l_{q-2} = \left(1 + \dfrac{\partial_t}{q-4}\right)\prod_{l=0}^{q-7}\left(1 + \dfrac{\partial_t}{q-2+l}\right)\left(1 + \sum_{j=1}^{3}\sigma_j\partial_{t^j}\right)$

$q \geqslant 6$, $\quad \nu_q = 2(2q-7)(2q-6)(2q-5)(2q-4)(2q-3)\sigma_3/(q-2)$

$\sigma_2 + [31 - 10q - 2(q-3)/(q-2)]\sigma_3 = 0$

$\sigma_1 + [4q + 14]\sigma_2 + \{18(q-2)^2 - 51(q-2) + 37 - (q-3)(6q-21) - [2(2q-3)/(q-2)][(q-3)^2 + 2(q-2)]\}\sigma_3 = 0$

$1 + (12 - 2q)\sigma_1 + (12-2q)^2\sigma_2 + [4(q-3)(2q-3)(2q-5)]\sigma_3 = 0$

for $q \geqslant 7$ and the associated l_{q-2}

(see Table III).

The method we give in Sec. IV is general and can generate for the KK_x nonlinearity other general classes of four $\Delta^{-i}, \cdots$. It essentially consists of building progressively the operator $l_{q-2}(\partial_t)$ in its factorized form $l_{q-2}(\partial_t) \equiv \Pi_{i=1}^{q-2}(1+\tau_i \partial_t)$. Let us remark that both $\tilde{l}_2 G(\Delta)$ and $(G_x)^2$ are polynomials in Δ^{-m}, but $\tilde{l}_2 G(\Delta)$ has lower Δ^{-m} terms which do not belong to $(G_x)^2$. Consequently, we proceed in two successive stages: First, we build a part $\Pi_{i=1}^{i_{\max}}(1+\tau_i \partial_t)$ in such a way that $\Pi_1^{i_{\max}}(1+\tau_i \partial_t)\tilde{l}_2 G$ has only Δ^{-m} powers which appear too in $(G_x)^2$; in a second step, coefficients τ_i of the last part $\Pi_{i_{\max}+1}^{q-2}(1+\tau_i \partial_t)$ of l_{q-2} are chosen so that they rearrange the coefficients of the remaining Δ^{-m} powers in order to be identical to $(G_x)^2$. It is worthwhile to notice a significant difference between the Burgers equation and all the other generated by our general scheme. The Burgers equation has only one linear constraint $a\Delta_t + b\Delta_x + c\Delta_{xx} = 0$ and, consequently, can lead to multisolitons or to an arbitrary number N of (γ_i, ρ_i) with denominator power of $(1 + \Sigma_1^N \omega_i)$. On the contrary, in all other cases we have another linear constraint $\Delta_t = \Delta - 1$ which restricts us to have only bisolitons or the γ_i cannot have more than two values. Coming back to the general problem of the possible bisolitons of (1.1) for $\lambda = 0$ or KK_x nonlinearity alone, we have verified in the first part of the paper for $q \leqslant 3$ (denominators powers of Δ) that they are all given by our general scheme developed in the second part of the paper. Is this property true for any q value? Unfortunately, we cannot verify it in the general case.

D. Possible generalizations for other quadratic nonlinearities?

Besides $K^2, KK_x, (\lambda + \mu\partial x)K^2$, is it possible that other quadratic nonlinearities carry out the same features for their bisolitons: do denominator powers of Δ lead to factorized linear operators? As a trivial remark, our results, which were expressed in potential formulation, show that this is true at least for $(K_x)^2$ with the same Burgers exception. As a curiosity we quote for $q \geqslant 3$ a general bisoliton solution for the nonlinearity KK_{xx} as well as a mixed nonlinearity $\lambda K^2 + 2\mu KK_{xx}$ which is going the same way as the present paper.

$$\prod_0^{q-3}\left(1+\frac{\partial_t}{q-2+p}\right)\left(\lambda + a\partial_t + b\partial_x + \frac{\mu(q-2)}{2q-3}\partial_{xx}\right)K = \lambda K^2 + 2\mu KK_{xx}, \quad K = \Delta^{2-q}$$

$$\Delta = 1 + \omega_1 + \omega_2, \quad \omega_i = \exp(t + \gamma_i x), \quad i = 1,2,$$

$$a\Delta_t + b\Delta_x + \frac{\mu(q-2)}{(2q-3)}\Delta_{xx} = 0,$$

$$a + b\gamma_i + \frac{\mu(q-2)}{2q-3}\gamma_i^2 = 0. \tag{1.13}$$

We have both cases $\lambda \neq 0$ and $\lambda = 0$. It is amusing to notice that for $\lambda = 0$ the linear operator still factorizes the linear Burgers operator and solution Δ^{2-q} as well as the same operator l_{q-2} were present in Eq. (1.10).

II. SOLITONS

The connection between the solutions of (1.1) with either of the three nonlinearities K^2, KK_x, or $\lambda K^2 + 2\mu KK_x$ can be seen first in the case of solitons, i.e., rational functions of a unique variable $\omega = e^{\xi}$, $\xi = \gamma x + \rho t$, γ and ρ being parameters to be determined. The relation between the cases

$$L_q K = \mu\partial_x K^2, \quad L_q = \sum_{i+j=1}^{q} a_{ij}\partial_{x^i t^j} \tag{2.1}$$

and

$$(\epsilon + L_{q-1})K = \lambda K^2, \quad L_{q-1} = \sum_{i+j=0}^{q-1} b_{ij}\partial_{x^i t^j} \tag{2.2}$$

is straightforward and is explained in the first subsection.

The link with the "mixed" case

$$(\epsilon + L_q)K = (\lambda + \mu\partial_x)K^2 \tag{2.3}$$

is less obvious. When $\lambda = 0$ or $\mu = 0$, we recover the solitons of KK_x and K^2; but when λ and μ are not zero, other solitons appear together to the previous ones. These new solitons are of two kinds. First, we get a continuous family, depending on ratio μ/λ. Secondly, we find "bizarre" solitons with unusual asymptotic behavior (they are not bounded when $\omega \to \infty$).

A. KK_x and K^2 nonlinearities

Equations (2.1) and (2.2) are rewritten, using variable ξ, on the same formal aspect

$$\sum_{k=1}^{q} A_k \frac{\partial^{k-1}K}{\partial\xi^{k-1}} = \bar{\lambda}K^2, \tag{2.4}$$

where

$$A_k = \sum_{i+j=k} a_{ij}\gamma^i\rho^j \quad \text{and} \quad \bar{\lambda} = \gamma\mu \tag{2.5}$$

in the first case (2.1) and

$$A_k = \sum_{i+j=k-1} b_{ij}\gamma^i\rho^j \quad \text{and} \quad \bar{\lambda} = \lambda \tag{2.6}$$

in the second case (2.2). This means that, in terms of the variable ω or ξ, solitons of (2.1) and (2.2) have the same intrinsic or canonical form. But the specific values of γ and ρ and the (possible) conditions on constants (a_{ij}) are different and the "dressed" solitons are not the same.

As an example, we treat completely case $q = 3$ for KK_x corresponding to L_2 for K^2. It is easy to see that $\lim_{\omega\to\infty} K(\omega)$ is 0 or A_1/λ, and it can be taken as zero through the invariance $K \to A_1/\lambda - K$. K is sought as an expansion[4] $K = \Sigma_{n\geq 0}\eta_n(-\omega)^n$ and η_n is a polynomial in n of degree 0 or 1. This leads to the four possible canonical forms:

$$\begin{aligned}
&\text{(I)} \quad K = \delta/(1+\omega), \quad A_3 = 0, \quad A_1 = A_2, \quad \delta\bar{\lambda} = A_2,\\
&\text{(II)} \quad K = \delta/(1+\omega)^2, \quad A_1 = 6A_3, \quad 5A_1 = 6A_2, \quad \delta\bar{\lambda} = A_1,\\
&\text{(III)} \quad K = \delta\omega/(1+\omega)^2, \quad A_2 = 0,\\
&\qquad A_1 = -A_3, \quad \delta\bar{\lambda} = 6A_1,\\
&\text{(IV)} \quad K = \delta(1+2\omega)/(1+\omega)^2, \quad A_1 = -6A_3,\\
&\qquad 5A_1 = 6A_2, \quad \delta\bar{\lambda} = A_1,
\end{aligned} \tag{2.7}$$

where we have explained the intrinsic relations [for $q = 2$,

we should have only soliton (I)]. Using definitions (2.5) and (2.6) for the A_k, we get the possible values of γ, ρ, and $\delta\bar{\lambda}$ in terms of the (a_{ij}). In the KK_x case, we get six "dressed" solitons for each intrinsic form (2.7), while in the K^2 *case* we get only two "dressed" solutions. No supplementary condition on the (a_{ij}) arises, but this is no more true for $q \geqslant 4$. Notice that in L_3 we note that (IV) contains the famous Bobylev–Krook–Wu soliton[7] of the Boltzmann equation[6] and (III) contains the classical soliton of the "completely integrable" KdV. More generally, if L_3 has no second-order derivatives (KdV or IKdV), $A_2 \equiv 0$, and only soliton (III) exists; $A_2 \not\equiv 0$ if, for instance, a dissipative term ∂_x^2 is present.

B. The complete nonlinearity $(\lambda + \mu\partial_x)K^2$

Going now to the solitons of the mixed equation (2.3), we get the intrinsic equation

$$\sum_{k=0}^{q} A_k \frac{\partial^k}{\partial \xi^k} K = (\lambda + \mu\gamma\partial_\xi)K^2, \tag{2.8}$$

where the A_k are given by (2.5) and $A_0 = \epsilon$.

We first consider the solitons which are bounded when $\omega \to \infty$, and, up to an additive constant, we write K as an expansion $\Sigma_{n\geqslant 0}\ \eta_n\ (-\omega)^n$ and η_n is a polynomial in n of degree $q-2$ at most. We restrict ourselves again to $q=3$. When $\mu \neq 1$, we get first the canonical forms (2.7) with new conditions on the A_k:

(I′) $A_3 = 0, \quad \lambda A_2 = \epsilon\mu = (A_1 - \epsilon)\lambda\ , \quad \lambda\delta = \epsilon,$

(II′) $\lambda A_3 = (\epsilon/6)\mu/\lambda, \quad \lambda A_2 = (\epsilon/6)(\lambda + 5\mu),$

$\lambda A_1 = \epsilon(5\lambda/6 + \mu), \quad \lambda\delta = \epsilon,$

(III′) $\lambda A_3 = -\lambda A_1 = \epsilon\mu, \quad \lambda A_2 = -\epsilon, \quad \delta\lambda = 6\epsilon,$

(IV′) $\lambda A_3 = -\frac{\epsilon}{6}\mu, \quad 6\lambda A_2 = \epsilon(-\lambda + 5\mu),$

$$\lambda A_1 = \epsilon\left(\frac{5\lambda}{6} + \mu\right), \quad \lambda\delta = \epsilon. \tag{2.9}$$

We get a supplementary bisoliton, not present in the previous cases, depending on the ratio μ/λ

$$\text{(V)} \quad K = \delta\{[\lambda + (\lambda - \mu)\omega]/(1+\omega)^2\} \tag{2.10a}$$

with the intrinsic equations

$$A_3 = (\epsilon/6)\mu^2/\lambda^2, \quad A_2 = \frac{\mu\epsilon}{\lambda}, \quad A_1 = \epsilon(1 + \mu/\lambda - \mu^2/6\lambda^2),$$
$$\delta\lambda^2 = \epsilon, \tag{2.10b}$$

from which we can determine parameters γ, ρ, and μ/λ with no supplementary condition on the a_{ij} and get at most 24 "dressed" bisolitons.

If $\lambda = 0$, we can verify in (2.9) and (2.10) that *necessarily* $\epsilon = 0$, which means that the nonlinearity KK_x has no exponential type soliton when the linear part has no term proportional to K. When both $\epsilon, \lambda \to 0$ in such a way that the ratio ϵ/λ is a constant, we recover the solitons (I) to (IV).

We recover them for $\mu = 0$ too. These solitons are the solitons of $L_2K = \lambda K^2$, and they are the solitons of $L_3K = \lambda K^2$ with denominator $(1+\omega)^{-1}$ and $(1+\omega)^{-2}$; the solitons of $L_3K = \lambda K^2$ having denominator $(1+\omega)^{-3}$ cannot be reached here.

It is easy to see that another class of solitons exists, not bounded for large ω because of the presence of *two* nonlinearities. They are presumably not so interesting both from the physical point of view (they have no corresponding term in the KK_x case) nor from the mathematical point of view (analyticity disappears near $\omega = 0$ in the exchange $\omega \leftrightarrow \omega^{-1}$). For nonlinearity $\lambda K^2 + 2\mu KK_x$, we find a finite number of solutions

$$k_1\omega + k_0 + H(\omega),$$

where $H(\omega)$ is a bounded rational function: $\lim_{\omega\to\infty} H(\omega) = 0$. We have the further relation $\lambda + 2\mu\gamma = 0$ and the rhs of (2.8) is always proportional to $(-2 + \partial_\xi)$. Their number depends on q and for $q = 2$, we have (up to the invariance $K \to \epsilon/\lambda - K$) the unique solution

$$\text{(VI)} \quad K = -(\epsilon/3\lambda)\omega^2/(1+\omega), \quad A_1 = 5\epsilon/6, \quad A_2 = \epsilon/6,$$

corresponding to the identity

$$\left(1 - \frac{\partial_\xi}{2}\right)\left(1 - \frac{\partial_\xi}{3}\right)\left(\frac{\omega^2}{1+\omega}\right) = (-2 + \partial_\xi)\left(\frac{\omega^2}{1+\omega}\right)^2.$$

When $q = 3$, we have soliton (VI) with $A_3 = 0$ plus the three independent solutions

$$\text{(VII)} \quad K = \frac{2}{23}\frac{\epsilon}{\lambda}\frac{(5\omega^3 + 6\omega^2)}{(1+\omega)^2}, \quad A_1 = -\frac{29}{6\times 23}\epsilon,$$
$$A_2 = -\frac{3}{23}\epsilon, \quad A_3 = -\frac{\epsilon}{6\times 23},$$

$$\text{(VIII)} \quad K = -\frac{\epsilon}{5\lambda}\frac{\omega^3}{(1+\omega)^2}, \quad A_1 = -\frac{47}{60}\epsilon,$$
$$A_2 = \frac{\epsilon}{5}, \quad A_3 = -\frac{\epsilon}{60},$$

$$\text{(IX)} \quad K = -\frac{\epsilon}{25\lambda}\frac{(6\omega^3 + 3\omega^2)}{(1+\omega)^2}, \quad A_1 = -\frac{4\epsilon}{5},$$
$$A_2 = \frac{17}{100}\epsilon, \quad A_3 = -\frac{\epsilon}{100}.$$

Besides these "bizarre" solitons exist an infinite number of binomial solutions $k_p\ \omega^p + k_0$, $p = 1,2,3,\ldots, k_0 = 0$ or ϵ/λ, which have no practical interest. We shall not consider them anymore.

III. NONTRIVIAL BISOLITONS OF $(\epsilon + L_q)K = (\lambda + \mu\partial_x)K^2$ FOR $q = 2,3$

In this section, we seek systematically bisolitons of the full nonlinearity, assuming that their denominators are powers of $1 + \omega_1 + \omega_2$,

$$(\epsilon + L_q)K = \lambda K^2 + 2\mu KK_x \tag{3.1}$$

for $q = 2,3$. We examine successively cases $\epsilon, \lambda \neq 0$ and $\epsilon = \lambda = 0$. For $\epsilon, \lambda \neq 0$, we shall verify that ϵ is factorized both on the lhs and rhs of (3.1), so it disappears in the balance and two cases may occur. Either λ and μ_i disappear too in the balance (3.1); then Eq. (3.1) contains no parameters, only numerical coefficients, and no limit exists when $\lambda \to 0$. We shall see that this happens for "bizarre" bisolitons (relaxing to the bizarre solitons of Sec. II). Or λ and μ_i still exist and we can go to the limit $\lambda = 0$; in that latter case, solutions of (3.1) for $\epsilon, \lambda \neq 0$ are prolonged into solutions for $\epsilon = \lambda = 0$.

At first sight, we expect bisolitons coming from four different directions:

(i) "Trivial" bisolitons, i.e., bisolitons for L_{q-1} $K = K^2$ with $\epsilon + L_q \equiv (\lambda + \mu\partial_x)L_{q-1}$. By construction, they yield a factorization of L_q; they are still present when taking the limit $\lambda = 0$.

(ii) Bisolitons of $(\epsilon + L_q)K = (\lambda + \mu\partial_x)K^2$ for $\epsilon, \lambda \neq 0$, which exist in the limit $\lambda = 0$. We shall see that $\epsilon + L_q$ is then factorized and that K is derived from a potential function G: $K = (\lambda + \mu\partial_x)G$.

(iii) Bisolitons of the KK_x case only. Except for the Burgers case, L_q is factorized. We shall see examples for any q in the last section.

(iv) Bisolitons specific of the mixed case $K^2 + KK_x$. In general, they relax to bizarre solitons (VI)-(IX) of Sec. II. They disappear when $\lambda = 0$, as they come from the presence of *two* nonlinearities. For $q = 2,3$, we proved that they lead to a factorization of $\epsilon + L_q$.

Other bisolitons perhaps exist, but, for $q = 2,3$, they are the only possibilities, up to the invariance $K \to \epsilon/\lambda - K$, $\omega_1 \leftrightarrow \omega_2 \cdots$ and they lead—Burgers case excepted—to a factorization of L_2 and L_3.

A systematic method was described in Refs. 4 and 6. It was recalled in the introduction [Eqs. (1.2)]. In order to simplify the formalism, we have symmetrized the nonlinearity. For that purpose, we take advantage of the invariance property of the linear part under linear transformations of the (x,t) variables. By the linear transformations $(x,t) \to (x_1,x_2)$, $x_i = \gamma_i x + \rho_i t$, $\omega_i = \exp x_i$, L_q becomes another differential operator

$$L_q = \sum_{i+j=1}^{q} a_{ij}\, \partial_{x_1^i x_2^j}, \tag{3.2}$$

where the new a_{ij} have a new meaning and are trivially deduced from the old ones; for simplicity, we keep the same notations in what follows. Equations (3.1) and (1.2b,c) are rewritten

$$(\epsilon + L_q)K = \left(\lambda + \sum_{i=1}^{2} \mu_i \partial_{x_i}\right) K^2 \tag{3.3}$$

and

$$\bar{F}_m(Z) = (\lambda + \mu_2 m + \mu_1 Z\partial_Z) \sum_{p=0}^{m} F_p(Z) F_{m-p}(Z) + (\mu_2 - \mu_1)\partial_Z \sum_{p=0}^{m-1} F_p(Z) F_{m-p-1}(Z), \tag{3.4}$$

respectively.

A. Case $q = 2$

When $\epsilon, \lambda \neq 0$, besides the trivial bisoliton $K = (\epsilon/\lambda)(1/\Delta)$ solution of $L_1 K = \lambda K^2$, we find no bisoliton common to $K^2 + KK_x$ and KK_x but only two bisolitons specific for the mixed nonlinearity. Their derivation is explained in the Appendix. We obtain

$$K = \epsilon/(-6\mu)\frac{\omega_2^2}{\Delta} \quad \text{and} \quad K = -\frac{\epsilon}{2\mu}\frac{\omega_1\omega_2}{\Delta},$$
$$\Delta = 1 + \omega_1 + \omega_2. \tag{3.5a}$$

These solutions have an unusual asymptotic behavior, and one of them relaxes to the bizarre soliton (VI). They require

$$\mu_1 = \mu_2 = \mu = -\lambda/2 \tag{3.5b}$$

and factorize on both sides of (3.1) the factor ϵ^2/μ, the rhs operator being proportional to $(-2 + \partial_{x_1} + \partial_{x_2})$; of course, we cannot go to the limit $\lambda = 0$.

As an example, we have for the first bisoliton, the identity

$$\epsilon\left(1 - \frac{\partial_{x_2}}{2}\right)\left(1 - \frac{\partial_{x_2}}{3}\right)\left[-\frac{\epsilon}{6\mu}\frac{\omega_2^2}{\Delta}\right] = \mu(-2 + \partial_{x_1} + \partial_{x_2})\left(-\frac{\epsilon}{6\mu}\frac{\omega_2^2}{\Delta}\right)^2,$$

and both bisolitons may be found in Table I (first and second lines) with the ordinary (x,t) variables.

When $\epsilon = \lambda = 0$, the bizarre bisolitons disappear, and we get one solution specific of the KK_x case

$$\left[(\mu_1^2\partial_{x_1} + \mu_2^2\partial_{x_2}) - (\mu_1\partial_{x_1} + \mu_2\partial_{x_2})^2\right](\mu_1\omega_1 + \mu_2\omega_2)/\Delta = (\mu_1\partial_{x_1} + \mu_2\partial_{x_2})[(\mu_1\omega_1 + \mu_2\omega_2)/\Delta]^2,$$

which is the Burgers solution with the change of variables $\partial_x = \mu_1\partial_{x_1} + \mu_2\partial_{x_2}$ and $\partial_t = \mu_1^2\partial_{x_1} + \mu_2^2\partial_{x_2}$. We rewrite it in the well-known form

$$(\partial_t - \partial_{x^2}^2)K = \partial_x K^2, \quad K = \partial_x \log\Delta$$

or

$$(\partial_t - \partial_{x^2}^2)G = (G_x)^2,$$

where we have put into evidence the potential function $G = \log\Delta$.

B. Case $q = 3$

Besides the trivial bisolitons, obtained with a monomial form[6] $(K = \omega_2 F_1)$, we get bisolitons in the three classes recalled at the beginning of the section. We have

1. A continuous family common to $K^2 + KK_x$ and KK_x

Both K and $\epsilon + L_3$ factorize parameter ϵ/λ^2. We may drop it in (3.1), and we get the identity

$$\left(1 + \sum_i \partial_{x_i}\right)\left[\lambda^2 + \lambda\sum\mu_i\partial_{x_i} - \tfrac{1}{6}\sum\mu_i^2\partial_{x_i} + \tfrac{1}{6}\left(\sum\mu_i\partial_{x_i}\right)^2\right] \times\left\{\left[\lambda + \sum(\lambda - \mu_i)\omega_i\right]/\Delta^2\right\} \equiv (\lambda + \sum\mu_i\partial_{x_i})\left\{\left[\lambda + \sum(\lambda - \mu_i)\omega_i\right]/\Delta^2\right\}^2 \tag{3.6}$$

and K may be derived from a potential function G

$$K = (\lambda + \sum\mu_i\partial_{x_i})G, \quad G = \Delta^{-1}. \tag{3.7}$$

We can take the limit $\lambda \to 0$, and the solution survives giving the new identity

$$\left(1+\sum \partial_{x_i}\right)\left[\sum \mu_i^2 \partial_{x_i} - \left(\sum \mu_i \partial_{x_i}\right)^2\right]\left(6\sum \mu_i \omega_i/\Delta^2\right)$$
$$= \left(\sum \mu_i\, \partial_{x_i}\right)\left(6\sum \mu_i\, \omega_i/\Delta^2\right)^2, \tag{3.8}$$
$$K = \left(\sum \mu_i\, \partial_{x_i}\right) G, \quad G = \Delta^{-1}.$$

2. A family specific to the KK_x case

When setting directly $\epsilon = \lambda = 0$ in (3.1), we find a bisoliton specific of KK_x without the corresponding solution for the mixed nonlinearity

$$K = (\mu_1\omega_1 + \mu_2\omega_2)(1+\Delta)/\Delta^2, \quad K = \left(\sum \mu_i \partial_{x_i}\right) G,$$
$$G = \log\Delta - 1/\Delta, \tag{3.9}$$

where $G \equiv G(\Delta)$ is the potential function. Solutions (3.6) and (3.8) are gotten with a restricted expansion $K = F_0 + \omega_2 F_1$, and the results can be checked with the formulae of the Appendix and using Table V, which gives the general expression of $(\epsilon + L_3)K$ when $F_2 \equiv F_3 \equiv 0$.

3. Bisolitons specific to the $K^2 + KK_x$ case

Up to all invariances (mainly $\omega_1 \leftrightarrow \omega_2$, $K \to \epsilon/\lambda - K$), we get six bisolitons which are not present in the pure K^2 or KK_x case. There is a supplementary condition

$$\mu_1 = \mu_2 = \mu = -\lambda/2$$

as for $q = 2$, so that the rhs operator is proportional to $(-2 + \partial_{x_1} + \partial_{x_2})$; both sides factorize $(\epsilon/\lambda^2)^2$, and we can drop this normalization factor. They are written in Table I (lines 3–8) in the (x,t) variables together with the associated linear operator $\tilde{L}_3 = 1 + L_3$; the parameter λ has completely disappeared and the limit $\lambda = 0$ cannot be taken. Like bizarre solitons, they have an unusual behavior when $\omega_i \to \infty$; moreover, four of them relax to bizarre solitons (VI)–(IX) when $\omega_1 = 0$. Their derivation is sketched at the end of the Appendix.

We have verified that no other bisoliton exists up to all invariances for $q = 2,3$. The calculation is long and tedious but can be done with the general formulae of the Appendix. Notice that solitons belonging to different kinds cannot be coupled to form a bisoliton and that each soliton appears in one bisoliton and one only.

IV. GENERAL SOLUTIONS IN THE KK_x CASE

The observation of the bi-solitons for $q = 2$ and $q = 3$ has shown that they belong to four kinds: (a) trivial bisolitons, solutions of $L_{q-1}K = K^2$, (b) bisolitons specific of the mixed case $\epsilon, \lambda \neq 0$ which may grow indefinitely for large ω_i and which disappear at the limit $\lambda \to 0$, (c) bisolitons common to KK_x and $K^2 + KK_x$, and (d) bisolitons which appear in KK_x only. We consider here bisolitons of kinds (c) and (d) only and generate classes of solutions for any q which are generalizations of solutions (3.6), (3.7), and (3.9). The key point is that they derive from a potential function $G(\Delta)$ of the unique variable $\Delta = 1 + \omega_1 + \omega_2$, $K = (\lambda + \mu\partial_x)G$ or $K = \partial_x G$ and the linear part L_q factorizes the second-order linear operator $\tilde{l}_2 = \lambda^2 + a\partial_t + (b + \lambda\mu)\partial_x + c\partial_{xx}^2$ of Eq. (1.8), which for $\lambda = 0$ reduces to the Burgers equation.

Potential $G(\Delta)$ is defined through Eqs. (1.7), and, as pointed out in the Introduction, there are only three possibilities for $H = l_{q-2}(t)G$: (i) $H \equiv G$; then $l_{q-2} \equiv 1$ and $\lambda = 0$ and we have the Burgers equation; (ii) $H \equiv G^2$, which gives *both for* $\epsilon, \lambda \neq 0$ *and* $\epsilon = \lambda = 0$ the unique solution $G = \Delta^{2-q}$, Eqs. (1.9) and (1.10); and (iii) $H \not\equiv G, G^2$. This latter case implies necessarily $\epsilon = \lambda = 0$, i.e., the *bisolitons (iii) are solutions for the KK_x case alone* and not for the mixed case. This is not a great restriction, as the physically interesting cases enter this category. This section is devoted to the study of some classes of solutions (iii) for any q, still assuming that the denominators are powers of Δ.

A. General features

We recall that the potential function $G(\Delta)$ is solution of the intrinsic equation (1.11).

$$\frac{\partial^2}{\partial\Delta^2} l_{q-2}(t) G(\Delta) = \frac{\partial^2}{\partial\Delta^2}\left(\sum_{i=0}^{q-2} b_i\, \partial_{t^i}\right) G(\Delta)$$
$$= \nu_q \left(\frac{\partial G}{\partial \Delta}\right)^2, \quad b_0 = 1, \tag{4.1}$$
$$l_{q-2}\,\tilde{l}_2 G = \nu_q (G_x)^2, \quad \tilde{l}_2 = a\partial_t + b\partial_x + c\partial_{xx}^2,$$
$$a\rho_i + b\gamma_i + c\gamma_i^2 = 0, \tag{4.2}$$

and the general solution has the form (1.12)

$$G = \sum_{k=1}^{q-2} \frac{\alpha_k}{\Delta^k} - \gamma \log\Delta \tag{4.3}$$

with $\alpha_{q-2} \neq 0$. When $q = 2$, we have the only solution $\log\Delta$ (Burgers) and, when $q = 3$, the two solutions $1/\Delta$ and $1/\Delta - \log\Delta$. Actually, their number increases very quickly with q as for $q = 4$ and $q = 5$, we find 6 and 20 solutions, respectively. A rapid estimation for $q = 6$ gives 35 solutions with $\gamma \neq 0$ and presumably the same number without the logarithmic terms! In the following, we shall investigate for any q the simplest cases where only α_{q-2}, or α_{q-2}, and α_{q-3}, or $\alpha_{q-2}, \alpha_{q-3}, \alpha_{q-4}$ ··· (subsections IVB and IVC) are present. Two methods are given, one of them was already used for nonlinearity K^2 in Ref. (4). For q small the existence of a nonvanishing logarithmic term requires that the (α_k) are all nonzero; when q increases, we may expect that most (α_k) are nonzero too. It is then impossible to give here all the solutions, but we can exhibit a peculiar one which generalizes both the Burgers solution for $q = 2$ and solution (3.9), $G \equiv -\log\Delta + 1/\Delta$ for $q = 3$, subsection IVD. All these solutions have been gathered in Tables II and III; we wrote them in a self-contained way so that the identities could be directly checked. Finally, in subsection IVE we give, for illustration, all solutions for $q = 4$.

We end this subsection with a more technical remark. There is some freedom in the choice of the time variable t as the rhs of $L_q K = \mu\partial_x K^2$ is invariant under any transformation $\partial_t \to \text{const}_1\partial_t + \text{const}_2\partial_x$. A convenient choice is

$$\omega_i = \exp(t + \gamma_i x), \quad i = 1,2, \quad \Delta_t = \Delta - 1, \tag{4.4}$$

and due to (4.2) we have always only two γ_i values. Then (4.1)

is a nonlinear differential equation *in variable Δ only*. We shall assume (4.4) from now on. Then, for any derivative $\partial_t^k F$ of any function $F(\Delta)$ we have

$$\partial_t^k F \equiv \sum_{l=1}^{k} (\Delta - 1)^l \mathscr{C}_k^l \partial_\Delta^l F, \tag{4.5}$$

where the $\mathscr{C}_k^l$ are the Stirling's numbers of the second kind,[8] which we have already encountered in the K^2 nonlinearity.[4]

B. Solutions with two Δ terms

We assume here that

$$G(\Delta) = 1/\Delta^{q-2} + \beta/\Delta^{q-3}, \tag{4.6}$$

all other terms being zero (for simplicity, we have set $\alpha_{q-2} = 1$, $\alpha_{q-3} = \beta$). For $\beta = 0$ and $q = 3$, we recover the solution $G = \Delta^{2-q}$, which is still valid when $\epsilon, \lambda \neq 0$ [see Eq. (1.10) and solution $G_{1,q}$ of Table II]. When $\beta \neq 0$, Eq. (4.1) splits into q numerical equations for the q variables $b_1,\ldots,b_{q-2}$, ν_q, and β and we may expect a finite number of solutions. We give here two methods; the first one was already used for deriving solutions with the K^2 nonlinearity.[4]

1. First method

Noticing that $\partial_t = Z\partial_Z$, where $Z = \omega_1 + \omega_2 = \Delta - 1$, Eq. (4.1) may be rewritten as

$$\left(\sum_{k=0}^{q-2} \mu_k Z^k \partial_{Z^{k+2}}\right) G = \nu_q G_Z^2,$$

where $\mu_0, \mu_1,\ldots,\mu_{q-2}$ are triangular linear combinations of

$$b_0 = 1, b_1,\ldots,b_{q-2} \quad [\mu_{q-2} = b_{q-2},\ \mu_{q-3} = b_{q-3} + 2(q-2)b_{q-2}\cdots].$$

Introducing the quantities

$$\Sigma_p = \sum_{k=p}^{q-2} \mu_k (-)^k (q-3)(q-2)\cdots(q-2+k) C_k^p,$$

$$\chi_p = \sum_{k=p}^{q-2} \mu_k (-)^k (q-2)\cdots(q-1+k)\, C_k^{p-1},$$

where $C_m^n = \binom{m}{n}$ is the binomial coefficient, we get $(q-3)$ linear equations without the rhs,

$$\begin{aligned} &\beta\Sigma_0 = 0, \\ &-\beta\Sigma_1 + \chi_0 = 0,\cdots, \\ &(-)^p(\beta\Sigma_p - \chi_{p-1}) = 0, \quad p = 1,2,\ldots,q-4, \end{aligned} \tag{4.7a}$$

and three non linear equations with the rhs,

$$\begin{aligned} &(-)^{q-3}(\beta\Sigma_{q-3} - \chi_{q-4}) = \beta^2(q-3)^2\nu_q, \\ &(-)^{q-2}(\beta\Sigma_{q-2} - \chi_{q-3}) = 2(q-2)(q-3)\beta\nu_q, \\ &(-)^q \chi_{q-2} = (q-2)^2\nu_q. \end{aligned} \tag{4.7b}$$

Notice that $\beta\Sigma_0 = 0$ excepted, Eqs. (4.7a) exist only when $q \geqslant 5$ while (4.7b) holds already when $q = 4$.

The resolution goes in the following way:

As $\beta \neq 0$, the first equation gives $\Sigma_0 = 0$, then, through some easy combinatorial manipulations on Stirling numbers $\mathscr{C}_k^l$, we get $\chi_0 = \Sigma_1/(q-3)$. Then, the second equation (4.7a) is rewritten $[\beta - 1/(q-3)]\Sigma_1 = 0$ and splits into two possibilities: either $\beta = 1/(q-3)$ or $\Sigma_1 = 0$. Choosing $\Sigma_1 = 0$, we get $\chi_1 = [2/(q-3)]\Sigma_2$, and the third equation gives in turn $[\beta - 2/(q-3)]\Sigma_2 = 0$ and another bifurcation: Either $\beta = 2/(q-3)$ or $\Sigma_2 = 0$, i.e., $\chi_2 = [3/(q-3)]\Sigma_3 \cdots$. After solving the $(q-3)$ first equations, we have found $q-4$ bifurcations $\beta = 1/(q-3), 2/(q-3),\cdots, (q-4)/(q-3)$ (provided $q \geqslant 5$) and, for $q \geqslant 4$, another way which corresponds to $\Sigma_0 = \Sigma_1 = \cdots = \Sigma_{q-4} = 0$ with $\chi_0 = \Sigma_1(q-3), \chi_1 = 2\Sigma_2/(q-3),\cdots, \chi_{q-4} = \Sigma_{q-3}$.

It is easy to see that the possibilities $\beta = 1/(q-3), 2/(q-3),\ldots, p/(q-3)\cdots\beta = (q-4)/(q-3)$ are actually solutions as system (4.7) is now linear and homogeneous in the (μ_k) and ν_q parameters and the determinant is zero. The first equation $\Sigma_o = 0$ fixes then ν_q as μ_0 is not independent of $\mu_1,\ldots,\nu_{q-2}$ (as $b_0 = 1$ is fixed). We get solutions $G_{3,q}$ of Table III with $p = 1,2,\ldots,q-4$ $(q \geqslant 5)$.

As to the last possible solution, we have to solve the three last equations (4.7b). Noticing that Σ_{q-2} and χ_{q-2} are proportional to μ_{q-2}, we are left with three variables, μ_{q-2}/ν_q, Σ_{q-3}/ν_q, and β. We get β as any of the roots of the second degree equation

$$\beta^2(q-2)(q-3) + \beta(2q^2 - 9q + 8) + (q-2)^2 = 0 \tag{4.8}$$

with $G = 1/\Delta^{q-2} + \beta/\Delta^{q-3}$ [Eq. (4.6)], and we obtain the two solutions $G_{2,q}$ of Table II.

2. Second method

The above calculation give all the possible solutions of (4.1), but we obtain the operator $l_{q-2}(t)$ or its coefficients b_i through a complicated linear (triangular) system. We may now directly construct $l_{q-2}(t)$ in its factorized form

$$l_{q-2}(t) = \prod_{i=1}^{q-2} (1 + \tau_i \partial_t). \tag{4.9}$$

We use the identities

$$\begin{aligned} &\tilde{l}_2(1/\Delta^{q-2} + \beta/\Delta^{q-3}) \\ &\quad = c(q-2)(\Delta_x)^2 [(q-1)/\Delta^q + \beta(q-3)/\Delta^{q-1}] \\ &\quad \equiv c(q-2)F[\beta(q-3), q-2], \end{aligned} \tag{4.10a}$$

$$\begin{aligned} &(1 + \tau\partial_t)(\Delta_x^2/\Delta^m) \\ &\quad = \Delta_x^2 \{[1 - \tau(m-2)]/\Delta^m + \tau m/\Delta^{m+1}\}, \end{aligned} \tag{4.10b}$$

$$\begin{aligned} &[1 + \partial_t/(m-1)] F(p,m) \\ &\quad = [(m+1)/(m-1)] F(p-1, m+1), \end{aligned} \tag{4.10c}$$

where

$$F(p,m) \equiv \Delta_x^2 [p/\Delta^{m+1} + (m+1)/\Delta^{m+2}]. \tag{4.10d}$$

As the final expression $L_q G = l_{q-2}\tilde{l}_2 G$ contains only the inverse powers $\Delta^{2-2q}, \Delta^{1-2q}, \Delta^{-2q}$ of the various Δ, it is necessary to kill, at each step $(1 + \tau_i\partial_t)$ but the last one, the lowest term in the powers of Δ^{-1}. Two cases may occur

(i) $\beta(q-3)$ is not an integer or $\beta(q-3)$ is an integer larger than $q-3$. Then from (4.10c)

$$c(q-2)\left\{\prod_{k=0}^{q-4}\left(1 + \frac{\partial_t}{q-3+k}\right) F[\beta(q-3), q-2)]\right\} = 2c(2q-5)F[\beta(q-3) - (q-3), 2q-5]$$

and

$$(1 + \tau_{q-2}\partial_t)F[(\beta-1)(q-3), 2q-5]$$

must be proportional to

$$(\Delta_x)^2[(q-2)/\Delta^{q-1}+\beta(q-3)/\Delta^{q-2}]\,.$$

We find

$$\tau_{q-2}=(q-2)/[(q-2)(3q-8)+\beta(q-3)(3q-4)]$$

and β is solution of (4.8). We recover the two complex conjugate solutions $G_{2,q}$ of Table II; we have set $p=\beta(q-3)$.

(ii) $\beta(q-3)=p$,p integer and $0<p<q-3$. This may happen only if $q\geqslant 5$. Then from (4.10c) again

$$c(q-2)\left[\prod_{k=0}^{p-1}\left(1+\frac{\partial_t}{q-3+k}\right)\right]F(p,q-2)$$
$$=\frac{(q-3+p)(q-2+p)}{(q-3)}\,cF(0,q-2+p)$$
$$=\frac{(q-3+p)(q-2+p)(q-1+p)}{(q-3)}\,c\frac{\Delta_x^2}{\Delta^{q+p}},$$

where at the pth step two terms were killed together. In the next $q-4-p$ operations, we go on killing the lower order term. Using (4.10b) with $\tau=1/(m-2)$, $m=q+p-1,\cdots$, we get

$$\prod_{j=0}^{q-5-p}\left(1+\frac{\partial_t}{p+q-2+j}\right)\frac{\Delta_x^2}{\Delta^{q+p}}$$
$$=\frac{(2q-6)(2q-5)}{(p+q-2)(p+q-1)}\,\frac{\Delta_x^2}{\Delta^{2q-4}}.$$

We choose the last two parameters τ_{q-3} and τ_{q-4} by requiring that

$$(1+\tau_{q-3}\,\partial_t)(1+\tau_{q-4}\,\partial_t)\Delta_x^2/\Delta^{2q-4}$$

is proportional to $(G_x)^2$.

It follows that τ_{q-3} and τ_{q-4} are the two (real) roots $\tau_\pm$ of the second degree equation

$$[2(q-2)(q-3)(2q-5)+2p(2q-3)(2q+p-6)]\tau^2$$
$$-[(q-2)(4q-11)+2p(2q-3)]\tau+(q-2)=0\,. \tag{4.11}$$

Factorization of L_q for $G_{3,q}$ follows (see Table II).

C. Three Δ terms

The derivation of the solutions become more and more tedious when the number of Δ terms increase. We just give two families of solutions with three Δ terms and use the second method explained in subsection IVB. We start with

$$G(\Delta)=\frac{\alpha}{\Delta^{q-2}}+\frac{\beta}{\Delta^{q-3}}+\frac{\beta'}{\Delta^{q-4}}$$
$$(\alpha=\alpha_{q-2},\beta=\alpha_{q-3},\ \beta'=\alpha_{q-4})$$

and use formula (4.10b) for $(1+\tau\partial_t)(\Delta_x^2/\Delta^m)$ together with

$$\tilde{l}_2\left(\frac{\alpha}{\Delta^{q-2}}+\frac{\beta}{\Delta^{q-3}}+\frac{\beta'}{\Delta^{q-4}}\right)$$
$$=c(\Delta_x)^2\left[\frac{\alpha(q-2)(q-1)}{\Delta^q}\right.$$
$$\left.+\frac{\beta(q-3)(q-2)}{\Delta^{q-1}}+\frac{\beta'(q-4)(q-3)}{\Delta^{q-2}}\right], \tag{4.12}$$

and we look for l_{q-2} under its factorized form (4.9).

The principle is the same: After having performed the transformation $(1+\tau_i\partial_t)$, we choose τ_i so that the lower term in Δ^{-1} vanishes. Actually, the experience of the previous case shows that most solutions appear by killing another Δ term along the sequence $\Pi_i(1+\tau_i\partial_t)$. Here, we may imagine that we may kill all terms but one and restore, in the last four operations, $\Pi_{q-5}^{q-2}(1+\tau_i\partial_t)$, the five terms Δ^{-2q}, $\Delta^{-2q+1},\ldots,\Delta^{-2q+4}$ which are present on the rhs $(G_\Delta)^2$ and impose $\tau_{q-5},\ldots,\tau_{q-2}$ so that we have proportionality. Or we may still kill two terms only and restore the five terms and proportionality to $(G_\Delta)^2$ in the last three operations by a convenient choice of $\tau_{q-4},\tau_{q-3},\tau_{q-2},\beta'$ so that $(\partial^2/\partial\Delta^2)1_{(q-2)}G$ would be again proportional to $(G_\Delta)^2$. In all cases, the requirement that L_qG be proportional to $(G_x)^2$ gives four equations for determining four variables, so that we may expect a finite number of solutions. The above processes are more numerous than in the preceding study as we may kill these terms in one time or successively in different steps $(1+\tau_i\partial_t)$. *A priori*, there are $(q^2-9q+22)/2$ processes and in general each of them leads to several solutions. We shall restrict to three solutions which we shall investigate successively.

(i) Killing all terms but one after the first step. This is possible if $q\geqslant 7$ only. We get solution $G_{5,q}$ of Table III.

(ii) Killing two terms after the first step and keeping two terms in the further operations. Then $q\geqslant 6$ and we get four solutions, one of them is solution $G_{6,q}$ of Table III.

(iii) Keeping three terms at each operation. We derive the solution for the minimum case $q=5$ only.

1. Case i

Using (4.10b) and (4.11), we get after one operation

$$(1+\tau_1\partial_t)\tilde{l}_2G=\frac{(q-1)q}{q-4}\,\frac{\Delta_x^2}{\Delta^{q+1}}$$

provided $\alpha(q-2)=1$, $(q-2)(q-3)\beta=2$, $(q-2)(q-3)(q-4)\beta'=2$, and $(q-2)\tau_1=1$ and $q\geqslant 7$. The possible solution is then already determined. The next $(q-7)$ derivations keep the monomial form and determine the (τ_i), $i=2,3,\ldots,q-6$:

$$\left(1+\frac{\partial_t}{q-4}\right)\prod_{l=0}^{q-8}\left(1+\frac{\partial_t}{q-1+l}\right)\tilde{l}_2G$$
$$=2(2q-7)\,\frac{(\Delta_x)^2}{\Delta^{2q-6}}.$$

Now, we perform $\Pi_{k=q-5}^{q-2}(1+\tau_k\partial_t)(\Delta_x)^2/\Delta^{2q-6}$ and impose that the result must be proportional to

$$(G_x)^2=\Delta_x^2\left[\frac{1}{\Delta^{q-1}}+\frac{2}{(q-2)\Delta^{q-2}}\right.$$
$$\left.+\frac{2}{(q-2)(q-3)\Delta^{q-3}}\right]^2.$$

The symmetric functions $\sigma_1=\Sigma\tau_i$, $\sigma_2=\Sigma\tau_i\,\tau_j$, $\sigma_3=\Sigma\tau_i\,\tau_j\,\tau_k$, and $\sigma_4=\tau_{q-2}\tau_{q-3}\tau_{q-4}\,\tau_{q-5}$ are solutions of a linear system which determinant $D=+4(2q-7)[2q^4-15q^3+45q^2-62q+36]/(q-2)$ is nonzero. A unique solution exists, solution $G=G_{5,q}$ of Table III, which we make explicit for $q=7$. We find $D=(56\times 122\times 6)/5$ and

$$\sigma_4 = 1/D, \quad \sigma_3 = 194/5D, \quad \sigma_2 = 4074/5D,$$
$$\sigma_1 = \tfrac{1}{6} + 6\sigma_2 - 36\sigma_3 + [(19\times 36)/5]\sigma_4.$$

2. Case ii

Here we kill two terms at the first step and keep the two others during the $(q-6)$ next operations. We have

$$(1+\tau_1\partial_t)\tilde{I}_2 G = [(q-1)/(q-4)] \times F[\beta(q-3)(q-2)-2, q-1],$$

where $\tau_1(q-4)=1$, $\alpha(q-2)=1$, $\beta'(q-4)=\beta$, $q\geqslant 6$, and function $F(p,m)$ is defined in (4.10d). With (4.10c), we perform the $(q-6)$ next derivations

$$\left(1+\frac{\partial_t}{q-4}\right)\prod_{l=0}^{q-7}\left(1+\frac{\partial_t}{q-2+l}\right) = 2\frac{(2q-7)}{(q-2)}F[\beta(q-3)(q-2)-q+4, 2q-7]$$

and determine the last three constants $\tau_{q-2}, \tau_{q-3}, \tau_{q-4}$, or rather their symmetric functions $\sigma_1,\sigma_2,\sigma_3$, and parameter $\beta(q-3)=B$ by performing the last three derivations. We get $\sigma_1,\sigma_2,\sigma_3$ as solutions of a fourth-order linear system the determinant of which must be zero. We find conditions for B:

—Either

$$B = \frac{(q-3)}{(q-2)},$$
$$G \equiv \frac{1}{(q-2)}\left[\frac{1}{\Delta^{q-2}} + \frac{1}{\Delta^{q-3}} + \frac{1}{(q-4)\Delta^{q-4}}\right] \quad (4.13)$$

is the solution $G_{6,q}$ of Table III and σ_1, σ_2, σ_3 are solutions of the linear system written on Table III.

—Or B is solution of an equation of degree 3

$$(q-2)^2(q-3)B^3 - B^2(q-3)(-q^2+9q-10) - B(q^3-3q^2-10q+18) - (q-1)(q-3)(q-4) = 0, \quad (4.14)$$

$$G \equiv \frac{1}{q-2}\frac{1}{\Delta^{q-2}} + \frac{B}{(q-3)\Delta^{q-3}} + \frac{B}{(q-3)(q-4)\Delta^{q-4}}.$$

As an illustration, for $q=6$, we have

$$G_{6,q=6}(\Delta) = \frac{1}{4\Delta^4} + \frac{1}{4\Delta^3} + \frac{1}{8\Delta^2}, \quad \nu_6 = \frac{360}{47}$$

with

$$l_4 = \left(1+\frac{\partial_t}{2}\right)\left[1+\frac{13}{42}\partial_t + \frac{61}{42\times 47}\partial_{t^2}^2 + \frac{1}{21\times 47}\partial_{t^3}^3\right].$$

In this example (4.13) we have assumed that after every step $1+\tau_i\partial_t$, $1\leqslant i\leqslant q-5$, we always remain with two Δ terms. It implies that $\beta(q-3)(q-2)$ is not an integer p with $2\leqslant p\leqslant(q-5)$. One easily sees that, when p in an integer, $2\leqslant p\leqslant q-5$ (and $q\geqslant 7$) another process takes place: After the $(p-1)$th step, we remain with one Δ term and we are left with a problem similar to that of subsection IVC1, if $p=2$, it is exactly case IVC1. More generally, for any integer p, $2\leqslant p\leqslant q-5$, it can be checked that

$$\left(1+\frac{\partial_t}{q-4}\right)\prod_{l=0}^{p-3}\left(1+\frac{\partial_t}{q-2+l}\right) \times \prod_{j=0}^{q-p-6}\left(1+\frac{\partial_t}{q+p-3+j}\right)\tilde{I}_2 G = \frac{2(2q-7)(q+p-4)c}{(q-2)}\frac{\Delta_x^2}{\Delta^{2q-6}},$$
$$G \equiv \frac{1}{(q-2)}\frac{1}{\Delta^{q-4}}\left[\frac{1}{\Delta^2} + \frac{p}{q-3}\frac{1}{\Delta} + \frac{p}{(q-3)(q-4)}\right].$$

It remains to determine, as in IVC1, the symmetric functions $\sigma_1,\sigma_2,\sigma_3,\sigma_4$ such that $(1+\sigma_1\partial_t+\sigma_2\partial_{t^2}+\sigma_3\partial_{t^3}+\sigma_4\partial_{t^4})(\Delta_x)^2/\Delta^{2q-6}$ is proportional to (G_Δ).[2]

3. Case iii

We briefly indicate the solution when three Δ terms are present during the $(q-4)$ first steps for the minimum case $q=5$.

We have

$$G = 1/\Delta^3 + \beta/\Delta^2 + \beta'/\Delta$$

and

$$\tilde{I}_2 G = c(\Delta_x)^2\left(\frac{12}{\Delta^5} + \frac{6\beta}{\Delta^4} + \frac{2\beta'}{\Delta^3}\right),$$
$$(1+\partial_t)\tilde{I}_2 G = c(\Delta_x)^2\left[\frac{60}{\Delta^6} + 24\frac{(\beta-1)}{\Delta^5} + 6\frac{(\beta'-\beta)}{\Delta^4}\right].$$

We must choose the last two parameters τ_2 and τ_3 and coefficients β and β' such that $(1+\tau_3\partial_t)(1+\tau_2\partial_t)(1+\partial_t)\tilde{I}_2 G$ is proportional to

$$(G_x)^2 = \Delta_x^2(3/\Delta^4 + 2\beta/\Delta^3 + \beta'/\Delta^2)^2.$$

Eliminating τ_2, τ_3, and β', we get for β an equation of the sixth degree and then express β' and $\tau_2+\tau_3=\sigma_1$ and $\tau_2\tau_3=\sigma_2$ in terms of β.

D. A solution with the logarithmic term

By inspection of the lower orders $q=2,3,4,5$ a peculiar solution appears which is connected to a peculiar form of the rhs,

$$G_\Delta \sim \sum_{k=1}^{q-1}\frac{1}{\Delta^k}$$

or, equivalently,

$$G(\Delta) = \sum_{k=1}^{q-2}\frac{1}{k\Delta^k} - \log\Delta, \quad (4.15)$$

denoted $G_{4,q}$ in Table II.

For $q=2$, it is the Burgers solution, and, for $q=3$, we recover solution (3.8). In the general case, we substitute this explicit solution in (4.1) and obtain

$$\sum_{j=0}^{q-3} b_j \partial_{t^j}\left(\frac{1}{\Delta^{q-1}}\right) = \sum_{m=0}^{q-3} \frac{(q-1-m)}{(q-1+m)(q+m)} \frac{1}{\Delta^{q-1+m}}.$$

In order to rewrite the lhs, we recall the identity $\partial_t K = \Sigma_{l=1}^{k} (\Delta - 1)^l \mathscr{C}_k^l \partial_{\Delta^l}$, Eq. (4.5), and introduce new coefficients $(\bar{b}_l)$, instead of (b_l):

$$\bar{b}_0 = b_1, \quad \bar{b}_l = \sum_{j=l}^{q-3} \mathscr{C}_j^l b_{j+1} \quad (l>0), \tag{4.16}$$

where the $(\bar{b}_l)$ are the linear combination of the higher-order coefficients $b_{l+1},\ldots,b_{q-2}$.

Then the rhs can be written as a sum of Δ^{q-1+m} terms and the coefficients give us $(q-2)$ numerical equations

$$\sum_{l=m}^{q-3} \bar{b}_l \frac{(q-2+l)!}{(q-2)!} C_l^m (-)^{l+m} = \frac{q-2-m}{(q-1+m)(q+m)}. \tag{4.17}$$

As (4.17) is a triangular system, there is one solution for the $(\bar{b}_l)$ and, as (4.16) is triangular too, we derive a unique solution for the (b_j) too. So the existence of the solution is proved but l_{q-2} is too complicated to be written down here in the general case. For $q=3$, we recover $b_1 = \frac{1}{6}$, $l_1 = 1 + \partial_t/6$, for $q=4$, $l_2 = 1 + \frac{13}{60}\partial_t + \frac{1}{60}\partial_{t^2}$, for $q=5$, $l_3 = 1 + \frac{101}{420}\partial_t + \frac{1}{35}\partial_{t^2} + \frac{1}{840}\partial_{t^3}$.

E. Solutions for $q=4$

Collecting all these results, we are able to list all the potential functions for $q=4$ with the factorized linear operator $L_4 = l_2(\partial_t)\tilde{l}_2$, $l_2 = 1 + b_1\partial_t + b_2\partial_{t^2}$, $\tilde{l}_2 = a\partial_t + b\partial_x + c\partial_{x^2}^2$, $(\partial^2/\partial\Delta^2)(1 + b_1\partial_t + b_2\partial_{t^2})G = \nu_4(G_\Delta)^2$ with $\nu_4 = cb_2(q-1)q(q+1)/(q-2)\alpha_{q-2}$ and $\Delta = 1 + \omega_1 + \omega_2$, $\omega_i = \exp(t + \gamma_i x)$, $a + b\gamma_i + c\gamma_i^2 = 0$.

We obtain six bisolitons, which are written in Table IV together with the associated L_4. We recover the solution $G_{1,q} = G_1$ the two complex conjugate solutions $G_{2,q} = G_\pm$ and the peculiar solution $G_{4,q} = G_4$, which were described in the preceding subsections. We obtain two supplementary potential functions G_5 and G_6 with a logarithmic term which were not considered above.

APPENDIX: RESEARCH OF THE BISOLITONS OF $L_3K = (\lambda + \mu\partial_x)\mathbf{K^2}$

Writing generally K as an expansion

$$K = \sum_{l=0}^{3} \omega_2^l F_l(Z),$$

we get seven coupled equations (1.2) in the variable Z, for F_0, F_1, F_2, F_3. We derive the general formula:

$$\begin{aligned} L_3(\omega_2^N F_N) = {} & \omega_2^N[(\epsilon + Na_{01} + N^2a_{02} + N^3a_{03}) + Z\partial_z(a_{10} + a_{20} + a_{30} + Na_{11} + Na_{21} \\ & + Z^2\partial_{Z^2}^2(a_{20} + 3a_{30} + Na_{21}) + a_{30}Z^3\partial_{Z^3}^3]F_N \\ & + \omega_2^{N+1}\{\partial_Z[-a_{10} - a_{20} - a_{30} + a_{01} - Na_{11} + (2N+1)a_{02} + (3N^2+3 \\ & + Z\partial_{Z^2}^2[a_{11} - 2a_{20} - 6a_{30} - (2N-1)a_{21} + (2N+1)a_{12}] + Z^2\partial_{Z^3}^3(-3 \\ & + \omega_2^{N+2}\{\partial_{Z^2}^2[-a_{11} + a_{20} + a_{02} + 3a_{30} + (N-1)a_{21} - (2N+1)a_{12} + 3 \\ & + Z\partial_{Z^3}^3(3a_{30} - 2a_{21} + a_{12})\}F_N + \omega_2^{N+3}(-a_{30} + a_{21} - a_{12} + a_{03})\partial_{Z^3}^3 F_N \end{aligned} \tag{A1}$$

(where we have used $\omega_i = e^{x_i}$). For example, Table V gives the coupled equations for F_0 and F_1 when $F_2 \equiv F_3 \equiv 0$.

By inspection, the only possible forms for F_0, F_1, F_2, and F_3 are

$$F_0(Z) = k_1 Z + k_0 + \frac{d}{\Delta} + \frac{c}{\Delta^2}, \quad F_1(Z) = k_2 + \frac{b}{\Delta} + \frac{a}{\Delta^2},$$
$$F_2(Z) = \frac{f}{\Delta} + \frac{e}{\Delta^2}, \quad F_3(Z) = \frac{g}{\Delta^2}, \tag{A2}$$

where $k_0, k_1, k_2, a, \ldots, g$ are ten parameters and the numerator N of $K = N/\Delta^2$ is a polynomial of degree 3 with ten parameters

$$N = (g + f + k_2)\omega_2^3 + (f + 2k_2)\omega_2^2\omega_1 + \ldots + (c + d + k_0).$$

In principle, one must put these ten parameters in Eqs. (1.2) and solve them. Great simplifications occur by noticing that

(i) A preliminary condition is

TABLE IV. Bisolitons for $q=4$ and the nonlinearity KK_x or $(G_x)^2$.

$l_2\tilde{l}_2 G = \nu(\partial_x G)^2$
$\tilde{l}_2 = a\partial_t + b\partial_x + c\partial_{xx}$, $G_x = K$, $\Delta = 1 + \omega_1 + \omega_2$, $\omega_i = e^{t+\gamma_i x}$, $a + b\gamma_i + c\gamma_i^2 = 0$

$G_1 = \Delta^{-2}$,	$l_2 = (1 + \partial_t/2)(1 + \partial_t/3)$, $\nu = 5c$
$G_\pm = \Delta^{-2} + \Delta^{-1}\beta_\pm$, $\beta_\pm = -1 \pm i$,	$l_2 = (1 + \partial_t)(1 \mp i\partial_t/4)$, $\nu = \frac{15}{2}ic$
$G_4 = -\log\Delta + \Delta^{-1} + \frac{1}{2}\Delta^{-2}$,	$l_2 = 1 + \frac{13}{60}\partial_t + \frac{1}{60}\partial_{t^2}$, $\nu = c$
$G_5 = 11(\Delta^{-2} - \Delta^{-1}) - \log\Delta$,	$l_2 = 1 + \frac{11}{30}\partial_t + \frac{11}{30}\partial_{t^2}$, $\nu = c$
$G_6 = -\log\Delta + \Delta^{-1} + \frac{1}{2}\Delta^{-2}$,	$l_2 = 1 + \frac{17}{60}\partial_t + \frac{1}{60}\partial_{t^2}$, $\nu = c$

Δ is a divisor of $N\sum \mu_i \Delta_{x^i} - 6 \sum_{i+j=3} a_{ij} (\Delta_{x_1})^i (\Delta_{x_2})^j$,

providing relations between a, c, e, g, and the highest order coefficients a_{30}, a_{21}, a_{12}, and a_{03}:

$$
\begin{aligned}
&g(\mu_2 - \mu_1) = 0,\\
&g\mu_1 - e(\mu_2 - \mu_1) = 6(a_{30} - a_{21} + a_{12} - a_{03}),\\
&e\mu_1 - a(\mu_2 - \mu_1) = 6(3a_{30} - 2a_{21} + a_{12}),\\
&a\mu_1 - c(\mu_2 - \mu_1) = 6(3a_{30} - a_{21}), \quad c\mu_1 = 6a_{30}.
\end{aligned}
\tag{A3}
$$

(ii) $F_0(Z)$ gives the soliton to which K relaxes when $\omega_2 = 0$. Coefficients a_{10}, a_{20}, a_{30} are known through the parameters $A_i \equiv a_{i0}$ of Sec. II. If $K(\omega_1, \omega_2 = 0) = 0$, we have simply $F_0 \equiv 0$. In the same way, a_{01}, a_{02}, a_{03} are known from the relaxation when $\omega_1 = 0$. It is then enough to consider successively the nine reported solitons of Sec. II, plus the case $F_0 \equiv 0$. We can divide the study into two cases.

1. Research of the bounded bisolitons

Numerator N has degree 2; we have $g = 0$ ($F_3 \equiv 0$) and $k_0 = k_1 = k_2 = 0$. By adding a constant, which is necessarily ϵ/λ if $\epsilon, \lambda \neq 0$ and is arbitrary when $\epsilon = \lambda = 0$, we can cancel the ω_1^2 term and get an equivalent bisoliton with $f = 0$. [When $\epsilon, \lambda \neq 0$ we can suppress too the ω_2^2 term and get $e + b = 0$].

It can be proved that:

—No bisoliton exist for $K = F_0(Z)$ ($F_1 \equiv F_2 \equiv 0$).

—Bisolitons exist for $K = \omega_2 F_1(Z)$ ($F_0 \equiv F_2 = 0$), but they are the trivial bisolitons of $L_2 K = K^2$ (see Ref. 6).

—Bisolitons exist for $K = F_0(Z) + \omega_2 F_1(Z)$. We recover trivial bisolitons, equivalent to the preceding ones, and the new bisolitons (5.11). These results can be checked by using the equations of Table V.

—No other bisoliton exists up to all invariances when adding the third term $\omega_2^2 F_2(Z)$.

—When $\epsilon = \lambda = 0$, a supplementary bisoliton of the Burgers type exists. As an example, we check that bisoliton (3.6) is a solution of the equations of Table V and do not show here that it is the only nontrivial binomial solution for $K = F_0 + \omega_2 F_1$ when $\epsilon, \lambda \neq 0$.

The relaxation to solitons [V] of Sec. II, both when $\omega_2 = 0$ and $\omega_1 = 0$, gives successively

$$
a_{30} = \frac{\epsilon}{6}\frac{\mu_1^2}{\lambda^2},
$$

$$
a_{20} = \epsilon\frac{\mu_1}{\lambda},
$$

$$
a_{10} = \epsilon\left(1 + \frac{\mu_1}{\lambda} - \frac{\mu_1^2}{6\lambda^2}\right),
$$

and similar results for a_{03}, a_{02}, a_{01} with exchange $\mu_1 \leftrightarrow \mu_2$. Moreover,

$$
c = \frac{\epsilon}{\lambda}\frac{\mu_1}{\lambda}, \quad d = \frac{\epsilon}{\lambda}\left(1 - \frac{\mu_1}{\lambda}\right),
$$

$$
a = \frac{\epsilon}{\lambda}\frac{\mu_1 - \mu_2}{\lambda}, \quad \text{and } b = 0.
$$

Relations (A3) determine the symmetric parameters a_{21} and a_{12},

$$
a_{21} = \frac{\epsilon}{6}\frac{\mu_1}{\lambda}\left(2\frac{\mu_2}{\lambda} + \frac{\mu_1}{\lambda}\right),
$$

$$
a_{12} = \frac{\epsilon}{6}\frac{\mu_2}{\lambda}\left(2\frac{\mu_1}{\lambda} + \frac{\mu_2}{\lambda}\right),
$$

and it remains to fix the only parameter a_{11}; for instance, in equation $\overline{F}_1(Z) = \overline{\overline{F}}_1(Z)$, at lowest order Δ^{-1}, we find

$$
a_{11} = \epsilon\left[2\left(\frac{\mu_2}{\lambda} + \frac{\mu_1}{\lambda}\right) - \frac{1}{3}\left(\frac{\mu_1}{\lambda} - \frac{\mu_2}{\lambda}\right)^2\right],
$$

and L_3 is written in the form (3.6). As a final point, we can verify either that all other equations of Table V are now identities or more simply that Eq. (3.6) is actually satisfied. For $\epsilon = \lambda = 0$, the supplementary solution corresponds to $d = 0$ and $a = b$ [$\overline{F}_3(Z) = \overline{\overline{F}}_3(Z)$].

2. Bizarre bisolitons for $q = 2,3$

In that case $\mu_1 = \mu_2 = \mu$, $\lambda = -2\mu$ so that the mixed nonlinearity equation becomes

$$
\begin{aligned}
L_q K &= \mu(-2 + \partial_{x_1} + \partial_{x_2})K^2\\
&(-2 + \partial_{x_1} + \partial_{x_2})K^2\\
&= (-2 + Z\partial_Z)F_0^2\\
&+ 2\omega_2(-1 + Z\partial_Z)F_0 F_1 + \omega_2^2 Z\partial_Z F_1^2 + 2\omega_2^2\\
&\times(1 + Z\partial_Z)(F_1 F_2 + F_0 F_3) + \omega_2^4(2 + Z\partial_Z)(F_2^2 + 2F_1 F_3)\\
&+ 2\omega_2^5(3 + Z\partial_Z)F_2 F_3 + \omega_2^6(4 + Z\partial_Z)F_3^2,
\end{aligned}
\tag{A4}
$$

whereas $L_2 K$ ($a_{ij} = 0$ if $i + j > 2$) and $L_3 K$ are sums of terms like (A1). In (A4) we call [l] the relation, telling that the coefficient of ω_2^l must be zero. As a result, L_q is proportional to ϵ and K to ϵ/μ, so that ϵ^2/μ factorizes in both sides of (A4). We throw away this unnecessary factor in the final results quoted in Table I. Although the calculations are performed in the x_i variables, in the table we come back to the original variables x and t by the changes $x_1 = x + t$ and $x_2 = x - t$. We have found two bisolitons for $q = 2$ and six for $q = 3$. We do not give all details in all cases. If we seek a bisoliton relaxing to a known bizarre soliton, we look at the factor $\epsilon + Na_{01} + N^2 a_{02} + N^3 a_{03}$ of F_N in (A1), which is specific to that soliton. We then know what are the necessary [l] linear (in F_m) relations. If the bisoliton relaxes to zero, we do not have this advantage. A relation [l] is said linear if only F_m terms are present and nonlinear if the rhs of (A4) gives a contribution.

$q = 2$

(i) $K_1 = \omega_2^2 F_2$ and $F_2 = f/\Delta$: the linear equation [2], [3], and [4] lead to $a_{10} = a_{11} = a_{20} = 0$, $a_{02} = \epsilon/6$, and $a_{01} = -5\epsilon/6$; the non linear equation [4] gives $f = \epsilon/6\mu$:

$$
K_1 = \frac{\epsilon}{6\mu}\frac{\omega_2^2}{\Delta}, \quad L_2 = \epsilon\left(1 - \frac{\partial_{x_2}}{2}\right)\left(1 - \frac{\partial_{x_2}}{3}\right).
$$

(ii) $K_2 = \omega_2 F_1 + \omega_2^2 F_2$, $F_1 = k_2 + b/\Delta$, and $F_2 = f/\Delta$ leads to $K_2 = (\epsilon/\mu)\omega_1\omega_2/\Delta$ and $L_2 = \epsilon(1 - \partial_{x_1})(1 - \partial_{x_2})$.

$q = 3$

(i) $K_1 = \omega_2^3 F_3$ and $F_3 = g/\Delta^2$: the relation (A3) and the

TABLE V. $(\epsilon + L_3)K = (\lambda + \mu\partial_x)K^2$ when $K = F_0(Z) + \omega_2 F_1(Z)$.

$$
\begin{aligned}
(\epsilon + L_3)(F_0 + \omega_2 F_1) = {} & [\epsilon + (a_{10} + a_{20} + a_{30})z\partial_z + (a_{20} + 3a_{30})z^2\partial_{z^2} + a_{30}z^3\partial_{z^3}]F_0 \\
& + \omega_2\{[(a_{01} + a_{02} + a_{03} - a_{10} - a_{20} - a_{30})\partial_z + (a_{11} - 2a_{20} - 6a_{30} + a_{21} + a_{12})z\partial_{z^2} + (a_{21} - 3a_{30})z^2\partial_{z^3}]F_0 \\
& + [\epsilon + a_{01} + a_{02} + a_{03} + (a_{10} + a_{20} + a_{11} + a_{30} + a_{21} + a_{12})z\partial_z + (a_{20} + 3a_{30} + a_{21})z^2\partial_{z^2} + a_{30}z^3\partial_{z^3}]F_1\} \\
& + \omega_2^2\{[(a_{02} + a_{20} - a_{11} + 3a_{30} - a_{21} - a_{12} + 3a_{03})\partial_{z^2} + (3a_{30} - 2a_{21} + a_{12})z\partial_{z^3}]F_0 \\
& + [(a_{01} - a_{10} + 3a_{02} - a_{20} - a_{11} - a_{30} - a_{21} - a_{12} + 7a_{03})\partial_z + (a_{11} - 2a_{20} - 6a_{30} - a_{21} + 3a_{12})z\partial_{z^2} + (a_{21} - 3a_{30})z^2\partial_{z^3}]F_1\} \\
& + \omega_2^3\{(a_{03} - a_{12} + a_{21} - a_{30})\partial_{z^2}F_0 + [(a_{20} + a_{02} - a_{11} + 3a_{30} - 3a_{12} + 6a_{03})\partial_{z^2} + (3a_{30} - 2a_{21} + a_{12})z\partial_{z^3}]F_1\} \\
& + \omega_2^4(a_{03} - a_{12} + a_{21} - a_{30})\partial_{z^2}F_1 \\
= {} & (\lambda + \mu_1 z\partial_z)F_0^2 + \omega_2[2(\lambda + \mu_2 + \mu_1 z\partial_z)F_0F_1 + (\mu_2 - \mu_1)\partial_z F_0^2] \\
& + \omega_2^2[(\lambda + 2\mu_2 + \mu_1 z\partial_z)F_1^2 + 2(\mu_2 - \mu_1)\partial_z F_0F_1] + \omega_2^3(\mu_2 - \mu_1)\partial_z F_1^2
\end{aligned}
$$

$$
\left(\epsilon + \sum_{i+j=1}^{3} a_{ij}\partial_{x_1^i x_2^j}\right)K = \left(\lambda + \sum \mu_i \partial_{x_i}\right)K^2, \quad K = F_0(z) + \omega_2 F_1(z), \quad z = \omega_1 + \omega_2, \quad \omega_i = e^{x_i}
$$

linear equations [3], [4], and [5] lead to $a_{30} = a_{21} = a_{12} = a_{20} = a_{11} = a_{10} = 0$, $g = \epsilon/10\mu$, $a_{03} = -\epsilon/60$, $a_{02} = \epsilon/5$, and $a_{01} = -47\epsilon/60$:

$$K_1 = \epsilon\omega_2^3/10\mu\Delta^2,$$

$$L_3 = \epsilon(1 - \partial_{x_2}/3)(1 - \partial_{x_2}/4)(1 - \partial_{x_2}/5).$$

(ii) $K_2 = \omega_2^2 F_2$ and $F_2 = e/\Delta^2 + f/\Delta$: (A3) and the linear relations [2] and [3] give $a_{30} = a_{21} = a_{20} = 0$, $a_{12} = a_{03} = \mu e/6$, $a_{11} = -5a_{03}$, $a_{10} = 6a_{03}$, $a_{01} = 6a_{03} - 5\epsilon/6$, and $a_{02} = -5a_{03} + \epsilon/6$. The nonlinear [4] gives $f = 5e$ and $a_{03} = \epsilon/6(23)$:

$$K_2 = \frac{\epsilon}{23\mu}\frac{\omega_2^2}{\Delta^2}(6 + 5\omega_1 + 5\omega_2),$$

$$L_3 = \epsilon\left(1 + \frac{1}{23}(\partial_{x_1} + \partial_{x_2})\right) \times \left(1 - \frac{\partial_{x_2}}{2}\right)\left(1 - \frac{\partial_{x_2}}{3}\right)$$

(iii) $K = \omega_2^3 F_2 + \omega_2^2 F_2$, $F_3 = g/\Delta^2$, and $F_2 = e/\Delta^2 + f/\Delta$: (A3) and linear [2] and [3] give $a_{30} = a_{21} = a_{20} = 0$, $\mu e = 6a_{12}$, $\mu g = 6(a_{12} - a_{03})$, $\epsilon + 2a_{01} + 4a_{02} + 8a_{03} = 0$, $a_{10} + 2a_{11} + 4a_{12} = 0$, $e(a_{11} + 5a_{12}) = 0$, and $e(a_{01} + 5a_{02} + 19a_{03}) = 0$. We have two subcases following $e = 0$ or $e \neq 0$. In the first case $e \neq 0$, the nonlinear equation [4] and [5] lead to $a_{03} = 0$, $e = g$, and $5f + 4e = \epsilon/\mu$, and still an equation $(e + f)(2e + f) = 0$. If $e \neq 0$ and $e + f = 0$, we find $6a_{12} = e\mu = -\epsilon$ and $K_3 = \epsilon\omega_2^2\omega_1/\mu\Delta^2$, $L_3 = \epsilon(1 - \partial_{x_1})(1 - \partial_{x_2}/2)(1 - \partial_{x_2}/3)$. If $e \neq 0$ and $2e + f = 0$, we find $6a_{12} = e\mu = -\epsilon$ and

$$K_4 = \frac{\epsilon}{6\mu}\frac{\omega_2^2}{\Delta^2}(1 + \omega_2 + 2\omega_1),$$

$$L_3 = \epsilon\left(1 - \frac{\partial_{x_1}}{6}\right)\left(1 - \frac{\partial_{x_2}}{3}\right)\left(1 - \frac{\partial_{x_2}}{2}\right).$$

In the second case $e = 0$, [3] gives $f = g$ and $a_{12} = 0$ and the nonlinear [4] and [5] give $a_{11} = a_{10} = 0$, $a_{02} + 17a_{03} = 0$, $a_{01} + 6a_{02} + 22a_{03} = 0$, $a_{05} = -4\epsilon/5$, $a_{02} = 17\epsilon/100$, $a_{03} = -\epsilon/100$, and $\mu f = 3\epsilon/50$:

$$K_5 = \frac{3\omega_2^2}{50\Delta^2}(1 + \omega_1 + 2\omega_2),$$

$$L_3 = \epsilon\left(1 - \frac{\partial_{x_2}}{2}\right)\left(1 - \frac{\partial_{x_2}}{5}\right)\left(1 - \frac{\partial_{x_2}}{10}\right)$$

(iv) $K = \omega_2 F_1 + \omega_2^2 F_2$, $F_1 = k_2 + b/\Delta + a/\Delta^2$, and $F_2 = f/\Delta + e/\Delta^2$ leads to $K_6 = 6\epsilon\omega_1\omega_2(6 + 5\omega_1 + 5\omega_2)/23\mu\Delta^2$ and $L_3 = \epsilon[1 + \frac{1}{23}(\partial_{x_1} + \partial_{x_2})](1 - \partial_{x_1})(1 - \partial_{x_2})$. Let us recall that for these eight bisolitons (two for L_2 and six for L_3) we can associate the same number by the transformation $K_i \to \tilde{K}_i = -\epsilon/2\mu - K_i$ satisfying $\tilde{L}_q\tilde{K}_i = \mu(-2 + \partial_{x_1} + \partial_{x_2})\tilde{K}_i^2$, with $\tilde{L}_q = 2\epsilon - L_q - \epsilon(\partial_{x_1} + \partial_{x_2})$ a nonfactorized linear operator.

[1] V. G. Makhankov, Phys. Rep. C **35**, 1 (1978).
[2] R. Hirota, *Direct methods in soliton theory*, *Solitons*, Vol. 17, edited by R. K. Bullough and P. J. Caudrey (Springer-Verlag, Berlin, 1980).
[3] M. Ito, J. Phys. Soc. Jpn. **49**, 771 (1980).
[4] H. Cornille, A. Gervois, Physica D **6**, 29 (1982).
[5] H. Cornille and A. Gervois, Phys. Lett. A **83**, 251 (1981).
[6] H. Cornille and A. Gervois, Physica D **6**, 1 (1982).
[7] A. V. Bobylev, Dokl. Akad. Nauk SSSR **225**, 1296 (1975); **231**, 571 (1976) [Sov. Phys. Dokl. **20**, 822 (1976); **21**, 632 (1976)]; M. Krook and T. T. Wu, Phys. Rev. Lett. **16**, 1107 (1976); Phys. Fluids **20**, 1589 (1979); H. Cornille and A. Gervois, J. Stat. Phys. **26**, 181 (1981).
[8] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical functions* N. B. S. Applied Math Series 55 (Natl. Bur. Stds., Washington, D. C., 1964), p. 824.

Application of nonlinear operator theory to the Edwards–Freed equations in the theory of polymer solutions

Tetz Yoshimura
Poste restante Paris 28, rue des Ecoles, F-75005 Paris 5e, France[a]

(Received 8 April 1982; accepted for publication 15 October 1982)

Mathematical structure of nonlinear integral equations for "screened hydrodynamic interaction" and "self-energy" in the static version of Edwards and Freed's theory of polymer solutions with finite concentrations are analyzed and it is shown that algorithms developed by Lika and Altman for nonlinear operator equations without Fréchet differentiability are applicable. Recipes for successive approximations are presented and questions to be investigated by those techniques are proposed.

PACS numbers: 02.30.Rz

I. INTRODUCTION

In the theory of polymer solutions with finite concentrations, Edwards and Freed[1–5] derived interesting nonlinear integral equations for the "screened interaction" and "self-energy." Here we neither repeat the derivation of those equations nor argue for or against assumptions and approximations used in the derivation. Though attempts have been made to express quantities such as frequency dependent viscosity, etc., in terms of "known" functions, so far no other closed system of equations have been derived for the "screened interaction" and "self-energy." So it seems worthwhile to investigate the mathematical structure of Edwards and Freed's integral equations more systematically than done so far, in spite of some shortcomings of the Edwards–Freed theory pointed out by Muthukumar.[6]

In Sec. 2, we present alternative forms of nonlinear integral equations to be considered. In Sec. 3, we present some relevant theorems from recent development of nonlinear operator theory, and recipes for application of those theorems to our problems is given in Sec. 4. In Sec. 5, some questions to be investigated further are raised.

2. NONLINEAR INTEGRAL EQUATIONS TO BE CONSIDERED

In Edwards and Freed's theory of polymer solutions,[1] the equation for the averaged hydrodynamics interaction $K(q,\omega)$ reads

$$K(q,\omega)=\left(\frac{3\rho_0}{2\pi^3}\right)^{1/2}\int\int d\tau\,du\,\exp(i\omega\tau+iuq)\times\left[4\rho_0 l\int\int d\omega'\,dq'\,\frac{\Omega(q')(1-\cos(q'u+\omega'\tau))}{q^2\{[\omega'-\omega_2(q',\omega')]^2+\omega_1(q',\omega')^2\}}+6\eta_0|\tau|\right]^{-3/2},\tag{2.1}$$

where

$$\omega_1(q,\omega)=3k_BTl^{-1}q^2\,\mathrm{Re}[K(q,\omega)(1+\tau K(q,\omega))^{-1}].\tag{2.2a}$$

$$\omega_2(q,\omega)=-3k_BTl^{-1}q^2\,\mathrm{Im}[K(q,\omega)(1+\tau K(q,\omega))^{-1}].\tag{2.2b}$$

It should be noticed that Eq. (2.1) has a structure very different from those for "propagators" in quantum field theory and the theory of condensed matters including the SCF approximation[7] in the theory of polymer chains.

For the hopping type model, Edwards and Freed[2] derived the following expression for the "self-energy"

$$\Sigma(k,\omega)=-\rho l\int\frac{d\omega'\,dq}{(2\pi)^2}(i\omega'\tau+3k_BTl^{-1}q^2)\mathfrak{G}(q,\omega')\times\int\int du\,dv\,\exp\{-iqu+i(\omega-\omega')v-\tfrac{1}{6}k^2B(u,v)\}.\tag{2.3}$$

In terms of the screened interaction

$$K(q,\omega)=:\frac{l}{b\zeta}\int\frac{d\omega'\,d^3k}{(2\pi)^4}[i\rho\omega'+\eta_0k^2-\Sigma(k,\omega')]^{-1},\tag{2.4}$$

$\mathfrak{G}$ is expressed as follows:

$$\mathfrak{G}(q,\omega)=[i\omega+K(q,\omega)(i\omega\tau+3k_BTl^{-1}q^2)]^{-1}.\tag{2.5}$$

Replacing $B(u,v)$ by its static value as in the Kirkwood–Riseman[8] type approximation, Edwards and Freed[2] got the following nonlinear integral equations for $K(q,c)$ and $\Sigma(k,c)$:

$$K(q,c)=\frac{l}{\zeta}+\frac{1}{9\pi^2}\int k^2\,dk\Gamma(k,q)[\eta_0k^2-\Sigma(k,c)]^{-1},\tag{2.6a}$$

$$\Sigma(k,c)=-\Re\int dq\,\Gamma(k,q)[K(q,c)]^{-1},\tag{2.6b}$$

where

[a] Alternative postal contact address: c/o Professor A. Vančura, FB Physik, University Kaiserslautern, D6750 Kaiserslautern, West Germany.

 0022-2488/83/082056-05$02.50

$$\Gamma(k,q) = \tfrac{1}{3}k^2 l\left[\tfrac{1}{36}(k^2)^2 l^2 + q^2\right]^{-1},$$
$$\mathfrak{N} = nNl/\pi V, \quad c = nNM_A/N_A V. \tag{2.7}$$

From now on we suppress the parameter c in K and Σ. Eliminating Σ from Eqs. (2.6), one gets

$$K(q) = \frac{l}{\xi} + \frac{1}{3\pi^2}\int_0^\infty k'^2 dk'\,\Gamma(k',q) \times\left\{\eta_0 k'^2 + \mathfrak{N}\int_0^\infty dq'\Gamma(k',q')[K(q')]^{-1}\right\} = :\mathfrak{K}(K,q). \tag{2.8}$$

Or "equivalently," eliminating K instead of Σ, one gets

$$\Sigma(k) = -\mathfrak{N}\int dq\,\Gamma(k,q) \times\left\{\frac{l}{\xi} + \frac{1}{3\pi^2}\int p^2\,dp\Gamma(p,q)[\eta_0 p^2 - \Sigma(p)]^{-1}\right\}^{-1} = :\mathfrak{S}(\Sigma,k). \tag{2.9}$$

Edwards and Freed[2,4,5] considered only negative definite Σ, but the situation is not so simple because of the nonlinearity of the equations. The screening constant

$$\kappa = :[-\Sigma(0)/\eta]^{1/2} \tag{2.10}$$

and the specific viscosity

$$\eta_{sp} = :(l^2 N_A/3\pi\kappa M_A)\int_{q_0}^\infty dq[q^2 K(q)]^{-1} \tag{2.11}$$

must be positive, however. Another "equivalent" equation is

$$L(q) = :[K(q)]^{-1} = \left\{\frac{l}{\xi} + \frac{1}{3\pi^2}\int k^2\,dk\Gamma(k,q) \times\left[\eta_0 k^2 + \mathfrak{N}\int_0^\infty dp\,\Gamma(k,p)L(p)\right]^{-1}\right\}^{-1} = :\mathfrak{L}(L,q). \tag{2.12}$$

In order to formulate existence theorems and successive approximation schemes, we regard Eqs. (2.8), (2.9), and (2.12) as operator equations in Banach spaces and try to apply recently developed methods[9–11] of nonlinear operator theory.

3. RELEVANT MATHEMATICAL THEOREMS

In this section we present several mathematical theorems about existence and uniqueness of solutions of nonlinear operator equations and convergence of successive approximations.

We write an equation abstractly,

$$x - \Phi(x) = : P(x) = 0, \tag{3.1}$$

and denote the Fréchet derivative of P at x operated upon x' as $P'(x)x'$. If the map Φ is Fréchet differentiable in a suitable domain, one may try to apply the Newton–Kantorovich type successive approximations.[12,13]

Recently, Lika[9] and Altman[10,11] developed successive approximation schemes for nondifferentiable maps. To formulate such successive approximation schemes one has to introduce the notion of Lipschitz approximation to a map. Map Ψ is said to be a Lipschitz approximation to map Φ, if $\Omega(x) = :\Phi(x) - \Psi(x)$ satisfies the Lipschitz condition

$$\|\Omega(x_1) - \Omega(x_2)\| \leqslant \lambda\,\|x_1 - x_2\| \quad \forall x_1, x_2 \in D(\Phi). \tag{3.2}$$

Now we have the following theorem adapted from Lika.[9]

Theorem 1: Suppose that the following conditions are satisfied:

(1) The Fréchet derivative Ψ' of the Lipschitz approximation Ψ to Φ satisfies the condition

$$\|\Psi'(x_1) - \Psi'(x_2)\| \leqslant \kappa\|x_1 - x_2\| \quad \forall x_1, x_2 \in M \subset X; \tag{3.3}$$

(2) $\Omega(x) = :\Phi(x) - \Psi(x)$ satisfies the condition

$$\|\Omega(x_1) - \Omega(x_2)\| \leqslant \lambda\,\|x_1 - x_2\| \quad \forall x_1, x_2 \in M \subset X; \tag{3.4}$$

(3) $G_0 = :[I - \Psi'(x_0)]^{-1}$ exists and

$$\|G_0\| \leqslant b_0,\ \|G_0(x_0 - \Phi(x_0))\| \leqslant \xi_0; \tag{3.5}$$

(4) $b_0\lambda < 1$; (3.6)

(5) $h_0 = :b_0\kappa\xi_0 < \tfrac{1}{2}(1 - b_0\lambda)^2$; (3.7)

(6) closed ball $S(x_0, r_0) \subset M$, where

$$r_0 = :[1 - b_0\lambda - \{(1 - b_0\lambda)^2 - 2h_0\}^{1/2}]\xi_0 h_0^{-1} = : R(h_0)\xi_0. \tag{3.8}$$

Then the equation $x - \Phi(x) = 0$ admits a solution $x^* \in S(x,r)$, to which the sequence $\{x_n\}$ defined by

$$x_{n+1} = x_n - [I - \Psi'(x_n)]^{-1}(x_n - \Phi(x_n)) \tag{3.9}$$

converges. The error estimate reads

$$\|x_0 - x^*\| \leqslant \{1 - B_n\lambda - [(1 - B_n\lambda)^2 - 2h_n]^{1/2}\}/B_n\kappa. \tag{3.10}$$

In the case when the existence of $[I - \Psi'(x_n)]^{-1}, n \geqslant 1$ is not guaranteed, we have the following theorem, which also asserts local uniqueness (Lika[9]).

Theorem 2: Suppose that the conditions (1)–(4) of Theorem 1 and the condition (5′), there exists a number $N \in ((1 - b_0\lambda)^{-1}, 2(1 - b_0\lambda)^{-1})$ such that the inequality

$$h_0 < 2N^{-2}\{(1 - b_0\lambda)N - 1\} \tag{3.11}$$

holds, are satisfied, then the sequence $\{x_n\}$ defined by

$$x_{n+1} = x_n - [I - \Psi'(x_0)]^{-1}(x_n - \Phi(x_n)) \tag{3.12}$$

converges to the unique solution x^* in $S(x_0, N\xi_0)$ of the equation $x - \Phi(x) = 0$, and the rate of convergence is

$$\|x_n - x^*\| \leqslant (Nh_0 - b_0\lambda)^n \xi_0 (1 - Nh_0 - b_0\lambda)^{-1}. \tag{3.13}$$

If $N > 2(1 - b_0\kappa)^{-1}$ but $h_0 < \tfrac{1}{2}(1 - b_0\lambda)^2$, a solution exists but uniqueness is not guaranteed.

We have also the following algorithm due to Altman.[10] Suppose that a Lipschitz approximation Ψ to Φ is Fréchet differentiable, the Fréchet derivative $\Psi(x)$ is continuous in $U = : D(\Phi) \cap \bar{S}(x_0, r)$, and for any $x \in U_0 = : D(\Phi) \cap S(x_0, r)$ there exists an element $h(x) \in X$ such that

$$[I - \Psi'(x)]h(x) - x + \Phi(x) = 0. \tag{3.14}$$

Then one can define an algorithm as follows. Given $x_0 \in D(\Phi)$, $\kappa C/\rho < \beta < 1$, $\kappa C < \rho < 1$, suppose that $x_1, \ldots, x_n$, $n \geqslant 1$ are already defined. Then put $\epsilon_n = 1$ if

$$\Xi(1, x_n, h_n) \leqslant \rho\|P(x_n)\|, \tag{3.15}$$

where

$$\Xi(\epsilon, x, h) = :\|P(x + \epsilon h) - (1 - \epsilon)P(x)\|\epsilon^{-1}. \tag{3.16}$$

If $\Xi(1, x_n, h_n) > \rho\|P(x_n)\|$, there exists a number $\epsilon_n \in (0,1)$ such that

$$\beta\rho\|P(x_n)\| \leqslant \Xi(\epsilon_n x_n, h_n) \leqslant \rho\|P(x_n)\|. \tag{3.17}$$

In either case, put

$$x_{n+1} = x_n + \epsilon_n h_n. \tag{3.18}$$

Now we have the following theorem (Altman[10]).

Theorem 3: If Ψ is a Lipschitz approximation to Φ in U_0, and $\Psi'(x)$ is continuous in U and satisfies the condition

$$\|\Phi(x) - \Psi(x) - \Phi(\bar{x}) + \Psi(\bar{x})\| < \kappa\|x - \bar{x}\| \tag{3.19}$$

in U_0 with radius

$$r \geqslant (1-\rho)^{-1} C \exp(1-\rho)\|P(x_0)\|, \tag{3.20}$$

then $\{x_n\} \subset U_0$, and there exists at least one x^* such that $x^* - \Phi(x^*) = 0$. The error estimate reads

$$\begin{aligned} &\|x_n - x^*\| \leqslant (1-\rho)^{-1} C b_n, \\ &b_n = \|P(x_0)\| \exp\{(1-\rho)(1-t_n)\}, \\ &t_0 = 0, \quad t_n = \sum_{i=0}^{n-1} \epsilon_i. \end{aligned} \tag{3.21}$$

Uniqueness is not guaranteed, however. Alternatively, we can apply the following theorem, also due to Altman,[10] which is useful when it is difficult to solve Eq. (3.14) "exactly" at each step of successive approximation.

Theorem 4: If there exists $\rho \in (0,1)$ such that for any $x \in U_0$ there exists $h(x) \in X$ such that

$$\|(I - \Psi'(x))h(x) - P(x)\| \leqslant \rho\|P(x)\|, \tag{3.22}$$

$$\|h(x)\| \leqslant C\|P(x)\|, \tag{3.23}$$

then define an algorithm as follows. Given $x_0 \in D(\Phi)$ and

$$\kappa C + \bar{\rho} < \rho < 1, \quad (\kappa C + \bar{\rho})\rho^{-1} < \beta < 1, \tag{3.24}$$

suppose that $x_1, \ldots, x_n$, $n \geqslant 1$ are already defined. Then put $\epsilon_n = 1$ or choose $\epsilon_n < 1$ in the same way as in the previous algorithm provided q and β are subject to the above conditions, and define $x_{n+1} = x_n + \epsilon_n h_n$. If the radius r of S is larger than $(1-\rho)^{-1}\exp(1-\rho)\|P(x_0)\|$, then $\{x_n\} \subset U$ and $x_n \to x^*$ as $n \to \infty$.

4. APPLICABILITY OF THE THEOREMS

Let us begin with Eq. (2.8). The Fréchet derivative $\mathfrak{R}'(K,\cdot,\cdot)$ of the map $\mathfrak{R}$ at K is expressed as follows:

$$\begin{aligned} \mathfrak{R}'(K,K',q) = \frac{\mathfrak{N}l}{9\pi^2} &\int dk\,(k^2)^2 \int dp' \\ &\times \left(\frac{1}{36}(k^2)^2 l^2 + p'^2\right)^{-1} [K(p')]^{-2} K'(p) \\ &\times \left\{\eta_0 k^2 + \frac{1}{3}\mathfrak{N}lk^2 \int dp \right. \\ &\left. \times \left(\frac{1}{36}(k^2)^2 l + p^2\right)^{-1} [K(p)]^{-1}\right\}^{-2}. \end{aligned} \tag{4.1}$$

Therefore, $\mathfrak{R}'(0,\cdot,\cdot)$ is not well defined and one cannot find a Lipschitz approximation to $\mathfrak{R}$ for K with small norm. So one cannot formulate an algorithm for Eq. (2.8). Now let us try Eq. (2.9) for Σ. Again, it can be easily seen that $\mathfrak{S}(\Sigma,0)$ is not well defined for almost all Σ, unless an "infrared" cutoff q_0 is introduced into the integration over q. In particular

$$\begin{aligned} \mathfrak{S}(0,k) = -\frac{1}{3}\mathfrak{N}lk^2 &\int_0^\infty dq \\ &\times \left(\frac{l}{\zeta} + \tau q^{-1/2}\right)^{-1} \left(\frac{1}{36}(k^2)^2 l^2 + q^2\right)^{-1}, \end{aligned} \tag{4.2}$$

where

$$\tau = (6l)^{-1/2}(\pi\eta_0)^{-1}. \tag{4.3}$$

Hence,

$$\begin{aligned} \mathfrak{S}(0,0) = -\frac{1}{3}\mathfrak{N}l \lim_{k\to 0} k^2 &\int_0^\infty dq \\ &\times \left(\frac{l}{\zeta} + \tau q^{-1/2}\right)^{-1} \left(\frac{1}{36}(k^2)^2 l^2 + q^2\right)^{-1} \end{aligned} \tag{4.4}$$

is not well defined. The situation is similar for any Σ such that $\Sigma(k) \sim_{k\to 0} c'k^2 + O(k^3)$. For a successive approximation to be applicable, the norm of $\mathfrak{S}(\Sigma,\cdot)$ with infrared cutoff must be sufficiently small for $\Sigma \in D(\mathfrak{S}) \ni 0$ but this is not guaranteed and may be sensitive to the cutoff. In other words, the theory is "singular" at the weak interaction limit.

Finally, we come to Eq. (2.12) for L. It can be easily seen that

$$\mathfrak{L}(0,q) = (l/\zeta + \tau q^{-1/2})^{-1}, \tag{4.5}$$

$$\begin{aligned} \mathfrak{L}(C,q) = \left\{\frac{l}{\zeta} + \frac{l}{3\pi^2}\int dk\,(k^2)^2(\eta_0 k^2 + 2\pi\mathfrak{N}C)^{-1}\right. \\ \left. \times (\tfrac{1}{36}(k^2)^2 l^2 + q^2)^{-1}\right\}^{-1}, \end{aligned} \tag{4.6}$$

so that $\mathfrak{L}(0,q)$ and $\mathfrak{L}(C,q), C > C_0 < 0$ with certain C_0 are bounded. On the other hand,

$$\begin{aligned} &\mathfrak{L}'(L,L',q) \\ &= -[\mathfrak{L}(L,q)]^2\left(\frac{\mathfrak{N}l^2}{27\pi^2}\right)\int (r^2)^3 dr \left(\frac{1}{36}(r^2)^2 l^2 + q^2\right)^{-1} \\ &\times \left\{\eta_0 r^2 + \frac{1}{3}\mathfrak{N}lr^2 \int ds \left(\frac{1}{36}(r^2)^2 l^2 + s^2\right)^{-1} L(s)\right\}^{-2} \\ &\times \int dt \left(\frac{1}{36}(r^2)^2 + t^2\right)^{-1} L'(t). \end{aligned} \tag{4.7}$$

In particular

$$\begin{aligned} \mathfrak{L}'(0,L',q) = -\frac{\mathfrak{N}l^2}{27\pi^2\eta_0^2}&\left(\frac{l}{\zeta} + \tau q^{-1/2}\right)^{-2} \\ &\times \int r^2 dr \left(\frac{1}{36}(r^2)^2 l^2 + q^2\right)^{-1} \\ &\times \int dt \left(\frac{1}{36}(r^2)^2 l^2 + t^2\right)^{-1} L'(t), \end{aligned} \tag{4.8}$$

so that $\mathfrak{L}'(0,L',0)$ is not well defined because of the "infrared" divergence, unless a cutoff is introduced, and may be sensitive to cutoff. Now, let us define a Lipschitz approximation Λ to $\mathfrak{L}$ by

$$\Lambda(L,q) =: \Big[\frac{l}{\zeta} + \frac{l^2}{27\pi^2}\int_0^\infty k^2 dk \theta_\rho(|k^2 - (k^\blacktriangle)^2| - \sigma)$$
$$\times \big(\tfrac{1}{36}(k^2)^2 l^2 + q^2\big)^{-1}$$
$$\times \Big\{ \eta_0 + \frac{1}{3}\mathfrak{N} l \int_0^\infty ds\, L(s)\theta_\omega(s - s_0)$$
$$\times \Big(\frac{1}{36}(k^2)^2 l^2 + s^2\Big)^{-1}$$
$$+ \frac{1}{3}\mathfrak{N} l \int_0^\infty ds\, \mathfrak{L}(L_0,s)(1 - \theta_\omega(s - s_0))$$
$$\times \Big(\frac{1}{36}(k^2)^2 + s^2\Big)^{-1} \Big\}^{-1}$$
$$+ \frac{l^2}{27\pi}\int_0^\infty k^2 dk \{1 - \theta_\rho(|k^2 - (k^\blacktriangle)^2| - \sigma)\} \Big\{ \eta_0$$
$$+ \frac{1}{3}\mathfrak{N} l \int ds\, \mathfrak{L}(L_0,s)\Big(\frac{1}{36}(k^2)^2 l^2 + s^2\Big)^{-1} \Big\}^{-1}$$
$$\times \Big(\frac{1}{36}(k^2)^2 l + q^2\Big)^{-1} \Big]^{-1}, \tag{4.9}$$

where $(k^\blacktriangle)^2$ is a (possible) real root of the transcendental equation

$$\varphi(k) =: \eta_0 + \frac{1}{3}\mathfrak{N} l \int ds \Big(\frac{1}{36}(k^2)^2 l^2 + s^2\Big)^{-1}$$
$$\times \{L(s)\theta_\omega(s - s_0) + \mathfrak{L}(L_0,s)(1 - \theta_\omega(s - s_0))\} = 0 \tag{4.10}$$

and the "smooth step function" θ_ρ is defined as follows:

$$\theta_\rho(u) =: \int_{-\infty}^{u} dv\, \chi_\rho(v),$$
$$\chi_\rho(v) =: \begin{cases} 0, & |v| \geqslant \rho, \\ \exp\{-\rho^2(\rho^2 - v^2)^{-1}\}, & |v| < \rho. \end{cases} \tag{4.11}$$

[If $\varphi(k) = 0$ has several real roots, the definition of Λ must be modified accordingly.] It is not difficult to verify that $\mathfrak{L}(L) - \Lambda(L)$ satisfies the Lipschitz condition for L in certain domain and for some combinations of values of the parameters ρ,ω,σ and s_0. This precaution is necessary even if one accepts positive definite solutions only, because $\Lambda'(L,\cdot,\cdot)$ must be defined for L in certain domain, not only for the solution.

The Fréchet derivative of Λ reads

$$\Lambda'(L,L',q)$$
$$= [\Lambda(L,q)]^2 \frac{\mathfrak{N} l^2}{27\pi}\int (k^2)^2 dk \theta_\rho(|k^2 - (k^\blacktriangle)^2| - \sigma)$$
$$\times \int dt\, L'(t)\theta_\omega(t - s_0)\Big(\frac{1}{36}(k^2)^2 l^2 + t^2\Big)^{-1}$$
$$\times \Big(\frac{1}{36}(k^2)^2 l^2 + q^2\Big)^{-1}$$
$$\times \Big\{ \eta_0 + \frac{1}{3}\mathfrak{N} l \int ds\, L(s)\theta_\omega(s - s_0)$$
$$\times \Big(\frac{1}{36}(k^2)^2 l^2 + s^2\Big)^{-1} + \frac{1}{3}\mathfrak{N} l \int ds\, \mathfrak{L}(L_0,s)$$
$$\times (1 - \theta_\omega(s - s_0))\Big(\frac{1}{36}(k^2)^2 l^2 + s^2\Big)^{-1} \Big\}^{-1}. \tag{4.12}$$

Now, one may take, for example, zeroth approximation of the form

$$L_0(q) = \sum_{n=0}^{n_0} \alpha_n q^n \Big(\sum_{m=0}^{m_0} \beta_m q^m\Big)^{-1} \exp\Big(\sum_{l=1}^{l_0} \gamma_l q^l\Big), \tag{4.13}$$

and choose parameters $\alpha_n, \beta_m, \gamma_l, s_0, \sigma, \rho$, and ω so that L_0 and $\mathfrak{L}(L_0)$ coincide for certain values of q. If the values of the parameters l, ζ, η_0, and $\mathfrak{N}$ are such that the conditions of Theorems 1, 2, 3, or 4 are satisfied, one can proceed, and the problem is reduced to solving a sequence of linear integral equations or finding $h(L_n)$ satisfying the following conditions:

$$\|\Lambda(L_n)h(L_n) - \mathfrak{L}(L_n)\| < \rho\|L_n - \mathfrak{L}(L_n)\|, \tag{4.14a}$$
$$\|h(L_n)\| < C\|L_n - \mathfrak{L}(L_n)\|. \tag{4.14b}$$

It can be easily seen that these conditions are more easily satisfied for large η_0 than for small η_0.

In order to ensure that the Lipschitz condition be satisfied, one may have to add some seminorms to the original norm so that undesirable functions are removed from the domain.

5. CONCLUDING REMARKS AND QUESTIONS TO BE INVESTIGATED

As has been seen above, neither Σ nor K but $L =: K^{-1}$ is the suitable function to be evaluated as a solution of the nonlinear integral equation in the static version of Edwards and Freed's theory of polymer solutions of finite concentrations. Though, until and unless we obtain an (approximate) solution of Eq. (2.12) explicitly (in practice numerically) for various combinations of values of the parameters, we cannot say anything definite, we can raise some questions concerning possible outcomes. How do asymptotic behaviors of $L(q)$ for $q \to 0$ and for $q \to \infty$ depend on the parameters? For $\eta_{sp} =: (l^2 N_A/3\pi\kappa M_A)\int_{q_0}^\infty dq\, L(q)q^{-2}$ not to be sensitive to the cutoff q_0 the solution $L^*(q)$ of Eq. (2.12) must behave asymptotically $\sim q^\beta$, $\beta > 1$. [It should be noticed that Eq. (2.12) without cutoff may have a solution.] Can it be interpreted as a phase transition if for certain combinations of the values of parameters the inequality $\beta > 1$ ceases to hold? It is an interesting problem to find critical combinations of values of the parameters at which L (and K) cease to be positive definite or Σ ceases to be negative definite. It may be also worthwhile to use the correlation function $\lambda^{-1}\int_0^\lambda \langle \exp ik\cdot(\vec{R}(s) - R(s'))\rangle e^{iq(\vec{s} - \vec{s}')} ds\, ds' = \lambda^{-1}\int_0^\lambda \times \exp(-k^2 l|s - s'|)\cos \vec{q}\cdot(\vec{s} - \vec{s}')ds\, ds'$ instead of $\Gamma(k,q)$ in Eq. (2.12) or introduce a cutoff.[14] (Here λ stands for the length of polymers.)

It may be also an interesting problem to solve Eq. (2.9) or (2.12) with a Γ drastically different from (2.7).

Anyway, even in the framework of the static theory, it seems to be relevant and interesting to find the k dependence of L (and K) and q dependence of Σ more accurately than done so far and compare the results with those based on qualitative arguments by Edwards and Freed.[2,4,5]

[1] S. F. Edwards and K. F. Freed, J. Chem. Phys. **61**, 1189 (1974).
[2] K. F. Freed and S. F. Edwards, J. Chem. Phys. **61**, 3626 (1974).
[3] K. F. Freed and S. F. Edwards, J. Chem. Phys. **62**, 4032 (1975).

[4]S. F. Edwards and K. F. Freed, J. Chem. Soc. Faraday, Trans. 1 **72**, 2025 (1975).
[5]K. F. Freed, "Polymer Solutions and Hydrodynamics of Polymer Solutions," in *Progress in Liquid Physics*, edited by C. A. Croxton (Wiley, New York, 1978), p. 343.
[6]M. Muthukumar, J. Phys. A **14** 2129 (1981).
[7]K. F. Freed, Adv. Chem. Phys. **22**, 1 (1972).
[8]J. G. Kirkwood and J. Riseman, J. Chem. Phys. **16**, 565 (1948).
[9]D. K. Lika, *Analiza funktsionale shi rezolvarya aproksimative a ekuatsiilor neliniare* (Functional Analysis and Approximate Solution of Nonlinear Equations) (Editura Shtiintsa, Kishineu, 1975).
[10]M. Altman, Nonlinear Analysis, Theor. Methods Appl. **4** 437, 761 (1980).
[11]M. Altman, *Contractors and Contractor Directions, Theory and Applications* (Dekker, New York, 1977).
[12]L. V. Kantorovich and G. P. Akilov, *Functional Analysis in Normed Spaces*. (Pergamon, New York, 1964)
[13]B. Jankó, *Rezolvarea ecuațiilor operaționale neliniare în spații Banach* (Solution of Nonlinear Operator Equations in Banach Spaces) (Editura Acad. R. S. România, București, 1968).
[14]M. Muthukumar and S. F. Edwards, Polymer **23**, 345 (1982).

Tetrads and arbitrary observers

John R. Urani
Physics Department, University of Missouri–Kansas City, Kansas City, Missouri 64110
Marilyn H. Kemp
Mathematics Department, University of Missouri–Kansas City, Kansas City, Missouri 64110

(Received 10 December 1982; accepted for publication 25 March 1983)

Recent results concerning globally isometric mappings for arbitrary observers in flat space-time are generalized to space-times admitting a time orientation. Critical to the method is the use of an orthonormal tetrad which, when it is defined globally, allows the construction of a global isometry which generalizes the pointwise boost on flat space-time. Connection coefficients are obtained, thereby defining acceleration covariant differentiation for both particle and tensor field equations. An application to orbiting observers in exterior Schwarzschild geometries is presented.

PACS numbers: 02.40. + m, 04.20. − q

I. INTRODUCTION

Recent work on accelerating observers in flat and gravitational space-time has centered on using invariant differential geometric methods[1] and, in some cases, presymmetry arguments[2–5] to obtain local expressions for the Fermi–Walker transport of an orthonormal tetrad associated with a differentiable timelike curve (observer trajectory). Similarly, Burghardt[6] used a covariant projection method to obtain a decomposition of the Einstein equations for rotating observers. Estabrook and Wahlquist[7] employed a dyadic formalism to obtain equations for acceleration near a general world line. A classic application of tetrads was the use of real and complex null tetrads by Newman and Penrose[8] in treating zero-mass particles and gravitational radiation.

This paper makes use of global space-time structures[9–13] to derive analytical methods for obtaining connection coefficients associated with an observer congruence on suitable space-times. The orthonormal tetrad field of a space-time is used to construct a close analog of the flat-space Lorentz boost. The tangent and contangent space maps are shown to be isometries (global if the tetrad field is globally defined). The tangent space map is then used to define connection coefficients for acceleration covariance of tensor equations for particles and fields.

An application to orbiting observers in exterior Schwarzschild space-times is presented as a local coordinate calculation. Exact observer maps are found which predict the correct asymptotic limits for spin precession (geodetic precession) along with a precession reversal at the photosphere radius.

II. GLOBAL STRUCTURES

We shall consider stably causal space-times[10] since they are time orientable by a global timelike vector field. Such space-times admit global observer congruences which are used to construct a simple isometric map associated with these congruences. The isometry is global if the space-time admits a global tetrad field.[9,10]

Let (M,g) be a stably causal space-time and let a tetrad field $K_a(x)$, $a = 1,2,3,4$ exist on some subset $U \subset M$. For an orthonormal tetrad, we have $g(K_a,K_b) = \eta_{ab}$ for each $x \in M$ where $\eta = (+\ +\ +\ -)$.

In a local coordinate chart, we have $K^{\mu}_{a}(x) g_{\mu\nu}(x) K^{\nu}_{b}(x) = \eta_{ab}$. Furthermore, let $v(x)$ denote a timelike vector field on M, whose integral curves form an observer congruence. Relative to a local tetrad basis, let $u_a(x) = g(K_a(x),v(x))$. That is, using local coordinates, $u_a(x) = K^{\mu}_{a}(x) v_{\mu}(x)$. It is this particular expression for the observer velocity that is used to parametrize the "boost" associated with the observer.

Clearly, due to the orthonormal nature of the tetrad field, $K_a(x)$, $a = 1,2,3,4$ is a local Lorentz frame.[14] Tensors which are expressed relative to the tetrad basis may then be boosted to the local Lorentz frame of $v(x)$ by a simple Loretnz transformation.

Choosing local coordinates again, $K^{\mu}_{a}(x)$ represents a nonsingular matrix which diagonalizes $g_{\mu\nu}$ by the congruence map $K^T g K = \eta$. An arbitrary (1,0) tensor $T(x)$ may be referred to tetrad components via

$$T_a(x) = K^{\mu}_{a}(x) T_{\mu}(x)$$

or

$$T^a(x) = K^{-1a}_{\mu}(x) T^{\mu}(x).$$

Using the observer velocity in tetrad form, $u(x)$, a Lorentz transformation matrix $\Lambda^{b}_{a}(u(x))$, may be constructed which boosts to the observer Lorentz frame. That is,

$$T'^{b}(x) = \Lambda^{b}_{a}(x) T^{a}(x).$$

The boosted tensor T' is then expressed relative to the tangent space basis derived from the local manifold coordinate chart using

$$\begin{aligned}\overline{T}^{\mu}(x) &= K^{\mu}_{b}(x) T'^{b}(x) \\ &= K^{\mu}_{b}(x) \Lambda^{b}_{a}(x) K^{-1a}{}_{\nu}(x) T^{\nu}(x).\end{aligned}$$

If the matrix $K \Lambda K^{-1} = J^{-1}(x)$, then $\overline{T}^{\mu}(x) = (J^{-1})^{\mu}_{\nu}(x) T^{\nu}(x)$, and (0,1) tensors boost according to $\overline{T}_{\mu}(x) = J^{\nu}_{\mu}(x) T_{\nu}(x)$, where $J(x) = K \Lambda^{-1} K^{-1}$. The construction extends to the entire tensor bundle over U by simple tensor products.

The isometric nature of the tangent space map is easily verified since $J^T g J = (K \Lambda^{-1} K^{-1})^T g K \Lambda^{-1} K^{-1} = g$ for each $x \in U$. This is, of course, due to the isometry $\Lambda^T \eta \Lambda = \eta$.

If the tetrad field is globally defined,[9,10] then the isometry associated with the observer congruence is also global. Also, the boosts $J^{-1}(x)$ and $J(x)$ are tangent and cotangent

 0022-2488/83/082061-04$02.50

space maps, respectively, parametrized by the observer congruence $v(x)$. But note that no manifold coordinate changes have been affected.

A given choice of a tetrad field $K_a(x)$, $a = 1,2,3,4$ on some $U \subset M$ may be taken to represent a reference rest frame at each $x \subset U$. Tensors are then boosted relative to that choice of rest frames by the (J,J^{-1}) mapping. The choice of tetrad field is, of course, arbitrary; hence no absolute rest frames are being defined. If an initial C^1 tetrad field $K_a(x)$ is changed smoothly to $K'_b(x)$ for each $x \in U$ by a map $L^a_b(x)$, this map is a pointwise proper Lorentz group basis change because these orthonormal tetrad fields are cross-sections of the orthonormal frame bundle $O(M)$ which has real $O(3,1)$ as its structure group.

Similarly, a given coordinate atlas for (M,g), inducing a basis ∂_μ for T_x for each x in some chart, then induces a cross section of $L(M)$ (the linear frame bundle) for the chart. The (J,J^{-1}) boosts are then changes of $L(M)$ cross section for the chart which are induced by the observer congruence $v(x)$, a cross section of the tangent bundle T_M. The nonsingular boost matrix $J^{-1}(x)$ on a chart is a cross section of $L(M)$ over the chart since $L(M)$ is a principle fiber bundle with structure group $GL(4,R)$. This frame bundle cross section defines a connection for acceleration covariance which is developed in the next section.

III. ACCELERATION COVARIANCE

The connection coefficients needed for acceleration covariance can be derived directly. Let $T(x)$ be a (1,0) tensor field which satisfies a covariant differential equation $\nabla T = B$, where B is a (1,1) tensor. In local coordinates, $T^\mu{}_{;\nu} = B^\mu{}_\nu$. The tetrad formalism is used to boost T, B, and the differential equation to the frame of an observer vector field $v(x)$. A derivative pseudoterm is generated since $\partial_\nu T^\mu$ becomes

$$\begin{aligned}\overline{\partial}_\nu \overline{T}^\mu &= J^\alpha{}_\nu \partial_\alpha (J^{-1\mu}{}_\beta T^\beta) \\ &= J^\alpha{}_\nu J^{-1\mu}{}_\beta \partial_\alpha T^\beta + J^\alpha{}_\nu (\partial_\alpha J^{-1\mu}{}_\beta) T^\beta \\ &= J^\alpha{}_\nu J^{-1\mu}{}_\beta \partial_\alpha T^\beta + J^\alpha{}_\nu (\partial_\alpha J^{-1\mu}{}_\beta) J^\beta_\lambda J^{-1\lambda}{}_\theta T^\theta \\ &= J^\alpha{}_\nu J^{-1\mu}{}_\beta \partial_\alpha T^\beta - J^\alpha{}_\nu (J^{-1\mu}{}_\beta (\partial_\alpha J^\beta{}_\lambda)) \overline{T}^\lambda.\end{aligned}$$

Acceleration covariance is obtained if the boosted equation is written

$$\overline{T}^\mu_{;\nu} = \overline{\partial}_\nu \overline{T}^\mu + \overline{\Gamma}^\mu_{\nu\beta} \overline{T}^\beta = \overline{B}^\mu{}_\nu,$$

with

$$\overline{B}^\mu{}_\nu = J^{-1\mu}{}_\beta J^\alpha{}_\nu T^\beta{}_{;\alpha}, \quad \overline{\partial}_\nu = J^\lambda{}_\nu \partial_\lambda$$

and

$$\overline{\Gamma}^\mu_{\alpha\beta} = J^{-1\mu}_\lambda J^\theta_\alpha J^\epsilon_\beta \Gamma^\lambda_{\theta\epsilon} + J^{-1\mu}_\lambda \overline{\partial}_\alpha J^\lambda{}_\beta. \tag{3.1}$$

Here Γ is the usual Levi-Cività connection which takes on the correct pseudoterm[12,14,15] under the boost. The advantage of the present method of treating observers via a congruence, which may then be restricted for the connection coefficients associated with arbitrary observers is obtained via simple methods[13,16] yielding exact results.

The isometric nature of the (J,J^{-1}) boost is advantageous too, since self-frame quantities can be easily obtained. For example, boosting the observer congruence to its own self-frame leads to a very simple result. If $v(x)$ is the observer congruence, $u(x) = K^{-1}(x)v(x)$ is its local Lorentz expression (relative to the local tetrad basis). Then, parametrizing the boost via $u(x)$ one obtains, in local coordinates,

$$u^a(x) = K^{-1}{}^a_\nu(x) v^\nu(x)$$

and

$$\hat{u}(x) = \Lambda^b_a(u(x)) u^a(x) = (0,0,0,c),$$

where $\hat{u}$ is the observer's self-velocity relative to the tetrad basis. Mapping back to the tangent space basis defined by local manifold coordinates gives

$$\overline{v}^\mu_{\text{self}} = K^\mu_b(x) \hat{u}^b(x) = K^\mu_4(x) c.$$

The observer self-frame velocity is then the same as the fourth tetrad field which is globally defined if (M,g) is at least stably causal.[10] In this case the fourth tetrad field can be used to define a time orientation and a global time function which naturally defines a local time axis for all observers (timelike smooth curves) relative to their self-frames.

A tetrad field, defined on a given local coordinate chart, is then useful in several ways. The isometric boost is easily constructed from the observer congruence and the local tetrad. The horizontal subspaces of $T_{L(M)}$ defined by the boost $J^{-1}(x)$ define the acceleration connection whose coefficients are simply given by $-J\partial J^{-1}$. Finally, self-frame calculations are easily done. An example follows in the next section.

IV. SCHWARZSCHILD GEOMETRY

The exterior Schwarzschild geometry is considered as an application of the methods presented above. This spacetime admits a global tetrad field. The standard Schwarzschild solution

$$\begin{aligned}ds^2 = &(1 - 2m^*/r)^{-1} dr^2 + r^2 d\Omega^2 \\ &- (1 - 2m^*/r) c^2 dt^2,\end{aligned}$$

where $m^* = Gm/c^2$, leads to the diagonalizing tetrad with components $k^1_1 = ((r - 2m^*)/r)^{1/2}$, $k^2_2 = k^3_3 = 1$, $k^4_4 = ((r - 2m^*)/r)^{-1/2}$, all other elements being zero.

The metric reduces to Lorentzian spherical coordinate form $\eta(r,\theta,\phi,ct)$ via the congruence transformation $k^T g k = \eta$. A spherical-to-cartesian coordinate Jacobian j is needed to take g to $\eta = (+\ +\ +\ -)$ form yielding $K = kj$ such that $K^T g K = \eta(+\ +\ +\ -)$.

Routine calculation shows that

$$K^1{}_1 = (x/r)((r - 2m^*)/r)^{1/2}, \quad K^2{}_1 = (y/r)((r - 2m^*)/r)^{1/2}$$

$$K^2{}_3 = -1/r, \quad K^3{}_1 = -y/r^2, \quad K^3{}_2 = x/r^2,$$

and

$$K^4{}_4 = (r/(r - 2m^*))^{1/2}$$

with all others zero. Here, $\theta = \pi/2$ $(z = 0)$ has been taken for simplicity. The inverse components (nonzero) are

$(K^{-1})^1{}_1 = (x/r)((r/(r - 2m^*))^{1/2}$, $(K^{-1})^1{}_3 = -y$,

$(K^{-1})^2{}_1 = (y/r)(r/(r - 2m^*))^{1/2}$, $(K^{-1})^2{}_3 = x$,

$(K^{-1})^3{}_2 = -r$, and $(K^{-1})^4{}_4 = ((r - 2m^*)/r)^{1/2}$.

For a circular orbit, with $r = (x^2 + y^2)^{1/2} > 2m^*$, the coordinate increment is $dx^\mu = (0,0,d\phi,c\,dt)$ and $c^2 d\tau^2 = ((r - 2m^*)/r)c^2 dt^2 - r^2 d\phi^2$. Then

$d\tau = dt\,((r-2m^*)/r - r^2\omega^2/c^2)^{1/2}$, where $\omega = d\phi/dt$ and the covariant velocity of the observer, $v = dx/d\tau$, is

$$v^\mu = (0,0,\zeta\omega,\zeta c), \tag{4.1}$$

where $\zeta = ((r-2m^*)/r - r^2\omega^2/c^2)^{-1/2}$. The corresponding Lorentzian (tetrad) form of v is $u = k^{-1}v$ which is, in component form,

$$u^a = (0,0,\zeta\omega,\zeta c((r-2m^*)/r)^{1/2}).$$

Since $u^4 = \gamma c$, $\gamma = (1-\beta^2)^{-1/2} = \zeta\,((r-2m^*)/r)^{1/2}$. However, the Lorentzian γ factor may be better expressed as

$$\gamma = (1 - (r^2\omega^2/c^2)(r/(r-2m^*)))^{-1/2}$$

from which $\beta = r|\omega|/c(r/(r-2m^*))^{1/2}$. Clearly for a timelike observer, $\beta < 1$, which reduces for large "r" to the flat-space condition $r|\omega| < c$.

The Lorentz boost matrix $\Lambda(\beta)$ is then constructed from $\beta_x = -\beta\sin\phi$ and $\beta_y = \beta\cos\phi$. The tangent space boost is $J^{-1} = K\Lambda K^{-1}$ and its dual is $J = K\Lambda^{-1}K^{-1}$, which is a global isometry from $J^T g J = g$. The structures of J and J^{-1} are quite simple. The nonzero components of J^{-1} are $(J^{-1})^1{}_1 = (J^{-1})^2{}_2 = 1$, $(J^{-1})^3{}_3 = (J^{-1})^4{}_4 = \gamma$, $(J^{-1})^3{}_4 = (-\gamma\beta/r)((r-2m^*)/r)^{1/2}$, and $(J^{-1})^4{}_3 = -\gamma\beta r(r/(r-2m^*))^{1/2}$. Similarly, $J^1{}_1 = J^2{}_2 = 1$, $J^3{}_3 = J^4{}_4 = \gamma$, $J^3{}_4 = -(J^{-1})^3_4$, and $J^4{}_3 = -(J^{-1})^4{}_3$.

This boost (J, J^{-1}) is expressed in manifold Schwarzschild coordinates and is a map of tangent-space (J^{-1}) and cotangent-space (J) coordinates.

Not unexpectedly, only ϕ and ct components mix under the boost since the observer three-velocity is in the ϕ direction. Also, the fact that ∂_ϕ is a Killing field for g greatly simplifies the (J, J^{-1}) matrices, but the isometric properties of the mappings are not dependent on any symmetries.

The limit of large r, the asymptotically flat region, leads to a Lorentzian boost with $\beta \simeq |\omega| r/c$ as expected, with $r|\omega| < c$ as a natural limit for a timelike observer.

The Levi-Cività connection in Schwarzschild coordinates has few nonzero components due to spherical symmetry (see Ref. 14). Under the observer boost (J, J^{-1}), the connection coefficients transform as

$$\overline{\Gamma}^\alpha{}_{\beta\gamma} = (J^{-1})^\alpha{}_\mu\, J^\nu{}_\beta\, J^\sigma{}_\gamma\, \Gamma^\mu{}_{\nu\sigma} + \Gamma^\alpha{}_{p\beta\gamma}, \tag{4.2}$$

where $\Gamma^\alpha{}_{p\beta\gamma} = J^{-1\,\alpha}{}_\theta \overline{\partial}_\beta J^\theta{}_\gamma$ is the pseudoterm form common to any affine connection's transformation.[15] Here, $\overline{\partial}_\beta = J^\alpha{}_\beta \partial_\alpha$ and since J and J^{-1} are functions of r only, $\overline{\partial}_\beta = J^1{}_\beta \partial_r = \delta^1{}_\beta \partial_r$. The only nonzero pseudoterms are $\Gamma^3_{p13} = -\Gamma^4_{p14} = (-\gamma^2\beta^2/r)(r-3m^*)/(r-2m^*)$.

The problem of the torque-free Fermi-transport of a classical spin S, carried by the observer in orbit, is examined via the Fermi-transport equation for S, namely, $\nabla S/\partial\tau = v g(S, a_c)$ in the observer frame, the self-frame of the spin. Here, $a_c = \nabla v/\partial\tau$, the covariant acceleration. The covariant velocity boosts to $\overline{v} = J^{-1}v$ which is, in components,

$$\overline{v}^\mu = k^\mu{}_4 c = (0,0,0,c(r/(r-2m^*))^{1/2}).$$

By inspection of Eq. (4.1), $dv/d\tau = 0$, so $a_c = \nabla v/\partial\tau = \Gamma(v,v)$. Using the tabulated Γ coefficients (Ref. 14), we obtain for the covariant acceleration, $a^\mu{}_c = (a^1{}_c, 0,0,0)$ with $a^1{}_c = (m^*c^2/r^2 - \omega^2 r)(1 - (r^2\omega^2/c^2)(r/(r-2m^*)))^{-1}$, which has the correct flat-space $(a^1{}_c = -\omega^2 r/(1 - r^2\omega^2/c^2))$ and Newtonian $(a^1{}_c = -\omega^2 r)$ limits.

Also $a_c = 0$ for a geodesic orbit and $a^1{}_c = 0$ in that case. Hence for geodesics, $\omega^2 r = m^*c^2/r^2$. Combining this result with the timelike restriction on β^2, namely $(r^2\omega^2/c^2)(r/(r-2m^*)) < 1$, results in $r > 3m^*$ for geodesic circles, which is the well-known photosphere limit for massive particle orbits. One then has the option of considering geodesic or nongeodesic cases for $r > 3m^*$; however, in the range $2m^* < r \leqslant 3m^*$, the orbits can be maintained only by an external force. The method applies to both types of orbit, however.

Boosting the acceleration a_c results in $\overline{a}_c = J^{-1}a_c = a_c$, since $(J^{-1})^1{}_\mu = \delta^1{}_\mu$. In the self-frame, $\overline{S}^4 = 0$ since $0 = \overline{g}(\overline{v},\overline{S}) = g(\overline{v},\overline{S}) = g_{44}(\overline{v}^4\overline{S}^4)$, where $g = \overline{g}$ from the boost isometry. Thus one needs to consider only parallel transport of $\overline{\mathbf{S}}$, viz., $\nabla\overline{S}^i/\partial\tau = 0$ in the self-frame case, since $\overline{S} = J^{-1}S = (\overline{\mathbf{S}},0)$ and $\overline{v} = (\mathbf{0},\overline{v}^4)$. The transport equation then simplifies to

$$\frac{d\overline{S}^i}{d\tau} + \overline{\Gamma}^i{}_{4j}(\overline{v}^4,\overline{S}^j) = 0. \tag{4.3}$$

No pseudoterms enter this problem since only $\Gamma_p{}^4{}_{14} = -\Gamma_p{}^3{}_{13} \neq 0$, and thus the Levi-Cività coefficients [first right-hand term of Eq. (4.2)] only need boosting.

From J, J^{-1} and Eq. (4.2), $\overline{\Gamma}^1_{41} = \overline{\Gamma}^1_{42} = 0$ and $\overline{\Gamma}^1_{43} = -\gamma^2\beta((r-2m^*)/r)^{1/2}((r-3m^*)/r)$ in $\nabla\overline{S}^1/\partial\tau = 0$, which then becomes

$$\frac{d\overline{S}^1}{d\tau} - \gamma^2\beta c\left(\frac{r-3m^*}{r}\right)\overline{S}^3 = 0. \tag{4.4}$$

All $\overline{\Gamma}^2_{4j} = 0$ for $\theta = \pi/2$, so $d\overline{S}^2/d\tau = 0$ as expected, since $\overline{S}^2$ is the θ component of $\overline{S}$ which is parallel to ω for the orbit.

Finally, only $\overline{\gamma}^3_{41} = (\gamma^2\beta/r^2)((r-3m^*)/(r-2m^*))((r-2m^*)/r)^{1/2}$ is nonzero among the $\overline{\Gamma}^3$ components for $\theta = \pi/2$, so

$$\frac{d\overline{S}^3}{d\tau} + \frac{\gamma^2\beta c}{r^2}\left(\frac{r-3m^*}{r-2m^*}\right)\overline{S}^1 = 0. \tag{4.5}$$

Taking second derivatives and combining Eqs. (4.4) and (4.5), the precession equations are $d^2\overline{S}^a/d\tau^2 + \omega_p^2\overline{S}^a = 0$ for $a = 1,3$. Here $\omega_p = -\gamma^2|\omega|(r-3m^*)/(r-2m^*)$ is the precession frequency (sign to agree with flat-space limits). All results are still in spherical Schwarzschild coordinates. At the photosphere radius $r = 3m^*$, there is no precession relative to the corotating spherical unit vectors, and the precession changes sign at that radius.

Writing $d\phi/d\tau$ in terms of $d\phi/dt = \omega$ and transforming to (x,y,z) coordinates, the proper-time precession frequency is

$$\omega_T = -\left[\frac{\gamma(r-3m^*)}{(r^2-2m^*r)^{1/2}} - 1\right]\frac{d\phi}{d\tau}. \tag{4.6}$$

The flat-space limit ($m^* = 0$) of Eq. (4.6) gives simple Thomas precession[17] and the Newtonian limit is then $\omega_T \simeq 0$ as expected. In the limiting case $r \gg 2m^*$ and $r\omega \ll c$, Eq. (4.6) becomes

$$\omega_T = -[1 + \tfrac{1}{2}(r^2\omega^2/c^2) - 2m^*/r - 1]\omega, \tag{4.7}$$

to first order in m^*/r and $r^2\omega^2/c^2$.

The separation of Eq. (4.7) into special and general relativistic terms is then obvious. For a geodesic orbit,

$\omega^2 r = m^* c^2/r^2$ and $\omega_T = \frac{3}{2} m^* \omega/r$ in Eq. (4.7). This is the well-known "geodetic precession" result.[18]

V. DISCUSSION

In this paper, a new method of treating arbitrary observers on space-times has been presented. The method makes extensive use of observer congruences which exist globally if the space-times are at least stably causal. Also of critical importance is the use of a tetrad field which will exist globally if, for example, the space-time is orientable and globally hyperbolic (a sufficient but not necessary condition).

The tetrad field is used to map tensors expressed relative to local tangent space bases defined by manifold coordinate charts to local tetrad bases, and vice versa. This is equivalent to expressing tensors with respect to cross sections of the linear frame bundle $L(M)$ or the orthonormal frame bundle $O(M)$, respectively.

The observer boost is isometric in all cases and is obviously parametrized by observer trajectories which is completely in accord with presymmetry requirements.[2–5]

Using an observer congruence initially and restricting to a single trajectory (integral curve) is advantageous because the connection for acceleration covariance can be obtained by simple calculations. The usual devices of Fermi propagated bases,[1,2] etc. may then be appended to any given observer trajectory.

ACKNOWLEDGMENTS

The authors wish to thank G. P. Barker, B. DeFacio, and J. K. Beem of the University of Missouri for valuable discussions.

[1] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).
[2] B. DeFacio, P. W. Dennis, and D. G. Retzloff, Phys. Rev. D **18**, 2813 (1978); **20**, 570 (1979). Many valuable references are cited herein.
[3] B. DeFacio and D. G. Retzloff, J. Math. Phys. **21**, 751 (1980).
[4] D. G. Retzloff, B. DeFacio, and P. W. Dennis, J. Math. Phys. **23**, 96 (1982); **23**, 105 (1982). These contain many useful citations.
[5] Y. Avishai and H. Ekstein, Found. Phys. **2**, 257 (1972); Commun. Math. Phys. **37**, 193 (1974).
[6] R. Burghardt, Acta Phys. Austriaca **48**, 181 (1978).
[7] F. B. Estabrook and H. D. Wahlquist, J. Math. Phys. **5**, 1629 (1964).
[8] E. Newman and R. Penrose, J. Math. Phys. **3**, 556 (1962).
[9] R. Geroch, J. Math. Phys. **9**, 1739 (1968); **11**, 343 (1970).
[10] J. K. Beem and P. E. Ehrlich, *Global Lorentzian Geometry* (Marcel Dekker, New York, 1981).
[11] C. T. J. Dodson, Categories, *Bundles and Space-time Topology* (Shiva, Orpington, 1980).
[12] J. R. Urani and M. H. Kemp, J. Math. Phys. **23**, 227 (1982).
[13] J. R. Urani and M. H. Kemp, J. Math. Phys. **23**, 423 (1982).
[14] S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972).
[15] D. Bleeker, *Gauge Theory and Variational Principles* (Addison-Wesley, Reading, MA, 1981), p. 31.
[16] The Lorentz group generators, which are put into the acceleration connection coefficients in the usual local treatment of an accelerating observer (see Ref. 1 and Ref. 3), will appear in Eq. (3.1) upon differentiation of the Lorentz transformation in the J matrix. Also, see Ref. 12 and 13 for flat space-time results.
[17] L. H. Thomas, Philos. Mag. **3**, 1 (1927).
[18] H. C. Ohanian, *Gravitation and Spacetime* (Norton, New York, 1976), p. 296.

Dynamical symmetries: An approach to Jacobi fields and to constants of geodesic motion

Giacomo Caviglia[a)]
Istituto Matematico dell'Università, Via L.B. Alberti, 4, 16132 Genova, Italy

(Received 23 December 1982; accepted for publication 18 February 1983)

It is shown that every dynamical symmetry (DS) of the Euler–Lagrange equations derived from the Lagrangian $L = \frac{1}{2} g_{ab}\dot{q}^a\dot{q}^b$ identifies a Jacobi field on each geodesic of the configuration manifold. Using the connections between Jacobi fields and DS's, it is proved that DS's always possess associated conserved quantities, whose expression is explicitly written down. An additional constant of motion concomitant with "pairs" of DS's, independently of the choice of L, is also determined. Applications to general relativity are emphasized in the course of the discussion.

PACS numbers: 02.40.Ky, 02.30.Jr, 04.20.Me

1. INTRODUCTION

As it is well known, Noether's theorem provides a well-defined procedure relating the constants of motion of a given Lagrangian system to its symmetry transformations.[1] Recently, there has been a renewed interest in the problem of enlarging the class of generators of conserved quantities. On the one hand, this has led to the introduction of general non-Noether symmetries, such as point symmetries and DS's (dynamical symmetries), which are essentially defined as vector fields on the extended tangent space that may be regarded as generators of transformations leaving the equations of motion invariant.[2–5] On the other hand, turning to the connections between symmetries and constants of motion, it has been shown that point symmetries identify first integrals (a brief survey of some basic results and definitions is given in Sec. 2); nevertheless, the relationships between DS's and conserved quantities seem to be so involved[2,4] as to require further investigation.

It is the main aim of this paper to analyze the properties of the DS's of the Euler–Lagrange equations generated by the regular Lagrangian $L = \frac{1}{2} g_{ab}\dot{q}^a\dot{q}^b$. Such a Lagrangian function is known to model the geodesic motion of freely falling particles in the general theory of relativity[6]; moreover, it characterizes the free motions of a mechanical system with a finite number of degrees of freedom in the field of analytical mechanics and may also be used to describe the evolution of a conservative Newtonian system, after a suitable conformal transformation has been performed.[7] With a view to future applications of the techniques described in this paper to the case of geodesic motion, we shall adopt the usual terminology of general relativity, even though our analysis is not restricted to the four-dimensional case.

Recalling that a practical procedure for the construction of DS's has already been proposed elsewhere,[8] we assume in Sec. 3 that an arbitrary DS, say Y, is given and we prove that Y identifies a Jacobi field along every geodesic of the configuration space. This result shows that DS's could possibly play a distinguished role in general relativity, in view of the well-known fundamental importance of Jacobi fields.[6]

To emphasize this viewpoint, we remark that a careful analysis of the properties of DS's leads to the conclusion that they yield the most natural framework for the extension of some recent results concerning the possibility of generating Jacobi fields.[9] Moreover, DS's also give a valuable insight into the geometrical interpretation of Killing tensors, generalized Killing tensors,[9] and the various families of geodesic collineations,[10] which may be regarded as generators of DS's.[8]

The connection between Jacobi fields and first integrals of motion is then exploited in order to associate conserved quantities to an arbitrarily assigned DS (see Sec. 4). Actually, besides giving rise to almost all classes of first integrals of geodesic motion that had been previously determined by using "ad hoc" methods, our approach yields also new conserved quantities by a very simple and easy procedure. In particular, we obtain a constant of motion that may *always* be associated with pairs of DS's, independently of the form of the Lagrangian function that generates the equations of motion.

Additional general comments can be found in Sec. 5.

2. SYMMETRIES OF A LAGRANGIAN SYSTEM

In this section, we shall briefly review some basic concepts concerning Lagrangian systems and possible definitions of symmetry.

Denote by M a real n-dimensional manifold representing the configuration space of a mechanical system, and consider its associated extended tangent space $R \times TM$, in which natural coordinates $(s,q^a,\dot{q}^a)$ $(a = 1,\ldots,n)$ have been introduced.

Suppose now that a system characterized by a regular Lagrangian L has been given. The Euler–Lagrange equations of motion can be written in the normal form

$$\frac{d\dot{q}^a}{ds} = g^{ab}\left(- \frac{\partial^2 L}{\partial\dot{q}^b\partial q^c}\dot{q}^c - \frac{\partial^2 L}{\partial\dot{q}^b\partial s} + \frac{\partial L}{\partial q^b}\right) \overset{\text{def}}{=} \Lambda^a(s,q,\dot{q}), \tag{2.1}$$

where g^{ab} denotes the inverse matrix of $\partial^2 L/(\partial\dot{q}^a\partial\dot{q}^b)$. It is a

[a)] Work done under the auspices of the National Group for Mathematical Physics of CNR (Consiglio Nazionale delle Ricerche).

 0022-2488/83/082065-05$02.50

straightforward result that the motions of the system are described by the projections onto M of the integral curves of the vector field Γ defined as

$$\Gamma = \frac{\partial}{\partial s} + \dot{q}^a \frac{\partial}{\partial q^a} + \Lambda^a \frac{\partial}{\partial \dot{q}^a}. \tag{2.2}$$

Moreover, if we introduce the Cartan form θ associated with the given L, namely,[2,11]

$$\theta = L\, ds + \frac{\partial L}{\partial \dot{q}^a}(dq^a - \dot{q}^a\, ds), \tag{2.3}$$

it is straightforward to verify[2,11] that the definition (2.2) of Γ is equivalent to

$$i_\Gamma\, d\theta = 0, \quad \langle \Gamma, ds \rangle = 1, \tag{2.4}$$

which imply that Γ is a characteristic vector field of $d\theta$.

A vector field Y on $R \times TM$ with local representation

$$Y = \tau(s,q,\dot{q}) \frac{\partial}{\partial s} + K^a(s,q,\dot{q}) \frac{\partial}{\partial q^a} + \eta^a(s,q,\dot{q}) \frac{\partial}{\partial \dot{q}^a} \tag{2.5}$$

is said to be a *dynamical symmetry* of Γ if and only if [2,4]

$$\mathcal{L}_Y \Gamma = [Y,\Gamma] = g\Gamma, \tag{2.6}$$

where g is a differentiable function and $\mathcal{L}$ denotes the Lie differentiation operator. In view of (2.2) and (2.5), (2.6) is equivalent to[2]

$$\eta^a = \Gamma(K^a) - \dot{q}^a \Gamma(\tau), \tag{2.7}$$

$$\Gamma(\eta^a) - \Lambda^a \Gamma(\tau) - Y(\Lambda^a) = 0, \tag{2.8}$$

$$g = -\Gamma(\tau). \tag{2.9}$$

A practical method for the construction of DS's, which works under the assumption $L = \frac{1}{2} g_{ab} \dot{q}^a \dot{q}^b$ has been described in Ref. 8. Namely, suppose that $K_{a_1 \cdots a_p} = K_{(a_1 \cdots a_p)}$ is a generalized Killing tensor, i.e., a tensor defined over M such that there exists another tensor field $k_{a_1 \cdots a_p} = k_{(a_1 \cdots a_p)}$ related to $K_{a_1 \cdots a_p}$ by the condition[9]

$$[(p+2)\, \nabla_{(b} \nabla_a K_{a_1 \cdots a_p)} - 2 \nabla_a \nabla_{(b} K_{a_1 \cdots a_p)}]/p = g_{a(b} k_{a_1 \cdots a_p)}, \tag{2.10}$$

where ∇_a is the covariant differentiation operator determined by the metric tensor g_{ab}. Then it may be shown[8] that there exists a DS Y such that K^a is given by

$$K^a = K^a_{a_2 \cdots a_p} \dot{q}^{a_2} \cdots \dot{q}^{a_p} \tag{2.11}$$

and η^a is defined by (2.7), provided τ has been determined by solving the equation

$$\Gamma\Gamma(\tau) = k_{a_1 \cdots a_p} \dot{q}^{a_1} \cdots \dot{q}^{a_p}. \tag{2.12}$$

This holds, in particular, if $K_{a_1 \cdots a_p}$ is a Killing tensor, i.e., if $\nabla_{(a} K_{a_1 \cdots a_p)} = 0$; indeed, in this case we may assume $\tau = 0$. It is also to be remarked for completeness that DS's may be related to the symmetries of the configuration manifold, e.g., to projective collineations and affine collineations,[10] because they belong to the class of generalized Killing tensors of order 1.[9]

The DS Y degenerates into a *point symmetry* of Γ iff the components τ and K^a depend only on the variables q and s.[3,5] Examples of point symmetries can be found in Refs. 3. In general, point symmetries may be related to first integrals of motion of the form[2,5]

$$F = f - \langle Y, \theta \rangle, \tag{2.13}$$

where the function f is suitably described in terms of Y and L.

A point relevant to our later discussion and more important for practical purposes is that there exists a conserved quantity of the form (2.13) corresponding to a given DS Y if and only if f is defined as a solution of[4]

$$\Gamma(f) = \Gamma\langle Y, \theta \rangle = \Gamma(\tau) L + Y(L), \tag{2.14}$$

as follows from (2.13), (2.3), (2.4), and (2.6). Of course, the problem of finding a solution for (2.14) is rather formidable. To simplify, it has been suggested either to look for a function f depending only on s and q, or to add the further restriction $Y(F) = 0$.[2,4] In Sec. 4, we shall find a canonical solution for Eq. (2.14) under the assumption $L = \frac{1}{2} g_{ab} \dot{q}^a \dot{q}^b$.

Finally, a *Noether symmetry* is generated by a vector field Y of the form (2.5) having the property

$$\mathcal{L}_Y\, d\theta = 0. \tag{2.15}$$

Every Noether symmetry is a DS.[2] Moreover, suitably defined classes of Noether symmetries can be put in one-to-one correspondence with arbitrary constants of the motion by a relation of the form (2.13), where f is now given by $\mathcal{L}_Y \theta = df$.[2,11]

3. JACOBI FIELDS ASSOCIATED WITH DS's

Let us restrict our investigation to a system described by a Lagrangian of the form $L = \frac{1}{2} g_{ab}(q)\, \dot{q}^a \dot{q}^b$, where g_{ab} is either an n-dimensional positive definite metric or a four-dimensional Lorentzian metric. The equations of motion (2.1) may be written in the form

$$\frac{d\dot{q}^a}{ds} = -\Gamma_{bc}{}^a \dot{q}^b \dot{q}^c = \Lambda^a(q,\dot{q}), \tag{3.1}$$

where the symbols $\Gamma_{bc}{}^a$ denote the connection coefficients of the metric form $g_{ab}\, dq^a\, dq^b$. Then it is easily seen that the integral curves of the field Γ project into the geodesics of M.

Consider now a DS Y of the form (2.5). If γ is an arbitrary geodesic of M, we define a vector field K over γ by the relation

$$K = K^a \frac{\partial}{\partial q^a} = K^a(s, q^b(s), \dot{q}^b(s)) \frac{\partial}{\partial q^a}. \tag{3.2}$$

It will be shown that K is strictly related to a Jacobi field on γ.

To this aim, let us note first that the restriction to γ of the scalar $\Gamma(\phi)$, where ϕ is an arbitrary function on $R \times TM$, coincides with the total derivative $d\phi(s,q^b(s),\dot{q}^b(s))/ds$. Then, substitution of (2.7) into (2.8) yields, in view of (3.1),

$$\frac{d^2K^a}{ds^2} - \Lambda^a \frac{d\tau}{ds} - \dot{q}^a \frac{d^2\tau}{ds^2} - \Lambda^a \frac{d\tau}{ds} - Y(\Lambda^a) = 0. \tag{3.3}$$

After substitution of the explicit expression for $Y(\Lambda^a)$, (3.3) reduces to

$$\frac{d^2K^a}{ds^2} + K^d \frac{\partial \Gamma_{bc}{}^a}{\partial q^d} \dot{q}^b \dot{q}^c + 2\Gamma_{bc}{}^a \frac{dK^b}{ds} \dot{q}^c = \dot{q}^a \frac{d^2\tau}{ds^2}, \tag{3.4}$$

that can be cast into the more significant form[12]

$$\frac{D^2K^a}{Ds^2} + R^a{}_{bcd}\dot{q}^b K^c \dot{q}^d = \dot{q}^a \frac{d^2\tau}{ds^2}, \tag{3.5}$$

where we have set $\dot{q}^a\nabla_a = D/Ds$ and $R^a{}_{bcd}$ denote, as usual, the components of the curvature tensor. If $d^2\tau/ds^2 = 0$, then (3.5) reduces to the well-known equation of geodesic deviation.[6]

Actually, an equation of the form (3.5) has already been dealt with in Ref. 9, where generalized Killing tensors have been characterized as the class of totally symmetric tensor fields over M that give rise to solutions of the form (2.11) for Eq. (3.5), along every geodesic of M. Indeed, taking also into account the connections between generalized Killing tensors and DS's (see Sec. 2), we can give a further contribution to the investigation made in Ref. 9; namely, we can reach the conclusion that DS's may be regarded as the most natural generators of solutions for Eq. (3.5).

As to the practical value of (3.5), it may be used to verify that the field $\hat{K} = \hat{K}^a\,\partial/\partial q^a = (K^a - \tau\dot{q}^a)\,\partial/\partial q^a$ is a Jacobi field along γ, i.e., a solution for the equation of geodesic deviation

$$\frac{D^2}{Ds^2}(K^a - \tau\dot{q}^a) + R^a{}_{bcd}\dot{q}^b(K^c - \tau\dot{q}^c)\,\dot{q}^d = 0. \tag{3.6}$$

We summarize the above discussion in the following theorem.

Theorem 3.1: Every DS Y of local expression (2.5) identifies a solution of the form (3.2) for the inhomogeneous equation of geodesic deviation (3.5) along every geodesic of M. In particular, the restriction of the vector field $\hat{K} = (K^a - \tau\dot{q}^a)\,\partial/\partial q^a$ to an arbitrary geodesic of M is a Jacobi field.

To comment on the meaning of the theorem, let us first remark that, in general, the class of DS's is not empty, since it contains at least the field $\partial/\partial s$. Unfortunately, the Jacobi field associated to $\partial/\partial s$ does always coincide, up to a sign, with the tangent vector to the fiducial geodesic.

Secondly, the connections between DS's and Jacobi fields will be used in the following section in order to find conserved quantities canonically associated with DS's. In so doing, we shall obtain a significant extension of recent results concerning the possibility of relating generalized Killing tensors to first integrals of geodesic motion.[9]

4. DS's AND CONSERVED QUANTITIES

It is now straightforward to write down the explicit expression for first integrals of geodesic motion derivable from Jacobi fields associated to DS's. Namely, multiplying (3.6) by $g_{ae}\dot{q}^e$, we deduce that the quantity

$$I_1 = g_{ab}\dot{q}^a \frac{DK^b}{Ds} - \Gamma(\tau)\,g_{ab}\,\dot{q}^a\dot{q}^b \tag{4.1}$$

is constant along each geodesic of M. Furthermore, if we notice that

$$I_1 = \frac{D}{Ds}(g_{ab}\dot{q}^a K^b - \tau g_{ab}\dot{q}^a\dot{q}^b), \tag{4.2}$$

we conclude that also

$$I_2 = \tau g_{ab}\dot{q}^a\dot{q}^b - g_{ab}\dot{q}^a K^b + s\left[g_{ab}\dot{q}^a\frac{DK^b}{Ds} - \Gamma(\tau)\,g_{ab}\dot{q}^a\dot{q}^b\right] \tag{4.3}$$

is constant along every geodesic of M.

The conserved quantities (4.1) and (4.3) depend only on the DS Y, in the sense that they can be determined once Y is known, no further integration being required.

If the DS Y is associated with a generalized Killing tensor through (2.11), then the first integrals I_1 and I_2 reduce to conserved quantities already described in Ref. 8. In addition, if Y is related to a Killing tensor, I_1 vanishes identically and I_2 yields, up to a sign, the well-known homogeneous polynomial that is known to be conserved along the geodesics.

Finally, when considering the case $Y = \partial/\partial s$, it turns out that I_1 vanishes and I_2 corresponds to the well-known energy integral, as was to be expected.

To clarify the meaning of I_1 and I_2, we remark that a straightforward calculation yields

$$I_1 = Y(L). \tag{4.4}$$

Recalling that $\Gamma(L) = 0$ and taking (4.4) into account, we may rewrite (2.14) in the form

$$\Gamma(f) = \Gamma(\tau L) + \Gamma[sY(L)],$$

from which it follows that

$$f = \tau L + sY(L) = Y(sL). \tag{4.5}$$

Accordingly, substitution of (4.5) into (2.13) yields

$$F = Y(sL) - \langle Y, \theta\rangle = I_2, \tag{4.6}$$

where the last equality can be proved by substitution of the expressions (2.3) and (2.5) for Y and θ into the pairing $\langle Y, \theta\rangle$, and by comparison with (4.5) and (4.3). Equation (4.6) shows that I_2 may be regarded as a Noether-type conserved quantity, in the sense of a definition recently given by Lutzky.[4]

Suppose now that another DS Z of local representation

$$Z = \eta(s,q,\dot{q})\frac{\partial}{\partial s} + H^a(s,q,\dot{q})\frac{\partial}{\partial q^a} + \lambda^a(s,q,\dot{q})\frac{\partial}{\partial\dot{q}^a} \tag{4.7}$$

is given, and define the vectors $H = H^a\,\partial/\partial q^a$ and $\hat{H} = \hat{H}^a\,\partial/\partial q^a = (H^a - \eta\dot{q}^a)\,\partial/\partial q^a$.

In view of the fact that both $\hat{K}$ and $\hat{H}$ satisfy (3.6), it may be shown that the quantities

$$I_3 = \frac{D\hat{K}^a}{Ds}\hat{H}_a - \frac{D\hat{H}^a}{Ds}\hat{K}_a, \tag{4.8}$$

$$I_4 = \left(\frac{D\hat{K}^a}{Ds}\hat{H}_b - \frac{D\hat{H}^a}{Ds}\hat{K}_b\right)g_{ac}\,\dot{q}^c\dot{q}^b \tag{4.9}$$

are constant along the geodesics of M. In order to obtain the proof of the above statement, it suffices to evaluate the D/Ds derivative of both sides of (4.8) and (4.9), and to take (3.6) into account. It is also to be noticed that the expression of I_3 and I_4 in terms of K^a, H^a, τ, and η is given by

$$I_3 = g_{ab}\left\{\left[\frac{DK^a}{Ds} - \Gamma(\tau)\,\dot{q}^a\right](H^b - \eta\dot{q}^b) - \left[\frac{DH^a}{Ds} - \Gamma(\eta)\,\dot{q}^a\right](K^b - \tau\dot{q}^b)\right\}, \tag{4.10}$$

$$I_4 = g_{ab}\, g_{cd} \left\{ \left[\frac{DK^a}{Ds} - \Gamma(\tau)\, \dot{q}^a \right] (H^c - \eta \dot{q}^c) \right.$$

$$\left. - \left[\frac{DH^a}{Ds} - \Gamma(\eta)\, \dot{q}^a \right] (K^c - \tau \dot{q}^c) \right\} \dot{q}^b \dot{q}^d. \tag{4.11}$$

As a first comment, let us remark that a rather straightforward calculation shows that I_4 is not independent of the remaining conserved quantities, since we have

$$I_4 = I_2 I_1' - I_1 I_2', \tag{4.12}$$

where I_1' and I_2' denote, respectively, the first integrals of the type (4.1) and (4.3) corresponding to Z.

Secondly, I_3 and I_4 reduce to the first integrals already described in Corollary 3.1 and in Theorem 3.3 of Ref. 9, respectively, under the assumption that both Y and Z are DS's associated with Killing tensors through (2.11).

Thirdly, let us consider the first integral I_5 defined by

$$I_5 = I_4 - g_{ab}\, \dot{q}^a \dot{q}^b I_3$$

$$= g_{ab}\, \dot{q}^a \dot{q}^b \left(K_d \frac{DH^d}{Ds} - H_d \frac{DK^d}{Ds} \right)$$

$$- \left(\frac{DH_a}{Ds} K_b - \frac{DK_a}{Ds} H_b \right) \dot{q}^a \dot{q}^b. \tag{4.13}$$

In view of (4.12), I_5 is essentially equivalent to I_3. Moreover, the detailed expression of I_5 shows that it does not depend on τ and η. Accordingly, in the particular case that both Y and Z are determined by generalized Killing tensors, it is possible to write down I_5 explicitly, without any need for a further integration of (2.12); furthermore, the conserved quantity (4.13) reduces to the one already found and commented on in Theorem 3.4 of Ref. 9. In addition, we want to emphasize that, under the more severe restriction that Y and Z correspond to Killing tensors, I_5 coincides with the first integral generated by the well-known Schouten–Nijenhuis bracket[13] of the given tensors, up to a factor $g_{ab}\, \dot{q}^a \dot{q}^b$, which, of course, is constant on every geodesic.

It follows from the above discussion that, besides giving rise to new classes of previously unknown conserved quantities, DS's also provide a unified approach yielding almost all known first integrals of geodesic motion. In addition, it is to be noticed that some results can be extended to general Lagrangian system, as will be shown below.

Theorem 4.1: Suppose that Y and Z are DS's of a generic Lagrangian system. Then the quantity

$$\langle Y \wedge Z, \quad d\theta \rangle \tag{4.14}$$

is a first integral of motion.

Proof: Making use of the identity $\Gamma \langle Y \wedge Z, d\theta > = \mathcal{L}_\Gamma \langle Y \wedge Z, d\theta \rangle$ and recalling the definition of DS, we have

$$\Gamma \langle Y \wedge Z, d\theta \rangle = \Gamma(\tau) \langle \Gamma \wedge Z, d\theta \rangle + \Gamma(\eta) \langle Y \wedge \Gamma, d\theta \rangle$$

$$+ \langle Y \wedge Z, \mathcal{L}_\Gamma\, d\theta \rangle. \tag{4.15}$$

The terms $\langle \Gamma \wedge Z, d\theta \rangle$ and $\langle Y \wedge \Gamma, d\theta \rangle$ vanish in consequence of (2.4a). Moreover, we have

$$\mathcal{L}_\Gamma\, d\theta = i_\Gamma\, d\, d\theta + d\, (i_\Gamma\, d\theta) = 0,$$

in view of (2.4a). The proof follows after substitution of the previous results into (4.15). ■

Coming back to the case $L = \frac{1}{2} g_{ab}\, \dot{q}^a \dot{q}^b$, it may be shown that

$$\langle Y \wedge Z, d\theta \rangle = \tfrac{1}{2} I_3, \tag{4.16}$$

where I_3 is given by (4.10). This result may be rephrased by saying that the first integral I_3, which is essentially equivalent to I_5 in the case of a quadratic Lagrangian, can be extended to arbitrary Lagrangian systems using the formula (4.14).

One last remark is in order now, before the end of this section. The vector field $[Y, Z]$ is a DS, because it is related to Γ by the condition

$$[[Y, Z], \Gamma] = \Gamma [Z(\tau) - Y(\eta)]\, \Gamma. \tag{4.17}$$

It follows that

$$I_2'' = [Y, Z](sL) - \langle [Y, Z], \theta \rangle \tag{4.18}$$

is a conserved quantity. The question now arises as to whether I_2'' is a new independent first integral or not. The answer is negative, because making use of the identity[14]

$$2\langle Y \wedge Z, d\theta \rangle$$

$$= Y \langle Z, \theta \rangle - Z \langle Y, \theta \rangle - \langle [Y, Z], \theta \rangle, \tag{4.19}$$

it may be shown that the following identity also holds:

$$Y(I_2') - Z(I_2) = I_2'' - 2\langle Y \wedge Z, d\theta \rangle. \tag{4.20}$$

Recalling that, in general, the derivative of a conserved quantity along the direction of a DS is a conserved quantity,[8] we conclude by inspection of (4.20) that I_2'' is linearly related to the constants of motion $Y(I_2')$, $Z(I_2)$, and $\langle Y \wedge Z, d\theta \rangle$.

5. COMMENTS

We hope to have shown in this paper the theoretical as well as practical advantages that can be achieved by the introduction of the concept of DS as a tool for the analysis of geodesic motion. We shall summarize here a number of arguments in order to support our claim.

(i) A DS identifies a Jacobi field along every geodesic of the space-time manifold (configuration space).

(ii) There exists a canonical correspondence showing that well-known classes of symmetry generators defined over the space-time manifold (configuration space), such as projective collineations, affine collineations, homothetic motions, and Killing vectors, can be viewed as DS's.[8]

(iii) Killing tensors and generalized Killing tensors, which are known to give rise to conserved quantities and to Jacobi fields,[9,15] can be reinterpreted as generators of DS's through (2.11) and (2.12). In particular, one can always associate quantities to generalized Killing tensors by means of the related DS.

(iv) If Y denotes a DS corresponding to a projective collineation or to a Killing tensor, then the constants of motion I_1 and I_2 [see (4.1) and (4.3)] reduce to the conserved quantities concomitant with the given projective collineation and Killing tensor, respectively.

(v) The first integral I_3 [see (4.10)], which is essentially equivalent to I_5, yields a proper extension of a family of conserved quantities that have been recently found in Ref. 9. In particular, it follows that when Y and Z are properly chosen, I_3 corresponds, via I_5, to the related integral theorem of Katzin and Levine.[10,16]

Coming to more general considerations, we want to remark that in the course of the paper we have not given specific examples, because they may be easily constructed using the connections between DS's and the projective collineations or the Killing tensors of the given space-time manifold. As yet, we have made no systematic attempt to find examples of DS's that do not come from tensor fields of the configuration space.

Similarly, the problem of extending the first integrals I_1 and I_2, or at least one of them, to arbitrary Lagrangian systems is still a matter of investigation. On the contrary, it has been shown that the quantity I_3 is conserved for arbitrary systems.

[1]D. Lovelock and H. Rund, *Tensors, Differential Forms, and Variational Principles* (Wiley, New York, 1975).
[2]W. Sarlet and F. Cantrijn, SIAM Rev. **23**, 467 (1981).
[3]G. E. Prince and C. J. Eliezer, J. Phys. A **13**, 815 (1980); G. E. Prince and C. J. Eliezer, J. Phys. A **14**, 587 (1981); M. Lutzky, J. Phys. A **11**, 249 (1978); P. G. L. Leach, J. Math. Phys. **22**, 679 (1981).
[4]M. Lutzky, J. Phys. A **12**, 973 (1979).
[5]W. Sarlet and F. Cantrijn, J. Phys. A **14**, 479 (1981); F. Cantrijn, J. Math. Phys. **23**, 1589 (1982); P. G. L. Leach, J. Austral. Math. Soc. Ser. B **23**, 173 (1981); F. Gonzales-Gascon, J. Phys. A **13**, 1965 (1980); M. Lutzky, Phys. Lett. A **75**, 8 (1979); M. Lutzky, Phys. Lett. A **72**, 86 (1979); M. Lutzky, J. Phys. A **15**, L87 (1982); G. H. Katzin and J. Levine, J. Math. Phys. **17**, 1345 (1976); L. Martinez Alonso, Lett. Math. Phys. **3**, 419 (1979).
[6]C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).
[7]B. C. Xanthopoulos, Lett. Math. Phys. **6**, 199 (1982).
[8]G. Caviglia, Internat. J. Theoret. Phys. (in press).
[9]G. Caviglia, F. Salmistraro, and C. Zordan, J. Math. Phys. **23**(12), 2346 (1982).
[10]G. H. Katzin and J. Levine, Colloq. Math. (Wroclaw) **26**, 21 (1972).
[11]M. Crampin, Internat. J. Theoret. Phys. **16**, 741 (1977).
[12]M. F. Shirokov, Gen. Relativ. Gravit. **4**, 131 (1973).
[13]P. Sommers, J. Math. Phys. **14**, 787 (1973); R. Geroch, J. Math. Phys. **11**, 1955 (1970).
[14]C. von Westenholz, *Differential Forms in Mathematical Physics* (North-Holland, Amsterdam, 1981).
[15]G. Caviglia, C. Zordan, and F. Salmistraro, Internat. J. Theoret. Phys. **21**, 391 (1982).
[16]G. H. Katzin and J. Levine, J. Math. Phys. **9**, 8 (1968); **14**, 1213 (1973); **15**, 1460 (1974).

The asymptotic evaluation of a class of path integrals. II

J. M. Luttinger
Department of Physics, Columbia University, New York, New York 10027

(Received 26 July 1982; accepted for publication 17 December 1982)

The asymptotic behavior of a class of Wiener-like path integrals (functions of the "local time") is determined. These integrals are of interest in themselves and also arise very naturally in the theory of disordered systems. We show that by making use of a Grassman algebra (i.e., a set of anticommuting variables), the earlier treatment of this problem can be greatly simplified. In particular, the previous use of "replica trick" (which involves a difficult to justify analytic continuation in the number of field components) is thus avoided.

PACS numbers: 02.50. + s, 02.30.Bi

1. INTRODUCTION

In a previous publication[1] we have shown that a class of Wiener path integrals (functionals of the "local time") may be asymptotically evaluated in the limit of large time by a version of the "replica trick." This involves a formal analytic continuation of an n-component field theory to the point $n = 0$. Whereas there is little doubt that the method is correct for a wide class of functionals, it has proved difficult to give more than a rather heuristic justification of it. In addition, the actual carrying out of the continuation process involves first doing the problem for a field theory with an arbitrary number of components. This is technically complicated, though straightforward, and makes the discussion of the corrections to the leading term almost prohibitive.

In this paper we approach the same problem from another point of view. The new approach involves no analytical continuation in n, and is also technically somewhat simpler. The method is due to Parisi and Sourlas[2] and McKane,[3] making use of anticommuting variables (a Grassman algebra) in a form familiar from field theory.[4] In Sec. 2 we give the new formulation of the problem. This seems interesting in its own right. (Although we shall not exploit it here, the theory has a very suggestive "supersymmetry.") In Sec. 3 we evaluate our expressions for large t using the Laplace method for integrals containing a large parameter, and recover the results of I.

2. GENERAL FORMULAS

We consider first the Green's function $G_t(x, x')$ associated with the Hermitian operator H in d-dimensional space. x, x' are points in the d-dimensional space, $t > 0$, and

$$H = H_0 + V(x), \tag{2.1}$$

$$\frac{\partial G_t(x, x')}{\partial t} + HG_t(x, x') = 0, \tag{2.2}$$

$$\lim_{t \to 0^+} G_t(x, x') = \delta(x - x'). \tag{2.3}$$

(In I we took $H_0 = -\frac{1}{2}\nabla_x^2$, which yields results for the usual Wiener integral. Our formulas will be slightly more general.) G_t has well-known representation

$$G_t(x, x') = \sum_\mu \phi_\mu(x)\, \phi_\mu^*(x')\, e^{-tE_\mu}, \tag{2.4}$$

where the ϕ_μ are the complete orthonormal set of eigenfunctions of H:

$$H\phi_\mu = E_\mu \phi_\mu, \tag{2.5}$$

$$\int dx\, \phi_\nu^*(x)\, \phi_\mu(x) \equiv (\phi_\nu, \phi_\mu) = \delta_{\mu\nu}. \tag{2.6}$$

We confine our system to a volume Ω in R^d and discretize the operator H on a set of N equispaced points in Ω. Ultimately, we will go to the limit $N \to \infty$, $\Omega \to \infty$. Now introduce (at each of these points x) $\eta(x)$ and $\eta^*(x)$ which are elements of a $2N$-dimensional Grassman algebra, i.e., all these $2N$ elements anticommute with each other. We define along with Berezin[4] an operation of "integration" on any elements η, η^* at the same point

$$\int d^2\eta(a_0 + a_1\eta + a_2\eta^* + a_3\eta\eta^*) = a_3 \frac{1}{\pi}. \tag{2.7}$$

We will now show that

$$G_t(x, x') = \int \mathscr{D}^2\phi\, \mathscr{D}^2\eta\eta(x)\, \eta^*(x')\, e^{-t[(\phi, H\phi) + (\eta, H\eta)]} \times \delta((\phi,\phi) + (\eta,\eta) - 1), \tag{2.8}$$

where

$$\mathscr{D}^2\phi\mathscr{D}^2\eta = \prod_{x=1}^{N} d(\mathrm{Re}\,\phi(x))\, d(\mathrm{Im}\,\phi(x))\, d^2\eta(x).$$

[The range of integration for Re $\phi(x)$ and Im $\phi(x)$ is $(-\infty, \infty)$.]

$$(\phi, H\phi) = \int \phi^*(x)\, H_{xy}\, \phi(y)\, dx\, dy,$$

$$(\phi,\phi) = \int \phi^*(x)\, \phi(x)\, dx,$$

$$(\eta, H\eta) = \int \eta^*(x)\, H_{xy}\eta(y)\, dx\, dy, \quad (\eta,\eta) = \int \eta^*(x)\, \eta(x)\, dx.$$

H_{xy} is the (discretized) matrix element $\langle x|H|y\rangle$ of H, and we write everything (as usual) as if we'd already gone to the continuum limit.

The δ-function in (2.8) is a finite sum of derivatives of δ-functions [obtained by expanding in (η,η), $(\eta,\eta)^N$ being the highest power that comes]. It is easy to see that we can manipulate this δ in all the usual ways.

Writing

 0022-2488/83/082070-04$02.50

$$\phi(x) = \sum_\nu b_\nu \phi_\nu(x),$$

$$\eta(x) = \sum_\nu a_\nu \phi_\nu(x), \tag{2.9}$$

$$\eta^*(x) = \sum_\nu a_\nu^* \phi_\nu^*(x).$$

The a_ν, a_ν^* again form a Grassman algebra. It is easy to see[4] that we can write

$$\mathscr{D}^2\phi\, \mathscr{D}^2\eta = \mathscr{D}^2 b\, \mathscr{D}^2 a \equiv \prod_\nu d(\mathrm{Re}\, b_\nu)\, d(\mathrm{Im}\,(b_\nu))\, d^2 a_\nu$$

so that the right-hand side of (2.8) becomes

$$\int \mathscr{D}^2 b\, \mathscr{D}^2 a \Big(\sum_{\nu,\nu'} \phi_\nu(x)\, \phi_{\nu'}^*(x')\, a_\nu\, a_{\nu'}^* \Big) \times \prod_\nu (e^{-tE_\nu (b_\nu^* b_\nu + a_\nu^* a_\nu)})\, \delta\Big(\sum_\nu (b_\nu^* b_\nu + a_\nu^* a_\nu) - 1 \Big)$$
$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega\, e^{i\omega} \int \mathscr{D}^2 b\, \mathscr{D}^2 a \times \Big(\sum_{\nu,\nu'} \phi_\nu(x)\, \phi_{\nu'}^*(x')\, a_\nu\, a_{\nu'}^* \Big) \times \prod_\nu (e^{-(tE_\nu + i\omega)(b_\nu^* b_\nu + a_\nu^* a_\nu)}). \tag{2.10}$$

Using the integration formulas

$$\int d^2 a_\nu\, e^{-\lambda a_\nu^* a_\nu} = \frac{\lambda}{\pi},$$
$$\int d^2 a_\nu\, a_\nu a_\nu^*\, e^{-\lambda a_\nu^* a_\nu} = \frac{1}{\pi}, \tag{2.11}$$

the right-hand side of (2.10) becomes[5]

$$\sum_\nu \phi_\nu(x')\, \phi_\nu^*(x) \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{e^{i\omega}}{tE_\nu + i\omega} = \sum_\nu \phi_\nu(x)\, \phi_\nu^*(x')\, e^{-tE_\nu} = G_t(x, x'). \tag{2.12}$$

Again, just as in I, we use for $G_t(x, x')$ the Feynman–Kac representation as an average over Brownian paths[6]

$$G_t(x, x') = G_t^{(0)}(x, x') \left\langle e^{-t\int L_t(y)\, V(y)\, dy} \,\middle|\, \begin{matrix} x(t) = x' \\ x(0) = x \end{matrix} \right\rangle, \tag{2.13}$$

$$G_t^{(0)}(x, x') = \frac{1}{(2\pi t)^{d/2}} \exp\left\{ -\frac{|x - x'|^2}{2t} \right\}, \tag{2.14}$$

$$L_t(y) = \text{"Local time"} \equiv \frac{1}{t} \int_0^t \delta(y - y(\tau))\, d\tau,$$

and the angular bracket represents the conditional average over all Brownian paths $y(\tau)$ starting at $y(\tau) = x$ for $\tau = 0$ and ending at $y(\tau) = x'$ for $\tau = t$. Therefore, we may write (2.8) as

$$\left\langle e^{-t\int L_t(y)\, V(y)\, dy} \,\middle|\, \begin{matrix} x(t) = x' \\ x(0) = x \end{matrix} \right\rangle = [G_t^{(0)}(x, x')]^{-1} \int \mathscr{D}^2\phi\, \mathscr{D}^2\eta\eta(x)\, \eta^*(x')\, \delta((\phi,\phi) + (\eta,\eta) - 1) \exp\left\{ -t \left[(\phi, H_0\phi) + (\eta, H_0\eta) + \int dy (\phi^*(y)\, \phi(y) + \eta^*(y)\, \eta(y))\, V(y)\, dy \right] \right\}. \tag{2.15}$$

Exactly the same reasoning as in I [discussion following (2.18)] now allows us to conclude that

$$\left\langle e^{-tF_t[L_t(\cdot)]} \,\middle|\, \begin{matrix} x' \\ x \end{matrix} \right\rangle = [G_t^{(0)}(x, x')]^{-1} \times \int \mathscr{D}^2\phi\, \mathscr{D}^2\eta\eta(x)\, \eta^*(x') \times \delta((\phi,\phi) + (\eta,\eta) - 1)\, e^{-tS}, \tag{2.16}$$

$$S \equiv (\phi, H_0\phi) + (\eta, H_0\eta) + F_t[\phi^*(\cdot)\, \phi(\cdot) + \eta^*(\cdot)\, \eta(\cdot)], \tag{2.17}$$

where $F_t[L_t(\cdot)]$ is an "arbitrary" functional of $L_t(y)$ such that both sides of (2.16) exist. We also note that the same method shows that (2.16) is also valid if $\eta(x)\, \eta^*(x')$ is replaced by $\phi(x)\, \phi^*(x')$ (this is related to the "supersymmetry" of the theory). For our purposes, however, (2.16) proves a little more convenient.

3. ASYMPTOTIC EXPANSION FOR LARGE t

Although we can easily consider more general cases we consider (as in I) for simplicity the case where F_t is independent of t, $F_t = F$. First, we make a slight change in variables. From (2.7) we see at once that if we replace $\eta \to \eta/t$, $\eta^* \to \eta^*\, d^2\eta \to t d^2\eta$ all η, η^* integrals are left invariant, so that (2.16) may also be written

$$G_t^{(0)}(x, x') \left\langle e^{-tF[L_t(\cdot)]} \,\middle|\, \begin{matrix} x' \\ x \end{matrix} \right\rangle = t^{N-1} \int \mathscr{D}^2\phi\, \mathscr{D}^2\eta\eta(x)\, \eta^*(x')\, \delta((\phi,\phi) + \frac{1}{t}(\eta,\eta) - 1)\, e^{-t\tilde{S}}, \tag{3.1}$$

$$\tilde{S} \equiv (\phi, H_0\phi) + (1/t)(\eta, H_0\eta) + F[\phi^*(\cdot)\, \phi(\cdot) + (1/t)(\eta^*(\cdot)\, \eta(\cdot)].$$

For large t the leading term of $\tilde{S}$ will be a functional of ϕ, ϕ^* alone and therefore (as in I) it is natural to use the Laplace method, i.e., put

$$\phi(x) = \psi_0(x) + \psi(x), \tag{3.2}$$

where ψ_0 is normalized to unity [the δ in (3.1)] and will be determined by the Laplace method. ψ will be of the order of $1/\sqrt{t}$, so we may expand in it as well as $\eta^*\eta/t$ terms. A simple calculation gives for the right-hand side of (3.1) the expression

$$t^{N-1} \int \mathscr{D}^2\phi\, \mathscr{D}^2\eta\eta(x)\, \eta^*(x')\, e^{-t[\tilde{S}_0 + \tilde{S}_1 + \tilde{S}_2]}\, \delta((\psi_0,\psi) + (\psi,\psi_0) + (\psi,\psi) + \frac{1}{t}(\eta,\eta)) \left[1 + O\left(\frac{1}{t}\right) \right],$$

$$\tilde{S}_0 = (\psi_0, H_0\psi_0) + F[\psi_0(\cdot)\, \psi_0(\cdot)], \tag{3.3}$$

$$\tilde{S}_1 = (\psi, H_0\psi_0) + (\psi_0, H_0\psi) + \int dx\, U(x)\, (\psi_0^*(x)\, \psi(x) + \psi^*(x)\, \psi_0(x)), \tag{3.4}$$

$$\tilde{S}_2 = (\psi, H_0\psi) + \frac{1}{t}(\eta, H_0\eta) + \int dx\, U(x)(\psi^*(x)\, \psi(x) + \frac{1}{t}\eta^*(x)\, \eta(x)) + \frac{1}{2}\int dx\, dx'\, U(x, x')\, (\psi_0^*(x)\, \psi(x) + \psi^*(x)\, \psi_0(x))(\psi_0^*(x')\, \psi(x') + \psi^*(x')\, \psi_0(x')), \tag{3.5}$$

where

$$U(x) \equiv \frac{\delta F[\psi_0^*(\cdot)\,\psi_0(\cdot)]}{\delta(\psi_0^*(x)\,\psi_0(x))}, \tag{3.6}$$

$$U(x,x') \equiv \frac{\delta^2 F[\psi_0^*(\cdot)\,\psi_0(\cdot)]}{\delta(\psi_0^*(x)\,\psi_0(x))\,\delta((\psi_0^*(x')\,\psi_0(x'))}. \tag{3.7}$$

ψ_0 is determined by making $\tilde{S}_0$ minimum subject to the constraint $(\psi_0,\psi_0) = 1$, i.e.,

$$H_0\psi_0(x) + U(x)\,\psi_0(x) = \epsilon_0\psi_0(x), \tag{3.8}$$

where ϵ_0 is the Lagrange multiplier corresponding to the constraint. Using (3.8) $\tilde{S}_1$ becomes

$$\tilde{S}_1 = \epsilon_0[(\psi,\psi_0) + (\psi_0,\psi)]$$

$$= -\epsilon_0\left[(\psi,\psi) + \frac{1}{t}(\eta,\eta)\right] \tag{3.9}$$

because of the δ-function in (3.1). Thus we may write for the right-hand side of (3.1)

$$t^{N-1}e^{-t\tilde{S}_0}\int \mathscr{D}^2\phi\,\delta((\psi_0,\psi) + (\psi,\psi_0))$$

$$\times e^{-tA}\int \mathscr{D}^2\eta\eta(x)\,\eta^*(x')\,e^{-B}\left[1 + O\left(\frac{1}{t}\right)\right], \tag{3.10}$$

$$A = (\psi, \mathcal{H}\psi) + \frac{1}{2}\int dx\,dx'\,U(x,x')\,(\psi_0^*(x)\,\psi(x)$$

$$+ \psi^*(x)\,\psi_0(x))(\psi_0^*(x')\,\psi(x') + \psi^*(x')\,\psi_0(x')), \tag{3.11}$$

$$B = (\eta, \mathcal{H}\eta), \tag{3.12}$$

$$\mathcal{H} = H_0 + U - \epsilon_0. \tag{3.13}$$

In Eq. (3.10) the η-integration is Gaussian and therefore trivial. To keep things as simple as possible let H_0 and F be real so that $\mathcal{H}$ is real, and let $\chi_\lambda(x)$ be a *real* complete orthonormal set of eigenfunctions of $\mathcal{H}$:

$$\mathcal{H}\chi_\lambda = \mathcal{H}_\lambda\,\chi_\lambda. \tag{3.14}$$

Clearly χ_0 satisfies (3.8) with $\mathcal{H}_0 = 0$. Put

$$\eta(x) = \sum_\lambda c_\lambda\,\chi_\lambda(x),$$

$$\eta^*(x) = \sum_\lambda c_\lambda^*\,\chi_\lambda(x). \tag{3.15}$$

c_λ, c_λ^* are again a Grassman algebra and

$$\int \mathscr{D}^2\eta\eta(x)\,\eta^*(x')\,e^{-B}$$

$$= \int \mathscr{D}^2 c\left(\sum_{\lambda,\lambda'}\chi_\lambda(x)\,\chi_{\lambda'}(x')\,c_\lambda\,c_{\lambda'}^*\right)\prod_\lambda e^{-\mathcal{H}_\lambda c_\lambda^* c_\lambda}$$

$$= \frac{1}{\pi^N}\sum_\lambda \chi_\lambda(x)\,\chi_\lambda(x')\prod_{\lambda'\neq\lambda}\mathcal{H}_{\lambda'}. \tag{3.16}$$

on using (2.11). If $\mathcal{H}_0 = 0$ is the only zero eigenvalue of $\mathcal{H}$ (which we assume—only in special cases will there be degeneracy) then (3.16) becomes

$$\int \mathscr{D}^2\eta\eta(x)\,\eta^*(x)\,e^{-B} = \frac{1}{\pi^N}\chi_0(x)\,\chi_0(x')\prod_{\lambda\neq 0}\mathcal{H}_\lambda$$

$$= \frac{1}{\pi^N}\chi_0(x)\,\chi_0(x')\,\mathrm{Det}'(\mathcal{H}), \tag{3.17}$$

where $\mathrm{Det}'(\mathcal{H})$ means the determinant of $\mathcal{H}$ calculated in the space orthogonal to χ_0.

The ϕ-integration in (3.10) is a bit more complicated for two reasons: the presence of the δ-function and of a degeneracy not previously mentioned. Any solution of (3.8) multiplied by an arbitrary constant phase factor $e^{i\theta}$ (θ real) is an equally good solution. Therefore the minimum of $\tilde{S}_0$ is not a point in function space but a line, and we must take this into account. (This is analogous to, but much simpler than a similar problem which arises in I.)

Expand $\phi(x)$ in the $\chi_\lambda(x)$,

$$\phi(x) = \sum_\lambda \alpha_\lambda\,\chi_\lambda(x) = \alpha_0\,\chi_0(x) + \sum_{\lambda\neq 0}\alpha_\lambda\,\chi_\lambda(x), \tag{3.18}$$

and put $\alpha_0 = \rho\,e^{i\theta}$ (ρ, θ real, $\rho > 0$, $0 \leqslant \theta \leqslant 2\pi$). Writing

$$\phi(x) = e^{i\theta}\chi_0(x) + (\rho - 1)\,e^{i\theta}\chi_0(x) + \sum_{\lambda\neq 0}\alpha_\lambda\,\chi_\lambda(x), \tag{3.19}$$

Eq. (3.19) is just (3.2) with the arbitrary phase factor $e^{i\theta}$ made explicit, i.e.,

$$\psi_0(x) = e^{i\theta}\chi_0(x), \tag{3.20}$$

$$\psi(x) = (\rho - 1)\,e^{i\theta}\chi_0(x) + \sum_{\lambda\neq 0}\alpha_\lambda\,\chi_\lambda(x). \tag{3.21}$$

Further since (3.18) is a unitary transformation we have

$$\mathscr{D}^2\phi = \prod_\lambda d^2\alpha_\lambda = d^2\alpha_0\prod_{\lambda=0} d^2\alpha_\lambda$$

$$= \rho\,d\rho\,d\theta\prod_{\lambda\neq 0} d^2\alpha_\lambda. \tag{3.22}$$

We also have for the argument of the δ-function in (3.10),

$$(\psi_0,\psi) + (\psi,\psi_0) = 2(\rho - 1). \tag{3.23}$$

Therefore the ϕ-integral in (3.10) becomes (after doing the ρ-integration)

$$\int \mathscr{D}^2\phi\delta((\psi_0,\psi) + (\psi,\psi_0))\,e^{-tA}$$

$$= \frac{1}{2}\int_0^{2\pi} d\theta\prod_{\lambda\neq 0} d^2\alpha_\lambda\,e^{-tA'}, \tag{3.24}$$

$$A' = \sum_{\lambda\neq 0}|\alpha_\lambda|^2\,\mathcal{H}_\lambda + \frac{1}{2}\sum_{\lambda,\lambda'\neq 0} L_{\lambda\lambda'}(\alpha_\lambda\,e^{-i\theta}$$

$$+ \alpha_\lambda^*\,e^{i\theta})(\alpha_{\lambda'}\,e^{-i\theta} + \alpha_{\lambda'}^*\,e^{i\theta}),$$

$$L_{\lambda\lambda'} = \int dx\,dx'\,U(x,x')\,\chi_0(x)\,\chi_0(x')\,\chi_\lambda(x)\,\chi_{\lambda'}(x'). \tag{3.25}$$

Set $\alpha_\lambda = e^{i\theta}\beta_\lambda$. This is just a rotation of the real and imaginary axis through θ, so that $d^2\alpha_\lambda = d^2\beta_\lambda$. Putting $\beta_\lambda = u_\lambda + iv_\lambda$ (u_λ, v_λ real), A' becomes

$$A' = \sum_{\lambda\neq 0}(u_\lambda^2 + v_\lambda^2)\,\mathcal{H}_\lambda + 2\sum_{\lambda,\lambda'\neq 0} L_{\lambda\lambda'}\,u_\lambda\,u_{\lambda'} \tag{3.26}$$

and

$$\prod_{\lambda\neq 0} d^2\beta_\lambda = \prod_{\lambda\neq 0} du_\lambda\,dv_\lambda.$$

The integral over θ is now trivial and the u_λ, v_λ integrals are Gaussian: the right-hand side of (3.24) becomes

$$(\pi^N/t^{N-1})\,[\det'(\mathcal{A})]^{-1/2}\,[\mathrm{Det}'(\mathcal{A}+2L)]^{-1/2}. \quad (3.27)$$

Combining all these results (3.1) becomes

$$G_t^{(0)}(x,x')\left\langle e^{-tF[L_t(\cdot)]}\Big|_{x}^{x'}\right\rangle = \chi_0(x)\,\chi_0(x')\,e^{-t\tilde{S}_0}\,\frac{[\det'(\mathcal{A})]^{1/2}}{[\det'(\mathcal{A}+2L)]^{1/2}} \times\left[1+O\left(\frac{1}{t}\right)\right] \quad (3.28)$$

which is exactly the result [(3.36) and (3.40)] of I.

It is not difficult to go to the next order [$O(1/t)$] by this method, but the resulting formulas are quite complicated and we won't give them here. We can also treat the case of translational invariance very easily by this same method giving of course the same result as I.

[1] J. M. Luttinger, J. Math. Phys. **23**, 1011 (1982). Henceforth we shall refer to this paper as I.

[2] G. Parisi and N. Sourlas, Phys. Rev. Lett. **43**, 744 (1979).

[3] A. J. McKane, Physics Lett. A **76**, 22 (1980).

[4] F. A. Berezin, *The Method of Second Quantization* (Academic, New York, 1966).

[5] We assume that $(\psi,H\psi)$ is bounded from below, so that without loss of generality we can take $E_\nu > 0$. This assumption is almost certainly not necessary, but simplifies the derivation of (2.8).

[6] This is the usual Wiener integral if $H_0 = -\frac{1}{2}\nabla^2$. For more general H_0 (representing, for example, a particle in a magnetic field) we get a generalization (not necessarily real) of the Wiener integral. From now on for concreteness we take the usual Wiener case.

The rotating harmonic oscillator eigenvalue problem. I. Continued fractions and analytic continuation[a)]

D. Masson
Department of Mathematics, University of Toronto, Toronto, M5S 1A1 Canada

(Received 12 January 1983; accepted for publication 11 March 1983)

The continued fraction approach to the solution of the rotating harmonic oscillator eigenvalue problem is examined in detail. It is shown how one may obtain eigenvalue information only from an analytic continuation of the continued fraction accomplished with the aid of modified approximants.

PACS numbers: 02.60.Lj, 02.30.Mv, 02.30.Lt, 02.30.Gp

1. INTRODUCTION

In a recent paper Singh *et al.*[1] claim to have obtained the eigenvalues λ of the rotating harmonic oscillator in terms of the poles of a convergent continued fraction. Further, by applying Worpitzky's theorem to the continued fraction they claim that in the strong coupling limit the eigenvalues are bounded below by the angular momentum l. This latter claim contradicts the phase integral results of Fröman and Fröman,[2] which coincide with the earlier results of Dunham[3] and Rosenthal and Motz[4] and are further substantiated by the calculations of Fröman *et al.*[5]

The present paper resolves this discrepancy by observing that the continued fraction must be analytically continued before it yields eigenvalue information. We show how this may be accomplished in a limited sense restricted to the weak coupling region.

The basic error in Ref. 1 is in their implicit assumption that the continued fraction is directly related to the physical eigenvalue problem. The coupling parameter is denoted by $\alpha > 0$ with $\alpha \to 0+$ representing the strong coupling limit. The reciprocal of their continued fraction converges for $|\arg \alpha| < \pi$ to a function $f(\lambda, \alpha, l)$ which (except for special values of λ) has a square root branch cut along the negative α axis. Their assumption that $f(\lambda, \alpha, l) = 0$ is the eigenvalue condition is not correct. The correct eigenvalue condition is given by $g(\lambda, \alpha, l) = 0$, where g is the analytic continuation of f onto the second sheet. For special values $\lambda = l + n$, $n = 1,2,\cdots$ one has f equal to a rational fraction, and the two conditions coincide since one then has $f = g$.

The boundedness proof of Ref. 1 is also in error. In order to use Worpitzky's theorem, they need a domain $D = \{\alpha: 0 < 8\alpha < |(n + l - \lambda)|^{-1}, n = 1,2,\cdots\} \supset (0, \epsilon)$ for some $\epsilon > 0$. Since $D = \varnothing$, this is clearly impossible. Their boundedness proof can, however, be corrected. The continued fraction in question [Eq. (28) of Ref. 1] can be shown to converge for $\lambda \leqslant l + 1$ and all finite α with $\arg \alpha = 0$ (in fact for $|\arg \alpha| < \pi$) by using the theory of real J-fractions (see Ref. 6 and Sec. 3). Thus the continued fraction can have no poles for $\lambda \leqslant l + 1$ on the first sheet in α cut along the negative real axis. One would conclude that all eigenvalues λ are bounded below by $l + 1$ for *any value* of $\alpha > 0$ unless one took into account the fact noted above: *In order to yield eigenvalue information, the continued fraction must be analytically continued to* $\arg \alpha = 2\pi$.

Also in Ref. 1 it is suggested that the values $\lambda = l + 1$ should be excluded from the eigenvalue spectrum. Now for the special values $\lambda = l + n$, $n = 2,3,\cdots$ there exist radial eigenfunctions which are Gaussian weighted polynomials (the $l = 0$ ones being the familiar Hermite functions) and the continued fraction terminates to become a rational fraction in α. For these special values of λ the poles of the rational fraction in α do yield values of α which produce those special values of λ as eigenvalues. These values of α are correctly given in Ref. 1. Their claim that $\lambda = l + 1$ should be excluded from the eigenvalue spectrum is based on these observations. The exclusion of $\lambda = l + 1$ is actually true only for $l = 0$. For $l > 0$ one does indeed obtain the eigenvalue $\lambda = l + 1$ for l values of α and each of the eigenvalues $\lambda = l + n$, $n = 2,3,\cdots$ for l additional values of α which are *not poles of the rational fraction in* α. These extra integer eigenvalues have corresponding eigenfunctions which are *not Gaussian-weighted polynomials*. Thus the condition $f(\lambda, \alpha, l) = 0$ is only a *sufficient* condition for eigenvalues *even* for the special values $\lambda = l + n$, $n = 1,2,\ldots,l \neq 0$.

In Sec. 2 we solve the $l = 0$ radial Schrödinger equation in terms of the parabolic cylinder function and obtain the eigenvalues in the weak and strong coupling limits.

In Sec. 3 we obtain an exact expression for the $l = 0$ continued fraction of Ref. 1 and prove that it converges for $|\arg \alpha| < \pi$ while the interval of physical interest is $\arg \alpha = 2\pi$ so that one requires an analytic continuation of the continued fraction.

In Sec. 4 we examine the possibility of modifying the continued fraction so as to accomplish this analytic continuation and show that, although this is possible theoretically, it is not of practical interest except in the weak coupling limit ($\alpha \to \infty$). With the modified continued fraction we are able to reproduce the weak coupling results of Sec. 2 and generalize them to arbitrary l.

In Sec. 5 we use differential equation techniques to prove that the eigenvalue trajectories $\lambda(\alpha)$ are monotonic increasing. Also an exact solution to the "undisplaced" oscillator problem with potential $\frac{1}{4}r^2 + l(l + 1)/r^2$ is obtained which allows one to solve the "displaced" oscillator problem, in principle, for unrestricted values of α.

[a)] Research partially supported by NSERC.

 0022-2488/83/082074-15$02.50

In Sec. 6 we use the above results to graph the $l = 0$ eigenvalue trajectories and give a schematic picture of the trajectories for arbitrary $l > 0$.

2. THE $l = 0$ SOLUTION

We use the notation of Refs. 1 and 2, where the reduced radial Schrödinger equation for the rotating harmonic oscillator is given by

$$\frac{d^2F}{dr^2} + \left[\frac{\lambda + \frac{1}{2}}{\alpha} - \frac{(r-1)^2}{4\alpha^2} - \frac{l(l+1)}{r^2}\right] F = 0 \quad (1)$$

with r the radial variable, λ the eigenvalue, $l = 0,1,2,\cdots$, the rotational quantum number, and $\alpha > 0$ a coupling parameter. One requires that the solution vanish at $r = 0$ and in the limit $r \to \infty$.

For $l = 0$ we put $z = (r-1)/\sqrt{\alpha}$ to obtain

$$\frac{d^2F}{dz^2} + (\lambda + \tfrac{1}{2} - \tfrac{1}{4}z^2)F = 0 \quad (2)$$

with boundary conditions $F = 0$ at $z = -1/\sqrt{\alpha}$ and $\lim_{z\to\infty} F = 0$.

Equation (2) may be solved exactly in terms of confluent hypergeometric functions.[5] We prefer to consider here the two linearly independent solutions given by $D_\lambda(\pm z)$, where D_λ is the parabolic cylinder function[7]

$$D_\lambda(z) = \frac{2^{-\lambda/2} e^{-z^2/2}}{\Gamma(-\lambda)} \sum_{m=0}^{\infty} (-1)^m \frac{\Gamma((m-\lambda)/2)}{m!} (z\sqrt{2})^m . \quad (3)$$

Thus

$$F = AD_\lambda(z) + BD_\lambda(-z), \quad (4)$$

where, in order to satisfy the $r = 0$ ($z = -1/\sqrt{\alpha}$) boundary condition, one may choose $A = CD_\lambda(1/\sqrt{\alpha})$ and $B = -CD_\lambda(-1/\sqrt{\alpha})$.

For large values of $|z|$ one has asymptotic expansions[7]

$$D_\lambda(z) \sim e^{-z^2/4} z^\lambda \times \left[1 - \frac{\lambda(\lambda-1)}{2z^2} + \frac{\lambda(\lambda-1)(\lambda-2)(\lambda-3)}{2\cdot 4z^4} + \cdots\right], \quad -\tfrac{3}{4}\pi < \arg z < \tfrac{3}{4}\pi, \quad (5)$$

and

$$D_\lambda(z) \sim e^{-z^2/4} \times z^\lambda \left[1 - \frac{\lambda(\lambda-1)}{2z^2} + \frac{\lambda(\lambda-1)(\lambda-2)(\lambda-3)}{2\cdot 4z^4} + \cdots\right] - \frac{\sqrt{2\pi}}{\Gamma(-\lambda)} e^{\pm\lambda\pi i} e^{z^2/4} z^{-\lambda-1} \left[1 + \frac{(\lambda+1)(\lambda+2)}{2z^2} + \frac{(\lambda+1)(\lambda+2)(\lambda+3)(\lambda+4)}{2\cdot 4z^4} + \cdots\right], \quad \tfrac{1}{4}\pi < \pm\arg z < \tfrac{5}{4}\pi. \quad (6)$$

(Note that in a standard reference, Ref. 8, Eq. (6) is given incorrectly on p. 123.)

From (4), (5), and (6) it follows that, in order to satisfy the $r \to \infty$ ($z \to \infty$) boundary condition, it is necessary and sufficient to have

$$D_\lambda(-1/\sqrt{\alpha}) = 0 \quad (7)$$

so that

$$F = CD_\lambda(1/\sqrt{\alpha})\, D_\lambda((r-1)/\sqrt{\alpha}). \quad (8)$$

The above formulas remain valid when $\lambda = n$, $n = 0,1,2,\cdots$, excepting (3) which is then replaced by the simpler and more familiar expression

$$D_n(z) = 2^{-n/2} e^{-z^2/4} H_n(z/\sqrt{2}), \quad (9)$$

where H_n is the Hermite polynomial of degree n.

Thus the $l = 0$ eigenvalues for the rotating harmonic oscillator are given by

$$\lambda = \lambda_m(\alpha), \quad m = 0,1,2,\cdots,$$

a countable set of zeros of Eq. (7) which we will order so that $\lambda_{m+1} > \lambda_m$. The corresponding eigenfunctions are given by (8). That is

$$F = C_m\, D_{\lambda_m(\alpha)}((r-1)/\sqrt{\alpha}), \quad (8')$$

where C_m is a suitable normalization constant.

In the weak coupling limit ($\alpha \to \infty$) one may solve the eigenvalue condition (7) using (3) to obtain

$$\lambda_m(\alpha) = 2m + 1 + (-1)^m \frac{2\sqrt{2/\alpha}}{m!\Gamma(-m-\frac{1}{2})} + O(\alpha^{-1}), \quad (10)$$

$m = 0,1,2,\cdots$.

In particular, one obtains

$$\lambda_0(\alpha) = 1 + a_0^{(1)}/\sqrt{\alpha} + a_0^{(2)}/\alpha + a_0^{(3)}/\alpha^{3/2} + O(\alpha^{-2}), \quad (10')$$

$$a_0^{(1)} = -\sqrt{2/\pi} = -0.797885,$$

$$a_0^{(2)} = (2/\pi)(1 - \ln 2) = 0.195349$$

$$a_0^{(3)} = -\sqrt{2/\pi}(\tfrac{1}{2} + 2/\pi - (6/\pi)\ln 2 + (3/\pi)(\ln 2)^2 - \pi/12) = -0.007824.$$

In the strong coupling limit ($\alpha \to 0+$) one may solve the eigenvalue condition (7) using (6) to obtain

$$\lambda_m(\alpha) = m + \frac{\alpha^{-m-1/2} e^{-1/(2\alpha)}}{\sqrt{2\pi}m!} [1 - \alpha(m^2 + m + 1) + O(\alpha^2)] + O(\alpha^{-2m-1} e^{-1/\alpha}). \quad (11)$$

We note that (11) is consistent with the results of Refs. 2–4 and is in good numerical agreement with the calculations of Ref. 5. For example, (11) yields $\lambda_0 \approx 7.65 \times 10^{-3}$ for $\alpha = 0.1$ as compared with the Ref. 5 value of 7.5×10^{-3}.

Equations (11) and (10) yield integer and odd integer eigenvalues in the limits $\alpha \to 0+$ and $\alpha \to \infty$, respectively. Integer eigenvalues are obtained also for intermediate values of α.[1] This can be seen from (7) and (9), from which it follows that one has $\lambda = n$ whenever $\alpha = 1/(2x_{n,i}^2)$, $i = 1,2,\ldots,[n/2]$, where $x_{n,i}$ is the ith negative zero of $H_n(x)$ the Hermite polynomial of degree n. Thus one has $\lambda = n$ for exactly $[n/2]$ distinct values of $\alpha > 0$ where $[\,\cdot\,]$ is the greatest integer function. This latter statement may be extended to more general values of $\lambda > 0$.

The zeros of $D_\lambda(z)$ appear to have been first investigated by Milne,[9] who showed that the number of real zeros was given by $[\lambda+1]$ and that a new negative zero makes an appearance at $z=-\infty$ as λ increases through integer values and an old negative zero disappears to the right of $z=0$ when λ increases through odd integer values. It follows from Milne's work that the eigenvalue $\lambda=2n-c$ is achieved for n distinct values of $\alpha>0$ while $\lambda=2n+c$ for $n+1$ distinct values of $\alpha>0$ with $0<c<1$. Also one has $\lambda=2m+1-\epsilon_m$ and $\lambda=m+\epsilon'_m$, where ϵ_m and ϵ'_m are arbitrarily small and positive for $\alpha>0$ sufficiently large and small, respectively. Equations (10) and (11) are a more precise formulation of this last statement.

3. THE CONTINUED FRACTION AND ANALYTIC CONVERGENCE

A. The continued fraction

One may attempt to solve (1) together with the boundary condition at $r=0$ by putting[1]

$$F=r^{l+1}\exp[-(r-1)^2/2\alpha]\,\chi. \tag{12}$$

One obtains

$$\frac{d^2\chi}{dr^2}+\left(\frac{2l+2}{r}-\frac{(r-1)}{\alpha}\right)\frac{d\chi}{dr}+\left(\frac{\lambda-l-1}{\alpha}+\frac{l+1}{\alpha r}\right)\chi=0 \tag{13}$$

and a convergent power series solution ($c_0\neq 0$)

$$\chi=\sum_{n=0}^{\infty}c_n r^n, \tag{14}$$

where the c_n satisfy the three-term recursion relation

$$\alpha(n+1)(n+2l+2)\,c_{n+1}+(n+l+1)\,c_n+(\lambda-l-n)\,c_{n-1}=0, \qquad n=0,1,2,\cdots. \tag{15}$$

From this one obtains

$$c_1/c_0=-1/(2\alpha) \tag{16}$$

and if c_{n-1}, c_n, and $\lambda-l-n$ are nonzero one may write (15) as

$$L_{n-1}=\frac{(l+n-\lambda)/(n+l+1)}{1+[\alpha(n+1)(n+2l+2)/(n+l+1)]\,L_n}, \tag{17}$$

where $L_n=c_{n+1}/c_n$.

Using (17) recursively together with (16), one obtains the equation

$$f(\alpha,\lambda,l)=0, \tag{18}$$

where f is the continued fraction

$$f(\alpha,\lambda,l)=1+\mathop{\mathrm{K}}_{n=1}^{\infty}\frac{\alpha a_n(\lambda,l)}{1} \tag{18'}$$

with

$$a_1=\frac{2(l+1-\lambda)}{(l+2)}, \qquad a_n=\frac{n(2l+n+1)(l+n-\lambda)}{(l+n+1)(l+n)}, \quad n=2,3,\cdots. \tag{19}$$

Equations (18), (18'), and (19) are equivalent to the vanishing of the infinite Hill determinant of the coefficients of the c_n.[10,11] In Ref. 1, Eq. (18) is wrongly interpreted as yielding the eigenvalues associated with the boundary condition $F\to 0$ as $r\to\infty$. That this is not the case can be seen by the following argument.

For λ not equal to an eigenvalue one has the large r behavior

$$\chi\sim Cr^{-\lambda-l-2}\exp[(r-1)^2/(2\alpha)], \tag{20}$$

and the eigenvalue condition must eliminate this exponential growth. Now the tail of the continued fraction in (18') is

$$\mathop{\mathrm{K}}_{i=n+1}^{\infty}\frac{\alpha a_i}{1}=\frac{\alpha(n+1)(n+2l+2)\,L_n}{(n+l+1)},$$

where $L_n=c_{n+1}/c_n$. If one assumes that $L_n\approx L_{n-1}$ for n large, then (17) has L_n satisfying the approximate equation

$$\frac{\alpha(n+1)(n+2l+2)}{(n+l+1)}L_n^2+L_n-\frac{(l+n-\lambda)}{(l+n+1)}\approx 0 \tag{21}$$

so that

$$L_n\approx L_n^{\pm}=\frac{-1\pm[1+4\alpha(n+1)(n+2l+2)(l+n-\lambda)/(n+l+1)^2]^{1/2}}{2\alpha(n+1)(n+2l+2)/(n+l+1)}. \tag{22}$$

In order to eliminate the exponential growth of (20), one must choose the negative sign in (22). This will then allow for a cancellation between the even and odd parts of (14) which would otherwise both contribute to this exponential growth. The continued fraction in (18'), however, chooses the wrong sign in the sense that its tail converges to the L_n associated with L_n^+. Thus (18) with f given by (18') cannot yield correct eigenvalues except for special cases where the continued fraction terminates, i.e., $\lambda=l+n$, $n=2,3,\cdots$.

One might hope to circumvent this restriction by modifying the continued fraction in (18'), that is, by replacing the tail L_n by an approximately correct tail such as L_n^-. This procedure of modifying a continued fraction is known to yield a correct analytic continuation for special types of limit periodic regular C-fractions.[12] We will show in Sec. 4 that, although such a modifying procedure is possible for (18'), it is, however, not practical, except in the weak coupling limit ($\alpha\to\infty$).

Our heuristic argument concerning the inappropriateness of (18) as an eigenvalue condition may be rigorously justified for the case $l=0$ by obtaining an explicit representation for the continued fraction (18') and comparing the condition (18) with the correct eigenvalue condition $D_\lambda(-1/\sqrt{\alpha})=0$ given by (7) in Sec. 2.

B. Analytic convergence

One has $D_\lambda(z)$ satisfying the recursion relation[7,8]

$$D_\lambda(z)-zD_{\lambda-1}(z)+(\lambda-1)\,D_{\lambda-2}(z)=0. \tag{23}$$

If $D_\lambda(z)$ and $D_{\lambda-1}(z)$ are not zero, one may write this as

$$\frac{D_{\lambda-1}(z)}{D_\lambda(z)} = \left(z + (1-\lambda)\frac{D_{\lambda-2}(z)}{D_{\lambda-1}(z)}\right)^{-1}. \tag{24}$$

Using this recursively yields

$$\frac{D_{\lambda-1}(z)}{D_\lambda(z)} = \frac{1}{z} + \frac{(1-\lambda)}{z} + \frac{(2-\lambda)}{z} + \cdots + \frac{(N-\lambda-1)}{z} \frac{D_{\lambda-N}(z)}{D_{\lambda-N+1}(z)} \tag{25}$$

and in the limit as $N \to \infty$ one obtains the formal continued fraction representation

$$\frac{D_{\lambda-1}(z)}{D_\lambda(z)} = \frac{1}{z} + \frac{1-\lambda}{z} + \frac{2-\lambda}{z} + \cdots. \tag{26}$$

The continued fraction in (26) is the J-fraction (or S-fraction) associated with the asymptotic series

$$\frac{D_{\lambda-1}(z)}{D_\lambda(z)} \sim z^{-1}\left(1 + \frac{\lambda-1}{z^2} + \frac{(\lambda-1)(2\lambda-3)}{z^4} + \cdots\right), \tag{27}$$

$$-\tfrac{3}{4}\pi < \arg z < \tfrac{3}{4}\pi$$

obtainable from (5). We are able to justify (26) for $\operatorname{Re} z > 0$ by establishing certain properties of $D_{\lambda-1}(z)/D_\lambda(z)$ and then applying the theory of the Hamburger moment problem.[6] An important property is the following.

Lemma 1: If $\nu < 1$, then $D_\nu(z)$ has no zeros in the half-plane $|\arg z| \leqslant \pi/2$.

Proof: $D_\nu(z)$ has no real nonnegative zeros if $\nu < 1$.[9] $D_\nu(z)$ has no pure imaginary zeros for $\nu < 1$ (more generally for $\operatorname{Im} \nu = 0$) since this would contradict the general Wronskian identity (Ref. 8, p. 117)

$$D_\nu(z)\frac{d}{dz}D_\nu(-z) - D_\nu(-z)\frac{d}{dz}D_\nu(z) = \sqrt{2\pi}/\Gamma(-\nu). \tag{28}$$

That is, $D_\nu(ia) = 0$ with a real and $a \neq 0$ implies that $D_\nu(-ia) = \bar{D}_\nu(ia) = 0$. Suppose now that $D_\nu(z)$ has a complex zero $z_0(\nu)$ in the region $|\arg z| < \pi/2$ for some $0 < \nu < 1$. $\lim_{\nu \to 0} D_\nu(z) = \exp(-z^2/4)$ certainly has no zeros. In order for $z_0(\nu)$ to have left the region $|\arg z| \leqslant \pi/2$, it would have had to either pass through the boundary of that region ($\operatorname{Re} z = 0$) or move out to $|z| = \infty$. The former is impossible since $D_\nu(ia) \neq 0$ for $\nu < 1$ and a real while the latter would contradict the asymptotic formula (5). Knowing the region to be free of zeros for $0 \leqslant \nu < 1$, it follows via the same argument that it will be free of zeros for all $\nu < 1$. Actually, the more general statement that for ν real $D_\nu(z)$ has only real zeros in the half-plane $|\arg z| \leqslant \pi/2$ also follows.

Since it is more convenient to deal with the variable $\zeta = iz$, we now consider the properties of the function

$$f_\nu(\zeta) = -iD_{\nu-1}(-i\zeta)/D_\nu(-i\zeta) \tag{29}$$

and the corresponding J-fraction

$$\frac{1}{\zeta} - \frac{1-\nu}{\zeta} - \frac{2-\nu}{\zeta} - \cdots. \tag{30}$$

Lemma 2: The function $f_\nu(\zeta)$ is asymptotically equal to the J-fraction (30) in the half-plane $\operatorname{Im} \zeta > 0$.

Proof: For the definition of asymptotic equality we refer the reader to Wall,[6] p. 316. The lemma follows immediately from (25) and (27).

Lemma 3: If ν and ζ are real, then

$$\operatorname{Im} f_\nu(\zeta) = -\pi\left[\sqrt{2\pi}\Gamma(1-\nu)|\, D_\nu(-i\zeta)|^2\right]^{-1}. \tag{31}$$

Proof: Equation (29) yields

$$2\operatorname{Im} f_\nu(\zeta) = \frac{-\left[D_{\nu-1}(-i\zeta)D_\nu(i\zeta) + D_{\nu-1}(i\zeta)D_\nu(-i\zeta)\right]}{|D_\nu(-i\zeta)|^2}. \tag{32}$$

Using the general formula (Ref. 8, p. 119)

$$D_\nu(z) = \frac{z}{2}D_{\nu-1}(z) - \frac{d}{dz}D_{\nu-1}(z), \tag{33}$$

one obtains the identity

$$D_{\nu-1}(z)D_\nu(-z) + D_{\nu-1}(-z)D_\nu(z) = W(D_{\nu-1}(z),\ D_{\nu-1}(-z)). \tag{34}$$

Equation (31) follows from (32), (34), and (28).

Lemma 4: If $\nu < 1$ and $\operatorname{Im} \zeta > 0$, then $f_\nu(\zeta)$ is analytic and is given by the Stieltjes transform

$$f_\nu(\zeta) = \int_{-\infty}^{\infty} \frac{\sigma_\nu(x)\,dx}{\zeta - x}, \tag{35}$$

where

$$\sigma_\nu(x) = (\sqrt{2\pi}\Gamma(1-\nu)|\, D_\nu(-ix)|^2)^{-1} > 0. \tag{36}$$

Proof: $D_\nu(z)$ is an analytic function of z and has no zeros in the half-plane $|\operatorname{Im} z| \leqslant \pi/2$ if $\nu < 1$ (Lemma 1). Hence $f_\nu(\zeta)$ is analytic in the half-plane $\operatorname{Im} \zeta \geqslant 0$ if $\nu < 1$. Using Cauchy's theorem and the fact that $|f_\nu(\zeta)| < c/|\zeta|$ [Eq. (27)] with a contour extending along the $\operatorname{Re} \zeta$ axis and out to ∞ in the half-plane $\operatorname{Im} \zeta \geqslant 0$, one obtains

$$f_\nu(\zeta) = \frac{1}{2\pi i}\int_{-\infty}^{\infty} \frac{f_\nu(x)\,dx}{x - \zeta} \tag{37}$$

and

$$0 = \frac{1}{2\pi i}\int_{-\infty}^{\infty} \frac{f_\nu(x)\,dx}{x - \bar{\zeta}} \tag{38}$$

valid for $\operatorname{Im} \zeta > 0$. From (37) and (38) one has

$$f_\nu(\zeta) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\operatorname{Im} f_\nu(x)\,dx}{x - \zeta}, \quad \operatorname{Im} \zeta > 0. \tag{39}$$

Equations (35) and (36) follow from (39) and (31).

Lemma 5: If $\nu < 1$ and $\operatorname{Im} \zeta \geqslant 0$, then $\operatorname{Im} f_\nu(\zeta) < 0$.

Proof: If $\operatorname{Im} \zeta = 0$, this follows from (31). If $\operatorname{Im} \zeta > 0$, one may use (35) to obtain

$$\operatorname{Im} f_\nu(\zeta) = -\operatorname{Im} \zeta \int_{-\infty}^{\infty} \frac{\sigma_\nu(x)\,dx}{|\zeta - x|^2}$$

which is negative definite from (36).

Theorem 1: If $\nu < 1$ and $\operatorname{Im} \zeta > 0$, then

$$f_\nu(\zeta) = \frac{1}{\zeta} - \frac{1-\nu}{\zeta} - \frac{2-\nu}{\zeta} - \cdots. \tag{40}$$

Proof: The continued fraction in (40) is a real J-fraction

(Ref. 6, p. 114) for which the determinate case holds since $\Sigma (n-\nu)^{-1/2} = \infty$ (Ref. 6, p. 109). Hence the J-fraction converges for $\mathrm{Im}\,\zeta \neq 0$ to a function $f(\zeta)$, which is uniquely determined by the properties that for $\mathrm{Im}\,\zeta > 0$ one has $f(\zeta)$ analytic, $\mathrm{Im}\,f(\zeta) < 0$, and f asymptotically equal to the J-fraction (Ref. 6, p. 320). Thus (40) follows from Lemmas 2, 4, and 5.

One should note that the continued fraction in (40) converges to the Stieltjes transform in (35) also for $\mathrm{Im}\,\zeta < 0$. However, $f_\nu(\zeta)$ is not represented by (35) [or (40)] for $\mathrm{Im}\,\zeta < 0$ but is instead an analytic continuation of the Stieltjes transform (or the J-fraction) into the lower half-plane.

Corollary 1: If λ is real, $\mathrm{Re}\,z > 0$ and z is not a zero of $D_\lambda(z)$, then

$$\frac{D_{\lambda-1}(z)}{D_\lambda(z)} = \frac{1}{z} + \frac{1-\lambda}{z} + \frac{2-\lambda}{z} + \cdots . \tag{41}$$

Proof: For λ real and $\lambda < N$, the tail of the S-fraction in (41) is given by

$$i(N-\lambda-1)\left(\frac{1}{\zeta} - \frac{1-\nu}{\zeta} - \frac{2-\nu}{\zeta} - \cdots\right)$$

with $\zeta = iz$ and $\nu = 1 - N + \lambda < 1$. For $\mathrm{Re}\,z > 0$ this tail converges to $-iD_{\lambda-N}(z)/D_{\lambda-N+1}$ from Theorem 1. Equation (41) follows from (25) since, although (25) was derived under the assumption that z was not a zero of $D_{\lambda-1}(z)$, $D_\lambda(z), \ldots, D_{\lambda-N+1}(z)$, it may be analytically continued to any value of z such that z is not a zero of either $D_\lambda(z)$ or $D_{\lambda-N+1}(z)$. That $D_{\lambda-N+1}(z)$ has no zeros for $\mathrm{Re}\,z > 0$ and $\lambda - N + 1 < 1$ follows from Lemma 1.

The connection between the eigenvalue condition (18) and the actual condition $D_\lambda(-1/\sqrt{\alpha}) = 0$ can now be made by noting that the reciprocal of (41) may be rewritten via an equivalence transformation to state that if $\mathrm{Re}\,z > 0$ and z is not a zero of $D_{\lambda-1}(z)$, then

$$z^{-1}\frac{D_\lambda(z)}{D_{\lambda-1}(z)} = 1 + \frac{(1-\lambda)z^{-2}}{1} + \frac{(2-\lambda)z^{-2}}{1} + \cdots . \tag{42}$$

If in (42) one puts $z = 1/\sqrt{\alpha}$, one obtains

$$\sqrt{\alpha}\,\frac{D_\lambda(1/\sqrt{\alpha})}{D_{\lambda-1}(1/\sqrt{\alpha})} = 1 + \frac{(1-\lambda)\alpha}{1} + \frac{(2-\lambda)\alpha}{1} + \cdots . \tag{43}$$

Since the continued fraction in (43) is identical to $f(\alpha, \lambda, 0)$ of (18'), it follows that (18) for $l = 0$ is the condition

$$\sqrt{\alpha}\,\frac{D_\lambda(1/\sqrt{\alpha})}{D_{\lambda-1}(1/\sqrt{\alpha})} = 0, \tag{44}$$

which, of course, is *not* the correct eigenvalue condition.

To obtain the correct $l = 0$ eigenvalue condition, one must first make an analytic continuation of $f(\alpha, \lambda, 0)$ [i.e., (43)] to $\arg\alpha = \pm 2\pi$ and then set it equal to zero. Since we are dealing with a square root branch point, either continuation will suffice. In the next section we consider how to obtain this analytic continuation by a modification of the tail of the continued fraction.[12]

4. ALGEBRAIC CONVERGENCE AND ANALYTICAL CONTINUATION

A. Convergence via the difference equation

Let f be the C-fraction

$$f = 1 + \frac{(1-\lambda)\alpha}{1} + \frac{(2-\lambda)\alpha}{1} + \cdots , \tag{45}$$

where unless stated otherwise we will consider λ real but noninteger and $\alpha > 0$. Let f_n be the tail of f given by

$$f_n = \frac{(n+1-\lambda)\alpha}{1} + \frac{(n+2-\lambda)\alpha}{1} + \cdots . \tag{46}$$

From (43) one has

$$f = \sqrt{\alpha}D_\lambda(1/\sqrt{\alpha})/D_{\lambda-1}(1/\sqrt{\alpha}) \tag{47}$$

and

$$f_n = \sqrt{\alpha}(n+1-\lambda)\,D_{\lambda-n-2}(1/\sqrt{\alpha})/D_{\lambda-n-1}(1/\sqrt{\alpha}) \tag{48}$$

with

$$f_n = (n+1-\lambda)\,\alpha/(1+f_{n+1}). \tag{49}$$

Consider the nth approximant of f given by

$$S_n = 1 + \frac{(1-\lambda)\alpha}{1} + \frac{(2-\lambda)\alpha}{1} + \cdots + \frac{(n-\lambda)\alpha}{1} \tag{50}$$

and the modified approximant

$$S_n(\omega) = 1 + \frac{(1-\lambda)\alpha}{1} + \frac{(2-\lambda)\alpha}{1} + \cdots + \frac{(n-1-\lambda)\alpha}{1} + \frac{(n-\lambda)\alpha}{1+\omega} . \tag{51}$$

One obviously has

$$S_n(0) = S_n, \quad \lim_{n\to\infty} S_n = f, \quad \text{and} \quad S_n(f_n) = f. \tag{52}$$

Let

$$g = -\sqrt{\alpha}D_\lambda(-1/\sqrt{\alpha})/D_{\lambda-1}(-1/\sqrt{\alpha}). \tag{53}$$

This g is the required analytic continuation of f. If we define

$$g_n = -\sqrt{\alpha}(n+1-\lambda)\,D_{\lambda-n-2}(-1/\sqrt{\alpha}) / D_{\lambda-n-1}(-1/\sqrt{\alpha}), \tag{54}$$

then (25) implies that

$$S_n(g_n) = g \tag{55}$$

and that g_n satisfies the same recursion relation as f_n, namely,

$$g_n = (n+1-\lambda)\,\alpha/(1+g_{n+1}). \tag{56}$$

The recursion relations (49) and (56) may be solved approximately as in (21). The approximate solutions are now given by

$$\omega_n^{\pm} = (-1 \pm \sqrt{1+4(n+1-\lambda)\alpha})/2. \tag{57}$$

Our remarks following (22) will be further justified when we show that for large n

$$g_n = \omega_n^- + O(1/\sqrt{n}) \tag{58}$$

while

$$f_n = \omega_n^+ + O(1/\sqrt{n}). \tag{59}$$

We will also examine the convergence of $S_n(\omega_n^{\pm})$ and find a necessary and sufficient condition on ω_n in order to have

$$\lim_{n\to\infty} S_n(\omega_n) = g. \tag{60}$$

An important estimate for (58), (59), and many later estimates is the following (Ref. 13, 19.9.1 and 19.9.2):

$$D_\nu(\mp 1/\sqrt{\alpha}) = \frac{2^{\nu/2}\sqrt{\pi}}{\Gamma(\frac{1}{2} - \frac{1}{2}\nu)} \exp[\pm\sqrt{-\nu/\alpha} + O(1/\sqrt{-\nu})] \tag{61}$$

provided that $\nu\alpha \ll -1$.

Using (61), (47), and (54), one has

$$f_n = \sqrt{\tfrac{1}{2}\alpha}(n+1-\lambda)\,[\Gamma(1+\tfrac{1}{2}(n-\lambda))/\Gamma(\tfrac{3}{2}+\tfrac{1}{2}(n-\lambda))] \times\exp[-1/2\sqrt{n\alpha} + O(n^{-3/2})] \tag{62}$$

and

$$g_n = -\sqrt{\tfrac{1}{2}\alpha}(n+1-\lambda)[\Gamma(1+\tfrac{1}{2}(n-\lambda))/\Gamma(\tfrac{3}{2}+\tfrac{1}{2}(n-\lambda))]\exp[1/2\sqrt{n\alpha} + O(n^{-3/2})]. \tag{63}$$

From (62), (63), and the Stirling series for Γ, one obtains (58) and (59). One may also verify that $\omega^{\pm}_{n-1/2}$ are actually better approximations since

$$g_n = \omega^{-}_{n-1/2} + O(n^{-3/2}), \tag{64}$$

$$f_n = \omega^{+}_{n-1/2} + O(n^{-3/2}). \tag{65}$$

It is also important to recall from the basic theory of continued fractions (Ref. 6, p. 15) the following formula for the modified approximant (51):

$$S_n(\omega) = (A_n + \omega A_{n-1})/(B_n + \omega B_{n-1}), \tag{66}$$

where A_n and B_n are the numerator and denominator respectively of S_n, the nth approximant of f. They may be obtained from the difference equations

$$A_{n+1} = A_n + (n+1-\lambda)\,\alpha A_{n-1}, \quad A_{-1} = A_0 = 1, \tag{67}$$

$$B_{n+1} = B_n + (n+1-\lambda)\,\alpha B_{n-1}, \quad B_{-1} = 0, \quad B_0 = 1. \tag{68}$$

Before proceeding with our main results, we emphasize the close connection between the difference equation [such as (67) and (68)] and the recursion relation [such as (49) and (56)] associated with a given C-fraction.

Lemma 6: If C_n, C_{n-1}, and $(n+1-\lambda)\,\alpha$ are nonzero, then there is a one-to-one correspondence between solutions to the difference equation

$$C_{n+1} = C_n + (n+1-\lambda)\,\alpha C_{n-1} \tag{69}$$

and the recursion relation

$$h_n = (n+1-\lambda)\,\alpha/(1+h_{n+1}) \tag{70}$$

given by

$$h_n = -C_n/C_{n-1}. \tag{71}$$

Proof: From (69) and (71) one has

$$-\frac{C_n}{C_{n-1}} + \frac{C_{n+1}C_n}{C_{n-1}C_n} = (n+1-\lambda)\,\alpha = h_n(1+h_{n+1}), \tag{72}$$

which on dividing by $(1+h_{n+1})$ yields (70). On the other hand, (72) follows from (70) and (71) and multiplication of (72) by C_{n-1} yields (69).

A remarkable feature of the difference equation (69) is that it may be solved exactly in terms of the parabolic cylinder function. One has:

Theorem 2: The general solution to (69) is given by

$$C_n = aC^d_n + bC^s_n, \tag{73}$$

where

$$C^d_n = 2^n\alpha^{n/2}\Gamma(1+\tfrac{1}{2}(n-\lambda)) \times\Gamma(\tfrac{3}{2}+\tfrac{1}{2}(n-\lambda))\,D_{\lambda-n-2}(-1/\sqrt{\alpha}) \tag{74}$$

and

$$C^s_n = (-1)^n 2^n\,\alpha^{n/2}\Gamma(1+\tfrac{1}{2}(n-\lambda)) \times\Gamma(\tfrac{3}{2}+\tfrac{1}{2}(n-\lambda))\,D_{\lambda-n-2}(1/\sqrt{\alpha}). \tag{75}$$

Proof: From (74), (75), (48), and (54) one has

$$g_n = -C^d_n/C^d_{n-1} \tag{76}$$

and

$$f_n = -C^s_n/C^s_{n-1}. \tag{77}$$

It follows from Lemma 6 that C^d_n and C^s_n satisfy the difference equation (69). They are obviously linearly independent.

One should note that C^d_n is the dominant solution for large n since from (73), (74), and(61) one has

$$C^d_n = \alpha^{n/2}2^{(n+\lambda)/2}\sqrt{\pi}\,\Gamma(1+\tfrac{1}{2}(n-\lambda)) \times\exp[\sqrt{n/\alpha} + O(1/\sqrt{n})] \tag{78}$$

and

$$C^s_n = (-1)^n\,\alpha^{n/2}2^{(n+\lambda)/2}\sqrt{\pi}\,\Gamma(1+\tfrac{1}{2}(n-\lambda)) \times\exp[-\sqrt{n/\alpha} + O(1/\sqrt{n})]. \tag{79}$$

Thus

$$C^d_n/C^s_n = (-1)^n\exp(2\sqrt{n/\alpha})[1 + O(1/\sqrt{n})]. \tag{80}$$

If one considers complex α, then C^d_n remains dominant for $|\arg\alpha| < \pi$. For $\alpha < 0$, C^d_n and C^s_n are competitive while on the second sheet $3\pi > \pm\arg\alpha > \pi$ the dominant and subdominant roles are interchanged.

Since A_n and B_n are also solutions to the difference equation (69), one has

$$A_n = aC^d_n + a'C^s_n \tag{81}$$

with

$$aC^d_{-1} + a'C^s_{-1} = 1, \quad aC^d_0 + a'C^s_0 = 1 \tag{82}$$

from (67) and

$$B_n = bC^d_n + b'C^s_n \tag{83}$$

with

$$bC^d_{-1} + b'C^s_{-1} = 0, \quad bC^d_0 + b'C^s_0 = 1 \tag{84}$$

from (68). From (80), (81), and (83) one has

$$A_n/B_n = (a/b)\{1 + (-1)^n(a'/a - b'/b) \times\exp(-2\sqrt{n/\alpha})[1 + O(1/\sqrt{n})]\}, \tag{85}$$

and one can check from (82), (84), (74), and (75) that

$$a/b = f = \sqrt{\alpha} D_\lambda(1/\sqrt{\alpha})/D_{\lambda-1}(1/\sqrt{\alpha}). \tag{86}$$

Thus (85) and (86) give one an alternative algebraic proof of results of Sec. 3 such as Eqs. (40)–(43) but with the additional feature that one now has an estimate of the rate of convergence of the approximants of the continued fraction.

One can also check that the analytic continuation which we are seeking is just [from (82), (84), (74), and (75)]

$$a'/b' = g = -\sqrt{\alpha} D_\lambda(-1/\sqrt{\alpha})/D_{\lambda-1}(-1/\sqrt{\alpha}) \tag{87}$$

with the a' and b' "hidden" in the subdominant parts of A_n and B_n, respectively. An analytic continuation of A_n and B_n to the second sheet accomplishes nothing, of course, since A_n and B_n are polynomials in α. It is, however, possible to gain access to these subdominant terms by using a suitably modified approximant. With respect to the modified approximants, the following identities are extremely useful:

$$S_n(\omega) - S_n(\omega') = (\omega - \omega') \frac{(A_{n-1}B_n - A_n B_{n-1})}{(B_n + \omega B_{n-1})(B_n + \omega' B_{n-1})} \tag{88}$$

and

$$S_n(\omega) - S_n(\omega'') = \frac{(\omega - \omega'')}{(\omega'' - \omega')} \frac{(B_n + \omega' B_{n-1})}{(B_n + \omega B_{n-1})} [S_n(\omega'') - S_n(\omega')], \tag{89}$$

where (88) follows from (66) and (89) follows from (88).

The principal results concerning the convergence of the modified approximants can now be obtained.

Theorem 3: The modified approximant (51) has the following convergence properties for fixed $\alpha > 0$, λ real but noninteger, and $n \to \infty$:

(i) $S_n(f_n) = f$;

(ii) $S_n(g_n) = g$;

(iii) $S_n(0) = f + O(e^{-2\sqrt{n/\alpha}})$;

(iv) $S_n(\omega_n^+) = f + O(n^{-1} e^{-2\sqrt{n/\alpha}})$;

(v) $S_n(\omega_n^-) = f + O(\sqrt{n}\, e^{-2\sqrt{n/\alpha}})$;

(vi) $S_n(\omega_{n-1/2}^+) = f + O(n^{-2} e^{-2\sqrt{n/\alpha}})$;

(vii) $S_n(\omega_{n-1/2}^-) = f + O(\sqrt{n}\, e^{-2\sqrt{n/\alpha}})$;

(viii) $\lim_{n\to\infty} S_n(\omega_n) = g$ if and only if

$$\omega_n - g_n = k_n \sqrt{n}\, e^{-2\sqrt{n/\alpha}} \tag{90}$$

with

$$\lim_{n\to\infty} k_n = 0. \tag{91}$$

Proof: Properties (i) and (ii) are by definition. Property (iii) has already been discussed [Eqs. (85) and (86)]. To prove (iv), we use (89) to obtain

$$S_n(\omega_n^+) - S_n(f_n) = [(\omega_n^+ - f_n)/f_n] \cdot [B_n/(B_n + \omega_n^+ B_{n-1})]\, [S_n(f_n) - S_n(0)]. \tag{92}$$

If one now uses the fact that

$$B_n/B_{n-1} = -g_n \{1 + (-1)^n (2b'/b) \times e^{-2\sqrt{n/\alpha}} [1 + O(1/\sqrt{n})]\}, \tag{93}$$

which follows from (83), (80), and (76), then (iv) follows from (i), (iii), (92), (93), (57), (58), and (59). The proof of (v), (vi), and (vii) is similar. To prove (viii), we write

$$S_n(\omega_n) - S_n(g_n) = [(\omega_n - g_n) B_n/g_n(B_n + \omega_n B_{n-1})]\, [S_n(g_n) - S_n(0)]. \tag{94}$$

Since $S_n(g_n) = g$ and $S_n(0) \to f$, it suffices to show that

$$\lim_{n\to\infty} \frac{(\omega_n - g_n) B_n}{g_n(B_n + \omega_n B_{n-1})} = 0 \tag{95}$$

if and only if one has (91). From (57) and (58) one may rewrite (93) to state that

$$B_n/B_{n-1} = -g_n + (-1)^n 2(b'/b)\sqrt{n}\, e^{-2\sqrt{n/\alpha}} \times [1 + O(1/\sqrt{n})] \tag{96}$$

so that (95) is equivalent to having

$$\lim_{n\to\infty} \frac{(\omega_n - g_n)}{[\omega_n - g_n + (-1)^n 2(b'/b)\sqrt{n}\, e^{-\sqrt{n/\alpha}}]} = 0. \tag{97}$$

Equation (91) is obviously both necessary and sufficient in order to have (97).

It is clear from Theorem 3 that there is no practical modification which will yield the desired analytic continuation. The use of $\omega_n = \omega_n^-$ merely "slows down" the convergence of $S_n(\omega_n)$ to f while an ω_n which will cause $S_n(\omega_n)$ to converge to g must be an exceptionally good approximation to g_n. The condition (90) on ω_n becomes, however, less stringent for larger values of α, and one can recover something useful by considering the limit $\alpha \to \infty$ in the sense described below.

B. Approximate analytic continuation

One may obtain approximations to the first few terms in the series expansion of g in powers of $\alpha^{-1/2}$ by the following scheme.

Consider the following expansions of g_n, g, ω_n, and $S_n(\omega_n)$ in powers of $\alpha^{-1/2}$:

$$g_n = \sqrt{\alpha}(g_n^{(0)} + g_n^{(1)}/\sqrt{\alpha} + g_n^{(2)}/\alpha + \cdots), \tag{98}$$

$$g = \sqrt{\alpha}(g^{(0)} + g^{(1)}/\sqrt{\alpha} + g^{(2)}/\alpha + \cdots), \tag{99}$$

$$\omega_n = \sqrt{\alpha}(\omega_n^{(0)} + \omega_n^{(1)}/\sqrt{\alpha} + \omega_n^{(2)}/\alpha + \cdots), \tag{100}$$

$$S_n(\omega_n) = \sqrt{\alpha}(S_n^{(0)}(\omega_n) + S_n^{(1)}(\omega_n)/\sqrt{\alpha} + S_n^{(2)}(\omega_n)/\alpha + \cdots). \tag{101}$$

If the coefficients $\omega_n^{(i)}$ are a sufficiently good approximation to the coefficients $g_n^{(i)}$ for large n, then the coefficients in the series expansion of $S_n(\omega_n)$ will yield the coefficients $g^{(i)}$ in the large n limit.

This scheme is of practical interest for our eigenvalue problem for instead of solving $g = 0$ and obtaining solutions

$\lambda = \lambda_m(\alpha)$ with

$$\lambda_m(\alpha) = 2m + 1 + a_m^{(1)}/\sqrt{\alpha} + a_m^{(2)}/\alpha + \cdots. \tag{102}$$

One may solve the approximate equation $S_n(\omega_n) = 0$ and obtain approximate eigenvalues $\lambda = \lambda_{mn}(\alpha)$ with

$$\lambda_{mn}(\alpha) = 2m + 1 + a_{mn}^{(1)}/\sqrt{\alpha} + a_{mn}^{(2)}/\alpha + \cdots, \tag{103}$$

where

$$\lim_{n\to\infty} a_{mn}^{(i)} = a_m^{(i)}. \tag{104}$$

The essence of this scheme is precisely stated and proved below for $i = 0,1,2$ and $m = 0$.

Theorem 4: Let

$$\omega_n^{(i)} = g_n^{(i)} + O(n^{-p_i}), \quad i = 0,1,2. \tag{105}$$

If λ is noninteger, then

$$S_n^{(i)}(\omega_n) = g^{(i)} + O(n^{-q_i}), \quad i = 0,1,2, \tag{106}$$

with

$$\begin{aligned} q_0 &= p_0 + \tfrac{1}{2}, \\ q_1 &= \min(p_0, p_1 + \tfrac{1}{2}), \\ q_2 &= \min(p_0 + \tfrac{1}{2}, p_1, p_2 + \tfrac{1}{2}). \end{aligned} \tag{107}$$

If $\lambda = 1$, then

$$S_n^{(0)} = g^{(0)} = 0, \tag{108}$$

$$S_n^{(1)} = g^{(1)} = 1, \tag{109}$$

$$S_n^{(2)} = g^{(2)} + O(n^{-q_0}), \tag{110}$$

$$\frac{dS_n^{(i)}}{d\lambda} = \frac{dg^{(i)}}{d\lambda} + O(n^{-q_i}), \quad i = 0,1, \tag{111}$$

$$\frac{d^2S_n^{(0)}}{d\lambda^2} = \frac{d^2g^{(0)}}{d\lambda^2} + O(n^{-q_0}). \tag{112}$$

Proof: From (89) one has

$$S_n(\omega_n) - g = \frac{(\omega_n - g_n)(B_n + f_nB_{n-1})(g - f)}{(g_n - f_n)(B_n + \omega_nB_{n-1})}. \tag{113}$$

Expanding this in powers of $\alpha^{-1/2}$ yields for n *even*

$$S_n^{(0)}(\omega_n) - g^{(0)} = [(\omega_n^{(0)} - g_n^{(0)})/2g_n^{(0)}]\, 2g^0, \tag{114}$$

$$\begin{aligned} S_n^{(1)}(\omega_n) - g^{(1)} = {} & [(\omega_n^{(1)} - g_n^{(1)})/g_n^{(0)}]\, g^{(0)} \\ & - (\omega_n^{(0)} - g_n^{(0)})(\omega_n^{(0)} + g_n^{(0)})\, B_{n-1}^{(0)}\, g^{(0)}/g_n^{(0)}B_n^{(0)}, \end{aligned} \tag{115}$$

and

$$\begin{aligned} S_n^{(2)}(\omega_n) - g^{(2)} = {} & [(\omega_n^{(2)} - g_n^{(2)})/g_n^{(0)}]\, g^{(0)} \\ & - (\omega_n^{(1)} - g_n^{(1)})(\omega_n^{(0)} + g_n^{(0)})\, B_{n-1}^{(0)}\, g^{(0)}/g_n^{(0)}B_n^{(0)} \\ & + 2(\omega_n^{(0)} - g_n^{(0)})(g^{(2)}g_n^{(0)} - g_n^{(2)}\, g^{(0)})/(g_n^{(0)})^2 \\ & - [(\omega_n^{(0)} - g_n^{(0)})(g_n^{(1)} + \omega_n^{(1)})\, B_{n-1}^{(0)}/g_n^{(0)}B_n^{(0)}]\, g^{(0)}. \end{aligned} \tag{116}$$

Here $B_n^{(0)}$ is the coefficient of the leading $\alpha^{n/2}$ term in B_n and $B_{n-1}^{(0)}$ is the coefficient of the leading $\alpha^{n/2-1}$ term in B_{n-1}. It follows from (68) that

$$B_n^{(0)} = (2 - \lambda)(4 - \lambda)\cdots(n - \lambda) \tag{117}$$

and

$$\frac{B_{n-1}^{(0)}}{B_n^{(0)}} = 1 - \frac{\Gamma(\frac{1}{2} + \frac{1}{2}(n - \lambda))\,\Gamma(1 - \frac{1}{2}\lambda)}{\Gamma(1 + \frac{1}{2}(n - \lambda))\,\Gamma(\frac{1}{2} - \frac{1}{2}\lambda)}. \tag{118}$$

Thus

$$B_{n-1}^{(0)}/B_n^{(0)} = 1 + O(1/\sqrt{n}). \tag{119}$$

The estimates (106) follow from (105), (115), (116), (117), (120), and the fact that

$$-g_n^{(0)}/\sqrt{n} \to 1, \quad g_n^{(1)} = O(1), \quad \text{and} \quad g_n^{(2)} = O(1/\sqrt{n}).$$

The proof for n odd is similar. When $\lambda = 1$, one has $g^{(0)} = 0$, $g^{(1)} = 1$, and $dg^{(0)}/d\lambda \neq 0$. Equations (109)–(113) then follow as before if one notes that $dg_n^{(0)}/d\lambda = O(1/\sqrt{n})$.

Corollary 2: If $\omega_n^{(i)} = g_n^{(i)} + O(n^{-p_i})$, $i = 0,1$, then $S_n(\omega_n) = 0$ yields a solution

$$\lambda = 1 + a_{0n}^{(1)}/\sqrt{\alpha} + a_{0n}^{(2)}/\alpha + \cdots \tag{120}$$

with

$$a_{0n}^{(1)} = -\sqrt{2/\pi} + O(n^{-r_1}), \tag{121}$$

$$a_{0n}^{(2)} = (2/\pi)(1 - \ln 2) + O(n^{-r_2}), \tag{122}$$

where

$$r_1 = p_0 + \tfrac{1}{2} \quad \text{and} \quad r_2 = \min(p_0 + \tfrac{1}{2}, p_1 + \tfrac{1}{2}). \tag{123}$$

Proof: If one expands $S_n(\omega_n) = 0$ about $\lambda = 1$ and $\sqrt{\alpha} = \infty$ using (108) and (109), one obtains

$$\begin{aligned} & \sqrt{\alpha}\left[(\lambda - 1)\frac{dS_n^{(0)}}{d\lambda} + (\lambda - 1)^2\frac{d^2S_n^{(0)}}{d\lambda^2} + O(\lambda - 1)^3\right] \\ & \quad + 1 + (\lambda - 1)\frac{dS_n^{(1)}}{d\lambda} \\ & \quad + O(\lambda - 1)^2 + \frac{1}{\sqrt{\alpha}}[S_n^{(2)} + O(\lambda - 1)] + \cdots = 0. \end{aligned} \tag{124}$$

This yields

$$\lambda = 1 + a_{0n}^{(1)}/\sqrt{\alpha} + a_{0n}^{(2)}/\alpha + \cdots$$

with

$$a_{0n}^{(1)} = -\left(\frac{dS_n^{(0)}}{d\lambda}\right)^{-1} \tag{125}$$

and

$$\begin{aligned} a_{0n}^{(2)} = {} & -\left(S_n^{(2)} + a_{0n}^{(1)}\frac{dS^{(1)}}{d\lambda} + (a_{0n}^{(1)})^2\frac{d^2S_n^{(0)}}{d\lambda^2}\right) \\ & \times\left(\frac{dS_n^{(0)}}{d\lambda}\right)^{-1}. \end{aligned} \tag{126}$$

The estimates (121) and (122) follow from (110), (111), (112), (125), (126), and (10′).

If one chooses $\omega_n = \omega_n^-$, then one obtains $a_0^{(1)}$ and $a_0^{(2)}$ to $O(n^{-1})$ since one has

$$\omega_n^{-(0)} = -\sqrt{n + 1 - \lambda} = g_n^{(0)} + O(n^{-1/2}) \tag{127}$$

and

$$\omega_n^{-(1)} = -\tfrac{1}{2} = g_n^{(1)} + O(n^{-1}). \tag{128}$$

The choice $\omega_n = \omega_{n-1/2}^-$ yields a better approximation since

$$-\sqrt{n + \tfrac{1}{2} - \lambda} = g_n^{(0)} + O(n^{-3/2}) \tag{129}$$

so that one then obtains $a_0^{(1)}$ to $O(n^{-2})$.

With $\omega_n^{(0)} = -\sqrt{n + \frac{1}{2} - \lambda}$ and $\omega_n^{(1)} = -\frac{1}{2}$ one obtains

from (125) and (126) the explicit formulas

$$a_{0n}^{(1)} = -\frac{1\cdot 3\cdots(n-2)\sqrt{n-\frac{1}{2}}}{2\cdot 4\cdots(n-1)}, \quad n \text{ odd},$$
$$= -\frac{1\cdot 3\cdots(n-3)(n-1)}{2\cdot 4\cdots(n-2)\sqrt{n-\frac{1}{2}}}, \quad n \text{ even}, \tag{130}$$

and

$$a_{0n}^{(2)} = (a_{0n}^{(1)})^2\left(\frac{1}{2} - \frac{1}{3} + \cdots + (-1)^{n-1}\frac{1}{n-1} + \frac{1}{2}(-1)^n\frac{1}{n-\frac{1}{2}}\right). \tag{131}$$

These values agree quite well with the exact values even for relatively small values of n. Some numerical values are displayed in Table I.

The above scheme can be improved upon since it is actually possible to calculate the coefficients in (98) to any desired degree of accuracy and thus obtain (105) with p_i as large as one wishes for $i = 0,1,\ldots,M$ using only the difference equation (69).

In (69) put

$$C_n = (2\alpha)^{n/2}\,\Gamma(1+\tfrac{1}{2}(n-\lambda))\,E_n \tag{132}$$

to get

$$E_{n+1} - E_{n-1} = \frac{1}{\sqrt{2\alpha}}\frac{\Gamma(1+\frac{1}{2}(n-\lambda))}{\Gamma(\frac{3}{2}+\frac{1}{2}(n-\lambda))}E_n. \tag{133}$$

Let

$$E_n = E_n^{(0)} + E_n^{(1)}/\sqrt{\alpha} + E_n^{(2)}/\alpha + \cdots. \tag{134}$$

Then (133) yields the following recursive difference formula for $E_n^{(m)}$:

$$\delta E_n^{(m)} = \frac{1}{\sqrt{2}}\frac{\Gamma(1+\frac{1}{2}(n-\lambda))}{\Gamma(\frac{3}{2}+\frac{1}{2}(n-\lambda))}E_n^{(m-1)}, \tag{135}$$

where δ is the central difference operator defined by

$$\delta y_n = y_{n+1} - y_{n-1}. \tag{136}$$

The difference equation (135) may be solved exactly while for a more general case (say $l \neq 0$) it would only be possible to solve the corresponding equation approximately to order n^{-p} for some desired p. Here, if one chooses $E_n^{(0)} = 2^{\lambda/2}\sqrt{\pi}$, then one generates in (132) the solution C_n^d [see (74)] while the choice $E_n^{(0)} = (-1)^n 2^{\lambda/2}\sqrt{\pi}$ generates the solution $C_n^s(\sqrt{\alpha}) = C_n^d(-\sqrt{\alpha})$.

TABLE I. Approximate eigenvalue coefficients in Eq. (120) using $\omega_n = \omega_{n-1/2}$.

n	$a_{0n}^{(1)}$	$a_{0n}^{(2)}$
3	−0.790 569	0.187 500
4	−0.801 783	0.198 979
5	−0.795 495	0.193 359
6	−0.799 502	0.196 603
7	−0.796 721	0.194 493
∞	−0.797 884	0.195 348

Solving (135) with $E_n^{(0)} = 1$ yields for the first few terms

$$E_n^{(1)} = \sqrt{2}\,\Gamma(\tfrac{3}{2}+\tfrac{1}{2}(n-\lambda))/\Gamma(1+\tfrac{1}{2}(n-\lambda)) + E^{(1)}, \tag{137}$$

$$E_n^{(2)} = \tfrac{1}{2}n + E^{(1)}\sqrt{2}\,\Gamma(\tfrac{3}{2}+\tfrac{1}{2}(n-\lambda))/\Gamma(1+\tfrac{1}{2}(n-\lambda)) + E^{(2)}, \tag{138}$$

$$E_n^{(3)} = \left[\sqrt{2}\,\Gamma(\tfrac{3}{2}+\tfrac{1}{2}(n-\lambda))/\Gamma(1+\tfrac{1}{2}(n-\lambda))\right] \times (n-3+2\lambda+6E^{(2)})/3! + \tfrac{1}{2}nE^{(1)} + E^{(3)}, \text{ etc.} \tag{139}$$

One may choose the constants $E^{(i)} = 0$ or determine them so as to satisfy some specific condition such as $E_0 = h(\alpha)$. One must, of course, avoid the homogeneous solution $(-1)^n E^{(i)}$ in order to keep a "pure" type C_n^d solution.

Knowing the $E_n^{(i)}$ exactly, one would obtain

$$g_n = -\sqrt{2\alpha} \times \frac{\Gamma(1+\frac{1}{2}(n-\lambda))(1+E_n^{(1)}/\sqrt{\alpha}+E_n^{(2)}/\alpha+\cdots)}{\Gamma(\frac{1}{2}+\frac{1}{2}(n-\lambda))(1+E_{n-1}^{(1)}/\sqrt{\alpha}+E_{n-1}^{(2)}/\alpha+\cdots)} \tag{140}$$

from (132) and (76). Having the expression (140), one could then use the property $S_n(g_n) = g$ and (66) to obtain

$$g = (A_n + g_n A_{n-1})/(B_n + g_n B_{n-1}) \tag{141}$$

with the A_n and B_n being polynomials in α calculated from the same difference equation (69) but starting recursively from $n = -1$ with the initial values in (67) and (68).

This method of solving the difference equation (69) both for large α and n on the one hand and small n on the other yields g in (141) as an exact ratio of two infinite series in the variable $1/\sqrt{\alpha}$. If, however, one were to solve the recursive difference equation (135) approximately for $m \leqslant M$, one would obtain an approximate g_n and a natural choice for ω_n so as to get

$$g \approx S_n(\omega_n) = (A_n + \omega_n A_{n-1})/(B_n + \omega_n B_{n-1}). \tag{142}$$

Equation (142) could then be described as giving g in terms of an approximate Padé approximant in the sense that the right-hand side of (142) is $\sqrt{\alpha}$ times a ratio of polynomials in $1/\sqrt{\alpha}$ with coefficients that approximate the exact coefficients in (141). This scheme is described more precisely with the following theorem and its corollary.

Theorem 5: Let λ be noninteger and consider

$$\omega_n = \sqrt{\alpha}\,\omega_n^{(0)}\frac{(1+F_n^{(1)}/\sqrt{\alpha}+F_n^{(2)}/\alpha+\cdots)}{(1+F_{n-1}^{(1)}/\sqrt{\alpha}+F_{n-1}^{(2)}/\alpha+\cdots)}. \tag{143}$$

If

$$\omega_n^{(0)} = g_n^{(0)} + O(n^{-p}), \tag{144}$$

$$F_{n-1}^{(m)} = E_{n-1}^{(m)} + O(n^{-p}), \tag{145}$$

$$F_n^{(m)} = E_n^{(m)} + O(n^{-p}), \tag{146}$$

$$m = 1,2,\ldots,M, \quad M < 2p+1,$$

then

$$S_n(\omega_n) = 1 + \omega_0^{(0)}\sqrt{\alpha}\,\frac{(1+F_0^{(1)}/\sqrt{\alpha}+F_0^{(2)}/\alpha+\cdots)}{(1+F_{-1}^{(1)}/\sqrt{\alpha}+F_{-1}^{(2)}/\alpha+\cdots)} \tag{147}$$

with

$$\omega_0^{(0)} = g_0^{(0)} + O(n^{-p-1/2}), \tag{148}$$

$$F_0^{(m)} = E_0^{(m)} + O(n^{-p-1/2+m/2}), \tag{149}$$

$$F_{-1}^{(m)} = E_{-1}^{(m)} + O(n^{-p-1/2+m/2}), \tag{150}$$

$m = 1,2,...,M$.

Proof: The proof is straightforward but tedious. It depends on the fact that $\omega_0^{(0)}$, $F_0^{(m)}$, and $F_{-1}^{(m)}$ may be obtained from the same recursion relations as $g_0^{(0)}$, $E_0^{(m)}$, and $E_{-1}^{(m)}$, namely,

$$\omega_{r-1}^{(0)} = (r - \lambda)/\omega_r^{(0)} \tag{151}$$

and

$$F_{r-2}^{(m)} = -F_{r-1}^{(m-1)}/\omega_r^{(0)} + F_r^{(m)} \tag{152}$$

plus the fact that

$$-g_n^{(0)}/\sqrt{n} \to 1 \quad \text{and} \quad E_n^{(m)} = O(n^{m/2}). \tag{153}$$

We give only a partial proof for the case n even and $m = 1$.

From (151) one has

$$\omega_r^{(0)} = g_r^{(0)}\,\omega_n^{(0)}/g_n^{(0)}, \quad n \text{ even}, \ r \text{ even}. \tag{154}$$

Equation (148) follows from (153), (154), and (144). From (152) and (153) one has

$$F_0^{(1)} = F_n^{(1)} - \left(\frac{1}{g_2^{(0)}} + \frac{1}{g_4^{(0)}} + \cdots + \frac{1}{g_n^{(0)}}\right)\frac{g_n^{(0)}}{\omega_n^{(0)}}. \tag{155}$$

Thus

$$F_0^{(1)} = F_n^{(1)} + (E_0^{(1)} - E_n^{(1)})\,g_n^{(0)}/\omega_n^{(0)}, \tag{156}$$

and (149) for $m = 1$ follows from (156), (148), (153), and (146).

Corollary 3: With the conditions in Theorem 5 one has

$$g - S_n(\omega_n) = \sqrt{\alpha}(e^{(0)} + e^{(1)}/\sqrt{\alpha} + e^{(2)}/\alpha + \cdots) \tag{157}$$

with

$$e^{(m)} = 0 + O(n^{-p-1/2+m/2}), \quad m = 0,1,...,M. \tag{158}$$

Proof: One has

$$g - S_n(\omega_n) = g_0^{(0)}\sqrt{\alpha} \times \frac{(1 + E_0^{(1)}/\sqrt{\alpha} + E_0^{(2)}/\alpha + \cdots)}{(1 + E_{-1}^{(1)}/\sqrt{\alpha} + E_{-1}^{(2)}/\alpha + \cdots)} - \omega_0^{(0)}\sqrt{\alpha}\,\frac{(1 + F_0^{(1)}/\sqrt{\alpha} + F_0^{(2)}/\alpha + \cdots)}{(1 + F_{-1}^{(1)}/\sqrt{\alpha} + F_{-1}^{(2)}/\alpha + \cdots)}. \tag{159}$$

Hence

$$\begin{aligned} e^{(0)} &= g_0^{(0)} - \omega_0^{(0)}, \\ e^{(1)} &= -\omega_0^{(0)}(F_0^{(1)} + E_{-1}^{(1)}) + g_0^{(0)}(E_0^{(1)} + F_{-1}^{(1)}) \\ &\quad - e^{(0)}(F_{-1}^{(1)} + E_{-1}^{(1)}), \\ e^{(2)} &= g_0^{(0)}(E_0^{(2)} + F_{-1}^{(2)} + E_0^{(1)}F_{-1}^{(1)}) \\ &\quad - \omega_0^{(0)}(F_0^{(2)} + E_{-1}^{(2)} + F_0^{(1)}E_{-1}^{(1)}) \\ &\quad - e^{(0)}(E_{-1}^{(2)} + F_{-1}^{(2)} + E_{-1}^{(1)}F_{-1}^{(1)}) \\ &\quad - e^{(1)}(F_{-1}^{(1)} + E_{-1}^{(1)}), \quad \text{etc.} \end{aligned} \tag{160}$$

The estimates (158) follow from (160) and the estimates (148)–(150).

The results obtained for $l = 0$ generalize to $l \neq 0$. The analysis is then, however, less transparent since one does not have closed form expressions for f, g, f_n, g_n, C_n^d, and C_n^s, and one must rely entirely on estimates of the large-n behavior of the solutions to the $l \neq 0$ difference equation which corresponds to (69). We will here only inform the reader that Theorem 3 remains valid as stated with the appropriate modification of (57). Theorems 4 and 5 and their corollaries also hold with, of course, appropriate changes such as the replacement of λ by $\lambda - l$. Thus one obtains the generalizations of (102) with the large-α expansion of the eigenvalues $\lambda = \lambda_{m,l}(\alpha)$ given by

$$\lambda_{m,l}(\alpha) = 2m + 1 + l + a_m^{(1)}(l)/\sqrt{\alpha} + a_m^{(2)}(l)/\alpha + \cdots. \tag{161}$$

A general theory of analytic continuation which will apply to C-fractions

$$f = b_0 + \frac{a_1\alpha}{b_1} + \frac{a_2\alpha}{b_2} + \cdots,$$

where a_n and b_n are suitable analytic functions of n will be published separately.

We may combine our knowledge of the large-α behavior of the eigenvalues given by (161) with the small-α behavior obtained in Refs. 2–4 to obtain a reasonable picture of the eigenvalues for the complete range of α if we have some idea of the behavior for intermediate values of α. This is the purpose of the next section where we show that $\lambda'_{m,l}(\alpha) > 0$ and provide a method for calculating $\lambda_{m,l}(\alpha)$, in principle, for any $l \geqslant 0$ and $\alpha > 0$.

5. DIFFERENTIAL EQUATION TECHNIQUES

The theory of second-order linear differential equations may be used to discuss the general properties of the solutions to Eq. (1), which is a singular Sturm–Liouville problem. Our analysis is modeled after the techniques which were applied to the problem of complex angular momentum by De Alfaro and Regge in potential scattering.[14]

A. Monotonicity

The eigenvalues $\lambda_{m,l}(\alpha)$ of the rotating harmonic oscillator will be called eigenvalue trajectories to denote their dependence on $\alpha > 0$ for fixed $l \geqslant 0$. These trajectories have the following properties:

(1) They are the zeros of an entire function of λ.

(2) The trajectories are distinct (simple zeros) with no accidental crossings.

(3) Each trajectory is a strictly monotonic increasing function of α.

To show property (1), let G be the solution to (1) which satisfies the boundary condition

$$G(\lambda, \alpha, l, r) = r^{l+1}[1 + O(r)]. \tag{162}$$

Since the coefficient of G in (1) is an entire function of λ and the boundary condition (162) is independent of λ, one has that G is an entire function of λ from Poincaré's theorem (Ref. 14, p. 9). One can check that the large-r behavior of G is given by

$$G \sim A(\lambda, \alpha, l)\,e^{-(r-1)^2/(4\alpha)} r^{\lambda} + B(\lambda, \alpha, l)\,e^{(r-1)^2/(4\alpha)} r^{-\lambda-1}. \tag{163}$$

It follows that $B(\lambda, \alpha, l)$ is an entire function of λ and that the eigenvalues are the zeros of B. It also follows that $A(\lambda, \alpha, l) = B(-\lambda - 1, -\alpha, l)$ since $G(-\lambda - 1, -\alpha, l, r) = G(\lambda, \alpha, l, r)$.

To show property (2), suppose that $\lambda_{m,l}(\alpha_0) = \lambda_{m',l}(\alpha_0)$. Then B has a multiple zero for $\alpha = \alpha_0$. Let $H_1(r) = G(\lambda_{m,l}(\alpha_0), \alpha_0, l, r)$ and $H_2(r) = (\partial G/\partial\lambda)(\lambda_{m,l}(\alpha_0), \alpha_0, l, r)$. One has

$$\lim_{r\to\infty} H_i(r) = 0, \tag{164}$$

$$H_i(0) = 0, \quad i = 1,2, \tag{164'}$$

$$\frac{d^2H_1}{dr^2} + \left[\frac{\lambda_{m,l}(\alpha_0) + \frac{1}{2}}{\alpha_0} - \frac{(r-1)^2}{4\alpha_0^2} - \frac{l(l+1)}{r^2}\right] H_1 = 0, \tag{165}$$

and

$$\frac{d^2H_2}{dr^2} + \left[\frac{\lambda_{m,l}(\alpha_0) + \frac{1}{2}}{\alpha_0} - \frac{(r-1)^2}{4\alpha_0^2} - \frac{l(l+1)}{r^2}\right] H_2 + H_1\alpha_0^{-1} = 0. \tag{166}$$

From (164'), (165), and (166) one obtains the Wronskian identity

$$W(H_2, H_1)(r) = \alpha_0^{-1}\int_0^r H_1^2(t)\,dt, \tag{167}$$

which contradicts (164).

To show property (3), we consider $H_3(z) = G(\lambda_{m,l}(\alpha), \alpha, l, r)$ and $H_4(z) = (\partial G/\partial\alpha)(\lambda_{m,l}(\alpha), \alpha, l, r)$ with $z = (r-1)/\sqrt{\alpha}$. One has

$$\frac{d^2H_3}{dz^2} + \left[\lambda_{m,l}(\alpha) + \frac{1}{2} - \frac{z^2}{4} - \frac{l(l+1)\alpha}{(1+\sqrt{\alpha}z)^2}\right] H_3 = 0 \tag{168}$$

and

$$\frac{d^2H_4}{dz^2} + \left[\lambda_{m,l}(\alpha) + \frac{1}{2} - \frac{z^2}{4} - \frac{l(l+1)\alpha}{(1+\sqrt{\alpha}z)^2}\right] H_4 + \left[\frac{d}{d\alpha}\lambda_{m,l}(\alpha) - \frac{l(l+1)}{(1+\sqrt{\alpha}z)^3}\right] H_3 = 0. \tag{169}$$

Hence

$$\frac{d}{dz}W(H_4, H_3) = \left[\frac{d}{d\alpha}\lambda_{m,l}(\alpha) - \frac{l(l+1)}{(1+\sqrt{\alpha}z)^3}\right] H_3^2. \tag{170}$$

Integrating (170) from $z = -1/\sqrt{\alpha}$ to ∞ and using the fact that, for $l > 0$,

$$H_i(-1/\sqrt{\alpha}) = 0 \tag{171}$$

$$\lim_{z\to\infty} H_i(z) = 0, \quad i = 3,4, \tag{172}$$

yields

$$\frac{d}{d\alpha}\lambda_{m,l}(\alpha) = l(l+1)\frac{\int_0^\infty G^2(\lambda_{m,l}(\alpha), \alpha, l, r)\, r^{-3}\,dr}{\int_0^\infty G^2(\lambda_{m,l}(\alpha), \alpha, l, r)\,dr} > 0. \tag{173}$$

for $l > 0$.

For $l = 0$ one may either take the limit of (173) as $l \to 0$ or again integrate (170) using the fact that for $l = 0$ one must replace (171) by the boundary conditions

$$H_3(-1/\sqrt{\alpha}) = 0, \quad H_4(-1/\sqrt{\alpha}) = -1/(2\alpha). \tag{174}$$

Either method of calculation yields

$$\frac{d}{d\alpha}\lambda_{m,0}(\alpha) = \left(2\int_0^\infty G^2(\lambda_{m,0}(\alpha), \alpha, 0, r)\,dr\right)^{-1} > 0. \tag{175}$$

B. A perturbation method

One may obtain solutions to (1) in terms of the solutions to the "undisplaced" rotating harmonic oscillator equation

$$\frac{d^2y}{dr^2} + \left[\frac{\lambda + \frac{1}{2}}{\alpha} - \frac{r^2}{4\alpha^2} - \frac{l(l+1)}{r^2}\right] y = 0 \tag{176}$$

by means of a Volterra integral equation.

Equation (176) may be solved exactly in terms of the confluent hypergeometric function ${}_1F_1$, where the general hypergeometric function is given by (Ref. 15, p. 1)

$${}_rF_s[(a); (b); t] = \sum_{n=0}^{\infty} \frac{(a_1)_n \cdots (a_r)_n t^n}{(b_1)_n \cdots (b_s)_n n!} \tag{177}$$

with $(a_0) = 1$ and $(a)_n = a(a+1)\cdots(a+n-1)$. Thus one may verify that two linearly independent solutions to (176) are given by

$$y_1(r) = r^{-l} e^{-r^2/4\alpha}\, {}_1F_1[-\tfrac{1}{2}(l+\lambda); \tfrac{1}{2} - l; r^2/(2\alpha)] \tag{178}$$

and

$$y_2(r) = r^{l+1} e^{-r^2/4\alpha}\, {}_1F_1[\tfrac{1}{2} + \tfrac{1}{2}(l-\lambda); \tfrac{3}{2} + l; r^2/(2\alpha)]. \tag{179}$$

Let G be the general solution to (1) which we rewrite as

$$\frac{d^2G}{dr^2} + \left[\frac{\lambda + \frac{1}{2}}{\alpha} - \frac{r^2}{4\alpha^2} - \frac{l(l+1)}{r^2}\right] G = -\frac{(2r-1)}{4\alpha^2} G. \tag{180}$$

It follows from (180) and the general theory of second-order linear differential equations that one has

$$G(r) = c_1 y_1(r) + c_2 y_2(r) - \int_a^r k(r, x)\frac{(2x-1)}{4\alpha^2} G(x)\,dx \tag{181}$$

with

$$k(r, x) = [y_2(r) y_1(x) - y_2(x) y_1(r)]/(2l+1) \tag{182}$$

where

$$W(y_1, y_2) = 2l + 1.$$

To obtain the boundary condition

$$G(r) = r^{l+1}[1 + O(r)],$$

one must choose $c_1 = 0$, $c_2 = 1$, and $a = 0$ so that

$$G(r) = y_2(r) - \int_0^r k(r, x)\frac{(2x-1)}{4\alpha^2} G(x)\,dx. \tag{183}$$

This latter equation may be solved by iteration with

$$G^{(0)}(r) = y_2(r)$$

and

$$G^{(n)}(r) = -\int_0^r k(r,x)\frac{(2x-1)}{4\alpha^2} \times G^{(n-1)}(x)\,dx, \quad n = 1,2,\cdots, \tag{184}$$

to obtain

$$G(r) = \sum_{n=0}^{\infty} G^{(n)}(r). \tag{185}$$

The method of solution may be justified by obtaining estimates which are sufficient to show that (185) is absolutely and uniformly convergent.

To obtain these estimates, it is necessary to know also the large-r behavior of $y_1(r)$ and $y_2(r)$. This may be obtained from (178), (179) and the asymptotic formula (Ref. 15, p. 60)

$$\begin{aligned} {}_1F_1[a;b;t] \sim {} & [\Gamma(b)/\Gamma(a)]\, e^t t^{a-b}\, {}_2F_0[b-a,1-a;1/t] \\ & + [\Gamma(b)/\Gamma(b-a)]\, e^{-i\pi a} t^{-a} \\ & \times {}_2F_0[a,1+a-b;-1/t]. \end{aligned} \tag{186}$$

It follows from (186) and (177)–(179) that one has (for $\lambda + 1 > 0$ and $l \neq \frac{1}{2}, \pm\frac{3}{2}, \pm\frac{5}{2}, \cdots$)

$$|y_1(r)| < \text{const}\,[r/(1+r)]^{-l} e^{r^2/4\alpha} \tag{187}$$

and

$$|y_2(r)| < \text{const}\,[r/(1+r)]^{l+1} e^{r^2/4\alpha}. \tag{188}$$

From (182), (187), and (188) one obtains

$$\begin{aligned} |k(r,x)| < {} & \text{const}\,[r/(1+r)]^{l+1} \\ & \times [x/(1+x)]^{-l} e^{(x^2+r^2)/(4\alpha)} \end{aligned} \tag{189}$$

for $x \leqslant r$, $l \geqslant 0$, $\lambda + 1 > 0$, and $l \neq \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \cdots$. Starting with the estimate (188) and using (184) and (189), one obtains by induction

$$\begin{aligned} |G^{(n)}(r)| < {} & (\text{const})^n [r/(1+r)]^{l+1} \\ & \times (2r+1)^n r^{2n} e^{(2n+1)r^2/(4\alpha)}/n!. \end{aligned} \tag{190}$$

Although this is an exceedingly crude estimate, it is clearly a sufficient justification for (185).

To solve the rotating oscillator eigenvalue problem and obtain the eigenfunction F and the corresponding eigenvalue λ, one must also consider the boundary condition $G \to 0$ as $r \to \infty$. In this connection it is important to note that there is a linear combination of y_1 and y_2 which also has a "good" large-r behavior.

It follows from (186) that

$$\begin{aligned} U(a;b;t) = {} & [\Gamma(1-b)/\Gamma(1+a-b)]\, {}_1F_1[a;b;t] \\ & + [\Gamma(b-1)/\Gamma(a)]\, t^{1-b}\, {}_1F_1[1+a-b;2-b;t] \end{aligned} \tag{191}$$

has the large-t behavior (Ref. 15, p. 60)

$$U(a;b;t) \sim t^{-a}\, {}_2F_0[a,1+a-b;-1/t]. \tag{192}$$

It follows from (178), (179), (191), and (192) that

$$\begin{aligned} y_3(r) &= \frac{\Gamma(\frac{1}{2}+l)}{\Gamma(\frac{1}{2}+\frac{1}{2}(l-\lambda))} y_1(r) + \frac{\Gamma(-\frac{1}{2}-l)\, y_2(r)}{\Gamma(-\frac{1}{2}(\lambda+l))(2\alpha)^{l+1/2}} \\ &= r^{-l} e^{-r^2/4\alpha} U(-\tfrac{1}{2}(l+\lambda); \tfrac{1}{2}-l; r^2/2\alpha) \end{aligned} \tag{193}$$

has the large-r behavior given by

$$y_3(r) = \frac{e^{-r^2/4\alpha} r^{\lambda}}{(2\alpha)^{(l+\lambda)/2}} [1 + O(r^{-2})]. \tag{194}$$

From (182) and (183) one has

$$\begin{aligned} G(r) = {} & y_2(r)\left(1 - \int_0^r \frac{y_1(x)(2x-1)\,G(x)\,dx}{(2l+1)(4\alpha^2)}\right) \\ & + y_1(r) \int_0^r \frac{y_2(x)(2x-1)\,G(x)}{(2l+1)(4\alpha^2)}\,dx. \end{aligned} \tag{195}$$

Hence, in order to have $G \to 0$ as $r \to \infty$, it would appear natural in lieu of (193), (194), and (195) to impose the condition

$$\begin{aligned} & \lim_{r\to\infty}\left(\int_0^r \frac{y_2(x)(2x-1)\,G(x)\,dx}{(2l+1)(4\alpha^2)}\right) \\ & \quad \times\left(1 - \int_0^r \frac{y_1(x)(2x-1)\,G(x)\,dx}{(2l+1)(4\alpha^2)}\right)^{-1} \\ & \quad = (2\alpha)^{l+1/2} \frac{\Gamma(\frac{1}{2}+l)\,\Gamma(-\frac{1}{2}(\lambda+l))}{\Gamma(\frac{1}{2}+\frac{1}{2}(l-\lambda))\,\Gamma(-\frac{1}{2}-l)} \end{aligned} \tag{196}$$

and use (193) to rewrite this as the condition

$$\frac{\Gamma(\frac{1}{2}+l)}{\Gamma(\frac{1}{2}+\frac{1}{2}(l-\lambda))} - \int_0^{\infty} \frac{y_3(x)(2x-1)\,G(x)\,dx}{(2l+1)(4\alpha^2)} = 0. \tag{197}$$

We are unable to justify (197) because of the long range nature of the perturbation term $(2r-1)/(4\alpha^2)$. Thus the integral in (197) diverges for any approximate G_N given by

$$G_N = \sum_{n=0}^{N} G^{(n)}. \tag{198}$$

Also the asymptotic behavior of G_N cannot correctly match the exact asymptotic behavior given by (163). Thus, although (185) gives one a convergent expression for G for $r < \infty$, it cannot be used in the limit $r \to \infty$.

One may circumvent this difficulty by noting that if one defines

$$Y_i(r) = y_i(r-1), \quad i = 1,2,3, \tag{199}$$

then one has from (176)

$$\frac{d^2 Y_i}{dr^2} + \left[\frac{\lambda+\frac{1}{2}}{\alpha} - \frac{(r-1)^2}{4\alpha^2} - \frac{l(l+1)}{(r-1)^2}\right] Y_i = 0. \tag{200}$$

Thus the general solution to (180) may also be written as (provided that $r, a > 1$)

$$\begin{aligned} G(r) = {} & c_1 Y_3(r) + c_2 Y_2(r) \\ & - \int_a^r K(r,x)\, l(l+1)\left(\frac{1}{(x-1)^2} - \frac{1}{x^2}\right) G(x)\,dx \end{aligned} \tag{201}$$

with

$$K(r,x) = [Y_3(r)\, Y_2(x) - Y_3(x)\, Y_2(r)]/W(Y_2,Y_3) \tag{202}$$

where

$$W(Y_2,Y_3) = -(2l+1)\,\Gamma(\tfrac{1}{2}+l)/\Gamma(\tfrac{1}{2}+\tfrac{1}{2}(l-\lambda)). \tag{203}$$

If we denote by J the solution which satisfies the boundary condition $J(r) \to 0$ as $r \to \infty$, then one has from (201), (199), and (194)

$$J(r) = Y_3(r) + \int_r^\infty K(r,x)\, l(l+1) \times \left(\frac{1}{(x-1)^2} - \frac{1}{x^2}\right) J(x)\, dx. \tag{204}$$

This may also be solved by iteration, yielding

$$J(r) = \sum_{n=0}^{\infty} J^{(n)}(r) \tag{205}$$

with

$$J^{(0)}(r) = Y_3(r) \tag{206}$$

and

$$J^{(n)}(r) = \int_r^\infty K(r,x)\, l(l+1) \left(\frac{1}{(x-1)^2} - \frac{1}{x^2}\right) \times J^{(n-1)}(x)\, dx, \quad n = 1,2,\cdots. \tag{207}$$

The fact that (205)–(207) yield a solution to (204) for $r > 1$ follows from the estimates

$$|Y_3(r)| < \text{const}\ [r/(r-1)]^l\, e^{-(r-1)^2/(4\alpha)}\, r^\lambda \tag{208}$$

and

$$|Y_2(r)| < \text{const}\ [(r-1)/r]^{l+1}\, e^{(r-1)^2/(4\alpha)}\, r^{-\lambda-1}, \tag{209}$$

which yield for $x \geqslant r$ and $\lambda \neq 2m + 1 + l$, $m = 0,1,2,\cdots$,

$$|K(r,x)| < \text{const}\ [r/(r-1)]^l\, [(x-1)/x]^{l+1} \times e^{[(x-1)^2 - (r-1)^2]/4\alpha}\, r^\lambda x^{-\lambda-1}. \tag{210}$$

With these estimates one obtains from (206) and (207) the estimate

$$|J^{(n)}(r)| < (\text{const})^n\, [r/(r-1)]^{l+n} \times e^{-(r-1)^2/(4\alpha)}\, r^{\lambda-3n}/n!\,. \tag{211}$$

One may obtain the eigenvalues λ by requiring that

$$W(J, G)(r_0) = 0 \tag{212}$$

with $r_0 > 1$ or equivalently by requiring that

$$J(r_1)\, G(r_0) = J(r_0)\, G(r_1) \tag{213}$$

with $r_1 \neq r_0$ and $r_0, r_1 > 1$.

Equation (213) is an exact eigenvalue condition which may be replaced by the approximate condition

$$J_N(r_1)\, G_M(r_0) = J_N(r_0)\, G_M(r_1) \tag{214}$$

with

$$J_N = \sum_{n=0}^{N} J^{(n)} \quad \text{and} \quad G_M = \sum_{n=0}^{M} G^{(n)}.$$

Although this method of solving the rotating oscillator eigenvalue problem may not be of practical interest (compared, for example, with standard variational methods), it does show how the problem may be solved in principle for unrestricted values of $\alpha > 0$ and $l \geqslant 0$ using the solutions to (176). In Paper II,[16] we again use the exact solutions to (176) as the basis for a Rayleigh–Schrödinger perturbation expansion in $1/\sqrt{\alpha}$ to obtain

$$\lambda_{m,l}(\alpha) = 2m + 1 + l + \sum_{n=1}^{\infty} \frac{a_m^{(n)}(l)}{(\sqrt{\alpha})^n} \tag{215}$$

together with an esimate for the radius of convergence of this expansion.

6. COMMENTS AND CONCLUSIONS

The rotating oscillator eigenvalue problem appears to have had a checkered history. The problem arose as a model for the rotational–vibrational energy spectrum for diatomic molecules (Ref. 17, p. 798). For this model the typical values of the parameter were exceedingly small ($\alpha \leqslant 0.01$) and interest was focused on an expansion for small α. Such an expansion based on the WKBJ or phase integral method was obtained by Dunham[3] in 1932 and duplicated by Rosenthal and Motz[4] in 1937 by a different approach based on a differential equation ansatz. At that time it was realized that the method of Wilson,[10] which relied on the continued fraction obtained from the three-term recursion relation (15), did not work for this problem.

In 1949 Langer,[18] critical of the level of rigor of previous work, proved that $\lambda_{m,l}(\alpha) = m + O(\alpha \ln \alpha)$. In 1978 Fröman and Fröman[2] (sceptical of the Langer result because of the nonanalytic nature of $\ln \alpha$ in the estimate) reexamined the problem and showed rigorously by the phase integral method that $\lambda_{m,l} = m + l(l+1)\,\alpha + 3l(l+1)\,\alpha^2(1+2m) + O(\alpha^3)$ in unknown agreement with the earlier nonrigorous work of Dunham, Rosenthal, and Motz.

In 1979, Flessas[19] claimed that the eigenvalues were, in fact, integers independent of α. This prompted Fröman *et al.*,[5] to make numerical calculations which clearly exhibited the α dependence of λ and confirmed their own previous results.

In 1982, a full circle was twice completed when Singh *et al.*,[1] having unknowingly rediscovered the Wilson method, drew false conclusions from it.

The present paper has concentrated on understanding the limitations of the continued fraction approach and attempting to overcome these limitations. We have pointed out the necessity of analytic continuation and described how this may be accomplished in a limited sense.

A reasonably clear picture of the general nature of the eigenvalues has emerged. Figure 1 displays portions of the eigenvalue trajectories $\lambda_{m,l}(\alpha)$ for $l = 0$ and $0 \leqslant m \leqslant 4$. Both the small α and large-α expansions [Eqs. (10′) and (11)] together with intermediate points calculated from tables of the parabolic cylinder function[13] have been used.

The behavior for $l > 0$ is similar and is depicted in Fig. 2. A minor difference occurs in the extremely-small-α region where the trajectories now rise linearly in accordance with the results of Fröman *et al.* and earlier work.

In Fig. 2 the region $\alpha \ll [l(l+1)(m+\frac{1}{2})]^{-1}$ is the large coupling region where the asymptotic expansion for small α is valid. The region $\alpha \geqslant l + 2m + \frac{5}{2}$ is the small coupling region where the large-α expansion (161) can be shown to converge.[16] We have assumed that nothing spectacular occurs in

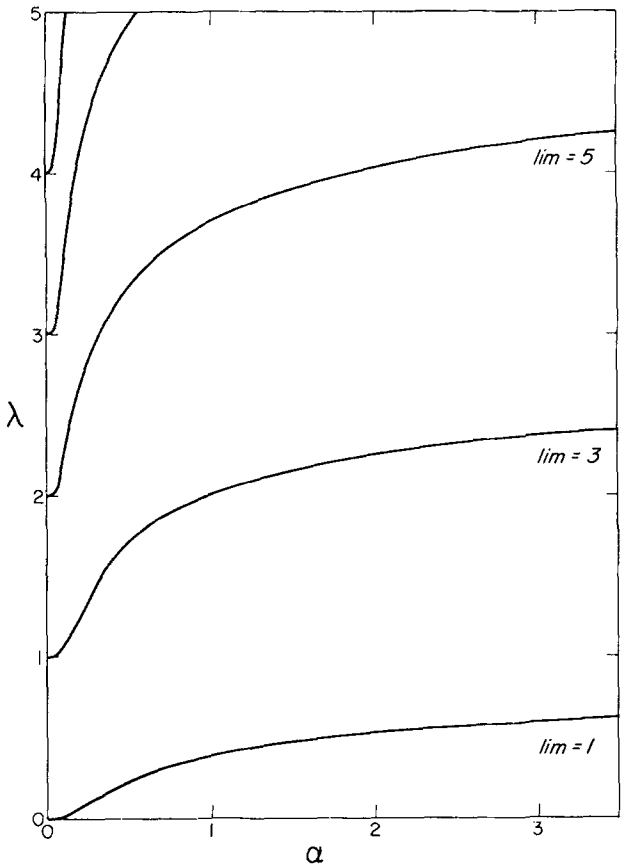


FIG. 1. Portions of the first five eigenvalue trajectories $\lambda_m(\alpha)$ for $l = 0$.

the intermediate region in analogy with the $l = 0$ calculations.

The present investigation has also revealed several facts which seem to be worthy of emphasis:

(a) Equation (34), which may be rewritten as

$$\frac{D_{\lambda-1}(z)}{D_\lambda(z)} = 2z\int_0^\infty \frac{\sigma_\lambda(x)\,dx}{x^2+z^2}, \quad \lambda < 1, \ \mathrm{Re}\, z > 0 \tag{216}$$

with

$$\sigma_\lambda(x) = \left[\sqrt{2\pi}\,\Gamma(1-\lambda)\,|\,D_\lambda(-ix)|^2\right]^{-1}$$

appears to be new. It may be further generalized by considering an analytic continuation in λ. The special case $\lambda = 0$ is a standard representation of the error function (Ref. 8, p. 137)

(b) Related to (a) is the continued fraction representation given in Eq. (41). Namely,

$$\frac{D_{\lambda-1}(z)}{D_\lambda(z)} = \frac{1}{z} + \frac{1-\lambda}{z} + \frac{2-\lambda}{z} + \cdots,$$
$$\mathrm{Im}\,\lambda = 0, \ \mathrm{Re}\, z > 0,$$

which we have found in the literature only for the special case $\lambda = 0$ (Ref. 6, p. 358).

(c)Related to (a) and (b) is the difference equation (69) whose general solution is given in Theorem 2 in terms of parabolic cylinder functions. This "model" difference equation and its exact solution provide an important insight into the general problem of analytic continuation for C-fractions since the results of Sec. 4 can be generalized to a large class of C-fractions.

(d) The exact solution of the "undisplaced" oscillator

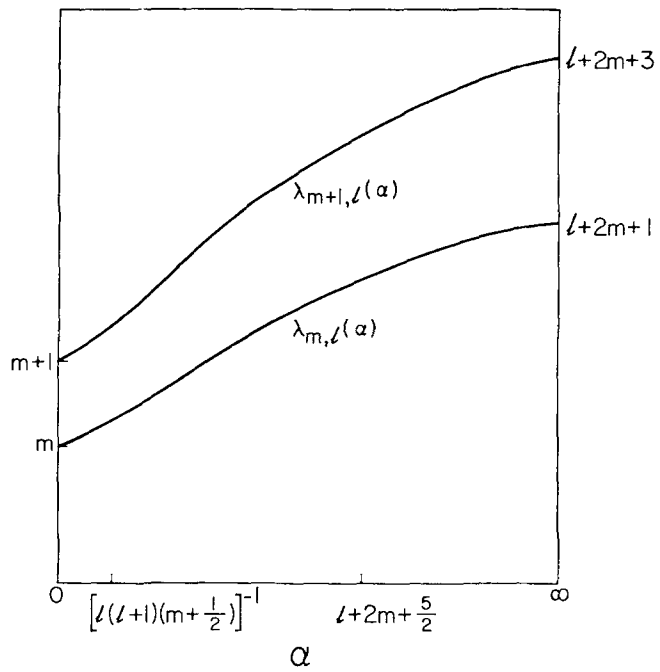


FIG. 2. Schematic picture of the eigenvalue trajectories $\lambda_{m,l}(\alpha)$ for $l > 0$.

equation

$$\frac{d^2y}{dr^2} + \left[\frac{\lambda+\frac{1}{2}}{\alpha} - \frac{r^2}{4\alpha^2} - \frac{l(l+1)}{r^2}\right]y = 0$$

in terms of confluent hypergeometric functions appears to have been overlooked in the literature. Its eigenvalue solution on the interval $[0,\infty)$ yields integer eigenvalues

$$\lambda = 2m + 1 + l, \quad m = 0,1,2,\cdots \tag{217}$$

and corresponding eigenfunctions

$$y = c_m\, r^{l+1} e^{-r^2/4\alpha}\, {}_1F_1\left[-m; \tfrac{3}{2}+l; r^2/2\alpha\right], \tag{218}$$

which may also be written in terms of generalized Laguerre polynomials (Ref. 13, p. 509). The solutions to the "undisplaced" oscillator problem are important for obtaining solutions to the "displaced" oscillator problem [see Eqs. (183) and (204)]. In Paper II[16] they form the basis for a Rayleigh–Schrödinger perturbation expansion which yields a convergent expansion of $\lambda_{m,l}(\alpha)$ in powers of $1/\sqrt{\alpha}$, thus generalizing (10), supporting our claim (161) and justifying our remarks concerning what is the weak coupling region in Fig. 2.

[1] V. Singh, S. N. Biswas, and K. Datta, J. Math. Phys. **23**, 1323 (1982).
[2] N. Fröman and P. O. Fröman, Ann. Phys. **115**, 269 (1978).
[3] J. L. Dunham, Phys. Rev. **41**, 721 (1932).
[4] J. E. Rosenthal and L. Motz, Proc. Natl. Acad. Sci. **23**, 259 (1937).
[5] N. Fröman, P. O. Fröman, and F. Karlsson, Phys. Lett. A **77**, 397 (1980).
[6] H. S. Wall, *Analytic Theory of Continued Fractions* (Van Nostrand, New York, 1948).
[7] I. Ye. Kireyeva and K. A. Karpov, *Tables of Weber Functions* (Pergamon, New York, 1961), Vol. 1.
[8] A. Erdelyi, Ed., *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. II.
[9] A. Milne, Proc. Edinburgh Math. Soc. **33**, 48 (1915).
[10] A. H. Wilson, Proc. Roy. Soc. London Ser. A **118**, 617 (1928).
[11] V. Singh, S. N. Biwas, and K. Datta, Phys. Rev. D **18**, 1901 (1978).
[12] W. B. Jones, W. J. Thron, and H. Waadeland, Eds., *Analytic Theory of Continued Fractions* (Springer-Verlag, Berlin, 1982), p. 38.

[13]M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions* (National Bureau of Standards, Washington, DC, 1966).
[14]V. De Alfaro and T. Regge, *Potential Scattering* (North-Holland Amsterdam, 1965).
[15]L. J. Slater, *Confluent Hypergeometric Functions* (Cambridge U.P., Cambridge, 1960).
[16]D. Masson, J. Math. Phys. **24**, 2089 (1983).
[17]A. Messiah, *Quantum Mechanics* (North-Holland, Amsterdam, 1962), Vol. II.
[18]R. E. Langer, Phys. Rev. **75**, 792 (1949).
[19]G. P. Flessas, Phys. Lett. A **71**, 315 (1979).

The rotating harmonic oscillator eigenvalue problem. II. Analytic perturbation theory[a)]

D. Masson
Department of Mathematics, University of Toronto, Toronto M5S 1A1, Canada

(Received 12 January 1983; accepted for publication 11 March 1983)

The theory of self-adjoint analytic families is applied to the rotating harmonic oscillator Hamiltonian in $L^2(0,\infty)$ to obtain weak and strong coupling expansions of the eigenvalues. Various estimates on the radius of convergence of the weak coupling expansion are obtained. The strong coupling expansion is shown to be an asymptotic series which, with the neglect of exponentially small terms, is expressible in terms of a simple formal perturbation of the ordinary harmonic oscillator Hamiltonian in $L^2(-\infty,\infty)$.

PACS numbers: 02.60.Lj, 02.30.Tb, 02.30.Lt, 02.30.Mv

1. INTRODUCTION

There has been a recent renewal of interest in the rotating harmonic oscillator eigenvalue problem[1-5] given by

$$\frac{d^2\phi}{dr^2}+\left[\frac{\lambda+\frac{1}{2}}{\alpha}-\frac{(r-1)^2}{4\alpha^2}-\frac{l(l+1)}{r^2}\right]\phi=0 \quad (1)$$

with boundary conditions $\phi=0$ at $r=0$ and $\phi\to 0$ as $r\to\infty$. It is assumed that $l\geqslant 0$.

Although this appears to be a simple and standard enough problem with a history dating back to the 1920's,[6-8] it remains not well understood. Recent claims made concerning the properties of the eigenvalues λ are certainly incorrect.[2,4] We examine the problem here from the point of view of analytic perturbation theory.[9,10]

In Sec. 2 we discuss the large-α expansion of λ and obtain various estimates for the radius of convergence of this expansion using the standard theory of analytic perturbations together with some modifications which yield improved estimates.

In Sec. 3 we obtain an asymptotic expansion of λ for small α which agrees with previous expansions.[5,7,8] However, with the present expansion we are able to discern the presence of nonanalytic terms which are exponentially small. The neglect of these exponentially small terms is shown to yield a simple, equivalent, but formal perturbation problem which allows one to easily calculate higher order terms in the expansion.

2. THE LARGE-α EXPANSION

Let $\alpha>0$ and put $x=r/\sqrt{\alpha}$. Then (1) may be written as

$$H\phi=E\phi, \quad (2)$$

where

$$H=-\frac{d^2}{dx^2}+\frac{x^2}{4}+\frac{l(l+1)}{x^2}-\frac{x}{2\sqrt{\alpha}} \quad (3)$$

and

$$E=\lambda+\tfrac{1}{2}-1/(4\alpha) \quad (4)$$

with H a self-adjoint operator on the Hilbert space $L^2(0,\infty)$, which is essentially self-adjoint on the core $C_0^\infty(0,\infty)$, the set of infinitely differentiable functions with compact support in $(0,\infty)$ (Ref. 9, p. 346).

H has a discrete nondegenerate spectrum of positive eigenvalues $\{E_m\}_{m=0}^\infty$, which we order so that $E_{m+1}>E_m$. It is known that for large α one has[1]

$$E_m=l+2m+\tfrac{3}{2}+O(1/\sqrt{\alpha}). \quad (5)$$

To obtain the large-α expansion of E_m, we consider H as a perturbation of H_0 with

$$H_0=-\frac{d^2}{dx^2}+\frac{x^2}{4}+\frac{l(l+1)}{x^2}. \quad (6)$$

Thus one may write $H=H(\beta)$ with

$$H(\beta)=H_0+\beta V, \quad V=x, \quad \text{and} \quad \beta=-1/(2\sqrt{\alpha}). \quad (7)$$

This decomposition is useful because the eigenvalue problem for the unperturbed self-adjoint operator H_0 is exactly solvable.[1] It has eigenvalues

$$E_m^{(0)}=l+2m+\tfrac{3}{2}, \quad m=0,1,2,\cdots, \quad (8)$$

and a complete set of corresponding eigenfunctions

$$\phi_m^{(0)}=c_m e^{-x^2/4}x^{l+1}{}_1F_1\left[-m;l+\tfrac{3}{2};x^2/2\right], \quad (9)$$

where ${}_1F_1$ is the confluent hypergeometric function. Also the operators $H(\beta)$ form a self-adjoint analytic family of type (A) having a constant domain $\mathscr{D}$ with $H(\beta)u$ analytic for $u\in\mathscr{D}$. More particularly, one has:

Lemma 1: Let H_0 be given by (6) with $V=x$. Then V is H_0-bounded with

$$\|Vu\|\leqslant(2/\sqrt{l+\tfrac{3}{2}})\|H_0u\|, \quad u\in\mathscr{D}(H_0)\subset\mathscr{D}(V). \quad (10)$$

Proof: One has $H_0\geqslant l+\frac{3}{2}$ from (8) and $H_0\geqslant x^2/4$ from (6) so that $H_0^2\geqslant(l+\frac{3}{2})H_0\geqslant(l+\frac{3}{2})x^2/4$.

Corollary 1: $H(\beta)$ is a self-adjoint analytic family of type (A) for $|\beta|<2/\sqrt{l+\frac{3}{2}}$.

Proof: See Kato,[9] pp. 378 and 386.

V is actually infinitesimally small with respect to H_0 because of the following:

Lemma 2: Let H_0 be given by (6), $V=x$ and $ab=2$ with $b>0$. Then

$$\|Vu\|\leqslant b\|H_0u\|+a\|u\|, \quad u\in\mathscr{D}(H_0)\subset\mathscr{D}(V). \quad (11)$$

[a)] Research partially supported by NSERC.

 0022-2488/83/082089-06$02.50

Proof: From $x^2/4 \leqslant H_0$ one obtains for any $u \in \mathscr{D}(H_0)$

$$\|Vu\| \leqslant [4(u,H_0u)]^{1/2} \leqslant 2(\|u\| \ \|H_0u\|)^{1/2} \leqslant a\|u\| + b\|H_0u\| \tag{12}$$

with $ab = 2$.

Corollary 2: $H(\beta)$ is a self-adjoint analytic family of type (A) for all β (an entire family).

An important aspect of analytic families is the justification of the Rayleigh–Schrödinger perturbation series for the eigenvalues and eigenfunctions of $H(\beta)$. One has

$$E_m(\beta) = E_m(0) + \beta E_m^{(1)} + \beta^2 E_m^{(2)} + \cdots \tag{13}$$

near $\beta = 0$ and estimates for β_0 the radius of convergence. The standard result is:

Theorem 1: Let H_0 be self-adjoint. Let V be symmetric with $(a,b \geqslant 0)$

$$\|Vu\| \leqslant b\|H_0u\| + a\|u\|, \quad u \in \mathscr{D}(H_0) \subset \mathscr{D}(V).$$

Then $H(\beta)$ is a self-adjoint analytic family for $|\beta| < 1/b$. Furthermore, if H_0 has an isolated nondegenerate eigenvalue E separated from the rest of its spectrum $\sigma(H_0)$ by an amount d, then $H(\beta)$ has a nondegenerate eigenvalue $E(\beta)$ such that $E(0) = E$ and $E(\beta)$ is analytic for $|\beta| < \beta_0$ with

$$\beta_0 \geqslant [b(2 + 2|E|/d) + 2a/d]^{-1}. \tag{14}$$

Proof: The proof is given in Kato (Ref. 9, pp. 288 and 384). The H_0-boundedness of V is used to obtain analyticity of the projection associated with $E(\beta)$ for $|\beta| < \beta_0$ with

$$\beta_0 \geqslant \min_{z \in \Gamma} [a\|(z - H_0)^{-1}\| + b\|H_0(z - H_0)^{-1}\|]^{-1} \tag{15}$$

and Γ a contour enclosing and separating E. With the choice of contour $\Gamma_0 = \{z: |z - E| = d/2\}$ one obtains for $z \in \Gamma_0$

$$\|(z - H_0)^{-1}\| \leqslant 2/d \quad \text{and} \quad \|H_0(z - H_0)^{-1}\| \leqslant 2 + 2|E|/d.$$

A different choice of contour will, of course, yield a different estimate.

In the present case one has $E = E_m(0) = l + 2m + \frac{3}{2}$, $d = 2$, and one may use either (from Lemmas 1 and 2)

$$a = 0, \quad b = 2/\sqrt{l + \tfrac{3}{2}} \tag{16}$$

or

$$ab = 2. \tag{17}$$

From (14) and (16) one obtains (13) for $|\beta| < \beta_0$ with

$$\beta_0 \geqslant \sqrt{l + \tfrac{3}{2}}/2(l + 2m + \tfrac{7}{2}), \tag{18}$$

while the use of (17) yields a family of estimates, with the sharpest one being

$$\beta_0 \geqslant 1/2\sqrt{2}(l + 2m + \tfrac{7}{2})^{1/2} \tag{19}$$

corresponding to the values $a = \sqrt{2}[2 + E_m(0)]^{1/2}$ and $b = \sqrt{2}[2 + E_m(0)]^{-1/2}$.

One should note that (19) is sharper than (18) for small l and m and large m while (18) is sharper for large l. Both these estimates may be improved on. In this connection it is useful to consider contours other than the standard one Γ_0.

Lemma 3: Let H_0 be a self-adjoint (or normal) operator with isolated eigenvalue E having isolation distance d. Let Γ_1 and Γ_2 be circular contours enclosing E defined by

$$\Gamma_1 = \{z: |z - E| = |E|d/(2|E| + d)\} \tag{20}$$

and

$$\Gamma_2 = \{z: |z - E| = \sqrt{|E|(|E| + d)} - |E|\}. \tag{21}$$

Then

$$\sup_{z \in \Gamma_1} \|H_0(z - H_0)^{-1}\| = 1 + 2|E|/d \tag{22}$$

and

$$\sup_{z \in \Gamma_2} \||H_0|^{1/2}(z - H_0)^{-1}\| = (\sqrt{|E| + d} + \sqrt{|E|})/d. \tag{23}$$

Proof: One has

$$\|H_0(z - H_0)^{-1}\| = \sup_{E' \in \sigma(H_0)} \frac{|E'|}{|z - E'|}.$$

If $E \in \sigma(H_0)$, $E' \neq E$, and $z \in \Gamma_1$, then

$$|z - E'| \geqslant |E - E'| - |z - E| \geqslant d - |E|d(2|E| + d)^{-1} = d(d + |E|)/(d + 2|E|).$$

Also

$$|E'| \ |z - E'|^{-1} = |E' - z + z| \ |z - E'|^{-1} \leqslant 1 + |z| \ |z - E'|^{-1}$$

and

$$|z| \leqslant |E| + |z - E| = 2|E|(|E| + d)(2|E| + d)^{-1}.$$

Thus,

$$|E'| \ |z - E'|^{-1} \leqslant 1 + 2|E|/d = |E| \ |z - E|^{-1}.$$

If $z \in \Gamma_2$, one similarly has

$$|z - E'| \geqslant |E| + d - \sqrt{|E|(|E| + d)}$$

and

$$|E'| \ |z - E'|^{-2} \leqslant (\sqrt{|E| + d} + \sqrt{|E|})^2/d^2 = |E| \ |z - E|^{-2}.$$

Corollary 3: If in Theorem 1 one has $a = 0$, then (14) may be replaced by

$$\beta_0 \geqslant [b(1 + 2|E|/d)]^{-1}. \tag{24}$$

Proof: If one uses the contour $\Gamma = \Gamma_1$, then (24) follows from (15) and (22).

Applying (24) to the present problem yields the estimate

$$\beta_0 \geqslant \sqrt{l + \tfrac{3}{2}}/2(l + 2m + \tfrac{5}{2}), \tag{25}$$

which is an improvement over (18). To improve (19), it is necessary to make a change in the assumptions of Theorem 1.

Corollary 4: If the H_0-boundedness condition (11) in Theorem 1 is replaced by the condition

$$\|Vu\| \leqslant c(\|u\| \ \|H_0u\|)^{1/2}, \quad u \in \mathscr{D}(H_0) \subset \mathscr{D}(V), \tag{26}$$

then the conclusions are unchanged excepting that one may now take $ab = c$ (so that b can be made arbitrarily small) and one may replace (14) by the estimate

$$\beta_0 \geqslant (d/2c)(d + |E|)^{-1/2}. \tag{27}$$

Proof: Equation (26) implies (11) with $ab = c$. Also in the proof of the estimate for β_0 it suffices to have

$$\beta_0 \geqslant \min_{z \in \Gamma} [\|V(z - H_0)^{-1}\|]^{-1} \tag{28}$$

with Γ a contour enclosing and separating E. From (26) one has

$$\|V(z-H_0)^{-1}\| \leqslant c[\,\|(z-H_0)^{-1}\|\ \|H_0(z-H_0)^{-1}\|\,]^{1/2},$$

and (27) then follows from (28) with the choice of contour $\Gamma = \Gamma_0$.

For the present problem one has (26) with $c = 2$ [see (12)], and (27) yields

$$\beta_0 \geqslant 1/2(l+2m+\tfrac{7}{2})^{1/2}, \tag{29}$$

which improves (19) and is sharper than (25) even for large l. However, an even sharper estimate can be obtained by using the inequality

$$x^2/4 \leqslant H_0. \tag{30}$$

Corollary 5: If the H_0-boundedness condition (11) in Theorem 1 is replaced by the condition

$$\|Vu\| \leqslant c(u,|H_0|u)^{1/2}, \quad u \in \mathscr{D}(H_0) \subset \mathscr{D}(V), \tag{31}$$

then the conclusions are unchanged excepting that one may now take $ab = c$ and replace (14) by the estimate

$$\beta_0 \geqslant (d/c)(\sqrt{|E|+d}+\sqrt{|E|})^{-1}. \tag{32}$$

Proof: See the proof of Corollary 4 noting that (31) implies (26) and that one now has

$$\|V(z-H_0)^{-1}\| \leqslant c\|\ |H_0|^{1/2}(z-H_0)^{-1}\|. \tag{33}$$

Thus

$$\beta_0 \geqslant \min_{z\in\Gamma}[c\|\ |H_0|^{1/2}(z-H_0)^{-1}\|]^{-1}. \tag{34}$$

If one chooses the contour $\Gamma = \Gamma_2$, then (32) follows from Eq. (23) of Lemma 3.

For the present problem (31) with $c = 2$ follows from (30) and estimate (32) then yields

$$\beta_0 \geqslant (\sqrt{l+2m+\tfrac{3}{2}}+\sqrt{l+2m+\tfrac{7}{2}})^{-1}, \tag{35}$$

which is sharper than all the previous estimates.

Thus the Rayleigh–Schrödinger perturbation expansion (13) converges for $|\beta| < \beta_0$, and (35) gives one a lower bound on the radius of convergence. In terms of the original variable α one then has $E_m(\alpha)$ expanded in a convergent Taylor series in powers of $1/\sqrt{\alpha}$ for

$$\alpha > (\sqrt{l+2m+\tfrac{3}{2}}+\sqrt{l+2m+\tfrac{7}{2}})^2/4, \tag{36}$$

which may be replaced by the more convenient estimate

$$\alpha \geqslant l+2m+\tfrac{5}{2} \tag{37}$$

without much loss of sharpness.

One final estimate will be derived which is not as sharp as (35) for small l and m or large m, but which is an improvement of (35) for large l.

Lemma 4: If $q \geqslant 27p^4/256$, then $x^4 - px^3 + q \geqslant 0$.

Proof: One has $x^4 - px^3 + 27p^4/256 = (x-\tfrac{3}{4}p)^2[(x+p/4)^2 + p^2/8] \geqslant 0$.

Corollary 6: If H_0 is given by (6) and $V = x$, then

$$V \leqslant bH_0 \tag{38}$$

with

$$b = (27)^{1/4}(2l+1)^{-1/2}. \tag{39}$$

Proof: On $C_0^\infty(0,\infty)$ one may use the basic inequality (Ref. 9, p. 345)

$$-\frac{d^2}{dx^2} \geqslant \frac{1}{4x^2} \tag{40}$$

to obtain

$$H_0 \geqslant \frac{x^2}{4} + \frac{l(l+1)+\frac{1}{4}}{x^2}. \tag{41}$$

The inequality

$$x \leqslant b\left(\frac{x^2}{4} + \frac{l(l+1)+\frac{1}{4}}{x^2}\right) \leqslant bH_0 \tag{42}$$

with b given by (39) then follows from Lemma 4. This extends to $\mathscr{D}(H_0)$ since $C_0^\infty(0,\infty)$ is a core for H_0.

To exploit the inequality (42), it is necessary to consider the theory of analytic families of type (B) or (C). One has the following result for type (C).

Theorem 2: If H_0 is self-adjoint and V is symmetric with $(b \geqslant 0,\ a + b\,|H_0| \geqslant 0)$

$$|(u,Vu)| \leqslant b\,(u,|H_0|u) + a\|u\|^2, \quad u \in \mathscr{D}(H_0) \subset \mathscr{D}(V), \tag{43}$$

then there exists a unique self-adjoint analytic family $H(\beta) \supset H_0 + \beta V$ for $|\beta| < b^{-1}$. In particular, if E is an isolated nondegenerate eigenvalue of H_0 with isolation distance d, then there exists an isolated nondegenerate eigenvalue $E(\beta)$ of $H(\beta)$ such that $E(0) = E$ and $E(\beta)$ is analytic for $|\beta| < \beta_0$ with

$$\beta_0 \geqslant [\,b(2+2|E|/d) + 2a/d\,]^{-1}. \tag{44}$$

Proof: The theorem as stated is a weakened version of one in Kato (Ref. 9, p. 413), where (43) is assumed to hold only on $\mathscr{D}(V) \subset \mathscr{D}(H_0)$ with $\mathscr{D}(V)$ a core of $|H_0|^{1/2}$. Uniqueness is concluded only if $\mathscr{D}(V)$ is also a core of H_0. For our purposes it suffices to take $\mathscr{D}(V) \supset \mathscr{D}(H_0)$. The estimate on β_0 is obtained by using the standard contour Γ_0 of Theorem 1.

Corollary 7: If in Theorem 2 one has $a = 0$, then one may replace (44) by the improved estimate

$$\beta_0 \geqslant [\,b(1+2|E|/d)\,]^{-1}. \tag{45}$$

Proof: the contour Γ_1 of Lemma 3 is used instead of Γ_0.

Using the estimate (45) for the present problem with $a = 0$ and b given by (39), one obtains

$$\beta_0 \geqslant \sqrt{2l+1}/(27)^{1/4}(l+2m+\tfrac{5}{2}), \tag{46}$$

which is sharper than (25) for $l > 1.353\cdots$ and sharper than (35) for sufficiently large l.

The large-l behavior of (46) can be further improved to obtain

$$\beta_0 \geqslant (1/\sqrt{2l})[1 + O(l^{-1})] \tag{47}$$

by modifying (38) so as to get

$$V \leqslant bH_0 + a$$

for suitable $b > 0$ and $a < 0$ and using the Theorem 2 estimate (44). Since the estimate one obtains is sharper than previous estimates only for sufficiently large l, we omit the details.

Having obtained various estimates on the radius of convergence of the Rayleigh–Schrödinger series (13), we would be remiss in not calculating at least the Born approximation. One has

$$E_m^{(1)} = \int_0^\infty x[\phi_m^{(0)}(x)]^2\,dx \Big/ \int_0^\infty [\,\phi_m^{(0)}(x)]^2\,dx \tag{48}$$

with $\phi_m^{(0)}$ given by (9).

The numerator in (48) can be obtained from the formula (Ref. 11, p. 54)

$$\int_0^\infty e^{-t}t^{l+1}\,{}_1F_1^2[-m;l+\tfrac{3}{2};t]\,dt$$
$$=\frac{\Gamma(l+2)\Gamma(l+\tfrac{3}{2})\Gamma(-\tfrac{1}{2}+m)}{\Gamma(l+\tfrac{3}{2}+m)\Gamma(-\tfrac{1}{2})}$$
$$\times{}_3F_2[-m,l+2,\tfrac{3}{2};\tfrac{3}{2}+l,\tfrac{3}{2}-m;1], \tag{49}$$

where ${}_3F_2$ is a generalized hypergeometric function.

To obtain the denominator in (48), we note that (Ref. 12, p. 509)

$${}_1F_1[-m;l+\tfrac{3}{2};t]=[m!\Gamma(l+\tfrac{3}{2}+m)/\Gamma(l+\tfrac{3}{2})]L_m^{(l+\frac{1}{2})}(t), \tag{50}$$

where $L_m^{(l+\frac{1}{2})}$ is the generalized Laguerre polynomial which satisfies (Ref. 12, p. 775)

$$\int_0^t t^{l+\frac{1}{2}}e^{-t}[L_m^{(l+\frac{1}{2})}(t)]^2\,dt=\frac{\Gamma(l+\tfrac{3}{2}+m)}{m!}. \tag{51}$$

From (9) and (48)–(51) it follows that

$$E_m^{(1)}=\frac{\sqrt{2}\Gamma(l+2)\Gamma(-\tfrac{1}{2}+m)}{m!\Gamma(l+\tfrac{3}{2})\Gamma(-\tfrac{1}{2})}$$
$$\times{}_3F_2[-m,l+2,\tfrac{3}{2};\tfrac{3}{2}+l,\tfrac{3}{2}-m;1], \tag{52}$$

which for $m=0$ reduces to

$$E_0^{(1)}=\sqrt{2}\Gamma(l+2)/\Gamma(l+\tfrac{3}{2}). \tag{53}$$

For $l=0$, Eq. (52) becomes

$$E_m^{(1)}=\frac{\sqrt{2}\Gamma(-\tfrac{1}{2}+m)}{m!\Gamma(\tfrac{3}{2})\Gamma(-\tfrac{1}{2})}\,{}_2F_1[-m,2;\tfrac{3}{2}-m;1]. \tag{54}$$

Using the formula (Ref. 13, p. 104)

$${}_2F_1[-m,2;\tfrac{3}{2}-m;1]$$
$$=\Gamma(\tfrac{3}{2}-m)\Gamma(-\tfrac{1}{2})/\Gamma(\tfrac{3}{2})\Gamma(-\tfrac{1}{2}-m), \tag{55}$$

(54) becomes

$$E_m^{(1)}=\frac{\sqrt{2}\Gamma(\frac{3}{2}+m)}{m!\Gamma^2(\frac{3}{2})}=\frac{(-1)^{m+1}4\sqrt{2}}{m!\Gamma(-\frac{1}{2}-m)}, \tag{56}$$

agreeing with the $l=0$ calculation of Ref. 1.

3. THE SMALL-α EXPANSION

Let $\alpha>0$ and put $x=(r-1)/\sqrt{a}$. Then (1) may be expressed as

$$H(1)\phi=E(1)\phi, \tag{57}$$

where

$$H(\beta)=H_0+\beta V, \tag{58}$$

$$H_0=-\frac{d^2}{dx^2}+\frac{x^2}{4}, \tag{59}$$

$$V=\frac{l(l+1)\alpha}{(1+\sqrt{\alpha}x)^2}, \tag{60}$$

and

$$E(1)=\lambda+\tfrac{1}{2} \tag{61}$$

with H, H_0, and V self-adjoint operators on the Hilbert space $L^2(-1/\sqrt{\alpha},\infty)$ which are essentially self-adjoint on $C_o^\infty(-1/\sqrt{\alpha},\infty)$.

The eigenvalue problem for H_0 is exactly solvable in terms of parabolic cylinder functions $D_\lambda(x)$.[1] One has

$$H_0\phi_m^{(0)}=E_m^{(0)}\phi_m^{(0)}, \tag{62}$$

where

$$E_m^{(0)}=\lambda_m(\alpha)+\tfrac{1}{2} \tag{63}$$

and

$$\phi_m^{(0)}=C_mD_{\lambda_m}(x) \tag{64}$$

with $\lambda_m(\alpha)$ a root of the equation

$$D_\lambda(-1/\sqrt{\alpha})=0. \tag{65}$$

From the estimate [see Sec. 2, Eq. (40)]

$$\alpha/4(1+\sqrt{\alpha}x)^2\leqslant H_0, \tag{66}$$

one obtains

$$V\leqslant\tfrac{1}{4}l(l+1)H_0, \tag{67}$$

which implies (using Theorem 2 and Corollary 7 of Sec. 2) that $H(\beta)$ has eigenvalue $E_m(\beta)$ with Rayleigh–Schrödinger series

$$E_m(\beta)=E_m^{(0)}+\beta E_m^{(1)}+\beta^2E_m^{(2)}+\cdots \tag{68}$$

convergent for $|\beta|<\beta_0$ with

$$\beta_0\geqslant 4d(\alpha)/l(l+1)[d(\alpha)+2E_m^{(0)}(\alpha)]. \tag{69}$$

It is known that[1]

$$m+\tfrac{1}{2}<E_m^{(0)}(\alpha)<2m+\tfrac{3}{2} \tag{70}$$

with $E_m^{(0)}\to 2m+\frac{3}{2}$ as $\alpha\to\infty$ and $E_m^{(0)}(\alpha)\to m+\frac{1}{2}$ as $\alpha\to 0+$ monotonically. Also one empirically has $1<d(\alpha)<2$. Thus, even in the most favorable case $\alpha\to 0+$, Eq. (69) yields the inadequate lower bound

$$\beta_0\geqslant 2/l(l+1)(m+1) \tag{71}$$

so that it is not known if (68) with $\beta=1$ converges even for $l=1$ and $m=0$ and $\alpha\to 0+$. Hence

$$E_m(1)=E_m^{(0)}+E_m^{(1)}+E_m^{(2)}+\cdots \tag{72}$$

with $E_m^{(i)}$ given by the Rayleigh–Schrödinger expressions

$$E_m^{(1)}=(\phi_m^{(0)},V\phi_m^{(0)})/(\phi_m^{(0)},\phi_m^{(0)}), \tag{73}$$

$$E_m^{(2)}=-\sum_{j\neq m}(E_j^{(0)}-E_m^{(0)})^{-1}V_{mj}V_{jm},\quad \text{etc.}, \tag{74}$$

where

$$V_{ij}=(\phi_i^{(0)},V\phi_j^{(0)})/(\|\phi_i^{(0)}\|\,\|\phi_j^{(0)}\|) \tag{75}$$

(see Ref. 9 and Ref. 10, p. 8) may only be a formal expansion evaluated outside the interval of convergence.

Equation (72) is certainly an asymptotic expansion for small α since we will show that

$$E^{(i)}(\alpha)=O(\alpha^{2i-1}),\quad i=1,2,\cdots, \tag{76}$$

and one has $E_m(\beta)$ analytic in a strip which includes the positive real axis. [Perturbing $H_0+\beta_1V$ by β_2V with $0<\beta_1<\beta_0$ and $\beta_2>0$ yields an analytic continuation of $E_m(\beta)$. By successive steps one may analytically continue $E_m(\beta)$ to any $0<\beta<\infty$. See also Ref. 10, Theorem XII. 10, p. 20].

To show (76), it is necessary to examine the Rayleigh–Schrödinger terms in more detail. One has

$$E_m^{(1)} = \frac{l(l+1)\alpha \int_{-1/\sqrt{\alpha}}^{\infty} D_{\lambda_m}^2(x)(1+\sqrt{\alpha}x)^{-2}\,dx}{\int_{-1/\sqrt{\alpha}}^{\infty} D_{\lambda_m}^2(x)\,dx}, \tag{77}$$

$$E_m^{(2)} = -[l(l+1)\alpha]^2 \sum_{j\neq m} (\lambda_j - \lambda_m)^{-1} \times \left[\int_{-1/\sqrt{\alpha}}^{\infty} D_{\lambda_j}(x) D_{\lambda_m}(x)(1+\sqrt{\alpha}x)^{-2}\,dx\right]^2 \Big/ \left(\int_{-1/\sqrt{\alpha}}^{\infty} D_{\lambda_j}^2(x)\,dx\right)\left(\int_{-1/\sqrt{\alpha}}^{\infty} D_{\lambda_m}^2(x)\,dx\right), \quad \text{etc.} \tag{78}$$

If one neglects terms that are exponentially small, then there is considerable simplification in the formulas for $E_m^{(i)}$. We illustate this in the expressions (77) and (78) for $E_m^{(1)}$ and $E_m^{(2)}$.

It is known that[1]

$$\lambda_m(\alpha) = m + O(\alpha^{-m+\frac{1}{2}} e^{-1/(2\alpha)}). \tag{79}$$

From (79) and the Taylor expansion

$$D_{\lambda_m}(x) = D_m(x) + \sum_{n=1}^{\infty} \frac{d^n}{(d\lambda)^n} D_\lambda(x) \Big|_{\lambda=m} \frac{(\lambda_m - m)^n}{n!}, \tag{80}$$

(D_λ is an entire function of λ), one obtains

$$D_{\lambda_m}(x) = D_m(x) + O(\alpha^{-m+\frac{1}{2}} e^{-1/(2\alpha)}). \tag{81}$$

The estimate (81) is, of course, not necessarily uniform in x. However, since $D_\lambda(x)$ is an entire function of x and $[d^n/(d\lambda)^n]D_\lambda(x) \to 0$ as $|x| \to \infty$ with $|\arg x| < \pi/4$ [see Ref. 1, Eq. (5)], the estimate is uniform for $x \in [-M, \infty)$. To obtain an estimate uniform on $[-1/\sqrt{\alpha}, \infty)$ with $\alpha \to 0$, one must consider the behavior of $[d^n/(d\lambda)^n]D_\lambda(x)|_{\lambda=m}$ as $x \to -\infty$. From Ref. 1, Eq. (6), one has

$$\frac{d^n D_\lambda(x)}{d\lambda^n}\Big|_{\lambda=m} \sim n(-1)^{m+n-1}\sqrt{2\pi}\, m!\, e^{x^2/4} \times (-x)^{-m-1}[\ln(-x)]^{n-1}, \quad n = 1,2,\cdots, \tag{82}$$

as $x \to -\infty$. Thus from (82) and (80)

$$D_{\lambda_m}(x) \sim D_m(x) + (-1)^m \sqrt{2\pi}\, m!\, e^{x^2/4}(\lambda_m - m)(-x)^{m-\lambda_m} \tag{83}$$

as $x \to -\infty$ so that

$$D_{\lambda_m}(x) = D_m(x) + O(\alpha^{-m+\frac{1}{2}} e^{-1/(4\alpha)}) \tag{84}$$

uniformly for $x \in [-1/\sqrt{\alpha}, \infty)$.

From (84) and the fact that

$$\int_{-\infty}^{\infty} |D_m(x)|\,dx < \infty$$

one obtains

$$\int_{-1/\sqrt{\alpha}}^{\infty} D_{\lambda_m}^2(x) = \int_{-1/\sqrt{\alpha}}^{\infty} D_m^2(x)\,dx + O(\alpha^{-m+\frac{1}{2}} e^{-1/(4\alpha)}). \tag{85}$$

Thus in all the denominator terms (normalization integrals) in the expressions for $E_m^{(i)}$ one may (with the neglect of exponentially small terms) replace

$$\int_{-1/\sqrt{\alpha}}^{\infty} D_{\lambda_j}^2(x)\,dx \quad \text{by} \quad \int_{-1/\sqrt{\alpha}}^{\infty} D_j^2(x)\,dx.$$

Noting also that (Ref. 12, p. 691)

$$D_m(x) = 2^{-m/2} e^{-x^2/4} H_m(x/\sqrt{2}), \tag{86}$$

where H_m is the Hermite polynomial of degree of m, one has

$$\int_{-1/\sqrt{\alpha}}^{\infty} D_m^2(x)\,dx = \int_{-\infty}^{\infty} D_m^2(x)\,dx + O(\alpha^{-m-\frac{1}{2}} e^{-1/(2\alpha)}). \tag{87}$$

From (85)–(87) and the fact that

$$\int_{-\infty}^{\infty} e^{-x^2/2} H_m^2(x/\sqrt{2})\,dx = 2^m m!\sqrt{2\pi} \tag{88}$$

one obtains

$$\int_{-1/\sqrt{\alpha}}^{\infty} D_{\lambda_m}^2(x)\,dx \approx \sqrt{2\pi}\, m! \tag{89}$$

to order $\alpha^{-m+\frac{1}{2}} e^{-1/(4\alpha)}$.

In the numerator integrals for $E_m^{(i)}$ one has quantities

$$I_{jk} = \int_{-1/\sqrt{\alpha}}^{\infty} D_{\lambda_j}(x) D_{\lambda_k}(x)(1+\sqrt{\alpha}x)^{-2}\,dx. \tag{90}$$

Here one may not replace D_{λ_j} and D_{λ_k} by D_j and D_k because of the presence of the singular term $(1+\sqrt{\alpha}x)^{-2}$. However, if one expands the singular term in powers of $\sqrt{\alpha}$, one obtains the asymptotic expansion

$$I_{jk} \sim \sum_{n=0}^{\infty} (-1)^n (n+1)\alpha^{n/2} \times \int_{-1/\sqrt{\alpha}}^{\infty} x^n D_{\lambda_j}(x) D_{\lambda_k}(x)\,dx, \tag{91}$$

and in each term of this asymptotic expansion one may now (with the neglect of exponentially small terms) replace D_{λ_j} and D_{λ_k} by D_j and D_k and take each integral over the interval $(-\infty, \infty)$. The calculations are further simplified by recalling (86) and using the recursion relation

$$xH_m(x/\sqrt{2}) = (1/\sqrt{2})H_{m+1}(x/\sqrt{2}) - \sqrt{2}mH_{m-1}(x/\sqrt{2}) \tag{92}$$

and the orthogonality relation

$$\int_{-\infty}^{\infty} H_j\left(\frac{x}{\sqrt{2}}\right) H_k\left(\frac{x}{\sqrt{2}}\right) e^{-x^2/2}\,dx = 0, \quad j \neq k. \tag{93}$$

Thus Eq. (76) follows from the general formula for $E^{(i)}$ (Ref. 9 or 10), (84), (86), (88), and (90)–(93).

As an illustration we now calculate $E_m^{(1)}$ and $E_m^{(2)}$ to order α^3. One has from (77), (84), (86), (88), and (91)

$$E_m^{(1)} = \frac{l(l+1)\alpha}{\sqrt{2\pi}\, m!}\left[m!\sqrt{2\pi} - 2\sqrt{\alpha}2^{-m}\int_{-\infty}^{\infty} xe^{-x^2/2} \times H_m^2\left(\frac{x}{\sqrt{2}}\right)dx + 3\alpha 2^{-m}\int_{-\infty}^{\infty} x^2 e^{-x^2/2} \times H_m^2\left(\frac{x}{\sqrt{2}}\right)dx - 4(\sqrt{\alpha})^3 2^{-m} \times \int_{-\infty}^{\infty} x^3 e^{-x^2/2} H_m^2\left(\frac{x}{\sqrt{2}}\right)dx + 5\alpha^2 2^{-m} \times \int_{-\infty}^{\infty} x^4 e^{-x^2/2} H_m^2\left(\frac{x}{\sqrt{2}}\right)dx \cdots\right]. \tag{94}$$

Using (92), (93), and (88) this becomes

$$E_m^{(1)} = l(l+1)\alpha[1 + 3\alpha(2m+1) + 15\alpha^2(2m^2+2m+1) + O(\alpha^3)]. \tag{95}$$

Similarly from (78) one obtains

$$E_m^{(2)} = -[l(l+1)\alpha]^2 \sum_{j\neq m} 2^{-(j+m)/2} \frac{(j-m)^{-1}}{2\pi j!m!} \times \left[\int_{-\infty}^{\infty} e^{-x^2/2} H_j(x) H_m(x) \times (1 - 2\sqrt{\alpha}x + 3\alpha x^2 + \cdots) dx\right]^2, \quad (96)$$

which [using (92) and (93)] yields

$$E_m^{(2)} = -[(l+1)\alpha]^2[4\alpha + O(\alpha^2)]. \quad (97)$$

(Since only $j = m \pm 1$ contributes to the order α term).

Thus from (63), (72), (76), (79), (95), and (97) one has

$$E_m(1) = m + \tfrac{1}{2} + l(l+1)\alpha[1 + 3\alpha(2m+1) + 15\alpha^2(2m^2 + 2m + 1)] - 4\alpha[l(l+1)\alpha]^2 + O(\alpha^4), \quad (98)$$

which agrees with previous calculations.[5,7,8]

One may circumvent much of the tediousness of these and higher-order calculations by considering the following formal perturbation problem.

In the Hilbert space $L^2(-\infty,\infty)$ let H_0 be the harmonic oscillator Hamiltonian

$$H_0 = -\frac{d^2}{dx^2} + \frac{x^2}{4}$$

with eigenvalues

$$E_m^{(0)} = m + \tfrac{1}{2}, \quad m = 0,1,2,\cdots \quad (99)$$

and corresponding normalized eigenfunctions

$$\phi_m^{(0)} = \frac{2^{-m/2} e^{-x^2/4}}{(2\pi)^{1/4}\sqrt{m!}} H_m\left(\frac{x}{\sqrt{2}}\right). \quad (100)$$

Let

$$H_F = H_0 + V_F \quad (101)$$

be a formal perturbation of H_0 with

$$V_F = l(l+1) \times \alpha(1 - 2\sqrt{\alpha}x + 3\alpha x^2 + \cdots + (-1)^n \times (n+1)\alpha^{n/2} x^n + \cdots). \quad (102)$$

Then we have shown the following:

Proposition: The formal Rayleigh–Schrödinger expansion of the eigenvalues of H_F in terms of the eigenvalues and eigenfunctions of H_0 in $L^2(-\infty,\infty)$ yields an asymptotic expansion of the eigenvalues $E_m(1)$ of the rotating harmonic oscillator in powers of α which neglects exponentially small terms.

One should note that, although the formal perturbation V_F contains powers of $\sqrt{\alpha}x$ the perturbation series for $E_m(1)$ will contain only powers of α. This follows from (92) and (93) since in order to have an odd power of $\sqrt{\alpha}$ one must have an nth order perturbation term of the form

$$V_{i_1 i_2} \cdots V_{i_p i_{p+1}} \cdots V_{i_n i_{n+1}}$$

with

p = odd integer,

$i_j - i_{j-1} = \pm$ odd integer, $j = 2,\ldots,p+1$,

$i_j - i_{j-1} = \pm$ even integer, $j = p+2,\ldots,n+1$,

where $i_{n+1} = i_1$. This yields a contradiction since $i_2 - i_1 = \pm$ odd integer while $i_1 - i_2 = \Sigma_{j=3}^{n+1} i_j - i_{j-1} = \pm$ even integer.

[1] D. Masson, J. Math. Phys. **24**, 2074 (1983).
[2] V. Singh, S. N. Biswas, and K. Datta, J. Math. Phys. **23**, 1323 (1982).
[3] N. Fröman, P. O. Fröman, and F. Karlsson, Phys. Lett. A **77**, 397 (1980).
[4] G. P. Flessas, Phys. Lett. A **71**, 315 (1979).
[5] N. Fröman and P. O. Fröman, Ann. Phys. **115**, 269 (1978).
[6] R. E. Langer, Phys. Rev. **75**, 792 (1949).
[7] J. E. Rosenthal and L. Motz, Proc. Natl. Acad. Sci. **23**, 259 (1937).
[8] J. L. Dunham, Phys. Rev. **41**, 721 (1932).
[9] T. Kato, *Perturbation Theory for Linear Operators* (Springer-Verlag, New York, 1966).
[10] M. Reed and B. Simon, *Methods of Modern Mathematical Physics IV: Analysis of Operators* (Academic, New York, 1978).
[11] L. J. Slater, *Confluent Hypergeometric Functions* (Cambridge U.P., Cambridge, 1960).
[12] M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions* (National Bureau of Standards, Washington DC, 1966).
[13] A. Erdélyi, Ed., *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. I.

Symmetries, conservation laws, and time reversibility for Hamiltonian systems with external forces

A. J. van der Schaft[a)]
Mathematics Institute, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands

(Received 8 June 1982; accepted for publication 1 October 1982)

A system theoretic framework is given for the description of Hamiltonian systems with external forces and partial observations of the state. It is shown how symmetries and conservation laws can be defined within this framework. A generalization of Noether's theorem is obtained. Finally a precise definition of time reversibility is given and its consequences are explored.

PACS numbers: 03.20. + i

I. INTRODUCTION

In the last century the mathematical formulation of classical mechanics has culminated in the elegant theory as described for instance in the books of Abraham and Marsden[1] and Arnold.[2] However, the emphasis in this approach has been put on analytical mechanics, i.e., Hamiltonian systems which can be described without external forces. If forces are present, they are assumed to come from a potential field, and therefore can be incorporated in the system by adding a potential function to the Hamiltonian function. Since external forces do come up at various places, for instance experimental devices and technical applications, and mostly cannot be derived from a potential function, this entails indeed quite a loss of generality.

In this paper we elaborate a framework, which can incorporate external forces on a conceptual level. At the same time we also formalize the idea that we may only partially observe the state of a system. Our basic notions stem from system theory, the discipline that explicitly deals with systems with inputs and outputs which in this context are called external forces and observations, respectively.

A simple example will make things more clear. Consider Newton's second law $F = m\ddot{q}$. Notice already that this law cannot be adequately formulated in a framework without external forces. We will look at it as a system with external force $u = F$,

$$\dot{q} = (1/m)p,$$
$$\dot{p} = u,$$

and an observation function y of the state (q,p) given by

$$y = q.$$

We will give, using the language of symplectic geometry, a general framework for the description of such systems. Furthermore we will show how in this framework symmetries, conservation laws and time reversibility can be defined and treated in an appealing way.

The paper is a further elaboration of previous work,[3–7] which was in turn much inspired by work of Brockett,[8] Takens,[9] and Willems.[10] The first definitions in Sec. II were basically given in,[3] while Sec. III is already partially contained in.[5] We have tried to make the paper more or less self-contained, at least with respect to definitions and statements of theorems, but for more background and detail we refer to the references mentioned (see also the survey[11]).

Some notation: For symplectic geometry we refer to Refs. 1 and 2. If (M,ω) is a symplectic manifold (of dimension $2n$) with symplectic form ω, and $H:M\to\mathbb{R}$ a smooth function, then the Hamiltonian vector field X_H on M is defined by $\omega(X_H, -) = dH$. Let $F,H:M\to\mathbb{R}$ be two functions, then the Poisson bracket $\{F,H\}$ is again a function on M defined by $\{F,H\} := \omega(X_F,X_H)$. Let $F_1,...,F_k$ be functions on M. Take all functions on M which are functions of the F_i, i.e., all functions $\varphi\circ(F_1,...,F_k): M\to\mathbb{R}$. This generates a linear subspace $\mathcal{F}$ of the space of functions on M. We call $\mathcal{F}$ a Poisson algebra (or "function group," see Ref. 12) if $\{f,g\}\in\mathcal{F}$ for all $f,g\in\mathcal{F}$. For every $x\in M$ we define $d\mathcal{F}(x)$ as the linear subspace of T_x^*M, given by $d\mathcal{F}(x) = \{df(x) | f\in\mathcal{F}\}$.

A submanifold $N\subset M$ is Lagrangian if $\omega|_N = 0$ and if the dimension of N is n. If $(q_1,...,q_n,p_1,...,p_n)$ are symplectic coordinates (i.e., $\omega = \Sigma_{i=1}^n dq_i\wedge dp_i$) and N can be parametrized by $q_1,...,q_n$, then there exists (locally) a function $S(q_1,...,q_n)$ such that $N = \{(q,p) | p_i = \partial S/\partial q_i, i = 1,...,n\}$. S is the generating function of N.

If f is a function on a manifold W, we can define the function $\dot{f}$ on TW by $\dot{f}(v) := df(v)$, for $v\in TW$. Therefore if $(x_1,...,x_n)$ are coordinates for W, then $(x_1,...,x_n,\dot{x}_1,...,\dot{x}_n)$ are coordinates for TW. Given an one-form α on W, in local coordinates given by $\alpha = \Sigma_{i=1}^n f_i\,dx_i$, with $f_i: W\to\mathbb{R}$, then we can define the one-form $\dot{\alpha}$ on TW by $\dot{\alpha} = \Sigma_{i=1}^n(\dot{f}_i\,dx_i + f_i\,d\dot{x}_i)$. Let $\omega = \Sigma_{i=1}^n dq_i\wedge dp_i$ be a symplectic form on M, then $\dot{\omega} := \Sigma_{i=1}^n(dq_i\wedge d\dot{p}_i + d\dot{q}_i\wedge dp_i)$ is a symplectic form on TM. All these constructions can also be done in a coordinate-free way.[12–14] Let X be a vector field on M with one-parameter group $X_t: M\to M$, $t\in\mathbb{R}$ and small. Then $(X_t)_*: TM\to TM$ is the one-parameter group of a vector field on TM, which we denote by $\dot{X}$. If $(y_1,...,y_m)$ are coordinates for a manifold Y then the natural coordinates $(y_1,...,y_m, u_1,...,u_m)$ for T^*Y are defined by letting $(\bar{y}_1,...,\bar{y}_m, \bar{u}_1,...,\bar{u}_m)$ correspond to the one-form $\Sigma_{i=1}^m \bar{u}_i\,dy_i$. We recall that cotangent bundles have a canonically defined symplectic form.[1,2]

II. HAMILTONIAN SYSTEMS WITH EXTERNAL FORCES AND PARTIAL OBSERVATIONS

The usual description of a Hamiltonian system (see for instance Refs. 1 and 2 is that of a triplet (M,ω,H), where M is a smooth symplectic manifold denoting the phase space and ω is a symplectic form on M which formalizes the typical structure of the phase space, namely the existence of "conju-

a) Current address: Department of Applied Mathematics, Twente University of Technology, P.O. Box 217, 7500 AE Enschede, The Netherlands.

 0022-2488/83/082095-07$02.50

gate" coordinates q_i and p_i. Finally, H is a smooth function on M which represents the energy of the system. The dynamical behavior of the system is in local coordinates x for M given by $\dot{x} = X_H(x)$, where X_H is the Hamiltonian vector field on M defined by

$$\omega(X_H, -) = dH. \tag{1.1}$$

Using Darboux's theorem we can take local coordinates $(q_1,...,q_n,p_1,...,p_n)$ for M such that $\omega = \Sigma_{i=1}^n dq_i \wedge dp_i$. Then (1.1) comes down to the familiar Hamiltonian equations

$$\dot{q}_i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial q_i}, \quad i = 1,...,n. \tag{1.2}$$

As mentioned in the Introduction, our starting point will be a generalization of this structure, incorporating the possible external forces on the system. Moreover, we will formalize the idea that we may not be able to observe the whole state of the system but only the values of some functions of the state. Therefore we introduce, apart from the state space manifold M, an observation (or output) manifold Y (dim $Y = m$). Then we define the space of external forces (or inputs) in every point of Y as the fiber of the cotangent bundle T^*Y in that point. This seems natural since an element α of a fiber of T^*Y is a linear function on the tangent vectors $\dot{y}$ of Y in that same point. Therefore $\alpha(\dot{y})$ (force times velocity) is defined and represents the external work performed on the system.

Recall that T^*Y has a canonically defined symplectic form, which we denote by ω^e. Since M has a symplectic form ω, we can also define (see the Introduction) a symplectic form $\dot{\omega}$ on TM. This enables us to define the symplectic form Ω on $TM \times T^*Y$ by setting $\Omega := \pi_1^*\dot{\omega} - \pi_2^*\omega^e$ (π_1 and π_2 are the projections of $TM \times T^*Y$ on TM, resp. T^*Y).

Definition 1.1 (Hamiltonian system): Let (M,ω) be a symplectic manifold. Let Y be an observation manifold. A Hamiltonian system $\Sigma(M,T^*Y,L)$ or shortly Σ is given by a submanifold $L \subset TM \times T^*Y$ such that

(i) L can be parametrized by the coordinates of M and the coordinates of the fibers of T^*Y.

(ii) L is a Lagrangian submanifold of $(TM \times T^*Y,\Omega)$.

(iii) The value of the Y-coordinates of a point on L is only a function of the M-coordinates of this point.

Proposition 1.2: Let $\Sigma(M,T^*Y,L)$ be a Hamiltonian system as above. Then in local coordinates the system is given by

$$\begin{aligned} \dot{x} &= X_H(x) + \sum_{i=1}^{m} u_i X_{C_i}(x), \\ y_i &= C_i(x), \quad i = 1,...,m, \end{aligned} \tag{1.3}$$

with x local coordinates for M, $y = (y_1,...,y_m)$ local coordinates for Y, and $u = (u_1,...,u_m)$ the corresponding natural coordinates for the fibers of T^*Y. We will call $C_i: M \to \mathbb{R}$ the observation (or output) functions.

Proof[3,6]: Because of (ii) L has a generating function. Because of (i) and (iii) this generating function has the form $H(x) + \Sigma_{i=1}^m u_i C_i(x)$, which we will abbreviate as $H + u^T C$. Therefore the $\dot{x}$-coordinates of a point on L are given by $\dot{x} = X_H(x) + \Sigma_{i=1}^m u_i X_{C_i}(x)$ and the y-coordinates by

$$y_i = C_i(x), \; i = 1,...,m.$$

Remark 1: If we drop the assumption that M is a symplectic and the conditions (ii) and (iii) in Definition 1.1, we call $\Sigma(M,T^*Y,L)$ just a system (see Refs. 15 and 4 and references therein).

Remark 2: Without condition (iii) we arrive at the more general class of Hamiltonian systems where the external forces (inputs) enter the equations in a nonlinear way. In this case we can also replace T^*Y by a general symplectic manifold (cf. Ref. 3).

Remark 3: From a mathematical point of view the above definition stresses again the importance of the concept of Lagrangian submanifolds, as already done before by many authors. The use of Lagrangian submanifolds in formulating "reciprocity" and "symmetry" in the description of static systems is successfully advocated in many works (see for some references Ref. 1, Sec. 5.3; a particularly nice account is given in Ref. 15). If we generalize Definition 1.1 in the direction given in Remark 2, such static systems correspond to a Hamiltonian system without state space M; i.e., a Lagrangian submanifold of T^*Y or a more general symplectic manifold (see Ref. 3). The description of a Hamiltonian vector field on M as a Lagrangian submanifold of $(TM,\dot{\omega})$ figures prominently in many works of Tulczyjew and coworkers.[13,16,15] Definition 1.1 (and its generalization indicated in Remark 2, see Ref. 3) combines both aspects and gives more rigor to idea of external forces and observations by using a system theoretic framework.

Another way to look at Eqs. (1.3) is to start from a triplet (M,ω,H), to add an observation map $C: M \to Y$, and to define the input vector fields (the directions in which we can exert external forces) as the Hamiltonian vector fields with Hamiltonian functions C_i, where in coordinates for $Y, C = (C_1,...,C_m)$. So we have formalized the idea that we may influence the system by adding to the internal energy H a Hamiltonian function $\Sigma_{i=1}^m u_i(t) C_i$, depending on the observations made on the system.

Examples:

1. Newton's second law as treated in the Introduction.

2. Consider the Euler–Lagrange equations

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) - \frac{\partial L}{\partial q_i} = F_i.$$

Assume that the Legendre transformation $\dot{q} \to p := \partial L / \partial \dot{q}$ is nondegenerate (otherwise see Ref. 3), giving the Hamiltonian function

$$H\left(q, \frac{\partial L}{\partial \dot{q}}\right) = \dot{q}\frac{\partial L}{\partial \dot{q}} - L(q,\dot{q}).$$

Then the Euler–Lagrange equations can be written as

$$\begin{aligned} \dot{q}_i &= \frac{\partial H}{\partial p_i}, \\ \dot{p}_i &= -\frac{\partial H}{\partial q_i} + F_i, \end{aligned}$$

together with the observation functions $y_i = q_i$.

3. Consider a capacitor C and inductance L, coupled in series, with an external voltage V_e. With q_c the charge on the capacitor and φ_L the magnetic flux on the inductance we obtain

$$\begin{pmatrix}\dot{q}_c\\ \dot{\varphi}_L\end{pmatrix}=\begin{pmatrix}0 & \frac{1}{L}\\ \frac{1}{C} & 0\end{pmatrix}\begin{pmatrix}q_c\\ \varphi_L\end{pmatrix}+\begin{pmatrix}0\\ V_e\end{pmatrix}$$

and the observation function $y=q_c$ (note that $\dot{q}_c$ is the extenral current).

As already announced, external work can be naturally defined in our framework. We need one more definition.

Definition 1.3 (External behavior): Let $\Sigma(M,T^*Y,L)$ be a Hamiltonian system locally given by

$$\dot{x}=X_H(x)+\sum_{i=1}^{m}u_iX_{C_i}(x), \tag{1.4a}$$

$$y_i=C_i(x),\quad i=1,...,m. \tag{1.4b}$$

Let $\bar{w}\colon[0,T]\to T^*Y$ be a curve in T^*Y which in natural coordinates (y,u) for T^*Y can be written as

$$\bar{w}(t)=(\bar{y}(t),\bar{u}(t)),\quad t\in[0,T]\quad \text{for a } T>0.$$

Then $\bar{w}$ belongs to the external behavior of the system, if there exists an $x_0\in M$ such that when we apply the force function $\bar{u}(\cdot)$ to the system (1.4a) with $x(0)=x_0$, Eq. (1.4b) yields the same observation function $\bar{y}(\cdot)$.

Let now $\bar{w}\colon[0,T]\to T^*Y$ belong to the external behavior. Then the performed external work is equal to

$$\int_0^T\sum_{i=1}^{m}\bar{u}_i\dot{\bar{y}}_i\,dt=\int_0^T\sum_{i=1}^{m}\bar{u}_i\left\{H+\sum_{j=1}^{m}\bar{u}_jC_j,C_i\right\}dt$$

$$=\int_0^T\sum_{i=1}^{m}\bar{u}_i\{H,C_i\}\,dt+\int_0^T\sum_{i,j=1}^{m}\bar{u}_i\bar{u}_j\{C_i,C_j\}\,dt, \tag{1.5}$$

where for the first identity we use

$$\dot{\bar{y}}_i=\left(X_H+\sum_{j=1}^{m}\bar{u}_jX_{C_j}\right)(\bar{y}_i)=\{H,C_i\}+\sum_{j=1}^{m}\bar{u}_j\{C_j,C_i\},$$

since $y_i=C_i(x)$. Because $\{C_i,C_j\}=-\{C_j,C_i\}$ the last term of (1.5) vanishes and we obtain

$$\text{External work}=\sum_{i=1}^{m}\int_0^T\bar{u}_i\{H,C_i\}dt.$$

Definition (1.1) formalizes that we can exert external forces in the direction of the Hamiltonian vector field of every observation function. This might be too strong. For instance, we should also like to cover the situation

$$\dot{x}=X_H(x),\quad y_i=C_i(x),\quad i=1,...,m,$$

i.e., a Hamiltonian system with partial observations of the state but without external forces. For this we give

Definition 1.4 (Degenerate Hamiltonian system): Let $\Sigma(M,T^*Y,L)$ be a Hamiltonian system. Let $P\subset T^*Y$ be a codistribution on Y (i.e., in every point y of Y a linear subspace of T^*_yY, denoting the possible forces), which is involutive. Assume that $L':=L\cap(TM\times P)$ is a submanifold of $TM\times T^*Y$. Then we call $\Sigma(M,T^*Y,L')$ a degenerate Hamiltonian system.

In local coordinates we obtain the easily proved analog of Proposition 1.2.

Proposition 1.5: Let $\Sigma(M,T^*Y,L'=L\cap(TM\times P))$ be a degenerate Hamiltonian system. Since P is involutive, there exist local coordinates $\{y_1,...,y_m\}$ for Y such that $P=\text{span}\{dy_1,...,dy_k\}$, $k\leqslant m$. In these coordinates for Y the system is locally given by

$$\dot{x}=X_H(x)+\sum_{i=1}^{k}u_iX_{C_i}(x),$$

$$y_i=C_i(x),\quad i=1,...,m. \tag{1.6}$$

Degenerative Hamiltonian systems are frequently encountered as the result of an interconnection of Hamiltonian systems.[3] Apart from interconnections also notions like "coupling Hamiltonians" and "interaction potentials" can be naturally described in our scheme. A time-varying function H_t on M of the form $H_t(x)=\sum_{i=1}^{m}\bar{u}_i(t)C_i(x)$, with $\bar{u}(\cdot)=(\bar{u}_1(\cdot),...,\bar{u}_m(\cdot))$ a certain force function, is sometimes called a coupling Hamiltonian. Consider two Hamiltonian systems $\Sigma(M_i,T^*Y_i,L_i)$, $i=1,2$. The product is again a Hamiltonian system $\Sigma(M_1\times M_2,T^*(Y_1\times Y_2),L_1\times L_2)$. If the generating function of L_1 is $H_1+u^{1\,T}C_1$, and of L_2, $H_2+u^{2\,T}C_2$, then $L_1\times L_2$ has generating function $H_1+H_2+u^{1\,T}C_1+u^{2\,T}C_2$. A coupling Hamiltonian or interaction potential is a function $V\colon Y_1\times Y_2\to\mathbb{R}$, which changes the generating function into

$$H_1(x)+H_2(x)+V(C_1(x),C_2(x))+u^{1\,T}C_1(x)+u^{2\,T}C_2(x).$$

We will now bring in some system theoretic concepts, the most important of which is minimality. Intuitively, one says that a system is minimal if the system cannot be reduced to a system living on a lower dimensional state space and with the same external behavior as the original system.

Definition 1.6 (Minimality): Let $\Sigma(M,T^*Y,L)$ be a (possibly degenerate) Hamiltonian system. Σ is called minimal, if, when there exists another system $\Sigma'(M',T^*Y,L')$ (not necessarily Hamiltonian) and a surjective submersion $\varphi\colon M\to M'$ such that $(\varphi_*,\text{id})(L)=L'$ (with id the identity mapping from T^*Y to T^*Y), then necessarily φ is a diffeomorphism.

A local version of minimality, called local minimality[4] has the following neat characterization, which we will frequently use in the sequel.

Proposition 1.7[4,6]: Let $\Sigma(M,T^*Y,L)$ be a Hamiltonian system with generating function $H+\sum_{i=1}^{m}u_iC_i$. Define the Poisson algebra (see the Introduction) $\mathcal{F}$ as the smallest Poisson algebra containing the functions $C_1,...,C_m$ and closed under taking Poisson brackets with H and C_i, $i=1,...,m$. Then Σ is locally minimal if and only if $\dim d\mathcal{F}(x)=2n$, for every $x\in M$.

The proposition above has an interesting system theoretic interpretation, since it implies that a locally minimal Hamiltonian system is "observable" as well as "controllable." Observability means grosso modo that from the knowledge of the external behavior on an interval $[0,T]$ we can deduce the value of the state on time 0. Controllability implies that if we look at Eq. (1.4a) for a certain $x(0)=x_0$ and consider the set of points in M which are reachable from x_0 by applying different force functions, this set has a nonempty interior in M.

In Ref. 6 it is proven that a Hamiltonian system which is not locally minimal can be reduced to a locally minimal system which is again Hamiltonian. Therefore we could also have formulated Definition 1.6 with a Hamiltonian system $\Sigma(M',T^*Y,L')$. Finally we note that we can give a characterization similar to Proposition 1.7 for local minimality of de-

generate Hamiltonian systems. However, we remark that if $\Sigma(M,T^*Y,L' = L\cap(TM\times P))$ is degenerate Hamiltonian, then the local minimality of $\Sigma(M,T^*Y,L)$ need not imply the local minimality of $\Sigma(M,T^*Y,L')$.

III. SYMMETRIES AND CONSERVATION LAWS

We recall the usual definition of a symmetry for the triplet (M,ω,H) (cf. Refs. 1 and 2): a diffeomorphism φ: $M\to M$ is Hamiltonian symmetry if (i) $\varphi^*\omega=\omega$, (ii) $\varphi^*H=H$.

This generalizes in our framework to

Definition 2.1 (Hamiltonian symmetry): Let $\Sigma(M,T^*Y,L)$ be a Hamiltonian system. A symmetry for Σ is a pair of diffeomorphisms (φ,ψ), with $\varphi: M\to M$, $\psi: T^*Y\to T^*Y$ such that $(\varphi_*,\psi)(L)=L$. A symmetry is called Hamiltonian if $\varphi^*\omega=\omega$ and $\psi^*\omega^e=\omega^e$.

Remark: Note that if (φ,ψ) is a symmetry, then the external behavior of the system is invariant under ψ.[5]

Example 2.2:

Consider a particle in $\mathbb{R}^3$ with mass m in a potential field V, and subject to an external force F. Then the system is given by

$$\left.\begin{aligned} m\ddot{q}_i &= \frac{\partial V}{\partial q_i}+F_i \\ y_i &= q_i \quad (y \text{ is the observation})\end{aligned}\right\} i=1,2,3.$$

Suppose the equations $m\ddot{q}_i=\partial V/\partial q_i$ are invariant under the rotation R around the e_1-axis (this is equivalent to $H=\frac{1}{2}m\dot{q}^2+V$ invariant). Then we know that $R^*: T^*\mathbb{R}^3\to T^*\mathbb{R}^3$ ($T^*\mathbb{R}^3$ is the phase space) is a Hamiltonian symmetry for the system without external force. In this case also $Y=\mathbb{R}^3$, and so $T^*Y=T^*\mathbb{R}^3$. The pair (R^*,R^*) is a Hamiltonian symmetry in the sense of Definition 2.1. It expresses the fact that for this system the observation corresponding to an external force which is rotated around the e_1-axis is obtained by rotating the observation in the same way. We can immediately prove the following

Theorem 2.3: Let $\Sigma(M,T^*Y,L)$ be a locally minimal Hamiltonian system. Let (φ,ψ) be a symmetry for Σ. Then (i) ψ is a fiber respecting bundle morphism and assuming $\psi^*\omega^e=\omega^e$, then, (ii) $\varphi^*\omega=\omega$, and hence (φ,ψ) is a Hamiltonian symmetry.

Proof (see also Refs. 17 and 5):

(i) follows from the structure of L.

(ii) Since $\dot{\omega}-\omega^e|_L=0$ and $(\varphi_*,\psi)(L)=L$, also $\dot{\omega}-\omega^e|_{(\varphi_*,\psi)(L)}=0$, and hence $(\varphi_*)^*\dot{\omega}-\psi^*\omega^e|_L=0$. Because $\psi^*\omega^e=\omega^e$ and $\dot{\omega}-\omega^e|_L=0$, it follows that $(\varphi_*)^*\dot{\omega}-\dot{\omega}|_L=0$.

Define $\alpha:=\varphi^*\omega-\omega$. Then $\dot{\alpha}|_L=0$. Let L have as a generating function $H+u^TC$. If follows that $\dot{\alpha}(\dot{X}_{C_i},-)=0$, $i=1,\ldots,m$, or equivalently, $\alpha(\dot{X}_{C_i},-)=0$, $i=1,\ldots,m$.

Since $L_{X_H}\omega=0$ and $L_{\varphi_*X_H}\omega=0$, and therefore $\varphi^*(L_{\varphi_*X_H}\omega)=L_{X_H}\varphi^*\omega=0$, it follows that $L_{X_H}\alpha=0$. Hence $0=L_{X_H}(\alpha(X_{C_i},-))=\alpha(L_{X_H}(X_{C_i}))=\alpha(\dot{X}_{\{H,C_i\}},-)$ for $i=1,\ldots,m$. In the same way we can prove by induction that for every function $f\in\mathcal{F}$ (see Proposition 1.7): $\alpha(X_f,-)=0$. Because Σ is locally minimal, $\dim d\mathcal{F}(x)=2n$ for every $x\in M$ and so $\alpha=0$, or $\varphi^*\omega=\omega$.

As is common[1,2] we will concentrate in the sequel on infinitesimal symmetries, in which case the analog of Definitio 2.1 becomes

Definition 2.4 (Infinitesimal Hamiltonian symmetry): Let $\Sigma(M,T^*Y,L)$ be a Hamiltonian system. An infinitesimal symmetry is a pair (S,T), with S vector field on M and T vector field on T^*Y, such that for every $z\in L$, $(\dot{S}(z),T(z))\in T_zL$. An infinitesimal symmetry (S,T) is an infinitesimal Hamiltonian symmetry if $L_S\omega=0$ and $L_T\omega^e=0$.

Analogous to Theorem 2.3 we can prove that the vector field T is necessary fiber respecting, and that if $\Sigma(M,T^*Y,L)$ is locally mininmal and (S,T) is an infinitesimal symmetry with $L_T\omega^e=0$, then also $L_S\omega=0$. We also note that because $L_T\omega^e=0$ and T is fiber respecting the Hamiltonian function G^e corresponding to T $[\omega^e(T,-)=dG^e]$ has the form $G^e(y,u)=\sum_{i=1}^m u_iK_i(y)+V(y)$ with K_i and V smooth functions on Y.[5]

Now we will give a generalization of Noether's theorem in our framework. Recall the setting of Noether's theorem for a triplet (M,ω,H). A vector field S on M is an infinitesimal Hamiltonian symmetry if $L_S\omega=0$ and $S(H)=0$. A function $G: M\to\mathbb{R}$ is a conservation law for (M,ω,H) if $X_H(G)=0$. Now let S be an infinitesimal symmetry, then since $L_S\omega=0$, there exists (locally) a $G: M\to\mathbb{R}$ such that $S=X_G$. Since $X_G(H)=S(H)=0$ it follows that $X_H(G)=0$. So G is a conservation law. Conversely, if G is a conservation law then $L_{X_G}\omega=0$ and $X_G(H)=-X_H(G)=0$, so X_G is an infinitesimal symmetry. We first derive

Theorem 2.5[5]: Let $\Sigma(M,T^*Y,L)$ be a Hamiltonian system with generating function $H+\sum_{i=1}^m u_iC_i$. Let (S,T) be an infinitesimal Hamiltonian symmetry. Let $G: M\to\mathbb{R}$ and G^e: $T^*Y\to\mathbb{R}$ with $G^e(y,u):=\sum_{i=1}^m u_iK_i(y)+V(y)$ the (locally defined) functions such that $\omega(S,-)=dG$ and $\omega^e(T,-)=dG^e$. Then $\{H(x)+u^TC(x),G(x)\}=G^e(C(x),u)$, for every u or equivalently

$$\{H,G\}=V\circ C,$$

$$\{C_i,G\}=K_i\circ C,\quad i=1,\ldots,m$$

with $\{\ ,\ \}$ the Poisson bracket on M.

The pair (G,G^e) as above can be called a conservation law in our framework. The derivative of the function G (the conserved quantity) along trajectories of the system is a function G^e of the behavior on the boundary of the system. Therefore we have proved in Theorem 2.5 that if (S,T) is an infinitesimal Hamiltonian symmetry, then (G,G^e), with $S=X_G$ and $T=X_{G^e}$, is a conservation law. Conversely, it can be easily seen that if (G,G^e) is a conservation law, i.e., $G^e=\sum_{i=1}^m u_iK_i+V$ and Eqs. (2.1) are satisfied, then (X_G,X_{G^e}) is an infinitesimal Hamiltonian symmetry.

Example 2.2 (continued): The group of rotations around the e_1-axis generates an infinitesimal Hamiltonian symmetry S on $M=T^*\mathbb{R}^3$ for the system without external forces. The corresponding conservation law is the angular momentum G around the e_1-axis. For zero external force we obtain $dG/dt=0$, with $G:=\langle\dot{q}\times mq,e_1\rangle$. However, for a nonzero external force F we obtain

$$\frac{dG}{dt}=\langle q\times F,e_1\rangle=\langle y\times u,e_1\rangle,\quad u=F.$$

Now $G^e(y,u) := \langle y \times u, e_1 \rangle$ is a function on $T^*Y = T^*\mathbb{R}^3$, and (G,G^e) is our conservation law.

Using minimality we can sharpen Theorem 2.5 in the following way.

Proposition 2.6: Let $\Sigma(M,T^*Y,L)$ be a locally minimal Hamiltonian system with generating function $H + \Sigma_{i=1}^m u_i C_i$. Let (S_i,T_i), $i = 1,...,k$ be infinitesimal Hamiltonian symmetries with corresponding conservation laws (G_i,G_i^e), $i = 1,...,k$. Let $\mathcal{G}$ be the Poisson algebra on M generated by G_i, $i = 1,...,k$, and let $\mathcal{G}^e$ be the Poisson algebra on T^*Y generated by G_i^e, $i = 1,...,k$. Then the map α: $\mathcal{G}$(modulo constant functions)$\to\mathcal{G}^e$, defined by α: $G(x)\to\{H(x) + u^T C(x), G(x)\} =: G^e(C(x),u)$ is a Poisson algebra isomorphism.

Proof: Let $G_1,G_2 \in \mathcal{G}$ and G_1^e, $G_2^e \in \mathcal{G}^e$, with $G_1^e(y,u) = \Sigma_{j=1}^m u_j K_j^i(y) + V^i(y)$ such that $\{H + u^T C, G_i\} = G_i^e \circ (C,\mathrm{id})$, $i = 1,2$ with $(G_1^e \circ (C,\mathrm{id}))(x,u) = G_i^e(C(x),u)$. The Jacobi identity implies

$$
\begin{aligned}
&\{H + u^T C,\{G_1,G_2\}\} \\
&\quad = \{\{H + u^T C,G_1\},G_2\} - \{\{H + u^T C,G_2\},G_1\} \\
&\quad = \{G_1^e \circ (C,\mathrm{id}),G_2\} - \{G_2^e \circ (C,\mathrm{id}),G_1\} \\
&\quad = \sum_{i,j=1}^m u_i \frac{\partial K_i^1}{\partial y_j}\{C_j,G_2\} + \sum_{j=1}^m \frac{\partial V^1}{\partial y_j}\{C_j,G_2\} \\
&\qquad - \sum_{i,j=1}^m u_i \frac{\partial K_i^2}{\partial y_j}\{C_j,G_1\} - \sum_{j=1}^m \frac{\partial V^2}{\partial y_j}\{C_j,G_1\} \\
&\quad = \sum_{i=1}^m u_i \left(\sum_{j=1}^m \frac{\partial K_i^1}{\partial y_j} K_j^2 - \frac{\partial K_i^2}{\partial y_j} K_j^1 \right) \\
&\qquad + \sum_{j=1}^m \left(\frac{\partial V^1}{\partial y_j} K_j^2 - \frac{\partial V^2}{\partial y_j} K_j^1 \right) \\
&\quad = \left\{ \sum_{i=1}^m u_i K_i^1 + V^1, \sum_{i=1}^m u_i K_i^2 + V^2 \right\}_{T^*Y} \\
&\quad = \{G_1^e,G_2^e\}_{T^*Y},
\end{aligned}
$$

where $\{\ ,\ \}_{T^*Y}$ means Poisson bracket on T^*Y.

Therefore the map α is a Poisson algebra morphism. It is immediate that constant functions are mapped to zero. Suppose that a function $G \in \mathcal{G}$ satisfies $\{H + u^T C, G\} = 0$. Then $\{H,G\} = 0$ and $\{C_i,G\} = 0$, $i = 1,...,m$. Therefore for every $f \in \mathcal{F}$ (see Proposition 1.7, the algebra generated by C_i under taking Poisson brackets with H and C_i), $\{f,G\} = 0$. Since Σ is locally minimal, this implies $G = \text{const}$. So α is an isomorphism. □

Proposition 2.6 implies that for locally minimal systems our definition of a Hamiltonian symmetry really covers the ususal one for a triplet (M,ω,H). Indeed, suppose that $S = X_G$ is an infinitesimal Hamiltonian symmetry, so $\{H,G\} = 0$, which cannot be observed, i.e., $\{C_i,G\} = 0$, $1,...,m$. Then G is constant and therefore $S = 0$.

A maybe unsatisfying feature of Theorem 2.5 is that we obtain $\{H,G\} = V \circ C$, instead of $\{H,G\} = 0$ as in the case without external forces. We will now show how by adding a potential P, only depending on the observations, to the Hamiltonian H we can change $\{H,G\} = V \circ C$ into $\{H + P \circ C,G\} = 0$.

Theorem 2.7: Let $\Sigma(M,T^*Y,L)$ be a Hamiltonian system. Let (S_i,T_i), $i = 1,...,k$, be infinitesimal Hamiltonian symmetries such that $\pi_* T_i$, $i = 1,...,k$, are independent vector fields on Y (π is the projection of T^*Y on Y), which are therefore nowhere zero. Let (G_i,G_i^e) be the corresponding conservation laws. Suppose that $\{G_i^e,G_j^e\}_{T^*Y} = 0$, $i,j = 1,...,k$. Then we can (locally) construct a function P: $Y\to\mathbb{R}$ such that $\{H + P \circ C, G_i\} = 0$, $i = 1,...,k$.

Proof: Since $\{G_i^e,G_j^e\} = 0$, also $[T_i,T_j] = 0$. This implies $[\pi_* T_i, \pi_* T_j] = 0$, $i,j = 1,...,k$. Therefore we can take local coordinates $\{y_1,...,y_m\}$ for Y such that $\pi_* T_i = \partial/\partial y_i$, $i = 1,...,k$. Denote $v_i := G_i^e$, $i = 1,...,k$. Then we have independent functions $y_1,...,y_m$ and $v_1,...,v_k$, $k \leqslant m$, such that

$$
\begin{aligned}
&\{y_i,y_j\} = 0, \quad i,j = 1,...,m, \\
&\{v_i,v_j\} = 0, \quad i,j = 1,...,k, \\
&\{y_i,v_j\} = \delta_{ij}, \quad i = 1,...,m, \quad j = 1,...,k.
\end{aligned}
$$

Therefore (cf. Ref. 2 Darboux's theorem) we can construct a complementary set of independent functions $v_{k+1},...,v_m$ such that

$$
\begin{aligned}
&\{v_i,v_j\} = 0, \quad i = 1,...,m, \quad j = k+1,...,m, \\
&\{y_i,v_j\} = \delta_{ij}, \quad i = 1,...,m, \quad j = k+1,...,m,
\end{aligned}
$$

or equivalently, $\{y_1,...,y_m,v_1,...,v_m\}$ are symplectic coordinates. The submanifold of T^*Y given by $v_1 = \cdots = v_m = 0$ is Lagrangian and has therefore (locally) a generating function $P: Y\to\mathbb{R}$. Since $\Sigma(M,T^*Y,L)$ has generating function $H + \Sigma_{i=1}^m u_i C_i$ in the old coordinates (y,u), it has generating function $H + P \circ C + \Sigma_{i=1}^m v_i C_i$ in the new coordinates (y,v). Because $G_i^e = v_i$ $i = 1,...,k$ it follows that $\{H + P \circ C, G_i\} = 0$, $i = 1,...,k$.

Remark: Notice that when we write $C = (C_1,...,C_m)$ corresponding to the y-coordinates constructed above we obtain

$$
\{C_j,G_j\} = \delta_{ij}, \quad i = 1,...,k, \quad j = 1,...,m.
$$

If $\dim T^*Y = \dim M = 2n$, and if we have n symmetries (S_i,T_i) satisfying the conditions of Theorem 2.7, we can construct (locally) a function $P: Y\to\mathbb{R}$ such that $\tilde{H} = H + P \circ C$ satisfies $\{\tilde{H}, G_i\} = 0$, $i = 1,...,n$. Moreover, Proposition 2.6 implies that since $\{G_i^e,G_j^e\}_{T^*Y} = 0$, also $\{G_i,G_j\}_M = c_{ij}$, with c_{ij} constants. Hence we are very near to the case of complete integrability (cf Refs. 1 and 2), for which we need n symmetries G_i satisfying $\{G_i,G_j\} = 0$. Therefore we cannot construct action-angle coordinates,[1,2] but using the remark above and assuming that $\ker dC$ is a Lagrangian submanifold of M (think for instance of observation of the positions), it follows from a result in[18] that in this case we can take symplectic coordinates $\{q_1,...,q_n,p_1,...,p_n\}$ for M such that

$$
\begin{aligned}
&C_i = q_i, \quad i = 1,...,n, \\
&G_i = p_i - \frac{1}{2}\sum_{j=1}^n c_{ij} q_j.
\end{aligned}
$$

Since $0 = \{\tilde{H}, G_i\} = \{\tilde{H},p_i\} - \frac{1}{2}\Sigma_{j=1}^n c_{ij}\{\tilde{H},q_j\}$ the system in these coordinates, after adding the potential P, is given by

$$
\dot{q}_i = \frac{\partial \tilde{H}}{\partial p_i}, \quad i = 1,...,n,
$$

$$\dot{p}_i = -\frac{\partial \tilde{H}}{\partial q_i} + F_i = -\frac{1}{2}\sum_{j=1}^{n} c_{ij}\dot{q}_j + F_i,$$

$y_i = q_i$ (F_i the external force).

For degenerate Hamiltonian systems we obtain the following definition of a Hamiltonian symmetry.

Definition 2.8: Let $\Sigma(M,T^*Y,L = L' \cap (TM \times P))$ be a degenerate Hamiltonian system. A pair of vector fields (S,T) is an infinitesimal Hamiltonian symmetry for $\Sigma(M,T^*Y,L)$ if (S,T) is an infinitesimal Hamiltonian symmetry for $\Sigma(M,T^*Y,L')$ and the vector field T is tangent to $P \subset T^*Y$, i.e., $T(z) \in T_z P$, for every $z \in P$.

Hence we have formalized the idea that the symmetry should only work on the possible external behavior. In local coordinates we obtain

Proposition 2.9: Let (S,T) be an infinitesimal Hamiltonian symmetry for the degenerate Hamiltonian system $\Sigma(M,T^*Y,L = L' \cap (TM \times P))$. Then we can find coordinates $y_1,...,y_m$ for Y such that $P = \text{span}\{dy_1,...,dy_k\}$, $k \leqslant m$, and such that G^e [with $\omega^e(T,-) = dG^e$] has the form

$$G^e(y,u) = \sum_{i=1}^{k} u_i K_i(y_1,...,y_k) + V(y_1,...,y_k) + \sum_{i=k+1}^{m} u_i K_i(y_1,...,y_m).$$

Proof: We know that $G^e = \Sigma_{i=1}^m u_i K_i(y_1,...,y_m) + V(y_1,...,y_m)$. Since T is tangent to P we must have

$$\Sigma_{i=1}^m u_i \frac{\partial K_i}{\partial y_j} + \frac{\partial V}{\partial y_j} = 0 \quad \text{for } u_{k+1} = \cdots = u_m = 0, \quad j = k+1,...,m,$$

or equivalently,

$$\frac{\partial K_i}{\partial y_j} = 0 \quad i = 1,...,k, \quad j = k+1,...,m$$

and

$$\frac{\partial V}{\partial y_j} = 0, \quad j = k+1,...,m. \qquad \square$$

We see that the vector field T in this case projects to a vector field on $T^*\tilde{Y}$, where $\tilde{Y}$ has coordinates $y_1,...,y_k$. It can also be seen that the additional potential P in Theorem 2.7 in this case only has to depend on $y_1,...,y_k$.

IV. TIME-REVERSIBLE HAMILTONIAN SYSTEMS

Time reversibility of a system is a widely used but many times rather vaguely defined notion. In our framework it can be defined in the following way.[19] We say that the external behavior of a system $\Sigma(M,T^*Y,L)$ is time reversible if, when $(\bar{y}(t),\bar{u}(t))$, $t \in \mathbb{R}$ belongs to the external behavior of the system; also the time-reversed signal $(\bar{y}(-t),\bar{u}(-t))$, $t \in \mathbb{R}$ is a feasible external behavior.

Let now $\Sigma(M,T^*Y,L)$ be a Hamiltonian system, locally given by

$$\begin{cases} \dot{x} = X_H(x) + \sum_{i=1}^{m} u_i X_{C_i}(x), & x \in M, \\ y_i = C_i(x), & i = 1,...,m. \end{cases} \tag{3.1}$$

If the external behavior of this system is time reversible then the Hamiltonian system

$$\dot{x} = -X_H(x) - \sum_{i=1}^{m} u_i X_{C_i}(x), \quad x \in M,$$
$$y_i = C_i(x), \quad i = 1,...,m \tag{3.2}$$

has the same external behavior.

If the system Σ is (locally) minimal it seems then reasonable to ask that there exists a diffeomorphism $\varphi: M \to M$ which carries Eqs. (3.1) over in (3.2). In the case of a linear Hamiltonian system this can actually be proven.[7]

Definition 3.1 (Time reversibility): Let $\Sigma(M,T^*Y,L)$ be a locally minimal Hamiltonian system, locally given by (3.1). Then Σ is time reversible if there exists a diffeomorphism $\varphi: M \to M$ such that

$$\varphi_* X_H = -X_H, \quad \varphi_* X_{C_i} = -X_{C_i},$$
$$\text{and } \varphi^* C_i = C_i, \quad i = 1,...,m.$$

In this case we will call φ a time-reversing symmetry.

Using local minimality we can prove two important properties.

Theorem 3.2: Let φ be a time-reversing symmetry for a locally minimal Hamiltonian system $\Sigma(M,T^*Y,L)$. Then

(i) φ is an involution, i.e., $\varphi^2 = \text{id}$.

(ii) φ is an anti-symplectomorphism, i.e., $\varphi^*\omega = -\omega$.

Proof: Denote $A := X_H$ and $B_i := X_{C_i}$.

(i) We will prove that every function f on M, generated by taking (repeated) Poisson brackets of C_i, $i = 1,...,m$, and H and C_i satisfies $\varphi^* f = \pm f$. By assumption $\varphi^* C_i = C_i$, and for instance

$$\varphi^*\{H,C_i\} = \varphi^*(L_A C_i) = L_{\varphi_* A} = 1_A \varphi^* C_i = L_{-A} C_i = -\{H,C_i\}.$$

By induction it follows that $\varphi^* f = \pm f$, for every f constructed as above. Hence $(\varphi^2)^* f = f$ for every $f \in \mathcal{F}$ (see Proposition 1.7). Since Σ is locally minimal, this implies $\varphi^2 = \text{id}$, if we assume that φ^2 has at least one fixed point.

(ii) Define $\alpha = \varphi^*\omega + \omega$. Since $\varphi^*\omega(B_i,-) = \omega(\varphi_* B_i, \varphi_* -) = -\omega(B_i, \varphi_* -) = -dC_i(\varphi_* -) = -\varphi_* dC_i = -d\varphi^* C_i = -dC_i = -\omega(B_i,-)$, we obtain that $\alpha(B_i,-) = 0$. $i = 1,...,m$.

Furthermore

$$L_A \alpha = L_A \omega + L_A \varphi^* \omega = \varphi^*(L_{\varphi_* A}\omega) = \varphi^*(L_{-A}\omega) = 0$$

and

$$0 = L_A(\alpha(B_i,-)) = \alpha(L_A B_i,-) = \alpha(X_{\{H,C_i\}},-),$$

and therefore by induction

$$\alpha(X_f,-) = 0$$

for every f constructed as above. Local minimality implies $\alpha = 0$, or $\varphi^*\omega = -\omega$.

Remark: Note that $\varphi_* X_H = X_H$ together with $\varphi^*\omega = -\omega$ implies $\varphi^* H = H + \text{const}$.

Maps $\varphi: M \to M$ with exactly these properties (i) and (ii) have been studied in Ref. 20 (see also the references cited there). It can be proved[20] that the points $p \in M$ such that $\varphi(p) = p$ form a Lagrangian submanifold Q of M. Furthermore we can find a neighborhood U of Q and local coordinates $(q_1,...,q_n,p_1,...,p_n)$ for U such that Q is the submanifold given by the equations $p_1 = \cdots = p_n = 0$, $\omega = \Sigma_{i=1}^n dq_i \wedge dp_i$ and φ is given by

$$\varphi:(q_1,\ldots,q_n,p_1,\ldots,p_n)\to(q_1,\ldots,q_n,-p_1,\ldots,-p_n).$$

Consequently if H and C_k, $k=1,\ldots,m$ are at most quadratic in the p_j-coordinates, it follows from $\varphi^*H = H + \text{const}$ and $\varphi^*C_k = C_k$ that H and C_k have the form

$$H(q,p) = \sum_{i,j=1}^{n} g_{ij}(q)p_ip_j + V(q),$$

$$C_k(q,p) = \sum_{i,j=1}^{n} h^k_{ij}(q)p_ip_j + W^k(q), \quad k=1,\ldots,m,$$

with $g_{ji} = g_{ij}$ and $h^k_{ij} = h^k_{ji}$. Especially the form of H is very appealing; it denotes a Hamiltonian consisting of the sum of a potential $V(q)$ and a kinetic energy given by a "Riemannian metric."

ACKNOWLEDGMENTS

I would like to thank Jan C. Willems for inspiring conversations, and Henk Nijmeijer for a careful reading of the manuscript.

[1]R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Benjamin/Cummings, Reading, MA, 1978).
[2]V. I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer, New York, 1978) (translation of the 1974 Russian edition).
[3]A. J. van der Schaft, "Hamiltonian Dynamics with External Forces and Observations," Math. Systems Theory **15**, 145–168 (1982).
[4]A. J. van der Schaft, "Observability and Controllability for Smooth Nonlinear Systems," SIAM J. Control Optimization **20**, 338–354 (1982).
[5]A. J. van der Schaft, "Symmetries and Conservation Laws for Hamiltonian Systems with Inputs and Outputs: A Generalization of Noether's Theorem," Systems Control Lett. **1**, 108-115 (1981).
[6]A. J. van der Schaft, "Controllability and Observability for Affine Nonlinear Hamiltonian Systems," IEEE Trans. Automatic Control **27**, 490–492 (1982).
[7]A. J. van der Schaft, "Time reversible Hamiltonian Systems," Systems Control Lett. **1**, 295–300 (1982).
[8]R. W. Brockett, "Control Theory and Analytical Mechanics" in *Geometric Theory*, edited by C. Martin and R. Hermann, Vol. VII of *Lie Groups: History, Frontiers and Applications* (Math. Sci. Press, Brookline, MA, 1977), pp. 1–46.
[9]F. Takens, "Variational and Conservative Systems," Report ZW-7603, Mathematics Institute, Groningen, 1976.
[10]J. C. Willems, "System Theoretic Models for the Analysis of Physical Systems," Richerche di Automatica **10**, 71–106 (1979).
[11]J. C. Willems and A. J. van der Schaft, "Modeling of Dynamical Systems using External and Internal Variables with Applications to Hamiltonian Systems," to appear in the *Proceedings of the International Seminar on Mathematical Theory of Dynamical Systems and Microphysics* (Academic, New York,1982).
[12]R. Hermann, *Geometric Theory of Non-linear Differential Equations, Bäcklund Transformations and Solitons, Part B, Interdisciplinary Mathematics*, Vol. XIV (Math. Sci. Press, Brookline, MA, 1976).
[13]W. M. Tulczyjew, "Hamiltonian Systems, Lagrangian Systems and the Legendre Transformation," Symposia Mathematica, **14**, 247–258 (1974).
[14]K. Yano and S. Ishihara, *Tangent and Cotangent Bundles* (Dekker, New York, 1973).
[15]J. Kijowski and W. M. Tulczyjew, "A Symplectic Framework for Field Theories," *Lecture Notes in Physics*, Vol. 107 (Springer, Berlin, 1979).
[16]M. R. Menzio and W. M. Tulczyjew, "Infinitesimal symplectic relations and generalized Hamiltonian dynamics," Ann. Inst. Henri Poincaré **28**, 349–367 (1978).
[17]J. Basto Goncalves, "Equivalence of gradient systems," Control Theory Centre Report No. 84, University of Warwick.
[18]J. Roels and A. Weinstein, "Functions whose Poisson Brackets are Constants," J. Math. Phys. **12**, 1482–1486 (1971).
[19]J. C. Willems, "Time reversibility in Deterministic and Stochastic Dynamical Systems" in *Recent Developments in Variable Structure systems, Economics and Biology*, edited by R. R. Mohler and A. Ruberti, *Springer Lecture Notes in Economics and Mathematical Systems*, Vol. 162 (Springer, Berlin, 1978).
[20]K. R. Meyer, "Hamiltonian Systems with a Discrete Symmetry," J. Diff. Equations **41**, 228–238 (1981).

Normal form and representation theory

Richard Cushman
Mathematisch Instituut, Rijksuniversiteit Utrecht, 3508 TA Utrecht, The Netherlands

André Deprit and Richard Mosak[a)]
National Bureau of Standards, Washington, D.C. 20234

(Received 19 July 1982; accepted for publication 28 January 1983)

Representation theory of Lie algebras is called upon to develop a procedure for normalizing a dynamical system with two degrees of freedom in the neighborhood of an equilibrium when the Hamiltonian $H(x,y,X,Y)$ in the coordinates (x,y) and their conjugate momenta (X,Y) is of the type $H=(X^2+Y^2)/2+V(x,y,X,Y)$, the potential energy V being a sum of homogeneous polynomials in the phase variables of degree strictly greater than 2. The fact that the resulting potential V' is a polynomial in the new coordinates (x',y') and the angular momentum $G'=x'Y'-y'X'$ implies that the normalization is a rotation in the configuration space from a fixed frame to an ideal frame. The technique is intended for normalizing an Hamiltonian in equilibrium at the origin when the Lie derivative associated with the quadratic part is not semisimple, e.g., the planar restricted problem of three bodies at the equilateral equilibrium L_4 when the basic frequencies are equal (Routh's singular case).

PACS numbers: 03.20. + i, 02.20. + b

1. INTRODUCTION

The literature about normalization deals mainly with semisimple systems in equilibrium at the origin. The Hamiltonian being a formal series

$$H\equiv H(x,y,X,Y)=\sum_{n\geq 0}\frac{1}{n!}H_n \tag{1}$$

whose terms H_n are homogeneous polynomials of degree $n+2$ in the coordinates (x,y) and their conjugate momenta (X,Y), it is generally assumed that the dominant term H_0 is a quadratic form reducible to the type

$$J=\tfrac{1}{2}(X^2+\epsilon_1\omega_1^2x^2)+\tfrac{1}{2}\epsilon(Y^2+\epsilon_2\omega_2^2y^2), \tag{2}$$

in which the frequencies ω_1 and ω_2 are real and >0, the factors ϵ, ϵ_1, and ϵ_2 being either $+1$ or -1. Such systems are called *semisimple* because their dominant term leads to a linear Hamiltonian vector field that is semisimple. Let

$$\mathbf{L}_J:F\to\mathbf{L}_J(F)=(F;J)$$

be the Lie derivative associated with the Hamiltonian linear vector field derived from J. For any $n\geq 2$, the restriction of the differential operator

$$\mathbf{L}_J=\left(X\frac{\partial}{\partial x}+\epsilon_1\omega_1^2x\frac{\partial}{\partial X}\right)+\epsilon\left(Y\frac{\partial}{\partial y}+\epsilon_2\omega_2^2y\frac{\partial}{\partial Y}\right)$$

to the vector space $\mathbf{P}_n$ of homogeneous polynomials of degree n in (x,y,X,Y) is an endomorphism of $\mathbf{P}_n$ that is semisimple; hence, with respect to $\mathbf{L}_J$, $\mathbf{P}_n$ may be decomposed into the direct sum

$$\mathbf{P}_n=\operatorname{Im}\mathbf{L}_J\oplus\operatorname{Ker}\mathbf{L}_J. \tag{3}$$

The concept of normalization for semisimple systems in equilibrium at the origin must be credited to Whittaker,[1] who created it first by adapting a method proposed by Delaunay[2] for eliminating periodic terms from the main problem of lunar theory. Later, Whittaker[3] carried out the normalization as a canonical transformation $(x,y,X,Y)\to(x',y',X',Y')$ defined through the implicit equations

$$X=\frac{\partial S}{\partial x},\quad Y=\frac{\partial S}{\partial y},\quad x'=\frac{\partial S}{\partial X'},\quad y'=\frac{\partial S}{\partial Y'}$$

derived from a generating function

$$S\equiv S(x,y,X',Y')=\sum_{n\geq 0}S_n,$$

where $S_0=xX'+yY'$ and, for any $n\geqslant 1$, the term S_n is a homogeneous polynomial of degree $n+2$ in (x,y,X',Y'). Although devised by Poincaré[4] as one of his "méthodes nouvelles," the procedure is referred to in some quarters of celestial mechanics as von Zeipel's method. Nowadays[5] the normalization is executed as a Lie transformation[6] to convert the formal power series (1) into a formal power series

$$H'\equiv H'(x',y',X',Y')=\sum_{n\geq 0}\frac{1}{n!}H'_n \tag{4}$$

such that (i) $H'_0=J(x',y',X',Y')$ and (ii), for each $n>0$, H'_n belongs to the kernel of $\mathbf{L}_J$. In action and angle variables (ϕ,ψ,Φ,Ψ), a polynomial in (x,y,X,Y) becomes a trigonometric sum in (ϕ,ψ), its component in $\operatorname{Im}\mathbf{L}_J$ consists of short-period terms whereas its component in $\operatorname{Ker}\mathbf{L}_J$ groups the terms which are either secular or of long period. The normalization is justified in this framework as a technique for removing short-period effects from the perturbation. The requirement that $(H';J)=(H'_0;H')=0$ implies at once that the dominant term H'_0 as a function of the normalizing variables is a (formal) integral of the system, and hence calls for a reduction of the Hamiltonian system from two to one degree of freedom.

Whittaker[7] in his theory of integration by series, and most textbooks following him, considers exclusively semisimple Hamiltonians in equilibrium. It must be realized[8] though that a nondegenerate quadratic form in (u,v,U,V) is

[a)] On leave from the University of Rochester, Rochester, N.Y. 14627.

 0022-2488/83/082102-16$02.50

reducible by a real symplectic linear transformation $(u, v, U, V) \to (x, y, X, Y)$ either to type (2) or to the type

$$J = \omega(xY - yX) - \tfrac{1}{2}\epsilon\omega^2(x^2 + y^2). \tag{5}$$

In the second case, the Lie derivative

$$\mathbf{L}_J = \omega\left(X\frac{\partial}{\partial X} - y\frac{\partial}{\partial Y}\right) + \omega\left[(Y - \epsilon\omega x)\frac{\partial}{\partial X} - (X + \epsilon\omega y)\frac{\partial}{\partial Y}\right] \tag{6}$$

will be decomposed into the sum

$$\mathbf{L}_J = \omega\mathbf{L}_G + \epsilon\omega^2\mathbf{L}_D, \tag{7}$$

where the differential operators

$$\mathbf{L}_G = x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x} + X\frac{\partial}{\partial Y} - Y\frac{\partial}{\partial X}, \tag{8}$$

$$\mathbf{L}_D = x\frac{\partial}{\partial X} + y\frac{\partial}{\partial Y} \tag{9}$$

are the Lie derivatives corresponding to the Hamiltonian linear vector fields derived from the functions

$$G = xY - yX \quad \text{and} \quad D = -\tfrac{1}{2}(x^2 + y^2). \tag{10}$$

The sum (7) realizes a Jordan decomposition of the endomorphism $\mathbf{L}_J: \mathbf{P}_n \to \mathbf{P}_n$, that is, $\mathbf{L}_G$ is semisimple, $\mathbf{L}_D$ is nilpotent, and these operators commute since

$$\mathbf{L}_D\mathbf{L}_G = \mathbf{L}_G\mathbf{L}_D = x\frac{\partial}{\partial Y} - y\frac{\partial}{\partial X}. \tag{11}$$

For any $n \geqslant 0$, the vector space $\mathbf{P}_n$ turns out to be the direct sum

$$\mathbf{P}_n = \operatorname{Im}\mathbf{L}_G \oplus \operatorname{Ker}\mathbf{L}_G,$$

and, on account of the commutativity relation (11), the restriction of $\mathbf{L}_D$ to the kernel of $\mathbf{L}_G$ is an endomorphism of $\operatorname{Ker}\mathbf{L}_G$. Hence the normalization of a nonsemisimple system in equilibrium at the origin could proceed in two steps. First a Lie transformation $\phi: (x, y, X, Y) \to (x', y', X', Y')$ normalizes the system with respect to the semisimple component $\mathbf{L}_G$, thereby changing (1) into a power series (4) in the kernel of $\mathbf{L}_G$. In the new phase variables, the angular momentum $G' = x'Y' - y'X'$ is a (formal) integral; hence the term $\omega G'$ may be omitted from the dominant part in the transformed Hamiltonian, and then reduced to $J' = \epsilon\omega^2 D'$. Rather than analyzing the partially normalized problem as a system reduced to one degree of freedom by means of the integral of angular momentum, van der Meer[9] proposes that the normalization be continued with a Lie tranformation ψ: $(x', y', X', Y') \to (x'', y'', X'', Y'')$ to convert (4) into a formal power series confined to a remarkable vector subspace of $\operatorname{Ker}\mathbf{L}_G$. Indeed, relative to the Lie derivative

$$\mathbf{L}_K = X\frac{\partial}{\partial x} + Y\frac{\partial}{\partial y} \tag{12}$$

associated with the Hamiltonian vector field derived from

$$K = \tfrac{1}{2}(X^2 + Y^2), \tag{13}$$

the kernel of $\mathbf{L}_G$ may be decomposed into the direct sum

$$\operatorname{Ker}\mathbf{L}_G = (\operatorname{Im}\mathbf{L}_D \cap \operatorname{Ker}\mathbf{L}_G) \oplus (\operatorname{Ker}\mathbf{L}_K \cap \operatorname{Ker}\mathbf{L}_G).$$

This makes it possible for the transformation ψ to convert (4) into a (formal) power series

$$H'' \equiv H''(x'', y'', X'', Y'') = \sum_{n \geqslant 0}\frac{1}{n!}H''_n$$

that belongs to $\operatorname{Ker}\mathbf{L}_G \cap \operatorname{Ker}\mathbf{L}_K$. As a combined effect of the transformations ϕ and ψ, the kinetic energy $H''_0 = \epsilon\omega^2 D''$ and the angular momentum G'' in the third set of variables come out formally as integrals. Because they admit two independent integrals in involution, *nonsemisimple systems with two degrees of freedom in equilibrium at the origin are (formally) integrable.*

van der Meer did not concern himself with developing an algorithm for generating the second normalization. As for ourselves, while engaged in designing such a procedure, we noticed that our techniques apply to a class of systems wider than the nilpotent part of a nonsemisimple system in equilibrium at the origin. In fact exchanging the coordinates and the momenta transposes the Hamiltonian of the latter system into one whose dominant term H'_0 is K' instead of D'. Such Hamiltonians represent motions of a particle subject to weak perturbations in the neighborhood of the origin. In that general context, the main contribution of this article is summed up in the following:

Theorem: Given a Hamiltonian system with two degrees of freedom represented by the formal series

$$H \equiv H(x, y, X, Y) = \tfrac{1}{2}(X^2 + Y^2) + \sum_{n \geqslant 1}\frac{1}{n!}H_n(x, y, X, Y), \tag{14}$$

where, for $n \geqslant 1$, the perturbation H_n is a homogeneous polynomial of degree $n + 2$, one may build formally a Lie transformation $(x, y, X, Y) \to (x', y', X', Y')$ to convert H into a formal power series

$$H' \equiv H'(x', y', X', Y') = \tfrac{1}{2}(X'^2 + Y'^2) + \sum_{n \geqslant 1}\frac{1}{n!}H'_n(x', y', X', Y'),$$

where, for each $n \geqslant 1$, H'_n is in the kernel of $\mathbf{L}_D$. More precisely,

$$H'_n = \sum_{\alpha + \beta + 2\gamma = n + 2} h_{\alpha,\beta,\gamma}x'^{\alpha}y'^{\beta}G'^{\gamma}.$$

The physical meaning of the normalization is exposed below in the corollary to the theorem. For the resulting potential

$$V' \equiv V'(x', y', G') = \sum_{n \geqslant 1}\frac{1}{n!}H'_n(x', y', G')$$

of the forces acting on the particle in the neighborhood of the origin, let the differential be written as the 1-form

$$dV' = \partial_1 V'dx' + \partial_2 V'dy' + \partial_3 V'dG'.$$

Also, let (C) be the original Cartesian frame of reference in the configuration plane (x, y) with $\mathbf{i}$ and $\mathbf{j}$ standing for the orthonormal unit vectors in the directions of the reference axes. Finally let (I) designate the moving frame obtained by rotating (C) about the origin at the adjusted angular velocity $-\partial_3 V'$, $\mathbf{i}'$ and $\mathbf{j}'$ denoting the orthonormal unit vectors in the directions of the coordinate axes. In these notations

Corollary: The particle's position $\mathbf{x}$, its velocity $\mathbf{v}$, and its acceleration $\mathbf{a}$ relative to the frame (C) are such that

$$\mathbf{x} = x\mathbf{i} + y\mathbf{j} = x'\mathbf{i}' + y'\mathbf{j}', \tag{15a}$$

$$\mathbf{v} = X\mathbf{i} + Y\mathbf{j} = X'\mathbf{i}' = Y'\mathbf{j}', \tag{15b}$$

$$\mathbf{a} = -\partial_1 V'\mathbf{i}' - \partial_2 V'\mathbf{j}'. \tag{15c}$$

Thanks to the normalization, the motion in the frame (C) is decomposed into a rotation about the normal to the plane at the variable rate $-\partial_3 V'(x', y', G')$, and a motion with respect to the moving frame (I). In the latter frame, whereas one would have expected Coriolis forces and centrifugal repulsion to compensate for the slow rotation, one finds that the forces are reduced to the gradient of the force function $U' = -V$. Thus the moving frame (I) constitutes what Hansen calls an *ideal* frame, and the normalization may be regarded as a procedure to extract from Hamiltonian (14) the instantaneous rate at which, along each particular orbit, the frame (C) should be set in rotation so that it becomes the ideal frame proper to that particular orbit.

Symmetry Lie algebras provide a natural framework in which to consider the normalization of Hamiltonian systems. A case in point is the class of semisimple systems in equilibrium which admit a 1:1 resonance. Credit for having discovered there the relationship between symmetry Lie algebras and normalization goes to Kummer.[10] As matter of fact, an algorithm can be set up to produce immediately the reduced Hamiltonian as a function over the Lie algebra su(2) spanned by the symmetry generators.[11] In that light, normalization of semisimple systems in 1:1 resonance is but an application of the reduction theorem.[12] A similar situation occurs for perturbed two-dimensional Keplerian problems where the Delaunay normalization[13] builds the reduced Hamiltonian as a function over the Lie algebra so(3). The present article shows that, for systems of type (14), the connection between symmetry Lie algebras and normalization is equally decisive, but of a different nature. For the symmetry Lie algebra is here solvable—and not semisimple. However, the Lie derivative $\mathbf{L}_K$ may be embedded in a simple Lie algebra, namely sl(2, **R**), whose representation specifies the requirements imposed by the normalization.

2. SYMPLECTIC SYMMETRIES

The task of developing a normalization algorithm for nilpotent systems of type (14) when K is given by (13) begins with studying the linear infinitesimally symplectic symmetries for the Hamiltonian K. The Lie algebra sp(4, **R**) of all infinitesimally symplectic linear maps $(x, y, X, Y) \to (x', y', X', Y')$ is isomorphic to the Lie algebra of quadratic Hamiltonians under the Poisson bracket[14]: To the infinitesimally symplectic matrix w in sp(4, **R**) corresponds the quadratic form

$$W(x, y, X, Y) = -(Jw\mathbf{v}, \mathbf{v}),$$

where

$$J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{v} = (x, y, X, Y),$$

in which case W is said to generate the infinitesimally symplectic linear map w. The latter is called a (linear infinitesimally symplectic) symmetry of K if there is a scalar ν such that $[w, k] = \nu k$ if k is the infinitesimally symplectic linear matrix generated by K, or equivalently

$$(W; K) = \nu K.$$

The definition adopted here for a symmetry is borrowed from Cartan[15]; it allows for a reparametrization of the integral curves of the vector field derived from K by the symmetries generated from W.

A quick evaluation shows that the infinitesimally linear symplectic symmetries form a five-dimensional solvable Lie algebra, a basis of which is generated by the sum

$$W = \epsilon_1 W_1 + \epsilon_2 W_2 + \epsilon_3 W_3 + \epsilon_4 W_4 + \epsilon_5 W_5,$$

whose terms are $W_1 = K$, $W_3 = G$,

$$W_2 = S = xX + yY, \quad W_4 = \tfrac{1}{2}(X^2 - Y^2), \quad W_5 = XY.$$

In G, one recognizes the generator of the rotations about the origin. From the infinitesimal symmetry s defined by the equations

$$x' = x, \quad y' = y,$$

$$X' = -X, \quad Y' = -Y,$$

derived from S, the exponential mapping $e^{\epsilon s}$ produces the symplectic symmetry

$$x' = e^{\epsilon}X, \quad y' = e^{\epsilon}Y, \quad X' = e^{-\epsilon}X, \quad Y' = e^{-\epsilon}Y$$

in finite form. It corresponds to the similarity due to the homogeneity in dimensions: Multiplying the coordinates by a constant λ $(= e^{\epsilon})$ and dividing the velocities by the same constant requires, for preserving the dimensional homogeneity, that the time itself be multiplied by λ^2. This rule of similarity can also be verified by checking that the symplectic symmetry changes Cartan's form

$$\omega = X\,dx + Y\,dy - K\,dt$$

into the 1-form

$$\omega = X'\,dx' + Y'\,dy' - \tfrac{1}{2}\lambda^2(X'^2 + Y'^2)\,dt,$$

thus suggesting that the time t be replaced by the independent variable $t' = \lambda^2 t$.

The symmetry generators K, G, W_4, and W_5 are evidently integrals for the free particle. The function K and G are respectively the particle's energy and its angular momentum. The integrals W_4 and W_5 restate the principle of conservation of linear momentum since they combine to yield that $X + Y$ and $X - Y$; hence the components X and Y of the velocity are integrals. The generator S itself is not an integral of the free particle. However, the symmetry condition $(S; K) = 2K$ gives rise to the differential relation $\dot{S} = (S; K) = 2K$, and thus to the integral $S - 2Kt = c_1$, from which, by yet another quadrature, is derived the so-called Jacobi integral

$$D + St - Kt^2 = c_0.$$

The infinitesimal symmetries w_i derived from the generators W_i $(1 \leqslant i \leqslant 5)$ span a subalgebra in the linear Lie algebra sp(4, **R**) of 4×4 symplectic matrices. The family is completed into a basis w_i $(1 \leqslant i \leqslant 10)$ of sp(4, **R**). Table I lists the generator and the nonzero elements $w_{i,j,k}$ (in row j and column k) for each symplectic matrix w_i in the basis.

For the basis of sp(4, **R**) established in Table I, one computes easily (at least on a home computer) the commutators

TABLE I. A basis in the algebra sp(4, **R**).

	Generator	Nonzero elements	Type n = nilpotent s = semisimple
w_1	$\frac{1}{2}(X^2+Y^2)$	$w_{1,1,3}=w_{1,2,4}=1$	n
w_2	$xX+yY$	$w_{2,1,1}=w_{2,2,2}=-w_{2,3,3}=-w_{2,4,4}=1$	s
w_3	$xY-yX$	$w_{3,2,1}=w_{3,4,3}=-w_{3,1,2}=-w_{3,3,4}=1$	s
w_4	$\frac{1}{2}(X^2-Y^2)$	$w_{4,1,3}=-w_{4,2,4}=1$	n
w_5	XY	$w_{5,1,4}=w_{5,2,3}=1$	n
w_6	$-\frac{1}{2}(x^2+y^2)$	$w_{6,3,1}=w_{6,4,2}=1$	n
w_7	$xX-yY$	$w_{7,1,1}=-w_{7,2,2}=-w_{7,3,3}=w_{7,4,4}=1$	s
w_8	$xY+yX$	$w_{8,1,2}=w_{8,2,1}=-w_{8,3,4}=-w_{8,4,3}=1$	s
w_9	$-xy$	$w_{9,3,2}=w_{9,4,1}=1$	n
w_{10}	$-\frac{1}{2}(x^2-y^2)$	$w_{10,3,1}=-w_{10,4,2}=1$	n

$[w_i, w_j]=w_iw_j-w_jw_i$ $(1\leqslant i,j\leqslant 10)$, the results of which have been entered in the respective ith row and jth column of Table II.

Basic to the normalization are the following facts which can be read immediately from Table II of the commutators.

(a) Each element a of a Lie algebra A over a field F determines an endomorphism ad $a: b\to[a,b]: A\to A$ of the F-vector space. The element a is said to be ad-nilpotent if $(\text{ad } a)^m=0$ for some integer $m>0$. A quick calculation using Table II shows that $(\text{ad } w_1)^3=0$, or that the symmetry w_1 is ad-nilpotent in the linear algebra sp(4, **R**).

(b) In a finite-dimensional semisimple Lie algebra A over the real or the complex field, every ad-nilpotent element may be embedded in a subalgebra B of A isomorphic to the *special* linear Lie algebra sl(2, **R**) of 2×2 matrices with zero trace.[16] More precisely, there exists a pair (b,c) of elements in A such that

$$[a,b]=2b,\quad [a,c]=-2c,\quad [b,c]=a.$$

In application of the embedding theorem to the ad-nilpotent matrix w_1 in sp(4, **R**), one reads from Table II that

$$[w_2,w_1]=2w_1,\quad [w_2,w_6]=-2w_6,$$
$$[w_1,w_6]=w_2. \tag{16}$$

(c) Each matrix w_i in Table I is, by construction, the Hamiltonian linear vector field derived from the corresponding generator W_i. But, given two vector fields α and β, the Lie derivative of their commutator $[\alpha,\beta]$ is usually[17] defined as the differential operator

$$\mathbf{L}_{[\alpha\beta]}=[\mathbf{L}_\alpha,\mathbf{L}_\beta]=\mathbf{L}_\beta\mathbf{L}_\alpha-\mathbf{L}_\alpha\mathbf{L}_\beta.$$

Hence the embedding relations (16) transposed in terms of Lie derivatives yield

Lemma 1:

$$[\mathbf{L}_S,\mathbf{L}_K]=2\mathbf{L}_K,\ [\mathbf{L}_S,\mathbf{L}_D]=-2\mathbf{L}_D,\ [\mathbf{L}_K,\mathbf{L}_D]=\mathbf{L}_S.$$

The lemma could have been obtained by evaluating long-hand the commutators of these Lie derivatives from their expressions as differential operators, or by evaluating the Poisson brackets $(W_i; W_j)$ since

$$[\mathbf{L}_{W_i},\mathbf{L}_{W_j}]=\mathbf{L}_{(W_i;\,W_j)}=\mathbf{L}_{[w_j,\,w_i]}.$$

While an algebraist interprets the results in Lemma 1 as saying that $\mathbf{L}_K$ and $\mathbf{L}_D$ are eigenvectors of ad $\mathbf{L}_S$ with eigenvalues respectively equal to 2 and -2, a physicist reads them as meaning that $\mathbf{L}_S$ is a symmetry of the vector field derived from K that *shortens* the time along the integral curves of K while, as a symmetry of the vector field corresponding to D, it *slows down* the time along the integral curves. This interpretation brings forth a close analogy between the normalization proposed in this article for perturbed free particles and the conventional averaging procedures applied to conditionally periodic systems: in both cases, the algorithm removes the short term effects caused by the perturbations.

From Tables I and II, the reader may collect the following triples:

TABLE II. Commutators of the basis matrices in sp(4, **R**).

	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}
w_1	0	$-2w_1$	0	0	0	w_2	$-2w_4$	$-2w_5$	w_8	w_7
w_2	$2w_1$	0	0	$2w_4$	$2w_5$	$-2w_6$	0	0	$-2w_9$	$-2w_{10}$
w_3	0	0	0	$2\omega_5$	$-2w_4$	0	$2w_8$	$-2w_7$	$-2w_{10}$	$2w_9$
w_4	0	$-2w_4$	$-2w_5$	0	0	w_7	$-2w_1$	0	$-w_3$	w_2
w_5	0	$-2w_5$	$2w_4$	0	0	w_8	0	$-2w_1$	w_2	w_3
w_6	$-w_2$	$2w_6$	0	$-w_7$	$-w_8$	0	$2w_{10}$	$2w_9$	0	0
w_7	$2w_4$	0	$-2w_8$	$2w_1$	0	$-2w_{10}$	0	$-2w_3$	0	$-2w_6$
w_8	$2w_5$	0	$2w_7$	0	$2w_1$	$-2w_9$	$2w_3$	0	$-2w_6$	0
w_9	$-w_8$	$2w_9$	$2w_{10}$	w_3	$-w_2$	0	0	$2w_6$	0	0
w_{10}	$-w_7$	$2w_{10}$	$-2w_9$	$-w_2$	$-w_3$	0	$2w_6$	0	0	0

$$(xX; \tfrac{1}{2}X^2) = X^2, \quad (xX; -\tfrac{1}{2}x^2) = x^2, \quad (\tfrac{1}{2}X^2; -\tfrac{1}{2}x^2) = xX,$$
$$(yY; \tfrac{1}{2}Y^2) = Y^2, \quad (yY; -\tfrac{1}{2}Y^2) = y^2, \quad (\tfrac{1}{2}Y^2; -\tfrac{1}{2}y^2) = yY,$$
$$(xX + yY; XY) = 2XY, \quad (xX + yY; -xy) = 2xy,$$
$$(XY; -xy) = xX + yY, \quad (xX - yY; Xy) = 2Xy,$$
$$(xX - yY; xY) = -2xY, \quad (Xy; xY) = xX - yY.$$

As will be seen in the next sections, to each of them corresponds a representation of the simple algebra sl(2, **R**), in the general linear algebra gl($\mathbf{P}_n$) relative to the vector space $\mathbf{P}_n$, hence a normalization scheme for a certain class of Hamiltonians. For example, the first triple concerns Hamiltonians of the type

$$H \equiv H(x, y, X, Y) = \frac{X^2}{2} + \sum_{n\geq 1} \frac{1}{n!} H_n(x, y, X, Y),$$

which may be normalized into Hamiltonians

$$H' \equiv H'(x', y', X', Y') = \frac{X'^2}{2} + \sum_{n\geq 1} \frac{1}{n!} H'_n(x', y', Y')$$

with the momentum X' eliminated from the perturbations.

3. DECOMPOSITION OF THE PERTURBATIONS

The objective of this section is to prove that any homogeneous polynomial p may be written in a unique way as the sum $p = p_K + p_D$ of two homogeneous polynomials of the same degree such that $\mathbf{L}_D p_D = 0$ and $p_K = \mathbf{L}_K q$ for some homogeneous polynomial q. It is to be noted that, in this section, $\mathbf{L}_K$ and $\mathbf{L}_D$ stand respectively for the restrictions of $\mathbf{L}_K$ and $\mathbf{L}_D$ to the vector space $\mathbf{P}_n$. The decomposition leads to a procedure for constructing a Lie transformation which will strip any perturbation term H_n in (14) of its component in Im $\mathbf{L}_K$.

Actually the decomposition of $\mathbf{P}_n$ as the direct sum of Ker $\mathbf{L}_D$ and Im $\mathbf{L}_K$ is but a particular case of a more general result about the representations of the special linear Lie algebra sl(2, **R**). To get at the essence of the problem, we need, however, to fix notations and terminology, and to recall a few basics from representation theory.

Let V be a real vector space of finite dimension; in the algebra of endomorphisms of V, $[a, b]$ denotes the commutator $ab - ba$. For readability, the image $\phi(x)$ of a vector x in V by an endomorphism ϕ will be written simply as ϕx. One says that V is an sl(2, **R**)-module or, equivalently that sl(2, **R**) is represented on V, if there exist three nonzero endomorphisms x, y, h of V satisfying the relations

$$[h, x] = 2x, \quad [h, y] = -2y, \quad [x, y] = h. \tag{17}$$

By virtue of Weyl's theorem, an sl(2, **R**)-module V may be decomposed into a direct sum of real vector subspaces

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_s, \tag{18}$$

whose summands are invariant and irreducible under the set of endomorphisms (x, y, h). There may be more than one way of accomplishing the decomposition, but the number s of summands and the equivalence classes of the irreducible representations are uniquely determined.

Proposition 1: Relative to a real vector space V of finite dimension that is an sl(2, **R**)-module for the three endomorphisms x, y, h satisfying the commutator relations (17):

(i) the endomorphisms x and y are nilpotent;
(ii) the endomorphisms xy and yx are semisimple;
(iii) Ker xy = Ker y and Im xy = Im x; likewise, Ker yx = Ker x and Im yx = Im y;
(iv) V = Ker $x \oplus$ Im y; likewise, V = Ker $y \oplus$ Im x;
(v) dim Ker x = dim Ker $y = s$, where s is the number of summands in any decomposition of V into a direct sum of irreducible vector subspaces.

The proposition is proved in two stages. First is considered the particular case when $s = 1$, that is, V itself is irreducible; then the results are extended to the general case where $s > 1$.

When it is irreducible, V admits, according to Humphreys,[18] a basis $(v_0, v_1, \ldots, v_m)$ such that, for $0 \leq i \leq m$,

$$hv_i = (m - 2i)v_i, \tag{19}$$
$$yv_i = (i + 1)v_{i+1}, \tag{20}$$
$$xv_i = (m - i + 1)v_{i-1}, \tag{21}$$

with the convention that $v_{-1} = v_{m+1} = 0$. As a consequence, $x^{m-1}v_i = y^{m-1}v_i = 0$, which proves that the endomorphisms x and y are nilpotent. There follows also that $(xy)v_i = (i + 1)(m - i)v_i$; hence the vectors (v_i) $(0 \leq i \leq m)$ are a basis of eigenvectors for the endomorphism xy (and likewise for the endomorphism yx), which means that xy and yx are semisimple. For that reason, V = Ker $xy \oplus$ Im xy, and also V = Ker $yx \oplus$ Im yx. More precisely, since, on the one hand, $(xy)v_i = 0$ if only $i = m$, and, on the other hand, $v_i = (xy)[(i + 1)^{-1}(m - i)^{-1}v_i]$ for $0 \leq i \leq m - 1$, it turns out that Ker xy is the one-dimensional vector subspace generated by v_m, while Im xy is the m-dimensional vector subspace generated by the vectors v_i $(0 \leq i \leq m - 1)$. In view of this decomposition of V, statements (iv) and (v) in the proposition are immediate corollaries of statement (iii). There remains thus to prove point (iii). To this end, observe that, on account of relation (21), v_i is in the image of x if and only if $0 \leq i < m$, while $xv_m = 0$; therefore, being the vector subspace generated by the vectors v_i $(0 \leq i \leq m - 1)$, Im x is identical to Im xy. Similarly, by virtue of (20), a linear combination $a_0v_0 + {}_1v_1 + \cdots + a_m v_m$ is in the kernel of y if and only if $a_i = 0$ for $0 \leq i \leq m - 1$; hence, being the one-dimensional subspace generated by v_m Ker y is identical to Ker xy.

The proof of Proposition 1 in the general case where V is completely reducible, although not irreducible, rests on decomposition (18) of V into a direct sum of irreducible submodules. It has just been proved that the restrictions of x and y to each of the summands are nilpotent, and that the restrictions of xy and yx are semisimple; hence the endomorphisms x and y themselves are nilpotent in V, whereas the endomorphisms xy and yx are semisimple in V. Demonstration of assertions (iii)–(v) involves the following:

Lemma 2: Let V be a vector space that is a direct sum of the subspaces (V_i) $(1 \leq i \leq n)$. Then, for any vector subspace W of V, the following statements are equivalent:

(i) W is the direct sum of the vector subspaces $(W \cap V_i)$ $(1 \leq i \leq n)$;

(ii) for any element w in W, the decomposition $w = w_1 + w_2 + \cdots + w_n$ such that w_i belongs to V_i for $1 \leq i \leq n$ implies that each component w_i belongs to W.

Intuitively speaking, the lemma says that W is the direct sum of its intersections with the V_i's if and only if the components of any vector in W along every "direction" V_i lie in W. Elementary as it may be, this lemma is not mentioned in the major textbooks in linear algebra; it is therefore in order to sketch its proof. Considering that V is the direct sum of the subspaces V_i, any element w in W may be decomposed into the sum $w = w_1 + w_2 + \cdots + w_n$ with w_i in V_i for each i. But, assuming that (i) holds, the same element w may be decomposed into the sum $w = u_1 + u_2 + \cdots + u_n$ with u_i in $W \cap V_i$ for each i. The decomposition of w in V being unique, there follows that $w_i = u_i$, and hence that w_i belongs to W for each i, which shows that (i) implies (ii). Conversely, if (ii) holds, then w_i belongs to $W \cap V_i$ for each i. Such a decomposition being unique, there results that W is the direct sum of the subspaces $W \cap V_i$, and hence that (ii) implies (i).

Lemma 2 is used to prove that

$$\text{Im } z = (V_1 \cap \text{Im } z) \oplus (V_2 \cap \text{Im } z) \oplus \cdots \oplus (V_s \cap \text{Im } z), \tag{22}$$

$$\text{Ker } z = (V_1 \cap \text{Ker } z) \oplus (V_2 \cap \text{Ker } z) \oplus \cdots \oplus (V_s \cap \text{Ker } z) \tag{23}$$

when the endomorphism z is either x, y, xy, or yx. Indeed, take w in Im z; in view of Weyl's theorem, it may be decomposed into a sum $w = w_1 + w_2 + \cdots + w_s$, where w_i belongs to V_i for $1 \leqslant i \leqslant s$. But $w = zv$ for some v in V which in turn may be decomposed into a sum $v = v_1 + v_2 + \cdots + v_s$ with v_i in V_i for $1 \leqslant i \leqslant s$; therefore, $zv = zv_1 + zv_2 + \cdots + zv_s$. Now, since z leaves each vector subspace V_i invariant and since the decomposition of w into its components in the V_i's is unique, there follows that $w_i = zv_i$ for $1 \leqslant i \leqslant s$, or that each component lies in Im z, hence formula (22) on account of Lemma 2. Similarly, for w in Ker z, there results that $0 = zw = zw_1 + zw_2 + \cdots + zw_s$, hence that $zw_i = 0$ for $1 \leqslant i \leqslant n$, since $0 = 0 + \cdots + 0$ is the unique way of decomposing the null vector in the direct sum (18). Because each component w_i belongs to Ker z, formula (23) results also from Lemma 2.

Proposition 1 having been proved when V is irreducible, there follows that Im $xy|V_i$ = Im $x|V_i$ and Ker $xy|V_i$ = Ker $y|V_i$ for $1 \leqslant i \leqslant s$. Then, by reason of the relations (22)–(23), one concludes that Im xy = Im x and Ker xy = Ker y, which thus proves statement (iii) in Proposition 1. The next statement is then a consequence of the fact that the endomorphism xy is semisimple. Finally, since Ker $x|V_i$ and Ker $y|V_i$ are one-dimensional for $1 \leqslant i \leqslant s$, one concludes from formula (23) that Ker x and Ker y are of dimension s. This completes the demonstration of Proposition 1.

Corollary: The vector space P_n is the direct sum Im $\mathbf{L}_K \oplus$ Ker $\mathbf{L}_D$.

If V is an sl(2, **R**)-module, the real number λ is called a *weight* of the representation when the vector subspace V_λ of elements x such that $hx = \lambda x$ is not the null space. In representation theory, it is proved that the number s of summands in the direct decomposition (18) is equal to dim V_0 + dim V_1. In the present case, the semisimple endomorphism involved in the representation of sl(2, **R**) is the differential operator

$$\mathbf{L}_S = x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y} - X\frac{\partial}{\partial X} - Y\frac{\partial}{\partial Y}.$$

From now on, in order to disencumber the notations, monomials like $x^\alpha y^\beta X^\gamma Y^\delta$ will be denoted $e(\alpha, \beta, \gamma, \delta)$. In those terms,

$$\mathbf{L}_S e(\alpha, \beta, \gamma, \delta) = (\alpha + \beta - \gamma - \delta) e(\alpha, \beta, \gamma, \delta);$$

hence the monomials $e(\alpha, \beta, \gamma, \delta)$ constitute a basis of the vector subspace V_k of $\mathbf{P}_n$ if and only if

$$\alpha + \beta + \gamma + \delta = n \quad \text{and} \quad \alpha + \beta - \gamma - \delta = k.$$

Therefore, when $n = 2m$, V_1 reduces to the null space whereas dim $V_0 = (m+1)^2$; otherwise, when $n = 2m+1$, it is V_0 that is reduced to the null space while dim $V_1 = (m+1)(m+2)$. Hence

Proposition 2: The dimension of Ker $\mathbf{L}_D$ in $\mathbf{P}_n$ is equal to $(m+1)^2$ when $n = 2m$ and to $(m+1)(m+2)$ when $n = 2m+1$.

By consulting Table III, the reader will gain a measure of appreciation for the extent to which the normalization simplifies the perturbed system (14). An arbitrary polynomial that is homogeneous of degree n in (x, y, X, Y) is the sum of $\binom{n+3}{3}$ monomials. Thus, from degree n to degree $(n+1)$, the perturbation term H_n grows in complexity by $(n+2)(n+3)/2$ terms; by contrast, Ker $\mathbf{L}_D$ increases only by $1 + (n+1)/2$ terms.

As the next section will indicate, there is, however, more to the normalization than a drastic reduction in algebraic complexity.

4. THE ANGULAR MOMENTUM INSIDE THE PERTURBATION

The decomposition vouchsafed by the corollary to Proposition 1 owes its physical interest to the algebraic nature of the kernel of $\mathbf{L}_D$ as one can see from

Proposition 3: With $\alpha + \beta + 2\gamma = n$, the polynomials $g(\alpha, \beta, \gamma) = x^\alpha y^\beta G^\gamma$ form a basis of Ker $\mathbf{L}_D$ in $\mathbf{P}_n$, and the polynomials $G(\alpha, \beta, \gamma) = X^\alpha Y^\beta G^\gamma$, a basis of Ker $\mathbf{L}_K$ in $\mathbf{P}_n$.

The proof of Proposition 3 rests on yet another decomposition of $\mathbf{P}_n$ into a direct sum of vector subspaces, which will be detailed first. In what follows, most of the time, a monomial $e(\alpha, \beta, \gamma, \delta)$ will be identified with the quadruple $(\alpha, \beta, \gamma, \delta)$ of its exponents. Clearly, an arbitrary quadruple $(\alpha, \beta, \gamma, \delta)$ of integers represents a monomial if and only if its elements are nonnegative, in which case the monomial belongs to $\mathbf{P}_n$ if and only if $\alpha + \beta + \gamma + \delta = n$; let $\mathbf{E}_n$ be the set of all quadruples satisfying these two conditions. The relation **R** defined by

$$(\alpha', \beta', \gamma', \delta')\mathbf{R}(\alpha, \beta, \gamma, \delta)$$

if $\exists$ an integer k ∋

$$(\alpha', \beta', \gamma', \delta') = (\alpha, \beta, \gamma, \delta) + (-k, k, k, -k)$$

is an equivalence among quadruples. For any quadruple $(\alpha, \beta, \gamma, \delta)$, let $E(\alpha, \beta, \gamma, \delta)$ designate the intersection with $\mathbf{E}_n$ of

TABLE III. Algebraic simplification accomplished by normalization.

Degree	3	4	5	6	7	8	...
dim $\mathbf{P}_n$	20	35	56	84	120	165	...
dim Ker $\mathbf{L}_D$	6	9	12	16	20	25	...

the class of quadruples equivalent to $(\alpha, \beta, \gamma, \delta)$ modulo $\mathbf{R}$; given the integer k, the quadruple $(\alpha - k, \beta + k, \gamma + k, \delta - k)$ is an element of $E(\alpha, \beta, \gamma, \delta)$ if and only if

$$\alpha - k \geqslant 0, \quad \beta + k \geqslant 0, \quad \gamma + k \geqslant 0, \quad \delta - k \geqslant 0,$$

or, equivalently, if and only if

$$-\min(\beta, \gamma) \leqslant k \leqslant \min(\alpha, \delta);$$

this shows in particular that the number of quadruples in the equivalence class $E(\alpha, \beta, \gamma, \delta)$ is equal to $1 + \min(\alpha, \delta) + \min(\beta, \gamma)$. The natural ordering on the integers k provides an ordering on $E(\alpha, \beta, \gamma, \delta)$. In that order, the lowest element in $E(\alpha, \beta, \gamma, \delta)$ is of the form $(\alpha_0, \beta_0, \gamma_0, \delta_0)$ with $\beta_0 \gamma_0 = 0$. Repeated addition of $(-1, 1, 1, -1)$ to the lowest element produces all the quadruples in the equivalence class $E(\alpha_0, \beta_0, \gamma_0, \delta_0)$ up to the highest element which is of the form $(\alpha_1, \beta_1, \gamma_1, \delta_1)$ with $\alpha_1 \delta_1 = 0$.

Now let $P(\alpha, \beta, \gamma, \delta)$ denote the vector space generated by those monomials whose exponents belong to the class $E(\alpha, \beta, \gamma, \delta)$. The monomials in $\mathbf{E}_n$ being a basis of $\mathbf{P}_n$, the preceding discussion establishes that $\mathbf{P}_n$ is the direct sum of the vector subspaces $P(\alpha, \beta, \gamma, \delta)$, and that $\dim P(\alpha, \beta, \gamma, \delta) = \min(\alpha, \delta) + \min(\beta, \gamma) + 1$. However, the vector subspaces $P(\alpha, \beta, \gamma, \delta)$ are not in general invariant under the operators $\mathbf{L}_K$ and $\mathbf{L}_D$; as a matter of fact, the way in which these operators act on $P(\alpha, \beta, \gamma, \delta)$ is given in the next statement.

Lemma 3: For any $(\alpha, \beta, \gamma, \delta)$ in $\mathbf{E}_n$ and with the convention that $P(\alpha, \beta, \gamma, \delta)$ designates the null vector space when $E(\alpha, \beta, \gamma, \delta)$ is empty,

(i) $\mathbf{L}_K$ maps $P(\alpha, \beta, \gamma, \delta)$ into $P(\alpha - 1, \beta, \gamma + 1, \delta)$,
(ii) $\mathbf{L}_D$ maps $P(\alpha, \beta, \gamma, \delta)$ into $P(\alpha + 1, \beta, \gamma - 1, \delta)$,
(iii) $\mathbf{L}_D \mathbf{L}_K$ maps $P(\alpha, \beta, \gamma, \delta)$ into itself.

Indeed, for any monomial $e(\alpha, \beta, \gamma, \delta)$,

$$\mathbf{L}_K e(\alpha, \beta, \gamma, \delta) = \alpha e(\alpha - 1, \beta, \gamma + 1, \delta) + \beta e(\alpha, \beta - 1, \gamma, \delta + 1) \tag{24}$$

by virtue of definition (12). But $(\alpha - 1, \beta, \gamma + 1, \delta) = (\alpha, \beta - 1, \gamma, \delta + 1) + (-1, 1, 1, -1)$, hence both monomials in the right-hand member of (24) are equivalent modulo $\mathbf{R}$, and (i) is thus proved. By an analogous argument, (ii) is a consequence of the identity

$$\mathbf{L}_D e(\alpha, \beta, \gamma, \delta) = \gamma e(\alpha + 1, \beta, \gamma - 1, \delta) + \delta e(\alpha, \beta + 1, \gamma, \delta - 1), \tag{25}$$

resulting from definition (9). Then (iii) follows by composing (i) and (ii).

Against this background information, Proposition 3 will now be proved, but for $\mathbf{L}_K$ only, since the result for $\mathbf{L}_D$ follows from swapping the coordinates (x, y) and the momenta (X, Y). Because $\mathbf{L}_K$ is a derivation, the relations $\mathbf{L}_K X = \mathbf{L}_K Y = \mathbf{L}_K (xY - yX) = 0$ imply that $\mathbf{L}_K G(\alpha, \beta, \gamma) = 0$; hence there remains to show that the polynomials $G(\alpha, \beta, \gamma)$ are linearly independent and that they span $\operatorname{Ker} \mathbf{L}_K$. On the one hand, the binomial expansion

$$\begin{aligned} X^\alpha Y^\beta (xY - yX)^\gamma &= X^\alpha Y^\beta \sum_{0 \leqslant k \leqslant \gamma} (-1)^k \binom{\gamma}{k} x^{\gamma - k} y^k X^k Y^{\gamma - k} \\ &= \sum_{0 \leqslant k \leqslant \gamma} (-1)^k \binom{\gamma}{k} e(\gamma - k, k, \alpha + k, \beta + \gamma - k) \end{aligned}$$

shows that $G(\alpha, \beta, \gamma)$ belongs to the vector subspace $P(\gamma, 0, \alpha, \beta + \gamma)$. The equivalence classes $E(\gamma, 0, \alpha, \beta + \gamma)$ and $E(\gamma', 0, \alpha', \beta' + \gamma')$ being disjoint when $(\alpha, \beta, \gamma, \delta) \neq (\alpha', \beta', \gamma', \delta')$, distinct polynomials $G(\alpha, \beta, \gamma)$ belong to distinct vector subspaces $P(\gamma, 0, \alpha, \beta + \gamma)$, which means that the polynomials $G(\alpha, \beta, \gamma)$ are linearly independent.

On the other hand, take a polynomial p in $\mathbf{P}_n$, and let $p = \Sigma_\iota p_\iota$ be its decomposition relative to the subspaces $P(\alpha, \beta, \gamma, \delta)$. Then $\mathbf{L}_K p = \Sigma_\iota \mathbf{L}_K p_\iota$; since, according toLemma 3, for $\iota \neq \kappa$, the images $\mathbf{L}_K p_\iota$ and $\mathbf{L}_K p_\kappa$ belong to distinct subspaces $P(\alpha, \beta, \gamma, \delta)$, the relation $\mathbf{L}_K p = 0$ implies that $\mathbf{L}_K p_\iota = 0$ for each index ι. There results by reason of Lemma 2 that $\operatorname{Ker} \mathbf{L}_K$ is the direct sum of the vector subspaces $\operatorname{Ker} \mathbf{L}_K \cap P(\alpha, \beta, \gamma, \delta)$. As will be seen, all of these summands are identical to the null subspace save the intersections $\operatorname{Ker} \mathbf{L}_K \cap P(\gamma, 0, \alpha, \beta + \gamma)$ which are of dimension 1 and are in fact generated by a polynomial $G(\alpha, \beta, \gamma)$ (see Lemma 4 below). This will prove that such special polynomials span the kernel of $\mathbf{L}_K$, and therefore constitute a basis of $\operatorname{Ker} \mathbf{L}_K$ as is announced in Proposition 3. There remains thus to examine the trace of $\operatorname{Ker} \mathbf{L}_K$ on each vector subspace $P(\alpha, \beta, \gamma, \delta)$, which will be done by studying in detail the action of $\mathbf{L}_K$ on each of them. Such an analysis will prove useful also in Sec. 5, where the normalization algorithm will be developed; it will show in particular that the decomposition of a polynomial into its components in $\operatorname{Im} \mathbf{L}_K$ and $\operatorname{Ker} \mathbf{L}_D$ reduces to inverting a few matrices of very low dimension.

Depending on the type presented by the lowest element $(\alpha_0, \beta_0, \gamma_0, \delta_0)$ in $P(\alpha, \beta, \gamma, \delta)$, the five cases mentioned in Table IV have to be considered. It is a question of setting proper bases for representing the linear map $\mathbf{L}_K : P(\alpha, \beta, \gamma, \delta) \to P(\alpha - 1, \beta, \gamma + 1, \delta)$ as a matrix. To this end, in each case, the monomials ordered from lowest to highest are

TABLE IV. Classes of $\mathbf{L}_K$ restricted to $P(\alpha, \beta, \gamma, \delta)$.

Class	$P(\alpha, \beta, \gamma, \delta)$	Matrix $\mathbf{L}_K$		Matrix $\mathbf{L}_D$	
	Lowest element	Dimensions	Rank	Nullity	Dimensions
I	$\beta_0 = 0, \quad \alpha_0 = 0$		0	1	
II	$\beta_0 = 0, \quad 1 \leqslant \alpha_0 \leqslant \delta_0$	$\alpha_0 \times (\alpha_0 + 1)$	α_0	1	$(\alpha_0 + 1) \times \alpha_0$
III	$\beta_0 = 0, \quad \alpha_0 > \delta_0$	$(\delta_0 + 1) \times (\delta_0 + 1)$	$\delta_0 + 1$	0	$(\delta_0 + 1) \times (\delta_0 + 1)$
IV	$\beta_0 \neq 0, \quad \gamma_0 = 0, \quad \alpha_0 \leqslant \delta_0$	$(\alpha_0 + 1) \times (\alpha_0 + 1)$	$\alpha_0 + 1$	0	$(\alpha_0 + 1) \times (\alpha_0 + 1)$
V	$\beta_0 \neq 0, \quad \gamma_0 = 0, \quad \alpha_0 > \delta_0$	$(\delta_0 + 2) \times (\delta_0 + 1)$	$\delta_0 + 1$	0	$(\delta_0 + 1) \times (\delta_0 + 2)$

adopted as a basis, the coefficients of a polynomial are regarded as a column vector, and the operator $\mathbf{L}_K$ is represented as a matrix acting by multiplication to the left.

Class I can be disposed of at once. For it is clear that $P(0, 0, \gamma_0, \delta_0)$ is of dimension 1, being generated as it is by the monomial $X^{\gamma_0}Y^{\delta_0}$, and that it is mapped by $\mathbf{L}_K$ onto the null space.

In class II, the monomials

$$e(\alpha_0, 0, \gamma_0, \delta_0), \ldots, e(0, \alpha_0, \gamma_0 + \alpha_0, \delta_0 - \alpha_0) \tag{26}$$

form a basis of $P(\alpha_0, 0, \gamma_0, \delta_0)$, and the monomials

$$e(\alpha_0 - 1, 0, \gamma_0 + 1, \delta_0), \ldots, e(0, \alpha_0 - 1, \gamma_0 + \alpha_0, \delta_0 - \alpha_0 + 1), \tag{27}$$

a basis of $P(\alpha_0 - 1, 0, \gamma_0 + 1, \delta_0)$. Restricted to $P(\alpha_0, 0, \gamma_0, \delta_0)$, $\mathbf{L}_K$ is, by virtue of (24), represented by the rectangular band matrix

$$\begin{bmatrix} \alpha_0 & 1 & 0 & \cdots & \cdots & 0 & 0 & 0 \\ 0 & \alpha_0 - 1 & 2 & \cdots & \cdots & 0 & 0 & 0 \\ 0 & 0 & \alpha_0 - 2 & \cdots & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \cdots & \alpha_0 - 2 & 0 & 0 \\ 0 & 0 & 0 & \cdots & \cdots & 2 & \alpha_0 - 1 & 0 \\ 0 & 0 & 0 & \cdots & \cdots & 0 & 1 & \alpha_0 \end{bmatrix}. \tag{28}$$

The first α_0 columns of (28) constitute a square matrix whose determinant ($= \alpha_0!$) is not zero. Hence (28) is of rank α_0 and of nullity 1, which means that $\mathrm{Ker}\, \mathbf{L}_K \cap P(\alpha_0, 0, \gamma_0, \delta_0)$ is generated by the special polynomial $G(\gamma_0, \delta_0 - \alpha_0, \alpha_0)$.

In class III, the restriction of $\mathbf{L}_K$ corresponds to the band matrix

$$\begin{bmatrix} \alpha_0 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \alpha_0 - 1 & 2 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \alpha_0 - 2 & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \alpha_0 - \delta_0 + 2 & \delta_0 - 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & \alpha_0 - \delta_0 + 1 & \delta_0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & \alpha_0 - \delta_0 \end{bmatrix} \tag{29}$$

when the vectors

$$e(\alpha_0, 0, \gamma_0, \delta_0), \ldots, e(\alpha_0 - \delta_0, \delta_0, \gamma_0 + \delta_0, 0) \tag{30}$$

are choosen as a basis in $P(\alpha_0, 0, \gamma_0, \delta_0)$ while the vectors

$$e(\alpha_0 - 1, 0, \gamma_0 + 1, \delta_0), \ldots, e(\alpha_0 - \delta_0 - 1, \delta_0, \gamma_0 + \delta_0 + 1, 0) \tag{31}$$

are taken for the basis of $P(\alpha_0 - 1, 0, \gamma_0 + 1, \delta_0)$. Matrix (29) is square, and it is clearly nonsingular. Therefore, $\mathbf{L}_K$ is an isomorphism of $P(\alpha_0, 0, \gamma_0, \delta_0)$ onto $P(\alpha_0 - 1, 0, \gamma_0 + 1, \delta_0)$.

Now, in class IV, the bases in $P(\alpha_0, \beta_0, 0, \delta_0)$ and $P(\alpha_0, \beta_0 - 1, 0, \delta_0 + 1)$ are chosen to be respectively

$$e(\alpha_0, \beta_0, 0, \delta_0), \ldots, e(0, \beta_0 + \alpha_0, \alpha_0, \delta_0 - \alpha_0) \tag{32}$$

and

$$e(\alpha_0, \beta_0 - 1, 0, \delta_0 + 1), \ldots, e(0, \beta_0 + \alpha_0 - 1, \alpha_0, \delta_0 - \alpha_0 + 1), \tag{33}$$

so that the square matrix representing the restriction of $\mathbf{L}_K$ is

$$\begin{bmatrix} \beta_0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \alpha_0 & \beta_0 + 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \alpha_0 - 1 & \beta_0 + 2 & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \beta_0 + \alpha_0 - 2 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 2 & \beta_0 + \alpha_0 - 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 & \beta_0 + \alpha_0 \end{bmatrix}, \tag{34}$$

which is obviously nonsingular.

Finally, in class V, the monomials

$$e(\alpha_0, \beta_0, 0, \delta_0), \ldots, e(\alpha_0 - \delta_0, \beta_0 + \delta_0, \delta_0, 0) \tag{35}$$

are chosen as the basis in $P(\alpha_0, \beta_0, 0, \delta_0)$, while the monomials

$$e(\alpha_0, \beta_0 - 1, 0, \delta_0 + 1), \ldots, e(\alpha_0 - \delta_0 - 1, \beta_0 + \delta_0, \delta_0 + 1, 0) \tag{36}$$

form the basis in $P(\alpha_0, \beta_0 - 1, 0, \delta_0 + 1)$. In this way, the operator $\mathbf{L}_K$ is given by the band matrix

$$\begin{bmatrix} \beta_0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \alpha_0 & \beta_0+1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \alpha_0-1 & \beta_0+2 & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \alpha_0-\delta_0+2 & \beta_0+\delta_0-1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & \alpha_0-\delta_0+1 & \beta_0+\delta_0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & \alpha_0-\delta_0 \end{bmatrix}. \tag{37}$$

The first (δ_0+1) rows of (37) have the nonzero product $\beta_0(\beta_0+1)\cdots(\beta_0+\delta_0)$ for determinant; therefore, matrix (37) has rank δ_0+1 and nullity 0.

That portion of the results just obtained which is needed in the proof of Proposition 3 is summed up in the following alternative:

Lemma 4: Let $(\alpha_0, \beta_0, \gamma_0, \delta_0)$ be the lowest quadruple in an equivalence class $E(\alpha, \beta, \gamma, \delta)$. If $\beta_0 = 0$ and $\alpha_0 \leqslant \delta_0$, then $\mathrm{Ker}\, \mathbf{L}_K \cap P(\alpha, \beta, \gamma, \delta) = 0$ is of dimension 1, and is generated by the special polynomial $G(\gamma_0, \delta_0 - \alpha_0, \alpha_0)$. Otherwise, $\mathrm{Ker}\, \mathbf{L}_K \cap P(\alpha, \beta, \gamma, \delta) = 0$.

Proposition 3 affords an easy way of decomposing $\mathbf{P}_n$ into a direct sum of irreducible weight spaces: Each polynomial $v_0 = g(\alpha, \beta, \gamma)$ generates a basis formed of the chain of polynomials v_k $(k \geqslant 0)$ such that $kv_k = \mathbf{L}_K v_{k-1}$ for $k \geqslant 1$. Such polynomials span a weight space V_λ of weight $\lambda = \alpha + \beta - 2\gamma$. Yet attempts at using the complete reduction (18) to decompose any polynomial into its constituents in $\mathrm{Im}\, \mathbf{L}_K$ and $\mathrm{Ker}\, \mathbf{L}_D$ have not resulted in clear and elegant software procedures. All the same, an algorithm based on the classes enumerated in Table V proved to be both expedient and easy to code.

5. THE DECOMPOSITION ALGORITHM

For most dynamical systems, normalization to degree 3 is sufficient; fortunately, at that minimal degree, the decomposition is readily executed by hand. Once Table V has been established, it becomes clear that the polynomial

$$P = \sum_{(\alpha, \beta, \gamma, \delta)} C_{\alpha,\beta,\gamma,\delta} x^\alpha y^\beta X^\gamma Y^\delta$$

may be written as the sum $p = p_D + \mathbf{L}_K q$, where

$$\begin{aligned} q = {} & \tfrac{1}{3}C_{2,0,1,0}x^3 + \tfrac{1}{3}(C_{1,1,1,0} + C_{2,0,0,1})x^2y \\ & + \tfrac{1}{3}(C_{0,2,1,0} + C_{1,1,0,1})xy^2 + \tfrac{1}{3}C_{0,2,0,1}y^3 \\ & + \tfrac{1}{2}C_{1,0,2,0}x^2X + \tfrac{1}{2}C_{0,1,2,0}xyX + \tfrac{1}{2}C_{0,1,1,1}y^2X \\ & + \tfrac{1}{2}C_{1,0,1,1}x^2Y + \tfrac{1}{2}C_{1,0,0,2}xyY + \tfrac{1}{2}C_{0,1,0,2}y^2Y \\ & + C_{0,0,3,0}xX^2 + C_{0,0,2,1}yX^2 \\ & + C_{0,0,1,2}xY^2 + C_{0,0,0,3}yY^2 \\ & - \tfrac{1}{2}C_{0,1,2,0}xG + \tfrac{1}{2}C_{1,0,0,2}yG \end{aligned} \tag{38}$$

and

$$\begin{aligned} p_D = {} & C_{3,0,0,0}x^3 + C_{2,1,0,0}x^2y + C_{1,2,0,0}xy^2 + C_{0,3,0,0}y^3 \\ & + (\tfrac{2}{3}C_{2,0,0,1} - \tfrac{1}{3}C_{1,1,1,0})xG - (\tfrac{2}{3}C_{0,2,1,0} - \tfrac{1}{3}C_{1,1,0,1})yG, \end{aligned}$$

which, by virtue of Proposition 3, is an element of $\mathrm{Ker}\, \mathbf{L}_D$.

However trivial the task appears to be at degree 3, Table III leaves one to gather that the calculations become rapidly voluminous past degree 4, even for a computer program prepared to handle sparse matrices. The next proposition is the foundation of a decomposition algorithm complete to the point of having been coded eventually in APL[19] and run on a DEC-20 at the National Institutes of Health in Bethesda, Md. (*N.B.*: The program is available upon request from the third author.)

Let V be a vector space; let also v and w be two endomorphisms of V, and set $u = vw$. For any x in V, a vector y in V satisfies the relation $v(x - wy) = 0$ if and only if $uy = vx$. Thus it is true that, *given any polynomial p in* $\mathbf{P}_n$, *a polynomial q in* $\mathbf{P}_n$ *satisfies the relation* $\mathbf{L}_D(p - \mathbf{L}_K q) = 0$ *if and only if it is a solution for q of the equation* $Tq = \mathbf{L}_D p$, *where T is the operator* $\mathbf{L}_D \mathbf{L}_K$.

Given an endomorphism u of the vector space V, one can always find an endomorphism u^- such that $uu^-u = u$. We call u^- a *generalized inverse* of u, but the reader should note that authors interested in classifying various species of generalized inverses associated with u would name u^- a $\{1\}$-*inverse*[20] of u or a *g-inverse*.[21] For any vector x in $\mathrm{Im}\, u$, $uu^-x = x$. In particular, assume that u is the product $u = vw$ of two endomorphisms and that $\mathrm{Im}\, v = \mathrm{Im}\, u$; then any vector y of the form $y = (u^-v)x$ is a solution of the equation $uy = vx$. For there exists by hypothesis a vector z such that $vx = uz$; hence $uy = (uu^-u)z = uz = vx$. Applied to the operator $T = \mathbf{L}_D \mathbf{L}_K : \mathbf{P}_n \to \mathbf{P}_n$, for which $\mathrm{Im}\, \mathbf{L}_D = \mathrm{Im}\, T$ (Proposition 1), the above considerations prove the following:

Lemma 5: Let T^- be a generalized inverse of T and p a polynomial in $\mathbf{P}_n$. Then the polynomial $q = T^- \mathbf{L}_D p$ is a solution of the equation $Tq = \mathbf{L}_D p$.

On account of Lemma 5, after a generalized inverse T^- has been produced, the problem of decomposing a polynomial into the sum of its constituents in $\mathrm{Im}\, \mathbf{L}_K$ and $\mathrm{Ker}\, \mathbf{L}_D$ will be solved by setting $p_K = \mathbf{L}_K q$ and $p_D = p - p_K$.

Among the many varieties of generalized inverses associated with T, preference should be given here to those which preserve the basic symmetry consisting in exchanging the coordinate x and its conjugate momentum X, respectively, with the coordinate y and its conjugate momentum Y. There is indeed no reason why the normalization should favor one coordinate more than the other. The symmetry requirement is best expressed by introducing the operator $Z : p \to Zp = p(x, y, X, Y) - p(y, x, Y, X) : \mathbf{P}_n \to \mathbf{P}_n$. A polynomial p is symmetric in the pairs (x, X) and (y, Y) if and only if $Zp = 0$. Furthermore, the differential operator

TABLE V. Image and coimage of $\mathbf{L}_K$ in the space P_3.

$\mathbf{L}_K \frac{1}{4}x^3 = x^2X$	$\mathbf{L}_K \frac{1}{2}x^2X = xX^2$	$\mathbf{L}_K xX^2 = X^3$
$\mathbf{L}_K \frac{1}{3}x^2y = xyX + \frac{1}{3}xG$	$\mathbf{L}_K(\frac{1}{2}xyX - \frac{1}{2}xG) = yX^2$	$\mathbf{L}_K yX^2 = X^2Y$
$\mathbf{L}_K \frac{1}{3}xy^2 = y^2X + \frac{2}{3}yG$	$\mathbf{L}_K \frac{1}{2}x^2Y = xXY$	$\mathbf{L}_K xY^2 = XY^2$
$\mathbf{L}_K \frac{1}{3}x^2y = x^2Y - \frac{2}{3}xG$	$\mathbf{L}_K \frac{1}{2}y^2X = yXY$	$\mathbf{L}_K yY^2 = Y^3$
$\mathbf{L}_K \frac{1}{3}xy^2 = xyY - \frac{1}{3}yG$	$\mathbf{L}_K(\frac{1}{2}xyY + \frac{1}{2}yG) = xY^2$	
$\mathbf{L}_K \frac{1}{4}y^3 = y^2Y$	$\mathbf{L}_K \frac{1}{2}y^2Y = yY^2$	

$$T = \mathbf{L}_D\mathbf{L}_K = \left(x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y}\right) + \left(xX\frac{\partial^2}{\partial x\partial X} + yY\frac{\partial^2}{\partial y\partial Y}\right) + \left(xY\frac{\partial^2}{\partial y\partial X} + yX\frac{\partial^2}{\partial x\partial Y}\right)$$

being invariant for the symmetry, the operators T and Z commute over $\mathbf{P}_n$. Our purpose thus is to find a generalized inverse of T with corresponding symmetry properties. It should map symmetric polynomials onto symmetric polynomials, that is, ZT^-p should be $= 0$ whenever $Zp = 0$; also, by exchanging the pairs (x, X) and (y, Y) in the image $q(x, y, X, Y) = T^-p(x, y, X, Y)$, one should have that $q(y, x, Y, X) = T^-p(y, x, Y, X)$, that is, $ZT^- = T^-Z$. It will be shown that a generalized inverse $T^\#$ called the *group inverse* of T satisfies these symmetry requirements.

For an endomorphism u of a vector space V, Erdelyi[22] calls the group inverse of u an endomorphism $u^\#$ such that

$$uu^\#u = u, \quad u^\#uu^\# = u^\#, \quad uu^\# = u^\#u.$$

If it exists, the group inverse of u is unique. If u is an isomorphism of V onto itself, then $u^\# = u^{-1}$; more generally, u admits a group inverse if and only if V may be decomposed[23] into the direct sum of Ker u and Im u. This is the case when u is semisimple; then the restriction $u_\#$ of u to Im u is bijective, and, in principle at least, the group inverse $u^\#$ may be built as follows:

$$u^\#p = \begin{cases} 0 & \text{for } p \text{ in Ker } u, \\ \mu_\#{}^{-1}p & \text{for } p \text{ in Im } u. \end{cases}$$

Lemma 6: Let V be a vector space, and u an endomorphism of V. If u admits a group inverse $u^\#$, then, for any endomorphism v of V, the relation $vu = uv$ implies the relation $vu^\# = u^\#v$. In particular, for any vector x of V, the relation $vx = 0$ implies the relation $(vu^\#)x = 0$.

The lemma is proved when it is shown that $(vu^\#)x = (v^\#u)x$ first for x in Ker u and then for x in Im u. In each case, the demonstration rests on the fact that, because u and v commute, v(Ker u) is contained in Ker u and v(Im u) in Im u.

If $ux = 0$, then, on the other hand, $u^\#x = 0$ because $u^\#$ admits Ker u as its null space,[22] and hence $(vu^\#)x = 0$; on the other hand, $u(vx) = (uv)x = 0$, which implies that $(u^\#v)x = u^\#(vx) = 0$. Therefore, when restricted to Ker u, $vu^\#$ and $u^\#v$ are identical, since they are both equal to the null endomorphism.

Now take x in Im u. There is a unique element y in Im u such that $uy = x$ and $u^\#x = y$. With x and y both in Im u, the elements vx and vy are also both in Im u. But $u(vy) = (uv)y = (vu)y = v(uy) = vx$, and vy is therefore, the unique element of Im u mapped by u onto vx. Hence $vy = u^\#(vx)$, and the latter relation implies that $(vu^\#)x = v(u^\#x) = vy = u^\#(vx) = (u^\#v)x$. This completes the proof of Lemma 5.

The operator $T = \mathbf{L}_D\mathbf{L}_K$ is semisimple (Proposition 1); hence it admits a group inverse $T^\#$. By virtue of Lemma 6, the group inverse $T^\#$ commutes with the symmetry operator Z, and it maps symmetric polynomials onto symmetric polynomials.

The construction proposed here for the group inverse $T^\#$ of T makes use of the decomposition of $\mathbf{P}_n$ into a direct sum of vector subspaces $P_\iota = P(\alpha, \beta, \gamma, \delta)$ specified in Sec. 4. Assume that, for each ι, the restriction T_ι of T to P_ι admits a group inverse; then, for each polynomial p of $\mathbf{P}_n$ decomposed into the sum $p = \Sigma_\iota\, p_\iota$ of its components in the summands P_ι, define the image $T^\#p = \Sigma_\iota\, T_\iota^\#p_\iota$. Manifestly $T^\#$ is a linear map $\mathbf{P}_n \to \mathbf{P}_n$, and it satisfies the three conditions $TT^\#T = T$, $T^\#TT^\# = T^\#$, $TT^\# = T^\#T$, which means that $T^\#$ is the group inverse of T. The decomposition of a general polynomial p into its components in Ker $\mathbf{L}_D$ and Im $\mathbf{L}_K$ is thereby reduced to the problem of building the group inverse for the restrictions of T on each of the vector subspaces in the classes enumerated in Table V. Statistics collected in Table VI for a homogeneous polynomial of degree 6 will convince the reader that, however tedious it may be, a careful discussion of each particular situation breaks up the general problem of producing the 84×84 matrix for the group inverse $T^\#$ into the solution of 42 linear systems in at most 3 unknowns, 22 of them being utterly trivial.

Class I is dealt with at once: T_ι being the null endomorphism, its group inverse $T_\iota^\#$ is also the null endomorphism.

In classes III–V, the factor $\mathbf{L}_K$ is injective; since Ker T_ι is equal to the kernel of $\mathbf{L}_K$ restricted to P_ι (Proposition 1), there follows that T_ι is an isomorphism of P_ι onto itself and hence that T_ι^{-1} is the group inverse of T_ι. Now a closer examination of classes III and IV will bring forth a straightforward procedure for inverting T_ι.

As was shown in Sec. 4 for class III, the map $\mathbf{L}_K$ is an isomorphism of $P(\alpha_0, 0, \gamma_0, \delta_0)$ onto $P(\alpha_0 - 1, 0, \gamma_0 + 1, \delta_0)$

TABLE VI. Count of matrix inversions sufficient to calculate the group inverse at degree 6.

Classes	Dimensions		
	1×1	2×2	3×3
II or V	5	3	1
III or IV	6	4	2

represented by matrix (29) for the bases (30) in $P(\alpha_0, 0, \gamma_0, \delta_0)$ and (31) in $P(\alpha_0 - 1, 0, \gamma_0 + 1, \delta_0)$. But, for the same bases, $\mathbf{L}_D : P(\alpha_0 - 1, 0, \gamma_0 + 1, \delta_0) \to P(\alpha_0, 0, \gamma_0, \delta_0)$ is represented by the $(\delta_0 + 1) \times (\delta_0 + 1)$ matrix:

$$\begin{bmatrix} \gamma_0 + 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \delta_0 & \gamma_0 + 2 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \delta_0 - 1 & \gamma_0 + 3 & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \gamma_0 + \delta_0 - 1 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 2 & \gamma_0 + \delta_0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 & \gamma_0 + \delta_0 + 1 \end{bmatrix}, \tag{39}$$

which is evidently nonsingular; hence $\mathbf{L}_D$ is an isomorphism of $P(\alpha_0 - 1, 0, \gamma_0 + 1, \delta_0)$ onto $P(\alpha_0, 0, \gamma_0, \delta_0)$. Because matrix (29) is upper triangular, and matrix (39) lower triangular, the equation $T_\iota q = p_\iota$ is solved readily first by forward substitution to obtain a polynomial r such that $\mathbf{L}_D r = p_\iota$, and then by backward substitution to find the polynomial q such that $\mathbf{L}_K q = r$.

One meets a similar situation in class IV. The factor $\mathbf{L}_K$ is an isomorphism of $P(\alpha_0, \beta_0, 0, \delta_0)$ onto $P(\alpha_0, \beta_0 - 1, 0, \delta_0 + 1)$ represented by the matrix (34) when bases (32) and (33) are selected in the subspaces $P(\alpha_0, \beta_0, 0, \delta_0)$ and $P(\alpha_0, \beta_0 - 1, 0, \delta_0 + 1)$, respectively. Further, for the same bases, $\mathbf{L}_D : P(\alpha_0, \beta_0 - 1, 0, \delta_0 + 1) \to P(\alpha_0, \beta_0, 0, \delta_0)$ is represented by the $(\alpha_0 + 1) \times (\alpha_0 + 1)$ matrix

$$\begin{bmatrix} \delta_0 + 1 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \delta_0 & 2 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \delta_0 - 1 & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \delta_0 - \alpha_0 + 3 & \alpha_0 - 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & \delta_0 - \alpha_0 + 2 & \alpha_0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & \delta_0 - \alpha_0 + 1 \end{bmatrix}, \tag{40}$$

which is nonsingular; thus $\mathbf{L}_D$ is an isomorphism of $P(\alpha_0, \beta_0 - 1, 0, \delta_0 + 1)$ onto $P(\alpha_0, \beta_0, 0, \delta_0)$. Matrix (34) being lower triangular and matrix (40) upper triangular, the equation $T_\iota q = p_\iota$ is solved first by backward substitution to obtain a solution of the equation $\mathbf{L}_D r = p_\iota$ and then by forward substitution to solve the equation $\mathbf{L}_K q = r$.

The solution is not that simple in class V. On the one hand, $\mathbf{L}_K$ is injective but not surjective. On the other hand, the linear map $\mathbf{L}_D : P(\alpha_0, \beta_0 - 1, 0, \delta_0 + 1) \to P(\alpha_0, \beta_0, 0, \delta_0)$ is surjective but not injective. Indeed, with (35) and (36) chosen as the bases in $P(\alpha_0, \beta_0, 0, \delta_0)$ and $P(\alpha_0, \beta_0 - 1, 0, \delta_0 + 1)$, respectively, $\mathbf{L}_D$ is represented by the band matrix

$$\begin{bmatrix} \delta_0 + 1 & 1 & 0 & \cdots & \cdots & 0 & 0 & 0 \\ 0 & \delta_0 & 2 & \cdots & \cdots & 0 & 0 & 0 \\ 0 & 0 & \delta_0 - 1 & \cdots & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \cdots & \delta_0 - 1 & 0 & 0 \\ 0 & 0 & 0 & \cdots & \cdots & 2 & \delta_0 & 0 \\ 0 & 0 & 0 & \cdots & \cdots & 0 & 1 & \delta_0 + 1 \end{bmatrix} \tag{41}$$

with $(\delta_0 + 1)$ rows and $(\delta_0 + 2)$ columns, and it is easily seen that matrix (41) is of maximum rank $\delta_0 + 1$ and of nullity 1. For the restrictions of T_ι in class V, there seems to be no quicker way of inverting T_ι than by actually multiplying matrices (41) and (37) row by column, thereafter calculating explicitly the inverse of their product.

In both classes I and II, the operator T_ι is not invertible. While, in class I, the group inverse of T_ι is the null endomorphism, in class II is encountered the case where a way of computing the group inverse $T_\iota^\#$ must be prescribed. But the conditions defining the group inverse characterize u, in the terminology of Drazin,[24] as a *pseudo-invertible* element in the associative algebra End V of the endomorphisms of V. Thus the group inverse $u^\#$ is a Drazin pseudo-inverse of u relative to which the index of u is equal to 1. In that context, recall the following statement, which is in fact a particular case of a general theorem proved by Cline.[25] Let U and V be finite-dimensional vector spaces; let also v be an injective linear map $U \to V$, and w a surjective linear map $V \to U$; if the product $u = vw : V \to V$ is pseudo-invertible in the sense of Drazin and if its index is equal to 1, then $wv : U \to U$ is bijective, and

$$u^\# = v(wv)^{-2}w. \tag{42}$$

The conditions under which Cline's formula (42) may be applied are satisfied by any endomorphism T_ι in class II. Indeed, from Sec. 4, it is already known that $\mathbf{L}_K P(\alpha 0, 0, \gamma_0, \delta_0) \to P(\alpha_0 - 1, 0, \gamma_0 + 1, \delta_0)$ is a surjective linear map; it remains to show that $\mathbf{L}_D : P(\alpha_0 - 1, 0, \gamma_0 + 1, \delta_0) \to P(\alpha_0, 0, \gamma_0, \delta_0)$ is injective. For the bases (27) and (26) in $P(\alpha_0 - 1, 0, \gamma_0 + 1, \delta_0)$ and $P(\alpha_0, 0, \gamma_0, \delta_0)$, respectively, $\mathbf{L}_D$ is represented by the band matrix

$$\begin{bmatrix} \gamma_0+1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \delta_0 & \gamma_0+2 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \delta_0-1 & \gamma_0+3 & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \delta_0-\alpha_0+3 & \gamma_0+\alpha_0-1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & \delta_0-\alpha_0+2 & \gamma_0-\alpha_0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & \delta_0-\alpha_0+1 \end{bmatrix} \tag{43}$$

with α_0+1 rows and α_0 columns. Clearly, the determinant made of the first α_0 rows of (43) is not zero; hence $\mathbf{L}_D$ restricted to $P(\alpha_0-1, 0, \gamma_0+1, \delta_0)$ is injective. In application of formula (42), the computer program computes the group inverse

$$T_i^{\#} = (\mathbf{L}_D\mathbf{L}_K)^{\#} = \mathbf{L}_D(\mathbf{L}_K\mathbf{L}_D)^{-2}\mathbf{L}_K$$

by multiplying matrix (28) by matrix (43), squaring the product, taking its inverse, and multiplying the result to the left by matrix (43) and to the right by matrix (28). Considering that the matrices involved have dimensions of the order of half the degree n, one will admit that the manipulations have been brought down to an elementary level. Furthermore, by taking advantage of the invariance with respect to the symmetry Z, the task of constructing the group inverse has been cut by almost a half. Alternatively, the invariance may be exploited to check the results coming out of the program.

The fundamental results obtained in the course of discussing the restriction classes mentioned in Table IV are gathered in the following:

Proposition 4: Given a polynomial p in $\mathbf{P}_n$, there exists a unique polynomial q in Im $\mathbf{L}_D$ such that $p-\mathbf{L}_K q$ belongs to Ker $\mathbf{L}_D$. Also Zq belongs to Im $\mathbf{L}_D$, and it is the unique polynomial r in Im $\mathbf{L}_D$ such that $Zp-\mathbf{L}_K r$ belongs to Ker $\mathbf{L}_D$. Whenever $Zp=0$, then $Zq=0$.

Indeed p may be decomposed in a unique way as the sum $p=p_K+p_D$ with p_k in Im $\mathbf{L}_K$ and p_D in Ker $\mathbf{L}_D$. Hence there exists a polynomial q such that $\mathbf{L}_D p=\mathbf{L}_D\mathbf{L}_K q$. Taking $q=T^{\#}\mathbf{L}_D p$ guarantees that q belongs to Im T which, by virtue of Proposition 1, is identical to Im $\mathbf{L}_D$. If there is another polynomial q' with the same property, then, on the one hand, $q-q'$ belongs to Im $\mathbf{L}_D$, while, on the other hand, $q-q'$ belongs to Ker $T=$ Ker $\mathbf{L}_K$; but $\mathbf{P}_n$ is the direct sum of Ker $\mathbf{L}_K$ and Im $\mathbf{L}_D$, and hence $q-q'=0$. In view of the fact that the expressions (9) and (12) for $\mathbf{L}_K$ and $\mathbf{L}_D$ are symmetric in the pairs (x, X) and (y, Y), $\mathbf{L}_D Z=Z\mathbf{L}_D$ and $\mathbf{L}_K Z=Z\mathbf{L}_K$; therefore, Zq belongs to Im $\mathbf{L}_D$ and $\mathbf{L}_D(Zp-\mathbf{L}_K Zq)=\mathbf{L}_D(Z(p-\mathbf{L}_K q))=Z(\mathbf{L}_D(p-\mathbf{L}_K q))=0$. From what has been proved already, it results that Zq is the unique element in Im $\mathbf{L}_D$ to have that property. Finally, the last part of Proposition 4 is an immediate consequence of Lemma 6.

Incidentally the computer program which implements the algorithm produced at degree 3 a polynomial $q=T^{\#}\mathbf{L}_D p$, which differs from the solution in (38) by the quantity

$$\tfrac{2}{3}G(C_{0,0,2,1}X-C_{0,0,1,2}Y).$$

The discrepancy is admissible since, in agreement with Proposition 1, it is an element of Ker $\mathbf{L}_K$.

6. ELIMINATION OF THE SHORT-TERM EFFECTS

A dynamical system described by an Hamiltonian of type (14) is said to be in *normal form* if, for each $n>2$, the term H_n belongs to Ker L_D, that is, $\mathbf{H}_n$ is a homogeneous polynomial in the coordinates (x, y) and the angular momentum G (Proposition 3). For instance, a Hamiltonian whose potential energy depends only on the coordinates is in normal form; so there is nothing the present normalization can contribute to further its solution. Such is the case for the *monkey saddle*.[26] But, if a Hamiltonian of type (14) is not in normal form, then one can construct a canonical transformation $(x,y,X,Y)\to(x',y',X',Y')$ which will convert (14) into a series in normal form. It is proposed to construct the normalization as a Lie transformation, that is, as the flow of an Hamiltonian vector field

$$\frac{dx}{d\epsilon}=\frac{\partial W}{\partial X},\quad \frac{dy}{d\epsilon}=\frac{\partial W}{\partial Y},\quad \frac{dX}{d\epsilon}=-\frac{\partial W}{\partial x},\quad \frac{dY}{d\epsilon}=-\frac{\partial W}{\partial y}$$

derived from a series

$$W\equiv W(x,y,X,Y)=\sum_{n>0}\frac{1}{n!}W_{n+1}.$$

It will be shown that each term W_n in the generator may be obtained as a homogeneous polynomial of degree $n+2$ in the phase variables (x, y, X, Y), so that the normalization may be pursued in a recursive fashion from one degree to the next.

Starting with the infinitesimal contact transformation that is the infinitesimal symplectic transformation tangent to the Lie mapping, the first-order terms H_1' and W_1 in the new Hamiltonian and in the generator respectively are linked by the identity

$$(K;W_1)+H_1=H_1',$$

equivalent, as a matter of fact, to the algebraic relation

$$\mathbf{L}_K W_1+H_1'=H_1. \tag{44}$$

With the requirement that $\mathbf{L}_D H_1'=0$, the problem of solving (44) amounts to decomposing the homogeneous polynomial H_1 into its constituents in the direct sum $\mathbf{P}_3=$ Ker $\mathbf{L}_D$ $\oplus$ Im $\mathbf{L}_K$.

A change in notation is helpful in following the recursion rules for constructing the normalization past the infinitesimal contact transformation: $H_{n,0}$ will stand for H_n in (14), and $H_{0,n}$ for H_n'. Now assume that all terms W_i $(1\leqslant i\leqslant n-1)$ and $H_{i,j}$ $(0\leqslant i,j\leqslant n-1,\ i+j\leqslant n-1)$ have been

obtained. Then one is in a position to calculate first

$$\tilde{H}_{n-1,1} = H_{n,0} + \sum_{1 \leqslant j \leqslant n-1} \binom{n-1}{j-1} (H_{n-j,0}; W_j),$$

and thereafter, by decreasing values of i, the terms

$$\tilde{H}_{i,j} = \tilde{H}_{i+1,j-1} + \sum_{0 \leqslant k \leqslant i} \binom{i}{k} (H_{i-k,j-1}; W_{k+1})$$

for $i+j=n$, $0 \leqslant i,j \leqslant n$, the last term in the recursive chain being $\tilde{H}_{0,n}$. By prescription of a perturbation algorithm involving a Lie transformation, the terms W_n in the generator and $H_{0,n}$ in the transformed Hamiltonian must satisfy the partial differential identity

$$(H_0; W_n) + \tilde{H}_{0,n} = H_{0,n}. \tag{45}$$

But, if p and q are homogeneous polynomials of degree l and m, respectively, then their Poisson bracket

$$(p; q) = \frac{\partial p}{\partial x}\frac{\partial q}{\partial X} - \frac{\partial p}{\partial X}\frac{\partial q}{\partial x} + \frac{\partial p}{\partial y}\frac{\partial q}{\partial Y} - \frac{\partial p}{\partial Y}\frac{\partial q}{\partial y}$$

is a homogeneous polynomial of degree $l+m-2$. Hence all terms $\tilde{H}_{i,j}$, for $i+j=n$, are homogeneous polynomials of degree $n+2$, and the partial differential equation (45) is equivalent to the algebraic equation

$$\tilde{H}_{0,n} = L_K W_n + H_{0,n}.$$

With the normalization requirement that $\mathbf{L}_D H_{0,n} = 0$, the latter is solved by decomposing $\tilde{H}_{0,n}$ into its constituents in the direct sum $\mathbf{P}_{n+2} = \mathrm{Im}\,\mathbf{L}_K \oplus \mathrm{Ker}\,\mathbf{L}_D$.

For readers interested either in calculating by hand some of the coefficients in the normalized Hamiltonian, or in checking their automated normalization procedures, listed below are the terms in $H_{0,n}$ $(3 \leqslant n \leqslant 6)$ emanating from the right-hand members

$$\tilde{H}_{0,n} = \sum_{(\alpha,\beta,\gamma,\delta)} C_{\alpha,\beta,\gamma,\delta} x'^{\alpha} y'^{\beta} X'^{\gamma} Y'^{\delta}:$$

Degree 4:

$$\begin{aligned} H_{0,4} = {} & C_{4,0,0,0}x'^4 + \tfrac{1}{4}(3C_{3,0,0,1} - C_{2,1,1,0})x'^2G' \\ & + C_{3,1,0,0}x'^3y' + \tfrac{1}{2}(C_{2,1,0,1} - C_{1,2,1,0})x'y'G' \\ & + C_{2,2,0,0}x'^2y'^2 - \tfrac{1}{4}(3C_{0,3,1,0} - C_{1,2,0,1})y'^2G' \\ & + C_{1,3,0,0}x'y'^3 + \tfrac{1}{6}(2C_{2,0,0,2} - C_{1,1,1,1} + 2C_{0,2,2,0})G'^2 \\ & + C_{0,4,0,0}y'^4; \end{aligned}$$

Degree 5:

$$\begin{aligned} H_{0,5} = {} & C_{5,0,0,0}x'^5 + \tfrac{1}{5}(4C_{4,0,0,1} - C_{3,1,1,0})x'^3G' \\ & + C_{4,1,0,0}x'^4y' + \tfrac{1}{5}(3C_{3,1,0,1} - 2C_{2,2,1,0})x'^2yG' \\ & + C_{3,2,0,0}x'^3y'^2 - \tfrac{1}{5}(3C_{1,3,1,0} - 2C_{2,2,0,1})x'y'^2G' \\ & + C_{2,3,0,0}x'^2y'^3 - \tfrac{1}{5}(4C_{0,4,1,0} - C_{1,3,0,1})y'^3G' \\ & + C_{1,4,0,0}x'y'^4 + \tfrac{1}{6}(3C_{3,0,0,2} \\ & - C_{2,1,1,1} + C_{1,2,2,0})x'G'^2 \\ & + C_{0,5,0,0}y'^5 + \tfrac{1}{6}(3C_{0,3,2,0} \\ & - C_{1,2,1,1} + C_{2,1,0,2})y'G'^2; \end{aligned}$$

Degree 6:

$$\begin{aligned} H_{0,6} = {} & C_{6,0,0,0}x'^6 + \tfrac{1}{6}(5C_{5,0,0,1} - C_{4,1,1,0})x'^4G' \\ & + C_{5,1,0,0}x'^5y' + \tfrac{1}{3}(2C_{4,1,0,1} - 2C_{3,2,1,0})x'^3y'G' \\ & + C_{4,2,0,0}x'^4y'^2 + \tfrac{1}{2}(C_{3,2,0,1} - C_{2,3,1,0})x'^2y'^2G' \\ & + C_{3,3,0,0}x'^3y'^3 - \tfrac{1}{3}(2C_{1,4,1,0} - 2C_{2,3,0,1})x'y'^3G' \\ & + C_{2,4,0,0}x'^2y'^4 - \tfrac{1}{6}(5C_{0,5,1,0} - C_{1,4,0,1})y'^4G' \\ & + C_{1,5,0,0}x'y'^5 + \tfrac{1}{20}(12C_{4,0,0,2} \\ & - 3C_{3,1,1,1} + 2C_{2,2,2,0})x'^2G'^2 \\ & + C_{0,6,0,0}y'^6 + \tfrac{1}{10}(3C_{3,1,0,2} \\ & - 2C_{2,2,1,1} + 3C_{1,3,2,0})x'y'G'^2 \\ & + \tfrac{1}{20}(12C_{0,4,2,0} - 3C_{1,3,1,1} + 2C_{2,2,0,2})y'^2G'^2 \\ & + \tfrac{1}{12}(3C_{3,0,0,3} - C_{2,1,1,2} + C_{1,2,2,1} - 3C_{0,3,3,0})G'^3. \end{aligned}$$

Let the relativistic corrections for a free particle illustrate the normalization. The kinetic energy

$$E = m_0c^2(1 - v^2/c^2)^{-1/2}$$

expanded in powers of v^2, after division by the mass at rest m_0 and omission of the constant energy at rest mc^2, gives rise to the Hamiltonian

$$H = \tfrac{1}{2}v^2 - \tfrac{3}{8}v^4/c^2 + \tfrac{5}{16}v^6/c^4 - \cdots$$

with $v^2 = X^2 + Y^2$. Since

$$\mathbf{L}_K(xX + yY)v^2 = v^4,$$

there will be no term of degree 4 in the normalized Hamiltonian, and the generator of the infinitesimal contact transformation will be

$$W_1 = -\tfrac{3}{8}(v^2/c^2)(xX + yY).$$

Hence, by definition of an infinitesimal contact transformation

$$x = x' + (x'; W_1), \quad X = X' + (X'; W_1),$$
$$y = y' + (y'; W_1), \quad Y = Y' + (Y'; W_1),$$

the relativistic corrections to the first order in v^2/c^2 are

$$\Delta x' = -\tfrac{3}{8}(v'^2/c^2)x' + \tfrac{3}{4}(G'/c^2)Y', \quad \Delta X' = \tfrac{3}{8}(v'^2/c^2)X',$$
$$\Delta y' = -\tfrac{3}{8}(v'^2/c^2)y' - \tfrac{3}{4}(G'/c^2)X', \quad \Delta Y' = \tfrac{3}{8}(v'^2/c^2)Y'.$$

A more substantial application of the present scheme for normalization has been made to the restricted problem of three bodies at the equilateral point L_4 for Routh's critical mass ratio; but this topic requires too much background information in celestial mechanics to be related here.

Unexpected as it comes after a long excursion in representation theory of Lie algebras, the physical meaning of the normalization achieved in the present section is in fact very simple. To elucidate this point, the notations will be revised. First the normalized Hamiltonian will be decomposed in the usual way as the sum

$$H' = \tfrac{1}{2}(X'^2 + Y'^2) - U(x', y', G')$$

of a kinetic energy and of a force function U, with $G' = x'Y' - y'X'$ designating the angular momentum. Next, in order to eliminate ambiguities concerning the partial derivatives, the differential of U will be written as the 1-form

$$dU = \partial_1 U dx' + \partial_2 U dy' + \partial_3 U dG'.$$

In these notations, the equations of normalized motions be-

come

$$\dot{x}' = \frac{\partial H'}{\partial X'} = X' + y'\,\partial_3 U,$$

$$\dot{y}' = \frac{\partial H'}{\partial Y'} = Y' - x'\,\partial_3 U,$$

$$\dot{X}' = -\frac{\partial H'}{\partial x'} = \partial_1 U + Y'\,\partial_3 U,$$

$$\dot{Y}' = -\frac{\partial H'}{\partial y'} = \partial_2 U - X'\,\partial_3 U.$$

Consider now a Cartesian frame of reference consisting of an orthonormal basis $(\mathbf{i}', \mathbf{j}', \mathbf{k})$ rotating at the angular velocity $\boldsymbol{\omega} = \partial_3 U\,\mathbf{k}$ about the normal to the plane of motion; assume also that (x', y') represent the Cartesian coordinates of the particle in the plane $(\mathbf{i}', \mathbf{j}')$, or that $\mathbf{x} = x'\mathbf{i}' + y'\mathbf{j}'$. Under these conditions, the particle's velocity in a frame (C) fixed in the plane is the vector

$$\dot{\mathbf{x}} = \dot{x}'\mathbf{i}' + \dot{y}'\mathbf{j}' + \boldsymbol{\omega}\times\mathbf{x},$$

equal, by virtue of the normalized equations of motion, to the sum

$$\dot{\mathbf{x}} = X'\mathbf{i}' + Y'\mathbf{j}'.$$

Thus the conjugate momenta X' and Y' are the components in the moving frame of the particle's velocity with respect to the fixed frame. For the same reason,

$$\ddot{\mathbf{x}} = \dot{X}'\mathbf{i}' + \dot{Y}'\mathbf{j}' + \boldsymbol{\omega}\times\dot{\mathbf{x}} = \partial_1 U\,\mathbf{i}' + \partial_2 U\,\mathbf{j}',$$

which exhibits $\partial_1 U$ and $\partial_2 U$ as the projections of the force on the axes of the rotating frame. The normalization appears now as a procedure to extract the rate at which the frame of reference should be rotated in order to confer the equations of motion the simple form

$$\ddot{\mathbf{x}} = \partial_1 U\,\mathbf{i}' + \partial_2 U\,\mathbf{j}'.$$

From that standpoint, the normalization is closely analogous to a method devised by Hansen[27] for handling perturbed Keplerian systems in three dimensions. A slow rate of rotation is imparted to the frame of reference; its axis and its rate are adjusted at each instant so that the rotating frame constitutes what Hansen calls an *ideal* (i.e., conceptual or virtual) reference system. In the ideal frame, the particle's motion appears to be planar; the forces acting on the mass point are expressed as a two-dimensional gradient, the coupling between the planar motion and the rotation of the ideal frame being accounted for by making the force function dependent explicitly on the angular momentum.

The kinematical interpretation given here to the normalization is further clarified by looking at the equations of motion in the polar variables defined by the canonical extension

$$x' = r'\cos\theta', \quad X' = R'\cos\theta' - (\Theta'/r')\sin\theta',$$

$$y' = r'\sin\theta', \quad Y' = R'\sin\theta' + (\Theta'/r')\cos\theta'.$$

It results at once from the Cartesian equations of motion that

$$\frac{d}{dt}r' = R',$$

$$\frac{d}{dt}R' = \frac{\Theta^2}{r'^3} + \partial_1 U\cos\theta' + \partial_2 U\sin\theta',$$

$$\frac{d}{dt}\theta' = \frac{\Theta'}{r'^2} - \partial_3 U,$$

$$\frac{d}{dt}\Theta' = r'(\partial_2 U\cos\theta' - \partial_1 U\sin\theta').$$

The angle σ of slow rotation being determined by the 1-form $d\sigma = \partial_3 U dt$ and the radial and transversal components of the force being

$$P = \partial_1 U\cos\theta' + \partial_2 U\sin\theta',$$

$$Q = \partial_2 U\cos\theta' - \partial_1 U\sin\theta',$$

the equations in polar coordinates are given the standard form

$$\frac{d^2}{dt^2}r' = \frac{\Theta'^2}{r'^3} + P, \quad \frac{d}{dt}(\theta' + \sigma) = \frac{\Theta'}{r'^2}, \quad \frac{d}{dt}\Theta' = Q.$$

Thanks to the normalization, the particle appears to move under the exclusive action of the gradient of U with respect to the normalizing coordinates.

APPENDIX: CONSTRUCTION OF A GROUP-INVERSE

It has been shown in Sec. 5 that the decomposition of a polynomial p of $\mathbf{P}_n$ into its components in Ker $\mathbf{L}_D$ and Im $\mathbf{L}_K$ reduces to the construction of the group inverse $T^{\#}$ of the semisimple operator $\mathbf{L}_D\mathbf{L}_K$. The algorithm given in Sec. 5 for constructing $T^{\#}$ depends on special properties enjoyed by the action of T on $\mathbf{P}_n$: $\mathbf{P}_n$ can be written as the direct sum of the spaces $P(\alpha, \beta, \gamma, \delta)$, each of which is left invariant by T, and on each subspace $P(\alpha, \beta, \gamma, \delta)$ either T is invertible or the group inverse of the restriction of T can be computed via an explicit factorization of its matrix. The decomposition scheme adopted in Sec. 5 owes its effectiveness to the fact that the subspaces $P(\alpha, \beta, \gamma, \delta)$ are much smaller than $\mathbf{P}_n$: The dimension of $\mathbf{P}_n$ is $\binom{n+3}{3}$, hence of order n^3, while the dimension of $P(\alpha, \beta, \gamma, \delta)$ is at most $n/2 + 1$ (by the discussion preceding Lemma 3). The purpose of this Appendix is to sketch an alternate algorithm for constructing the group inverse of T. This algorithm hinges on the fact, to be proved next, that T is a *diagonalizable* operator whose eigenvalues are simple to find; it does not make use of any other properties of T or of $\mathbf{P}_n$.

Lemma: T is a diagonalizable operator whose eigenvalues consist of the products $(n - 2\gamma - i)(i + 1)$, with $0 \leqslant \gamma \leqslant [n/2]$ and $0 \leqslant i \leqslant n - 2\gamma$. The number of distinct eigenvalues is at most $1 + [(n + 1)/2][(n + 3)/2]/2 < 1 + (n + 2)^2/8$.

Proof: $\mathbf{P}_n$ may be written, as in (18), as the direct sum of certain subspaces V_k $(1 \leqslant k \leqslant s)$, each of which is invariant and irreducible under the set of operators $\{\mathbf{L}_D, \mathbf{L}_K, \mathbf{L}_S\}$. Furthermore, in each V_k, there is a basis $v_0 = v_0^{(k)}, \ldots, v_m = v_{m_k}^{(k)}$ satisfying relations (19)–(21) (with x, y, h replaced by $\mathbf{L}_D, \mathbf{L}_K, \mathbf{L}_S$, respectively). It follows from (20) and (21) that the v_i's $(0 \leqslant i \leqslant m_k)$ are a basis of eigenvectors for $T = \mathbf{L}_D\mathbf{L}_K$, with $Tv_i = (m_k - i)(i + 1)\,v_i$. Thus the eigenvalues of T on V_k consist precisely of the products $(m_k - i)(i + 1)$ (for $0 \leqslant i \leqslant m_k$). To see which m_k's arise, one may observe, on the one hand, that in each V_k the maximal weight vector $v_0^{(k)}$ is an eigenvector of $\mathbf{L}_S$ of eigenvalue m_k. Furthermore, since the s linearly independent vectors $v_0^{(k)}$ $(1 \leqslant k \leqslant s)$ form a basis of Ker $\mathbf{L}_D$ [by relation (21) and Proposition 1], the set $\{m_k:$

TABLE VII. Bounds of eigenvalues.

N	Number of eigenvalues	$1+[(N+1)/2]*[(N+3)/2]/2$	$1+[(N+2)^2/8]$
3	4	4	4
4	4	4	5
5	7	7	7
6	6	7	9
7	11	11	11
8	10	11	13
9	14	16	16
10	13	16	19
99	830	1 276	1 276
100	714	1 276	1 301
999	66 908	125 251	125 251
1000	55 918	125 251	125 501

$1\leqslant k\leqslant s\}$ consists of the eigenvalues which $\mathbf{L}_S$ takes on Ker $\mathbf{L}_D$. On the other hand, the nonomials $g(\alpha,\beta,\gamma)=x^\alpha y^\beta G^\gamma$ with $\alpha+\beta+2\gamma=n$ are, by Proposition 3, also a basis of Ker $\mathbf{L}_D$, and are eigenvectors of $\mathbf{L}_S$ as the following formula shows:

$$\mathbf{L}_S g(\alpha,\beta,\gamma)=(\alpha+\beta)g(\alpha,\beta,\gamma)=(n-2\gamma)g(\alpha,\beta,\gamma).$$

Thus the numbers m_k $(1\leqslant k\leqslant s)$ are exactly the numbers $n-2\gamma$ $(0\leqslant\gamma\leqslant[n/2])$. Finally, for a fixed γ, there are at most $[(n-2\gamma+1)/2]=[(n+1)/2]-\gamma$ distinct nonzero products among the numbers $(n-2\gamma-i)(i+1)$ (for $0\leqslant i\leqslant n-2\gamma$), so, in all, T has at most $1+\Sigma_{0<\gamma<(n/2)}$ $[(n+1)/2]-\gamma=1+[(n+1)/2][(n+3)/2]2$ distinct eigenvalues. It is easy to check that this number is always at most $1+(n+2)^2/8$.

A glance at Table VII indicates that the bound given in the lemma is far too generous.

It is perhaps worthwhile to note that a slight extension of the argument just presented yields a description of the decomposition of $\mathbf{P}_n$ into its irreducible components: For each γ satisfying $0\leqslant\gamma\leqslant[n/2]$, the irreducible sl(2, **R**)-module of highest weight $n-2\gamma$ occurs with the multiplicity $n-2\gamma+1$, and no other occur. This can be derived without difficulty from the fact that the polynomials $g(\alpha,\beta,\gamma)$ generate irreducible subspaces of highest weight $n-2\gamma$.

In order to state the main result of the appendix, it will be convenient to have some more notation. If λ is a scalar, let $\lambda^{\#}$ denotes λ^{-1} if $\lambda\neq 0$, and 0 otherwise. If $\lambda_0,\ldots,\lambda_t$ is a sequence of distinct scalars, then there is a unique polynomial $f(x)$ of degree $\leqslant t$ satisfying $f(\lambda_k)=\lambda_k^{\#}$ for $0\leqslant k\leqslant t$. If $f[\lambda_0,\ldots,\lambda_k]$ denotes the k th divided difference (constructed, for example, from the prescription $f[\lambda_0]=f(\lambda_0)$ and

$$f[\lambda_0,\ldots,\lambda_k]=\frac{f[\lambda_1,\ldots,\lambda_k]-f[\lambda_0,\ldots,\lambda_{k-1}]}{\lambda_k-\lambda_0}$$

for $k\geqslant 1$), then, as is well known, $f(x)$ can be written as

$$f(x)=f[\lambda_0]+\sum_{1<k<t}f[\lambda_0,\ldots,\lambda_k](x-\lambda_{k-1})\cdots(x-\lambda_0).$$

A number of authors[28] have observed that the group inverse $T^{\#}$ can be obtained as a polynomial in T. In the proof of the next proposition, it will be seen that, for the case considered here,

$$T^{\#}=f(T)=f[\lambda_0]I+\sum_{1<k<t}f[\lambda_0,\ldots,\lambda_k](T-\lambda_{k-1},I)\cdots(T-\lambda_0 I).$$

Proposition: Let T be a diagonalizable operator on a finite-dimensional vector space, and suppose the distinct eigenvalues of T are $\lambda_0,\ldots,\lambda_t$. For any vector v in V, $w=T^{\#}v$ can be computed by the following algorithm:

Set $v_0=v$, $w_0=f[\lambda_0]v_0$;

For $k=1$ **to** t,

set $v_k=(T-\lambda_{k-1})v_{k-1}$,

$w_k=w_{k-1}+f[\lambda_0,\ldots,\lambda_k]v_k$;

Then $w=w_t$.

[*N.B.*: The cycle in k may be stopped as soon as it encounters a k for which $v_k=0$.]

Proof: It is clear that w_k is just the k th partial sum of the sum

$$f(T)v=f[\lambda_0]v+\sum_{1<k<t}f[\lambda_0,\ldots,\lambda_k](T-\lambda_{k-1}I)\cdots(T-\lambda_0)v,$$

so $w_t=f(T)v$. It only remains to see that $f(T)v=T^{\#}v$ for all v in V. Since T is diagonalizable, V may be written as the direct sum $V_0+\cdots+V_t$ of the eigenspaces of T, where $V_k=\{v\in V: Tv=\lambda_k v\}$ for $0\leqslant k\leqslant t$. But the endomorphism of V which sends $u=u_0+\cdots+u_t$ $(u_k\in V_k$ for each $k)$ to $\lambda_0^{\#}u_0+\cdots+\lambda_t^{\#}u_t$ satisfies the relations for the group inverse of T, since $Tu=\lambda_0u_0+\cdots+\lambda_tu_t$. Thus by uniqueness of the group inverse, $T^{\#}u=\lambda_0^{\#}u_0+\cdots+\lambda_t^{\#}u_t$. On the other hand, $f(T)u=\Sigma f(T)u_t$, and the relation $Tu_k=\lambda_k u_k$ implies easily that $f(T)u_k=\lambda_k^{\#}u_k$ for each k. Hence $f(T)u=\Sigma\lambda_k^{\#}u_k=T^{\#}u$ for all u in V, and the proof is complete.

It is worthwhile to place this algorithm in a somewhat wider context. For, clearly, the method just outlined may be used to compute any polynomial in T, not just the specific polynomial which yields the group inverse. The range of applicability of this algorithm is thus determined exactly by the following theorem in linear algebra[29]:

Theorem: Let S and T be linear operators in a finite-dimensional vector space V. Then S may be written as a polynomial in T if and only if S commutes with every linear operator which commutes with T, that is, if and only if for every linear operator L in V, the relation $LT=TL$ implies $LS=SL$.

If S and T are diagonalizable, the theorem is quite easy to prove directly. The special case when T is diagonalizable and $S=T^{\#}$ has been confirmed in the proposition of the Appendix and in Lemma 6.

[1]E. T. Whittaker, Proc. London Math. Soc. **34**, 206–21 (1902).

[2]C. Delaunay, *Théorie du mouvement de la lune*, Mémoires de l'Institut Impérial des Sciences (Bachelet, Paris, 1863 and 1867).

[3]E. T. Whittaker, Proc. Royal Soc. Edinburgh **37**, 95–116 (1916). For an automated development of the generating function, see, e.g., F. G. Gustavson, Astron. J. **71**, 670–86 (1966).

[4]H. Poincaré, Bull. Astron. **3**, 57–61 (1886); C. R. Acad. Sci. Paris **108**, 21–4 (1889); *Les méthodes nouvelles de la mécanique céleste* (Gauthier-Villars,

Paris, 1893), Vol. 2.
[5]A. Deprit, J. Henrard, and A. Rom, Icarus **6**, 381–406 (1967).
[6]A. Deprit, Cel. Mech. **1**, 12–21 (1969); A. H. Nayfeh, *Perturbation Methods* (Wiley, New York, 1973); J. R. Cary, Phys. Rep. **79**, 131–59 (1981).
[7]E. T. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies* (Cambridge U. P., Cambridge, 1937).
[8]J. Williamson, Am. J. Math. **58**, 141–63 (1936).
[9]J.-C. van der Meer, Cel. Mech. **27**, 131–50 (1982).
[10]M. Kummer, Commun. Math. Phys. **48**, 53–79 (1976).
[11]A. Deprit, "Méthodes nouvelles de simplification en mécanique céleste," Publications du Département de Mathématique, Facultés Universitaires Notre Dame de la Paix, Namur, Belgium, 1980.
[12]R. Cushman and D. Rod, Physica D **6**, 105–12 (1982).
[13]A. Deprit, Cel. Mech. **26**, 9–21 (1982); Cel. Mech. **29**, 229–46 (1983).
[14]See, e.g., R. Cushman, Symp. Math. **14**, 323–42 (1974).
[15]E. Cartan, *Leçons sur la théorie des invariants intégraux* (Hermann, Paris, 1922).
[16]The theorem was first stated by V. V. Morozov in the C. R. (Dokl.) Acad. Sci. USSR **36**, 83–6 (1942), but proved later on by N. Jacobson in the Proc. Am. Math. Soc. **2**, 105–37 (1951), finally restated and proved again by V. V. Morozov in the Usp. Mat. Nauk **15**, 137–9 (1960).
[17]V. Arnold, *Mathematical Methods of Classical Mechanics* (Springer, New York, 1978), p. 211.
[18]J. E. Humphreys, *Introduction to Lie Algebras and Representation Theory* (Springer, New York, 1980), 3rd printing, rev.
[19]K. E. Iverson, *A Programming Language* (Wiley, New York, 1962).
[20]A. Ben-Israel and T. E. N. Greville, *Generalized Inverses: Theory and Applications* (Wiley, New York, 1974).
[21]C. R. Rao, *Linear Statistical Inference and Its Applications* (Wiley, New York, 1973), 2nd ed.
[22]I. Erdelyi, J. Math. Anal. Appl. **17**, 119–32 (1967).
[23]P. Robert, J. Math. Anal. Appl. **22**, 658–69 (1968).
[24]P. Drazin, Am. Math. Mon. **65**, 506–13 (1958).
[25]R. E. Cline, "An application of representations for the generalized inverse of a matrix," Technical Report 592, U. S. Army Mathematics Research Center, University of Wisconsin-Madison, 1965.
[26]D. Rod, J. Diff. Eq. **14**, 129–70 (1973).
[27]P. A. Hansen, *Fundamenta nova Investigationis Orbitae verae quam Luna perlustrat* (Glaser, Gotha, 1838); E. W. Brown, *An Introductory Treatise on the Lunar Theory* (Cambridge U. P., Cambridge, 1896); H. Andoyer, *Cours de mécanique céleste* (Gauthier-Villars, Paris, 1923); A. Deprit, J. Res. Nat. Bur. Std. Sec. B, **79**, 1–15 (1975).
[28]S. Campbell and C. Meyer, *Generalized Inverses of Linear Transformations* (Pitman, London, 1979), pp. 130–6.
[29]N. Jacobson, *Lectures in Abstract Algebra, Vol. II, Linear Algebra* (Van Nostrand, Princeton, NJ, 1953), p. 113.

A simple proof of a particular case of C. Siegel's center theorem[a)]

R. de la Llave[b)]
Institute for Mathematics and its Applications, University of Minnesota, Vincent Hall, 206 Church Street SE, Minneapolis, Minnesota 55455

(Received 1 June 1982; accepted for publication 4 March 1983)

We give an elementary proof of a particular case of C. Siegel's center theorem, based on a method of M. Herman. Even if the proof has less generality than the standard one, it is simpler and provides sharper bounds.

PACS numbers: 03.20. + i, 02.30. + g

I. INTRODUCTION

Recently, Herman[1] has introduced a method to treat several "small denonimator" problems.

This type of difficulty appears very frequently in perturbation expansions and the standard technique to overcome it systematically has become known—rather loosely—as K.A.M. theory.

One of the outstanding problems for these phenomena[2] is to obtain good estimates of the strength of the perturbations that make the results of the theory no longer hold. (Of course, the mechanisms of proof break down earlier than the conclusions.)

In that respect, we believe Herman's method is highly relevant. The proofs are much simpler and it is, therefore, easier to discuss optimality. Even without any attempt to optimization, the proofs yield much better constants that the ones obtained by applying straightforwardly the standard argument. (However, by slightly nonstandard ways of doing the estimates, Chierchia has obtained good constants in a particular example.[3]) The main shortcoming of the method of proof is that it seems to go through only under (much) stronger assumptions than the usual one. However, the way physicists have been looking at the problem is the analysis of particular examples. It turns out that the method applies to the ones that physicists have been considering as most relevant for the breakdown problem.

We observed that the same methods apply also to Siegel's center theorem and that, in that case, it was possible to make further modifications of the argument so that the proof became quite explicit and only used elementary techniques.

Even if this results in a certain loss of mathematical elegance, we have strived for elementarity and explicitness. This is consistent with the philosophy that there is a lot to learn from the analysis of concrete cases; we introduced a little bit of redundancy so that given a particular example, it should be possible to discuss the modifications that yield the best constants *and compute* them. The more mathematically inclined reader, interested only in qualitative theorems, without explicitly computing the constants, may advantageously substitute a good part of the computations in the first part by an invocation to Sobolev inequalities.

[a)] Supported in part by NSF PHY 8117463.

[b)] In absentia from Department of Mathematics, Princeton University, Princeton, New Jersey 08544.

II. SIEGEL'S CENTER THEOREM

Suppose we were given a function $f:C\to C$ leaving the origin fixed and analytic in a neighborhood of it.

$$f(z) = az + \sum_{k\geq 2} f_k z^k.$$

If we think of f as defining a dynamical evolution, it is natural to try to analyze the stability of the origin. It turns out[4] that the origin is stable (both backwards and forward in time) if and only if $|a| = 1$ and moreover, there is an analytic function conjugating f to its linear part and leaving the origin fixed.

$$\begin{aligned} &f\circ\phi(z) = \phi(az),\\ &\phi(0) = 0,\\ &\phi'(0) = 1. \end{aligned} \tag{1}$$

The existence of such conjugating ϕ is interesting in itself and we just remark that it is very easy to show it exists if $|a| \neq 1$. (So that in that case, dynamics is determined by the linear approximation.)

The following theorem is much more delicate.

Theorem 1 (Siegel): If a satisfies $|a^k - 1|^{-1} \leq \gamma k^\nu$, $\gamma, \nu > 0$, then Eq. (1) has a unique analytic solution.

The standard proof of this theorem can be found in (Ref. 4), where you can also find the reference to the original paper by Siegel (which was based on a different idea) and other applications. There are also generalizations to several complex variables[5,6] that, however, we will not discuss, since the methods we are going to use do not seem to carry over to several variables.

What we are going to prove in this paper is a *weaker version* of this theorem.

Theorem 1': If a satisfies $|a^k - 1|^{-1} \leq \gamma k$, then Eq. (1) has a unique analytic solution.

We point out that uniqueness follows from identification of coefficients, so that the only nontrivial part to prove is convergence of the series thus obtained.

We also want to point out that there is no a, with modulus 1, which satisfies the inequality in Theorem 1 if $\nu < 1$. (We state the theorem that way to emphasize that the proof would, obviously, go through.) When $\nu > 1$, the numbers satisfying an inequality of this form have full measure on the unit circle. For the critical case $\nu = 1$, even if it is a set of zero Lebesgue measure, it is nonempty and, indeed, contains very interesting numbers; it consists precisely of the rotations

 0022-2488/83/082118-04$02.50

with winding number having a bounded continued fraction expansion. See Refs. 7 and 8 and references therein for a more detailed discussion of these numbers and their abundance.

Proof of Theorem 1': We first remark that, once we assume the boundary conditions, the conjugacy equation is equivalent to the one we obtain taking derivatives, namely

$$f'\circ\phi(z)\phi'(z) = a\phi'(az).$$

Since $\log f'$ is an analytic function in a neighborhood of the origin, we may require as well

$$\log f'\circ\phi(z) - \log a = \log\phi'(az) - \log\phi'(z).$$

Calling $\log f'(z) - \log a$, $h(z)$ we are led to the study of

$$h\circ\phi(z) = \log\phi'(az) - \log\phi'(z).$$

The second remark is that we can modify our problem so that we only have to consider h's defined on arbitrarily big neighborhoods and being there arbitrarily small with all their derivatives.

In effect, if we can solve our problem for $f_\lambda(z) = (1/\lambda) f(\lambda z)$ (λ any number) in place of f, we have also solved the original problem (in another neighborhood).

Taking $|\lambda|$ small enough, we may assume $\log f'_\lambda(z) - \log a$ is analytic in any ball we please, and it satisfies there any smallness conditions we want. We will state the ones needed in the place of the proof where they are used.

The strategy of the proof is, as in Herman's method, to study the operator τ_h that sends the function ψ into the function ϕ satisfying

$$\begin{aligned} &h\circ\psi = \log\phi'(az) - \log\phi'(z),\\ &\phi(0) = 0,\\ &\phi'(0) = 1, \end{aligned} \tag{2}$$

and to show that, under suitable smallness assumptions on h, it maps a compact convex set into itself and is continuous on it so that, by the Schauder–Tychonov theorem, it has a fixed point that solves our problem.

A convenient choice for the set on which to study τ_h is

$$A = \Big\{\psi \mid \psi(z) = \sum_{k\geqslant 0}\psi_k z^k,\ \psi_0 = 0,\ \psi_1 = 1,\ \sum_{k\geqslant 2}|\psi_k|^2 k^2(k-1)^2 \leqslant \tfrac{1}{4}\Big\}.$$

Of course, there is nothing special about $\frac{1}{4}$; we picked it not to clutter the proof with choices.

We first prove several properties of functions in A, and afterwards, will construct τ_h step by step, keeping good track of the estimates.

Lemma 1: All ψ in A have radius of convergence bigger or equal to 1 and satisfy

$$\sup_{|z|<1}|\psi(z)| \leqslant \frac{1}{2}\left(\sum_{k\geqslant 2}\frac{1}{k^2(k-1)^2}\right)^{1/2} + 1 \equiv C_1,$$

$$\sup_{|z|<1}|\psi'(z)| \leqslant \frac{1}{2}\left(\sum_{k\geqslant 2}\frac{1}{(k-1)^2}\right)^{1/2} + 1 \equiv C_2,$$

The proof is an obvious application of Cauchy–Schwartz inequality.

From this, we can deduce the helpful representation

$$\sum_{k\geqslant 2}|\psi_k|^2 k^2(k-1)^2 = \sup_{r<1}\frac{1}{2\pi}\int_{|z|=r}|\psi''(z)|^2\,dz,$$

which is proved by expanding the integrand in powers and taking into account the orthogonality relations between them.

By a very slight, using H_2 theory, abuse of notation, I will denote the rhs simply by $\int_{|z|=1}$.

We endow A with the usual topology of sets of analytic functions. In this topology A is a compact set, since by Lemma 1 it is contained in a compact (uniformly bounded set) and is closed. (The two first conditions are obviously closed and the third one is also closed from the integral representation for it.) A is also convex.

In the light of Lemma 1, $h\circ\psi$ will always be defined provided h is defined in $\{|z|\leqslant C_1\}$. That is the first smallness condition in h to be imposed.

The integral representation renders elementary the proof of (uniform for $\psi\in A$) estimates for $h\circ\psi$.

$$(h\circ\psi)'' = h''\circ\psi(\psi')^2 + h'\circ\psi\psi'',$$

$$\begin{aligned}\sum_{k\geqslant 2}|(h\circ\psi)_k|^2 k^2(k-1)^2 &= \frac{1}{2\pi}\int_{|z|=1}|(h\circ\psi)''|^2\,ds\\ &\leqslant\Big[C_2^2\sup|h''(z)|\\ &\quad+\frac{1}{2}\sup_{|z|<1}|h'(z)|\Big]^2 \leqslant \epsilon_1^2,\end{aligned} \tag{3}$$

$$|(h\circ\psi)_1| = |h_1| \leqslant \epsilon_2,$$

$$(h\circ\psi)_0 = 0.$$

The second smallness conditions to be imposed in h are those stated above. ϵ_1, ϵ_2 are small numbers which depend only on a; the explicit form of the dependence will be given later.

Lemma 2: If g is analytic in the ball of radius 1, $g(0) = 0$, there is one and only one η analytic in the ball of radius 1 (we will write $\eta = \Gamma g$) satisfying

$$g(z) = \eta(az) - \eta(z) \quad \forall|z|<1,$$

$$\eta(0) = 0.$$

Moreover, if $\Sigma|g_k|^2 k^2(k-1)^2 \leqslant \epsilon_1^2$, $|g_1| \leqslant \epsilon_2$, then

$$\sum_{k\geqslant 2}|\eta_k|^2(k-1)^2 \leqslant \gamma^2\epsilon_1^2,$$

$$|\eta_1| \leqslant \gamma\epsilon_2.$$

Proof: Identifying coefficients we have

$$\eta_k = g_k/(a^k - 1).$$

Use the assumption on a to estimate absolute values. ■

Therefore, in the same way as before, we obtain

$$\sup_{|z|<1}|\eta(z)| \leqslant \gamma\epsilon_1\left(\sum_{k\geqslant 2}\frac{1}{(k-1)^2}\right)^{1/2} + \gamma\epsilon_2,$$

$$\frac{1}{2\pi}\int_{|z|=1}|\eta'(z)|^2\,ds = \sum_{k\geqslant 1}|\eta_k|^2 k^2 \leqslant \gamma^2\epsilon_2^2 + 4\gamma^2\epsilon_1^2.$$

It is clear that

$$\tau_h\psi(z) = \int_0^z \exp(\Gamma(h\circ\psi))(W)\,dW.$$

So that we have automatically

$$(\tau_h\psi)_0 = 0,$$

$$(\tau_h\psi)_1 = 1.$$

The only thing we still have to check to prove $\tau_h(A)\subset A$ is

$$\frac{1}{2\pi}\int_{|z|=1} |(\tau_h\psi)''(z)|^2\, ds \leqslant \tfrac{1}{4} \quad \forall\psi\in A.$$

But, for such ψ,

$$|(\tau_h\psi)''(z)| = |(\tau h\circ\psi)'(z)|\,|\exp(\tau h\circ\psi)(z)|,$$

$$\leqslant |(\tau h\circ\psi)'(z)|\exp\gamma(\epsilon_1\pi/\sqrt{6}+\epsilon_2).$$

So that our goal is achieved whenever

$$\tfrac{1}{4}\geqslant(\gamma^2\epsilon_2^2+4\gamma^2\epsilon_1^2)\exp 2\gamma[\epsilon_2+(\pi/\sqrt{6})\epsilon_1],$$

which can obviously be satisfies for certain ϵ's bigger than 0, therefore, when h satisfies Eq. (3) with these ϵ's in the rhs, τ_h maps A into A.

As we remarked in the beginning, these smallness conditions for h can always be adjusted by scaling our original function f.

Once we have that τ_h maps a compact set into itself, it is easy to show it is continuous. When this is satisfied, continuity is the same as closedness of the graph, but the points in the graph are those pairs (ψ,ϕ) satisfying Eq. (2), which is obviously a closed condition.

Remark 1: There are other choices of sets A which are also acceptable for the proof. The one that was used was selected because it resembles the Sobolev spaces used in other theorems, but has the integral representation which makes elementary and explicit some of the steps. There are, however, other possibilities, e.g.,

$$A = \left\{\psi(z)\,|\,\psi_0 = 0,\ \psi_1 = 1,\ \sum_{k\geqslant 2}|\psi_k|k\leqslant\frac{1}{2}\right\}.$$

Let us sketch briefly how you can adapt the steps of the proof.

The only uniform estimate we are going to use is

$$\sup_{|z|\leqslant 1}|\psi(z)|\leqslant\tfrac{2}{3}\quad\forall\psi\in A.$$

A clearly is convex and compact with the usual topology of analytic functions of the unit disc. By the previous estimate it is equibounded and, therefore, contained in a compact set. It is also closed since

$$A = \bigcap_{n=3}\left\{\psi(z)\,|\,\psi_0 = 0,\ \psi_1 = 1\sum_{n\geqslant k\geqslant 2}|\psi_k|k\leqslant\tfrac{1}{2}\right\},$$

which are closed conditions.

If h is an analytic function in a ball of radius bigger than $\frac{3}{2}$, then h is analytic in a ball of radius 1. Identifying coefficients and using the triangle inequality in all the sums, we obtain

$$n|(h\psi)_n|$$

$$\leqslant n|h_1|\psi_n| + |h_2|\sum_{m_1+m_2=n}|\psi_{m_1}||\psi_{m_2}|(m_1+m_2)$$

$$+\cdots+|h_n|\sum_{m_1+\cdots+m_n=n}|\psi_{m_1}|\cdots|\psi_{m_n}|(m_1+\cdots+m_n).$$

Since

$$\sum_{n\geqslant 1}\sum_{m_1+\cdots+m_j=n}|\psi_{m_1}|\cdots|\psi_{m_j}|(m_1+\cdots+m_j)$$

$$= j\left(\sum_m m|\psi_m|\right)\left(\sum_m|\psi_m|\right)^{j-1},$$

we clearly have

$$\sum_{n\geqslant 1}n|(h\circ\psi)_n|\leqslant\sum_{n\geqslant 1}n|h_n|(\tfrac{3}{2})^n.$$

From that, we easily obtain

$$\sum_{n\geqslant 1}|\eta_n|\leqslant\gamma\sum_{n\geqslant 1}n|h_n|(\tfrac{3}{2})^n.$$

By the same method, we obtain also

$$\sum_{n\geqslant 0}|(\exp\eta)_n|\leqslant\exp\gamma\sum_{n\geqslant 1}n|h_n|(\tfrac{3}{2})^n,$$

since

$$\sum_{n\geqslant 0}|(\exp\eta)_n| = \sum_{n\geqslant 1}n|\phi_n|.$$

It suffices to impose the following smallness condition to h:

$$\exp\gamma\sum_{n\geqslant 1}n|h_n|(\tfrac{3}{2})^n\leqslant\tfrac{3}{2}.$$

The argument for continuity is, obviously, the same.

Let me discuss briefly how to treat a simple case, namely,

$$f(z) = az + z^2$$

with $a = \exp 2\pi i\frac{1}{2}(\sqrt{5}-1)$.

The reason why we picked such an a is because, in some well-defined sense, it is "the most irrational number" and it is very convenient for small denominator problems. Besides, it has many remarkable properties, one being that we can compute the best acceptable value of γ.

In effect, the convergents of the continued fraction expansion are the quotient of two consecutive Fibonacci numbers. Moreover, they converge exponentially fast. The first means that in order to compute the best constant, we only have to worry about the Fibonacci numbers. The second means that it suffices to check a finite number of them.

Doing the argument in detail, it turns out that the worst resonance is the first so that

$$\gamma = \sup_k(1/k)|a^k-1|^{-1} = |a-1|^{-1}\approx 0.54.$$

In this case, everything can be computed.

$$f'_\lambda(z) = a + 2\lambda z,$$

$$h(z) = \sum_{n\geqslant 1}\frac{(-1)^{n+1}}{n}\left(\frac{2\lambda}{a}\right)^n z^n.$$

The condition given at the end becomes

$$\exp(3\lambda/(1-3\lambda))\leqslant\tfrac{3}{2}.$$

It suffices to take

$$\lambda\leqslant\frac{\log\frac{3}{2}}{2\times\frac{3}{2}(\gamma+\log\frac{3}{2})}\approx 0.143.$$

Of course, we could have chosen other numbers instead

of $\frac{3}{2}$ in the definition of A. Call them e^u. Then, the condition becomes

$$\lambda = ue^{-u}/2(\gamma + u).$$

It is not very difficult to find the optimal choice of u. It also gives that

$$\lambda = 0.143$$

suffices.

Of course, this can be improved even further by a more sophisticated tailoring of the proof like using the estimates (more refined than the one used here) in Ref. 8.

The reason why it is interesting to choose λ as big as possible is because it gives us information about the size of the domain of stability. This domain (Siegel domain) for f_λ is the range of ϕ. But since ϕ is injective (this is true, in general, from results in Ref. 4, but for our case it suffices to remark that it is a perturbation of the identity with Lipschitz constant less than 1), the area of the domain can be computed as follows:

$$\text{meas}(\text{Range } \phi) \geqslant \int_{|z|<1} |\phi'|^2 \, dx \, dy = \pi \sum_k k \, |\phi_k|^2 \geqslant 1,$$

so that the area of the Siegel domain for f is bigger or equal than $\pi\lambda^2$.

We can also obtain other pieces of information about the Siegel domain. For example, if the set A is chosen as in the proof given in Remark 1,

$$A = \left\{ \psi(z) \mid \psi_0 = 0, \ \psi_1 = 1, \ \sum_{k>2} |\psi_k| k < e^u - 1 \right\}.$$

It can be readily proved that all the functions in this set are such that their range contains a ball around the origin of radius $r = 1 - \frac{1}{2}(e^u - 1) = \frac{3}{2} - e^u/2$. Therefore, the Siegel domain for f should contain a ball around the origin of radius

$$\lambda r = (3 - e^u)ue^{-u}/4(\gamma + u).$$

Taking u = 0.33, we obtain, for the particular case we are discussing, that the Siegel domain contains a ball of radius

$$\lambda r = 0.110.$$

On the other hand, since the Siegel domain cannot contain in its interior any point in the orbit of the critical point, we can see that the Siegel domain does not contain any ball centered in the origin and with radius bigger than $\frac{1}{4}$. (The image of the critical point is $-a^2/4$.) Actually, taking a few iterations with the help of a pocket calculator (nine suffice), this upper bound can be improved to 0.22.

Notice the fact that the Siegel domain cannot contain a ball of radius $\frac{1}{4}$ about 0 remains valid for all the functions of the form $f(z) = az + z^2$, where a is any number of modulus 1. If we put that together with the lower estimate of the radius given γ, we obtain that all possible γ have to satisfy

$$\gamma \geqslant u(3e^{-u} - 2).$$

For $u = 0.2$, the rhs becomes 0.0912. This is a rather surprising way of establishing a number theoretic results. In Ref. 9, p. 164, a related number is computed. Their results are enough to establish

$$\gamma \geqslant \sqrt{5}/2\pi \approx 0.355.$$

In any case, we know that $= 0.54$ was possible for the number we considered before, and this is also a rather satisfactory confirmation that the bounds we obtained are fairly good.

Remark 2: After the preceding proof was completed, I received a letter from M. Herman pointing out that he was aware of the existence of such simple proofs and had even written notes for a seminar. He also mentioned that he and R. Douady observed that it is not necessary to use the Schauder–Tychonov theorem; using stronger smallness conditions in h, it suffices to use the contraction mapping principle. This observation also applies to other proofs.

ACKNOWLEDGMENTS

This work would have not been a possible if I had not attended the Les Houches session of July 1981. I profited very much from conversations with M. Herman, J. Mather, and D. Goroff. R. McKay gave a very useful critical reading to the manuscript.

[1] M. Herman, lectures given at Les Houches, July 1981.
[2] W. Thirring, *A Course in Mathematical Physics, Vol. I* (Springer, New York, 1978), p. 139.
[3] L. Chierchia, private communication with G. Gallavotti.
[4] C. L. Siegel and J. Moser, *Lectures on Celestial Mechanics* (Springer, New York, 1971).
[5] E. Zehnder, in *Lecture Notes in Mathematics*, Vol. 597 (Springer, New York, 1977).
[6] V. Arnold, *Chapitres Suplementaires de la Theorie des Equations Differetielles Ordinaires* (Mir, Moscow, 1980).
[7] M. Herman, IHES **49**, 5–233 (1979).
[8] H. Rüsmann, in *Lecture Notes in Physics*, Vol. 38 (Springer, New York, 1975).
[9] G. Hardy and E. Wright, *An Introduction to the Theory of Numbers* (Oxford U. P., Oxford, 1979).

Measure for measure: Covariant skeletonizations of phase space path integrals for systems moving on Riemannian manifolds

Karel Kuchař
Department of Physics, The University of Utah, Salt Lake City, Utah 84112

(Received 13 October 1982; accepted for publication 25 February 1983)

We define phase space path integrals for systems moving on a Riemannian manifold and subject to a generalized potential by a skeletonization procedure which is manifestly covariant under point transformations. We achieve this goal by introducing a natural analog $S(x'',t''|x',p',t')$ of the Hamilton principal function with phase space initial data. One class of such functions is based on the parallel transport of momentum, a second class is obtained by a modification of the first class, and a third class is based on the geodesic deviation transport of momentum. The third class of principal functions is geometrically privileged. We skeletonize the canonical action integral by replacing it by a manifestly covariant chain of phase space principal functions. Different functions lead to the same functional as we infinitely refine the skeletonization along a smooth path. Our phase space path integral is always taken with the natural Liouville measure, but the integration over momentum variables brings down a nontrivial measure to the remaining configuration space path integral. Because nondifferentiable rather than smooth paths dominate the path integral, different phase space principal functions generate different configuration space path measures. Such measures lead to quantum propagators which satisfy Schrödinger's equations with all possible scalar curvature terms $\sim\hbar^2R$. The geometrically privileged phase space principal function (based on the geodesic deviation transport) leads to the Schrödinger equation without any curvature term.

PACS numbers: 03.65.Ca, 02.40.Ky

1. INTRODUCTION

Path integrals as introduced by Feynman[1] were originally able to handle only systems moving in flat space. It lasted almost a decade before DeWitt took the decisive step and generalized the method to Newtonian systems moving on a curved manifold.[2] In a clear and elegant manner, DeWitt showed how to skeletonize the configuration space path integrals by a prescription which is manifestly covariant under transformation of coordinates. There are two ingredients which enter into DeWitt's prescription: One must decide on the skeletonization of the action integral, and one must choose a measure in the space of skeletonized paths. The first decision is easy, because it follows naturally from the Hamilton–Jacobi theory: The vertebral points of the skeletonized path are connected by segments which are actual classical paths of the system, and the action integral is thereby approximated by a chain of Hamilton's principal functions between the vertebral points. Each of these functions is a biscalar in its arguments. This enables one to keep track of the covariance throughout the procedure.

The second decision is more difficult. It involves the measure which must be assigned to each segment of the skeletonized path. DeWitt followed Pauli's reformulation[3] of Feynman's method and chose the square root of the Van Vleck–Morette determinant[4] as his measure. With this measure he derived the Schrödinger equation for the quantum propagator and discovered a surprising fact that the Hamilton operator in this equation contains an extra term $\frac{1}{12}\hbar^2R$ proportional to the scalar curvature. DeWitt also generalized Feynman's measure to curved Riemannian manifolds and found that under this measure the path integral yields a different term, $\frac{1}{6}\hbar^2R$, in the Hamilton operator. He observed that, to obtain the standard Hamilton operator without any curvature term, one should modify the Lagrangian by a potential which amounts to $-\frac{1}{12}\hbar^2R$ for the Pauli measure and $-\frac{1}{6}\hbar^2R$ for the Feynman measure. Unfortunately, such a compensation potential brings the Planck constant into classical Lagrangian and takes thus much away from the intuitive appeal of Feynman's theory. An alternative attitude is to accept extra terms in the Hamilton operator as standard ambiguities depending on the factor ordering and live happily with them ever after.

In the past few years, another way was found out of this blind alley. In 1979, Parker[5] investigated path integrals with measure proportional to an arbitrary power of the Van Vleck–Morette determinant and discovered that he can generate in this way an arbitrary term $\sim\hbar^2R$ in the Hamilton operator. In particular, he discovered that for the first power of the determinant the curvature term drops out from the Hamilton operator. At the same time, DeWitt–Morette, *et al.*[6] introduced this kind of measure by a reasoning based on stochastic differential equations, prodistributions, and the development mapping and argued that it follows naturally from this framework.

In recent times, phase space path integrals became increasingly popular due to the influence of Fadeev's investigations.[7] Because there is an unquestionably privileged measure on the phase space, namely, the Liouville measure $d^n x d^n p$, one can ask whether this resolves in a natural way the ambiguity discovered by DeWitt. Unfortunately, while the notation falsely suggests that phase space path integrals are covariant under canonical transformations, this is merely an illusion.[8] Skeletonization spoils covariance, or at least

 0022-2488/83/082122-20$02.50

its manifest exhibition. It is even difficult to maintain manifest covariance of phase space path integrals under point transformations.[9] Of course, one can always start from the evolution operator whose Hamilton operator is covariantly ordered and, by inserting the decomposition of unity in the momentum eigenstates, cast the quantum propagator into the form of a phase space path integral.[10] However, the classical Hamiltonian which appears in the canonical action functional of this path integral does not look covariant at all; it contains terms with the Planck constant which are not scalars under point transformations.[11]

One would like to generalize DeWitt's skeletonization to phase space path integrals, but it is not obvious how to do it. The Hamilton principal function $S(x'',t''|x',t')$ is defined as the action integral evaluated along the actual path between x',t' and x'',t''. In phase space, there is in general no actual path connecting x',p',t' with x'',p'',t'' and we cannot define $S(x'',p'',t''|x',p',t')$ as the canonical action integral evaluated along nonexistent actual path. There are, of course, principal functions $S(x'',t''|p',t')$ and $S(p'',t''|x',t')$ with mixed boundaries, but they do not have simple properties under point transformations. For example, the $S(x'',t''|p',t')$ function differs from the biscalar function $S(x'',t''|x',t')$ by a boundary term $x^{a'}p_{a'}$, which is obviously not a scalar. Intuitively, it is difficult to apply the point transformation to the momentum boundary because we do not know the position at which the momentum is sitting. For such reasons, it is hard to make an often used alternating $x-p$ skeletonization manifestly covariant.

Our main goal is to define phase space path integrals for systems moving on a Riemannian manifold by a skeletonization procedure which is manifestly covariant under point transformations. We achieve this goal by finding a natural analog $S(x'',t''|x',p',t)$ of the Hamilton principal function with phase space initial data. In fact, there are several classes of such functions: one class based on the parallel transport of momentum, the second class obtained from the first one by certain modifications, and still another class based on the geodesic deviation transport of momentum. Any such phase space principal function allows a manifestly covariant skeletonization of the canonical action functional. As we refine the skeletonization along a smooth phase space path, we recover the canonical action functional. In this respect, all principal functions are equal. However, when it comes to the method of their construction, the principal function based on the true geodesic deviation transport of momentum is, like an Orwellian pig, more equal than the others. We explain why is the geodesic deviation transport privileged in Sec. 6.

Our phase space path integral is always taken with the natural measure $\Pi_{(K)}\,[d^{n}x_{(K)}\,d^{n}p_{(K)}/(2\pi\hbar)^{n}]$, but the integration over momentum variables brings down a nontrivial measure to the remaining configuration space path integral. Because the paths which contribute most to the path integral are nondifferentiable paths rather than smooth paths, different phase space principal functions bring down different configuration space path measures. By taking different principal functions, we generate quantum propagators which satisfy Schrödinger's equations with all possible curvature terms $\sim\hbar^{2}R$. The geometrically privileged phase space principal function (based on the geodesic deviation transport), however, leads naturally to the Hamilton operator without any curvature term. Within the phase space path skeletonization approach, it is the counterpart of the measure introduced by DeWitt–Morette, *et al.*[6]

Two other points which we want to mention in this introduction belong to the realm of esthetics rather than to the realm of facts. First, DeWitt's procedure is manifestly covariant, except in one minor detail: The expansions of biscalars like the Hamilton principal function or the Van Vleck–Morette determinant are carried in differences of coordinates rather than in powers of a covariant object. We want to maintain the covariance throughout our argument, and we thus use covariant expansions in powers of the first derivatives of Synge's "world function." This enables us also to use the Riemann normal coordinates without actually using the Riemann normal coordinates, i.e., without working in a special coordinate system. The final results are always reached by taking coincidence limits.

Second, Feynman's original proof that the propagator represented by a path integral satisfies the Schrödinger equation relies heavily on the specific properties of the classical Lagrangian. On the other hand, Pauli's presentation of Feynman's method derives the same result entirely from the Hamilton–Jacobi equation for the Hamilton principal function. In the spirit of Pauli, all our results are derived from the Hamilton–Jacobi equation for the principal function expanded in powers of the time interval.

Ultimately, we want to find a heuristically acceptable definition of phase integrals in canonical geometrodynamics. Canonical geometrodynamics is an infinitely dimensional system with an inherent Riemannian structure. However, unlike a Newtonian system treated in this paper, the canonical data are restricted by constraints associated with gauge and by other constraints associated with presence of an intrinsic time. As the first step in this direction, we shall treat in forthcoming papers[12] simple model systems with constraints associated either with gauge or with time parametrization.

2. HAMILTON'S PRINCIPAL FUNCTION. INTERPOLATION, SKELETONIZATION

The actual path of a dynamical system extremizes the action

$$S[x(t)]=\int_{t'}^{t''}dt\,l(x,\dot{x}) \tag{2.1}$$

in comparison with all neighboring kinematically possible paths which pass through the same initial position x' at the same initial time t' and through the same final position x'' at the same final time t'',

$$x^{a}(t'')=x^{a''},\quad x^{a}(t')=x^{a'},\quad a=1,\ldots,n. \tag{2.2}$$

We use primes to distinguish points and to indicate points to which various tensor or bitensor indices are referring. As a rule, we are often suppressing indices, especially in arguments of functions. As we vary the path, the action changes

by

$$\delta S\,[x(t)] = p_{a''}\,\delta x^{a''} - p_{a'}\,\delta x^{a'} + \int_{t'}^{t''} dt\,\frac{\delta S}{\delta x^a(t)}\,\delta x^a(t). \tag{2.3}$$

Here, $p_a(x,\dot{x})$ is the momentum

$$p_a = l_{,\dot{a}}(x,\dot{x}) \tag{2.4}$$

and $\delta S/\delta x^a(t)$ is the variational derivative of the action integral (2.1),

$$\frac{\delta S}{\delta x^a(t)} = l_{,a} - d_t l_{,\dot{a}}. \tag{2.5}$$

From the variational formula (2.3) we learn that the actual path between the fixed boundary points (2.2) satisfies the Euler–Lagrange equations

$$\frac{\delta S}{\delta x^a(t)} = l_{,a} - d_t l_{,\dot{a}} = 0. \tag{2.6}$$

The actual path passing through sufficiently close boundary points is uniquely determined by the system (2.6) of second-order differential equations with boundary conditions (2.2). The action integral evaluated along the actual path and considered as a function of the boundary data is the Hamilton principal function $S(x'',t''|x',t')$. As we vary the boundary points x' and x'' at fixed times t' and t'', but always connect them by the actual path (2.6), we conclude from Eq. (2.3) that

$$\begin{aligned} S_{,a''}\,\delta x^{a''} + S_{,a'}\,\delta x^{a'} &\equiv \delta S(x'',t''|x',t') \\ &= p_{a''}\,\delta x^{a''} - p_{a'}\,\delta x^{a'}. \end{aligned} \tag{2.7}$$

This enables us to obtain the momenta at the boundary points of the actual path from the Hamilton principal function,

$$p_{a''} = S_{,a''}(x'',t''|x',t'), \quad p_{a'} = -S_{,a'}(x'',t''|x',t'). \tag{2.8}$$

Let us move now the boundary points x',x'' along an actual path $x(t)$ while evaluating the Hamilton principal function,

$$S(x''(t''),t''|x'(t'),t') = \int_{t'}^{t''} dt\, l(x(t),\dot{x}(t)). \tag{2.9}$$

By differentiating Eq. (2.9), we see that

$$d_{t''} S(x''(t''),t''|x',t') = l(x'',\dot{x}'')$$

and (2.10)

$$d_{t'} S(x'',t''|x'(t'),t') = -l(x',\dot{x}').$$

On the other hand,

$$d_{t''} S(x''(t''),t''|x',t') = S_{,a''}\,\dot{x}^{a''} + \partial_{t''} S$$

and (2.11)

$$d_{t'} S(x'',t''|x'(t'),t') = S_{,a'}\,\dot{x}^{a'} + \partial_{t'} S.$$

From Eqs. (2.8)–(2.11) we conclude that the Hamilton principal function satisfies the Hamilton–Jacobi equation

$$\partial_{t''} S + H(x^{a''}, S_{,a''}) = 0 \tag{2.12}$$

in the final arguments and a similar equation

$$-\partial_{t'} S = H(x^{a'}, -S_{,a'}) = 0 \tag{2.13}$$

in the initial arguments. In both of these equations, H is the Hamiltonian

$$\begin{aligned} H(x,p) &\equiv [p_a\dot{x}^a - l(x,\dot{x})]_{\dot{x}=\dot{x}(x,p)}, \\ p_a &= l_{,\dot{a}}(x,\dot{x}(x,p)) \end{aligned} \tag{2.14}$$

of the system.

Let now the final position x'' approach the initial position x' along an arbitrary smooth path $x(t)$, $x(t') = x'$, not necessarily the actual path. As the actual path between x',t' and $x(t)$, t is shrinking to a point, the final velocity $v^a(t)$ on this path is approaching $\dot{x}^{a'}$. Consequently,

$$\lim_{t\to t'} S(x(t),t|x',t') = 0 \tag{2.15}$$

and

$$\lim_{t\to t'} d_t S(x(t),t|x',t') = l(x(t'),\dot{x}(t')). \tag{2.16}$$

From here,

$$S(x(t),t|x',t') = l(x(t'),\dot{x}(t'))\cdot\tau + o(\tau^2), \tag{2.17}$$

where

$$\tau = t - t'.$$

Extremization of the action functional (2.1) yields the actual path $x(t)$ between the boundary points x' and x''. It may happen, however, that we are not interested in the whole path, but only in the actual position x which the system assumes at a time $t\in[t',t'']$ on its motion from x' to x'' (the interpolation problem). To answer this question, we do not need to extremize a functional of the path $x^a(t)$, but only a function of the point x^a, namely,

$$S(x'',t''|x,t|x',t') \equiv S(x'',t''|x,t) + S(x,t|x',t'). \tag{2.18}$$

Indeed, $S(x'',t''|x,t|x',t')$ is the action functional evaluated along a broken path consisting from the actual path from x',t' to x,t followed by the actual path from x,t to x'',t''. This possible path becomes the actual path when the interpolated point x^a extremizes the action, i.e., when

$$S_{,a}(x'',t''|x,t|x',t') \equiv S_{,a}(x'',t''|x,t) + S_{,a}(x,t|x',t') = 0. \tag{2.19}$$

This condition has a simple meaning: The momenta (2.8) must match at the interpolated point,

$$p_a(x'',t''|x,t) = p_a(x,t|x',t'). \tag{2.20}$$

Continuing the interpolation process, we can ask what are the actual positions $x_{(1)}, x_{(2)}, \ldots, x_{(N-1)}$, which the system assumes at times $t_{(1)}, t_{(2)}, \ldots, t_{(N-1)}$ on its motion from $x_{(0)} \equiv x'$ at $t_{(0)} \equiv t'$ to $x_{(N)} \equiv x''$ at $t_{(N)} \equiv t''$. The answer is a simple extension of Eq. (2.15): The actual positions extremize the function

$$S(x_{(A)},t_{(A)}) \equiv \sum_{K=0}^{N-1} S(x_{(K+1)},t_{(K+1)}|x_{(K)},t_{(K)}) \tag{2.21}$$

with respect to the interpolated position $x_{(I)}$, $I = 1,\ldots,N-1$ while the boundary positions $x_{(0)}$ and $x_{(N)}$ are kept fixed.

The action functional (2.1) along a smooth path $x(t)$ can be recovered as a continuous limit of the function (2.21). Take a division $t_{(A)}$ of the interval $[t',t'']$ and skeletonize the path $x(t)$, connecting each point $x(t_{(K)})$ with the successive point $x(t_{(K+1)})$ by the actual path. To the skeletonized path

there belongs the skeletonized action

$$S(x(t_{(A)}),t_{(A)}) = \sum_{K=0}^{N-1} S(x(t_{(K+1)}),t_{(K+1)}|x(t_{(K)}),t_{(K)}). \tag{2.22}$$

Any link $S(x(t_{(K+1)}), t_{(K+1)}|x(t_{(K)}), t_{(K)})$ of this action can be expanded in the time interval $\tau_{(K)} = t_{(K+1)} - t_{(K)}$ by Eq. (2.17). We refine the skeletonization of the path $x(t)$ so that $\tau_{\max} = \max_{K=0,\ldots,N-1} \tau_{(K)} \to 0$. In this limit,

$$\lim_{\tau_{\max}\to 0} S(x(t_{(A)}),t_{(A)}) = \lim_{\tau_{\max}\to 0} \left\{\sum_{K=0}^{N-1} (l(x(t_{(K)}),\dot{x}(t_{(K)}))\tau_{(K)} + o(\tau^2_{(K)}))\right\} \equiv \int_{t'}^{t''} dt\, l(x(t),\dot{x}(t)) \equiv S[x(t)]. \tag{2.23}$$

The action functional is thereby defined through the Hamilton principal function.

The configuration space formalism is manifestly covariant under transformation of coordinates,

$$x^{\bar{a}} \to x^{\bar{a}} = x^{\bar{a}}(x^b). \tag{2.24}$$

Under such transformations, the Lagrangian $l(x,\dot{x})$ is a scalar, the action functional is an invariant, the momentum (2.4) and the variational derivative (2.5) are covectors, the Hamilton principal function $S(x'',t''|x',t')$ is a biscalar, the functions (2.18) and (2.21) are multiscalars, and all our equations are tensor equations holding in an arbitrary system of coordinates.

Let us pass now from the configuration space to the phase space of the system by adjoining the momenta p_a to the configuration variables x^a. First, we replace the action (2.1) by a new action

$$S[x(t),p(t),v(t)] \equiv \int_{t'}^{t''} dt\,(l(x,v) + p_a(\dot{x}^a - v^a)), \tag{2.25}$$

in which the configuration variables x^a are varied under fixed boundary conditions (2.2), while the Lagrange multipliers p_a and v^a have their boundaries free. By varying p_a, we learn that $v^a = \dot{x}^a$. When we substitute this constraint into the new action (2.25), we reduce it back to the Lagrangian action (2.1). On the other hand, when we vary v^a, we get the connection $p_a = l_{,\dot{a}}(x,v)$ between the velocity v^a and the momentum p_a. Using this connection, we can eliminate v^a from the action (2.25) and arrive thus at the canonical action functional

$$S[x(t),p(t)] = \int_{t'}^{t''} dt\, L(x,\dot{x},p), \tag{2.26}$$

$$L(x,\dot{x},p) \equiv p_a\dot{x}^a - H(x,p). \tag{2.27}$$

The canonical Lagrangian $L(x,\dot{x},p)$ is related to the ordinary Lagrangian $l(x,\dot{x})$ by the Legendre dual transformation

$$L\left(x, \frac{\partial H}{\partial p}, p\right) = l\left(x, \frac{\partial H}{\partial p}\right), \qquad L\left(x,\dot{x}, \frac{\partial l}{\partial \dot{x}}\right) = l(x,\dot{x}). \tag{2.28}$$

The actual motion of the system in phase space extremizes the canonical action in comparison with neighboring motions passing through the same initial and final configuration (2.2). The variations of x and p lead to the Hamilton equations of motion

$$\dot{x}^a = \frac{\partial H}{\partial p_a}, \tag{2.29}$$

$$\dot{p}_a = -\frac{\partial H}{\partial x^a}. \tag{2.30}$$

The elimination of p^a reduces the Hamilton equations back to the Lagrange equations (2.5) and (2.6).

The actual path in the phase space is determined by the initial and final configurations x' and x''. When these configurations are given, the initial and final momenta p' and p'' cannot be freely specified, but are determined from Eqs. (2.8). The actual momentum $p_a(t)$ along the actual path $x^a(t)$ is fixed by Eq. (2.29). In general, there is no actual path $x(t)$, $p(t)$ in the phase space which would connect the overdetermined initial data x',p',t' with the overdetermined final data x'',p'',t'' or even with the final data x'',t''. Here is another description why an attempt to find such a path fails: Evaluate the canonical action integral (2.26) along the actual path $x(t)$ between x',t' and x'',t'' for a test distribution $p(t)$ of momentum which matches the given values p' and p'' at the ends. We know that the action (2.26) is extremized by the momentum distribution (2.29) which, at the ends, has different values of p, given by Eq. (2.8). To extremize the action, the test distribution $p(t)$ should slip as fast as it can from the specified values p', p'' to the actual values (2.29) and run through them for the rest of the configuration path $x(t)$. However, there is no fastest slip in the space of continuous functions $p(t)$, and thus the attempt fails.

Because there is no actual path in the phase space connecting overdetermined boundary data, there is no obvious analog $S(x'',p'',t''|x',p',t')$, or even $S(x'',t''|x',p',t')$, of the Hamilton principal function. Still, we need such a function to solve the interpolation and skeletonization problem in the phase space. Indeed, our goal is to construct an algorithm with the following properties:

I. Interpolation: The actual position x^a and momentum p_a which the system assumes at a time $t \in [t',t'']$ on its motion through the phase space from the initial configuration $x^{a'}$ at t' to the final configuration $x^{a''}$ at t'' is obtained by extremizing a function

$$S(x'',t''|x,p,t|x',t'). \tag{2.31}$$

This function is a triscalar in the arguments $x^{a''}; x^a, p_a; x^{a'}$ under point transformations

$$x^{\bar{a}} = x^{\bar{a}}(x^b), \quad p_{\bar{a}} = \frac{\partial x^b(x^{\bar{c}})}{\partial x^{\bar{a}}} p_b \tag{2.32}$$

induced by the coordinate transformations (2.24).

II. Skeletonization: The actual positions $x_{(I)}$ and momenta $p_{(I)}$, $I = 1, 2, \ldots, N-1$ which the system assumes at times $t_{(I)}$ on its motion through the phase space from the initial configuration $x_{(0)} \equiv x'$ at $t_{(0)} \equiv t'$ to the final configuration $x_{(N)} \equiv x''$ at $t_{(N)} \equiv t''$ is obtained by extremizing a func-

tion

$$S(x_{(N)},t_{(N)}|x_{(I)},p_{(I)},t_{(I)}|x_{(0)},t_{(0)}) \tag{2.33}$$

in the canonical variables $x_{(I)}, p_{(I)}$. This function is a multiscalar in the arguments $x_{(0)}; x_{(I)}, p_{(I)}; x_{(N)}$ under point transformations (2.32).

III. Continuous limit: The canonical action functional (2.26)–(2.27) evaluated along a smooth path $x(t), p(t)$ in the phase space leading from the configuration x' at t' to the configuration x'' at t'' is obtained as a limit

$$S[x(t),p(t)] = \lim_{\tau_{max}\to 0} S(x'',t''|x(t_{(I)}),p(t_{(I)}),t_{(I)}|x',t') \tag{2.34}$$

of the skeletonized action (2.33) when $N\to\infty$ while $\tau_{\max} = \max_{K=0,\ldots,N-1} \tau_{(K)}\to 0$.

We find such an algorithm for a dynamical system described by a quadratic Lagrangian

$$l = \tfrac{1}{2}g_{ab}(x)\dot{x}^a\dot{x}^b + A_a(x)\dot{x}^a - V(x) \tag{2.35}$$

or a corresponding Hamiltonian

$$H = \tfrac{1}{2}g^{ab}(x)(p_a - A_a)(p_b - A_b) + V(x). \tag{2.36}$$

For simplicity, we assume that the nondegenerate metric $g_{ab}(x)$, the vector potential $A_a(x)$, and the scalar potential $V(x)$ do not depend on time. It is, however, straightforward to generalize the formalism to time-dependent metrics and potentials.

A prototype of our dynamical system is a system of nonrelativistic particles subject to holonomic scleronomic constraints and moving under the influence of forces possessing a generalized potential $V - A_a\dot{x}^a$. The potentials A_a and V may include a contribution from "ficticious" forces due to a noninertial character of the reference frame. The masses of the particles are absorbed into the metric g_{ab} and the charge parameters are absorbed into the potentials A_a and V.

To build the algorithm, we find first a phase space principal function $S(x'',t''|x',p',t')$ with overdetermined initial data. The functions (2.31) and (2.33) are then constructed from this principal function. The phase space principal function is not unique. There are many such functions, each of them leading to an algorithm with the desired properties I–III. Classically, it does not matter which one we are using to skeletonize the canonical action integral, because the differences among them disappear in the continuous limit. However, these differences become important in quantum mechanics, because the paths which contribute most to a path integral are not differentiable. When we use the skeletonized action (2.33) with the natural phase space measure $\Pi_{(K)}(d^n x_{(K)}\, d^n p_{(K)}/(2\pi\hbar)^n)$ for the definition of the phase space path integral, different choices of the principal function lead to different quantum propagators $\langle x'',t''|x',t'\rangle$. Each of these propagators satisifies a Schrödinger equation, but the Hamilton operator may contain a potential term $\sim\hbar^2 R$ proportional to the scalar curvature. However, there is a geometrically natural phase space principal function for which the scalar curvature term drops out from the Hamilton operator.

The construction of various phase space principal functions is most transparent for a free system ($A_a = 0 = V$). Such a system follows a geodesic in a curved space $g_{ab}(x)$. We shall discuss a free system first in the following four sections.

3. GEODESIC MOTION: COVARIANT EXPANSIONS

For a free system,

$$l(x,\dot{x}) = \tfrac{1}{2}g_{ab}(x)\dot{x}^a\dot{x}^b. \tag{3.1}$$

The actual motion of the system, by the Euler–Lagrange equations (2.6), is a geodesic

$$\frac{\delta S}{\delta x^a(t)} = \nabla_t\dot{x}_a = 0. \tag{3.2}$$

The system moves along the geodesic (3.2) with a constant velocity

$$v = (g_{ab}\dot{x}^a\dot{x}^b)^{1/2}, \quad \nabla_t v = 0. \tag{3.3}$$

This allows us to express the Hamilton principal function in terms of the geodesic separation $\sigma(x'',x')$ between the points x' and x'' and the time interval $\tau = t'' - t'$:

$$S(x'',t''|x',t') = \tfrac{1}{2}v^2\tau = \tfrac{1}{2}\sigma^2\tau^{-1}. \tag{3.4}$$

We introduce the biscalar

$$\Sigma(x''|x') \equiv \tfrac{1}{2}\sigma^2(x''|x') \tag{3.5}$$

called the world function[13] and write the Hamilton principal function as the ratio

$$S(x'',t''|x',t') = \Sigma(x''|x')/\tau. \tag{3.6}$$

Time and position variables separate in the expression (3.6). The Hamilton–Jacobi equations (2.12)–(2.13)

$$\begin{aligned} &\partial_{t''}S + \tfrac{1}{2}g^{a''b''}S_{,a''}S_{,b''} = 0, \\ &-\partial_{t'}S + \tfrac{1}{2}g^{a'b'}S_{,a'}S_{,b'} = 0 \end{aligned} \tag{3.7}$$

then imply the equations

$$\tfrac{1}{2}g^{a''b''}\Sigma_{,a''}\Sigma_{,b''} = \Sigma = \tfrac{1}{2}g^{a'b'}\Sigma_{,a'}\Sigma_{,b'} \tag{3.8}$$

for the world function. By differentiation

$$\Sigma_{,a'b''}\Sigma^{,b''} = \Sigma_{,a'}, \quad \Sigma_{,a''b'}\Sigma^{,b'} = \Sigma_{,a''} \tag{3.9}$$

and also

$$\Sigma_{|a'b'}\Sigma^{,b'} = \Sigma_{,a'}, \quad \Sigma_{|a''b''}\Sigma^{,b''} = \Sigma_{,a''}. \tag{3.10}$$

As the initial position x' approaches the final position x'', $x'\to x''$, any bitensor $W_{a'\ldots b''\ldots}$ at x' and x'' passes into a corresponding tensor at x'',

$$[W_{a'\ldots b''\ldots}] = \lim_{x'\to x''} W_{a'\ldots b''\ldots}(x''|x'), \tag{3.11}$$

called the coincidence limit of $W_{a'\ldots b''\ldots}(x''|x')$. In this notation, any index of a bitensor enclosed in the [ ] bracket becomes a tensor index at x''.

From its definition (3.5) in terms of the geodesic separation, it is obvious that

$$[\Sigma] = 0. \tag{3.12}$$

Further, from Eqs. (2.4), (2.8), and (3.6) we can express the velocity

$$\dot{x}_{a'} = p_{a'} = -S_{,a'} = -\tau^{-1}\Sigma_{,a'} = -v\cdot\sigma^{-1}\Sigma_{,a'} \tag{3.13}$$

in terms of its mgnitude v and the unit vector

$$u^{a'} = -[d_s x^a]_{s=s'} = \sigma^{-1}\Sigma^{,a'}, \quad g_{a'b'}u^{a'}u^{b'} = 1 \tag{3.14}$$

tangent to the geodesic and pointing from x'' to x'. Similar equations hold at the final boundary,

$$\dot{x}_{a''} = -vu_{a''}, \quad u^{a''} = -[d_s x^a]_{s=s''} = -\sigma^{-1}\Sigma^{,a''}. \tag{3.15}$$

From Eqs. (3.14) and (3.15) we conclude that the coincidence limits

$$[\sigma^{-1}\Sigma_{,a'}] = -[\sigma^{-1}\Sigma_{,a''}] = u_{a''} \tag{3.16}$$

stay finite; $u_{a''}$ in Eq. (3.16) can be interpreted as the unit vector tangent to the curve along which x' approaches x''. The derivatives $\Sigma_{,a'}$ and $\Sigma_{,a''}$ are thus of the same order as σ and

$$[\Sigma_{,a'}] = 0 = [\Sigma_{,a''}]. \tag{3.17}$$

When we divide Eq. (3.9) by σ and take the coincidence limit, we learn that

$$[\Sigma_{,a'b''}]u^{b''} = -u_{a''}. \tag{3.18}$$

On the assumption[13] that the coincidence limit $[\Sigma_{,a'b''}]$ does not depend on the direction in which x' approaches x'', we infer from Eq. (3.18) that

$$[\Sigma_{,a'b''}] = -g_{a''b''}. \tag{3.19}$$

Similarly,

$$[\Sigma_{|a'b'}] = g_{a''b''} = [\Sigma_{|a''b''}]. \tag{3.20}$$

We can take covariant derivatives of Eqs. (3.9) and (3.10) to any order and pass then to the coincidence limits.[13] By this recursive procedure we conclude that all coincidence limits in which Σ gets differentiated three times vanish,

$$[\Sigma_{|a'b'c'}] = [\Sigma_{|a'b'c''}] = [\Sigma_{|a'b''c''}] = [\Sigma_{|a''b''c''}] = 0, \tag{3.21}$$

while those in which it is differentiated four times can be expressed through the Jacobi curvature tensor J_{abcd}:

$$[\Sigma_{|a''b''c''d''}] = [\Sigma_{|a'b'c'd'}] = [\Sigma_{|a'b'c''d''}] = J_{a''b''c''d''}, \tag{3.22}$$

$$[\Sigma_{|a''b'c'd'}] = -J_{a''d''b''c''}, \quad [\Sigma_{|a''b''c''d'}] = J_{a''b''c''d''}.$$

Our conventions are

$$R^{a}{}_{bcd} \equiv -\Gamma^{a}{}_{bc,d} + \Gamma^{a}{}_{bd,c} - \Gamma^{i}{}_{bc}\Gamma^{a}{}_{id} + \Gamma^{i}{}_{bd}\Gamma^{a}{}_{ic} \tag{3.23}$$

and

$$J_{abcd} \equiv -\tfrac{1}{3}(R_{acdb} + R_{adbc}). \tag{3.24}$$

When we have a biscalar $F(x''|x')$ and know the coincidence limits of its covariant derivatives $F_{|a'}, F_{|a'b'}, F_{|a'b'c'}, \cdots$, we can write a covariant expansion of $F(x''|x')$ in powers of the geodesic separation $\sigma(x''|x')$, i.e., we can express $F(x''|x')$ in the form

$$F(x''|x') = F(x'') + F^{a''}(x'')\Sigma_{,a''}(x''|x') + F^{a''b''}(x'')\Sigma_{,a''}(x''|x')\Sigma_{,b''}(x''|x') + \cdots. \tag{3.25}$$

To do that, we replace the coordinates $x^{a'}$ by the new coordinates

$$y_{b''} = -\Sigma_{,b''}(x''|x'). \tag{3.26}$$

By Eq. (3.15), $y_{b''}$ are nothing else but the Riemann normal coordinates of the point x' based on the origin x''. The Jacobi matrix of the transformation (3.26) is

$$\frac{\partial y_{b''}}{\partial x^{a'}} = -\Sigma_{,a'b''}. \tag{3.27}$$

In a finite neighborhood of x'', $\Sigma_{,a'b''}$ is a regular matrix, and we can thus introduce its inverse,

$$\Sigma^{a'c''}\Sigma_{,b'c''} = \delta^{a'}_{b'}, \quad \Sigma^{c'a''}\Sigma_{,c'b''} = \delta^{a''}_{b''}. \tag{3.28}$$

Of course,

$$\frac{\partial x^{a'}}{\partial y_{b''}} = -\Sigma^{a'b''}. \tag{3.29}$$

We can now calculate the partial derivatives of the function $F(x''|x'(x'',y''))$ with respect to y''. First of all,

$$\frac{\partial F}{\partial y_{a''}} = F_{,b'}\cdot\frac{\partial x^{b'}}{\partial y_{a''}} = -F_{|b'}\Sigma^{b'a''}. \tag{3.30}$$

Further,

$$\begin{aligned}\frac{\partial^2 F}{\partial y_{b''}\,\partial y_{a''}} &= -\frac{\partial(F_{|c'}\Sigma^{c'a''})}{\partial y_{b''}} \\ &= (F_{|c'}\Sigma^{c'a''})_{,d'}\Sigma^{d'b''} = (F_{|c'}\Sigma^{c'a''})_{|d'}\Sigma^{d'b''} \\ &= F_{|c'd'}\Sigma^{c'a''}\Sigma^{d'b''} + F_{|c'}\Sigma^{c'a''}{}_{|d'}\Sigma^{d'b''}.\end{aligned} \tag{3.31}$$

By differentiating Eq. (3.25), we obtain

$$\Sigma^{c'a''}{}_{|d'} = -\Sigma^{c'i''}\Sigma_{|i''j''d'}\Sigma^{j''a''}. \tag{3.32}$$

From here,

$$\frac{\partial^2 F}{\partial y_{b''}\,\partial y_{a''}} = F_{|c'd'}\Sigma^{c'a''}\Sigma^{d'b''} - F_{|c'}\Sigma^{c'i''}\Sigma_{|i''j''k'}\Sigma^{j''a''}\Sigma^{k'b''}. \tag{3.33}$$

In this manner, we can calculate the y'' derivatives to an arbitrary order.

From Eq. (3.26) we see that putting $y'' = 0$ is equivalent to taking the coincidence limit $x' \to x''$. Equations (3.30) and (3.33) thus give

$$\left[\frac{\partial F}{\partial y_{a''}}\right]_{y''=0} = [F^{|a'}] \tag{3.34}$$

and

$$\left[\frac{\partial^2 F}{\partial y_{b''}\,\partial y_{a''}}\right]_{y''=0} = [F^{|a'b'}]. \tag{3.35}$$

Therefore,

$$F = [F] + [F^{|a'}]y_{a''} + \tfrac{1}{2}[F^{|a'b'}]y_{a''}y_{b''} + o(\sigma^3), \tag{3.36}$$

where, returning to the original variables, we can write $-\Sigma_{,a''}(x''|x')$ in place of $y_{a''}$:

$$F(x''|x') = [F] - [F^{|a'}]\Sigma_{,a''} + \tfrac{1}{2}[F^{|a'b'}]\Sigma_{,a''}\Sigma_{,b''} + o(\sigma^3). \tag{3.37}$$

The procedure remains unchanged when F is a tensor at x'', $F_{m'_1\cdots m'_y}(x''|x')$. When F carries any indices at x', one must first convert them into indices at x'' by the parallel propagator $g^{a'}_{b''}$ and only then expand the homogenized tensor in the powers of σ. We shall not need to consider such a case in this paper.

Apply now the expansion formula (3.37) to the tensor $\Sigma_{|c''d''}(x''|x')$:

$$\Sigma_{|c''d''} = [\Sigma_{|c''d''}] - [\Sigma^{|a'}{}_{c''d''}]\Sigma_{,a''} + \tfrac{1}{2}[\Sigma^{|a'b'}{}_{c''d''}]\Sigma_{,a''}\Sigma_{,b''} + o(\sigma^3). \tag{3.38}$$

Using the coincidence limits (3.19)–(3.22), we get

$$\Sigma_{|c''d''} = g_{c''d''} - \tfrac{1}{3}R_{c''}{}^{a''}{}_{d''}{}^{b''}\Sigma_{,a''}\Sigma_{,b''} + o(\sigma^3). \tag{3.39}$$

By contracting this equation, we learn that

$$\Delta''\Sigma - n = -\tfrac{1}{3}R^{a''b''}\Sigma_{,a''}\Sigma_{,b''} + o(\sigma^3). \tag{3.40}$$

In flat space, $\Sigma = \frac{1}{2}\delta_{ab}(x^{a''} - x^{a'})(x^{b''} - x^{b'})$ and $\Delta''\Sigma - n = 0$. Equation (3.40) shows that the deviation of $\Delta''\Sigma$ from its flat space value n depends in the second order of σ on the Ricci tensor $R^{a''b''}$.

Later on, we shall also need the expanded form of the determinants

$$\Sigma_{\cdot\cdot} \equiv \det \Sigma_{|a''b''}, \quad \Sigma_{\cdot\cdot} \equiv \det \Sigma_{|a'b'}, \tag{3.41}$$

$$\Sigma_{\cdot\cdot} \equiv \det(-\Sigma_{,a'b''}). \tag{3.42}$$

In particular, the last determinant is the Jacobian of the transformation (3.26) to the Riemann normal coordinates. The determinants (3.41)–(3.42) are biscalar densities with the weights (2'',0'), (0'',2') and (1'',1'). Sometimes, it is more convenient to replace them by the corresponding biscalars

$$D_{\cdot\cdot} \equiv g'^{-1}\Sigma_{\cdot\cdot}, \quad D_{\cdot\cdot} \equiv g'^{-1}\Sigma_{\cdot\cdot},$$
$$D_{\cdot\cdot} \equiv g'^{-1/2}g''^{-1/2}\Sigma_{\cdot\cdot}. \tag{3.43}$$

By taking the determinant of Eq. (3.39) and using the trace rule,

$$\Sigma_{\cdot\cdot} = g''(1 - \tfrac{1}{3}R^{a''b''}\Sigma_{,a''}\Sigma_{,b''}) + o(\sigma^3). \tag{3.44}$$

Mutatis mutandis,

$$\Sigma_{\cdot\cdot} = g'(1 - \tfrac{1}{3}R^{a'b'}\Sigma_{,a'}\Sigma_{,b'}) + o(\sigma^3). \tag{3.45}$$

The factor $R^{a'b'}\Sigma_{,a'}\Sigma_{,b'}$ is a biscalar; as such, it can be expanded in powers of $\Sigma_{,a''}$ according to Eq. (3.37). Quite predictably,

$$R^{a'b'}\Sigma_{,a'}\Sigma_{,b'} = R^{a''b''}\Sigma_{,a''}\Sigma_{,b''} + o(\sigma^3). \tag{3.46}$$

We can thus also write

$$\Sigma_{\cdot\cdot} = g'(1 - R^{a''b''}\Sigma_{,a''}\Sigma_{,b''}) + o(\sigma^3). \tag{3.47}$$

To expand the last determinant $\Sigma_{\cdot\cdot}$, we expand first the bitensor

$$F_{c''d''}(x''|x') \equiv \Sigma_{,m'c''}g^{m'n'}\Sigma_{,n'd''}. \tag{3.48}$$

From the coincidence limits (3.19)–(3.22),

$$[F_{c''d''}] = g_{c''d''}, \quad [F_{c''d''|a'}] = 0,$$
$$[F_{c''d''|a'b'}] = -J_{a''b''c''d''}, \tag{3.49}$$

and hence

$$F_{c''d''} = g_{c''d''} + \tfrac{1}{3}R^{a''}{}_{c''}{}^{b''}{}_{d''}\Sigma_{,a''}\Sigma_{,b''} + o(\sigma^3). \tag{3.50}$$

By taking the determinant of Eqs. (3.48) and (3.50), we learn that

$$\Sigma_{\cdot\cdot} = g'^{-1/2}g''^{1/2}(1 + \tfrac{1}{6}R^{a''b''}\Sigma_{,a''}\Sigma_{,b''}) + o(\sigma^3). \tag{3.51}$$

We can summarize these results by listing the biscalars (3.43):

$$D = 1 + \tfrac{1}{3}\lambda R^{a''b''}\Sigma_{,a''}\Sigma_{,b''} + o(\sigma^3),$$
$$\lambda_{\cdot\cdot} = -1 = \lambda_{\cdot\cdot}, \quad \lambda_{\cdot\cdot} = \tfrac{1}{2}. \tag{3.52}$$

4. GEODESIC MOTION: PHASE SPACE PRINCIPAL FUNCTION FROM PARALLEL TRANSPORT OF MOMENTUM

We return now to the phase space interpolation problem for a free system. From the Hamiltonian

$$H = \tfrac{1}{2}g^{ab}(x)p_a p_b \tag{4.1}$$

we obtain the Hamilton equations

$$\dot{x}^a = p^a \tag{4.2}$$

and

$$\dot{p}_a = -\tfrac{1}{2}g^{bc}{}_{,a}(x)p_b p_c. \tag{4.3}$$

The first equation is covariant under point transformations (2.26), while the second equation is not, unless taken in conjunction with the first one. Together, the Hamilton equations (4.2)–(4.3) imply the geodesic equation

$$\nabla_t \dot{x}_a = 0 \tag{4.4}$$

for the actual path in configuration space. The second Hamilton equation can be then replaced by a covariant equation obtained by covariantly differentiating the first Hamilton equation,

$$\nabla_t p_a = 0. \tag{4.5}$$

The actual momentum (4.2) is certainly propagated along the actual path (4.4) by the parallel transport (4.5). However, Eq. (4.5) enables us to propagate any momentum p_a, not only the actual one, from its initial value $p_{a'}$ at x',t' along the actual path (4.4) to x'',t''. This is possible because the new system of equations, (4.4)–(4.5), is not strictly equivalent to the old system of equations, (4.2)–(4.3): Hamilton's equations (4.2)–(4.3) imply Eqs. (4.4)–(4.5), but Eqs. (4.4)–(4.5) imply the Hamilton equations only under the intital condition

$$p^{a'} = \dot{x}^{a'}. \tag{4.6}$$

Trivial as this change may be, the new system of equations (4.4)–(4.5) enables us to define a phase space principal function $S(x'',t''|x',p',t')$. We take the actual path in the configuration space—the geodesic (4.4)—connecting the configuration x' at t' with the configuration x'' at t''. We then parallel propagate an arbitrarily assigned initial momentum p' along this geodesic by Eq. (4.5). We define the phase space principal function $S(x,'',t''|x',p',t')$ as the canonical action functional (2.26)–(2.27) evaluated for the $x(t), p(t)$ path which we have just described.

Equations (4.4) and (4.5) imply that the canonical Lagrangian (2.26) is conserved along our path $x(t), p(t)$:

$$d_t L(x,\dot{x},p) = \nabla_t L(x,\dot{x},p) = 0. \tag{4.7}$$

Therefore,

$$S(x'',t''|x',p',t') = L(x',\dot{x}',p')\cdot\tau$$
$$= (p_{a'}\dot{x}^{a'} - \tfrac{1}{2}g^{a'b'}p_{a'}p_{b'})\cdot\tau. \tag{4.8}$$

Here, the initial velocity $\dot{x}^{a'}$ must be expressed as a function of the boundary configurations [Eq. (3.14)],

$$\dot{x}^{a'} = -\Sigma^{,a'}\tau^{-1}. \tag{4.9}$$

This operation yields the final form of the phase space principal function,

$$S(x'',t''|x',p',t') = (-p_{a'}\Sigma^{,a'}(x''|x')\tau^{-1} - \tfrac{1}{2}g^{a'b'}(x')p_{a'}p_{b'})\tau. \tag{4.10}$$

From the definition of $S(x'',t''|x',p',t')$, it is obvious that the expression (4.10) must be extremized by the actual initial momentum

$$p_{a'} = -S_{a'} = -\Sigma_{,a'}\tau^{-1}. \tag{4.11}$$

It is sufficient to realize that (1) the canonical action integral is extremized by the actual distribution (4.2) of momentum and (2) this actual distribution is inevitably developed from the initial value (4.11) by the parallel transport (4.5). The same fact can also be verified by direct calculation; indeed,

$$0 = \frac{\partial S}{\partial p_{a'}} = (-\Sigma^{,a'}\tau^{-1} - p^{a'})\tau \tag{4.12}$$

amounts to Eq. (4.11).

It is also obvious that the phase space principal function (4.10) must reduce for the actual value (4.11) of momentum to the Hamilton principal function,

$$S(x^{a''},t''|x^{a'},p_{a'} = -S_{,a'} = -\Sigma_{,a'}\tau^{-1},t') = S(x'',t''|x',t') = \Sigma(x''|x')\tau^{-1}. \tag{4.13}$$

Intuitively, when the actual distribution of momentum (4.2) ensured by the initial value (4.11) and the transport equation (4.5) is substituted into the canonical Lagrangian (2.27), this Lagrangian reduces by the Legendre transformation (2.28) to the ordinary Lagrangian $l(x,\dot{x})$ evaluated along the actual configuration path. The canonical action integral (2.26) thus reduces to the action integral (2.1) along the actual path, i.e., to the Hamilton principal function. Formally, Eq. (4.13) is a consequence of the Hamilton–Jacobi equation (3.8) for the world function.

From here it is easy to infer that we obtain the actual initial momentum p' together with the actual position x and the actual momentum p at a time $t\in[t',t'']$ on the phase space path from x',t' to x'',t'' by extremizing the function

$$S(x'',t''|x,p,t|x',p',t') = S(x'',t''|x,p,t) + S(x,t|x',p',t') \tag{4.14}$$

in the variables x, p, and p'. Indeed,

$$\frac{\partial S}{\partial p_a}(x'',t''|x,p,t|x',p',t') = 0 = \frac{\partial S}{\partial p_{a'}}(x'',t''|x,p,t|x',p',t') \tag{4.15}$$

yields, as in Eq. (4.12), the actual momenta $p_{a'}$ and p_a. Instead of extremizing the function (4.14) directly in x, we use the well-known trick and substitute into it first the actual momenta $p_{a'}$ and p_a as functions of the remaining variables. Under this operation, the function (4.14) reduces, by Eq. (4.13), to our old configuration space function (2.18). Here we already know that by extremizing the expresion (2.18) in the position variable x^a we obtain the actual position x^a at t. The function (4.14) thus correctly solves the interpolation problem. If we are not interested in the actual initial value $p_{a'}$ of momentum, but only in the interpolated value, we can simply substitute the actual initial value (4.11) into the function (4.14), reducing it thereby to the function

$$S(x'',t''|x,p,t|x',t') = S(x'',t''|x,p,t) + S(x,t|x',t'). \tag{4.16}$$

The phase space principal function (4.10) is manifestly a biscalar under the point transformations (3.25). So are the functions (4.14) and (4.15) constructed from two such functions. We can thus conclude that the function (4.16) satisfies requirement I we have imposed on the function (2.31).

It is easy to generalize the procedure to a finite chain of interpolated positions and momenta and find the function (2.33) which satisfies requirement II. We simply take

$$S(x_{(N)},t_{(N)}|x_{(K)},p_{(K)},t_{(K)}) = \sum_{K=0}^{N-1} S(x_{(K+1)},t_{(K+1)}|x_{(K)},p_{(K)},t_{(K)}),$$
$$K = 0, 1, ..., N-1, \tag{4.17}$$

and extremize it in the variables $x_{(I)}$, $p_{(I)}$, $I = 1, ..., N-1$ and in $p_{(0)}$. When we are not interested in the initial momentum $p_{(0)}$, we get rid of it by substituting its actual value into the function (4.16), i.e., we replace the first function, $K = 0$, in the sum by the Hamilton principal function $S(x_{(1)},t_{(1)}|x_{(0)},t_{(0)})$. This gives us the function (2.33) which solves the chain interpolation problem.

In the last step, we investigate the continuous limit of the function (4.16). Note first how to expand (4.8) in the powers of τ on any path $x(t)$ from t' to t''. In fact, the only term which needs expansion is the initial velocity (4.9), and we have already mentioned several times that in the limit $t''\to t'$ this velocity goes over into the velocity $\dot{x}^a(t')$ on the path along which x'' approaches x'. With this change in the meaning of the symbol $\dot{x}^{a'}$, we can write

$$S(x''(t''),t''|x',p',t') = (p_{a'}\dot{x}^{a'} - \tfrac{1}{2}g^{a'b'}p_{a'}p_{b'})\tau + o(\tau^2). \tag{4.18}$$

We now take any smooth path $x(t)$, $p(t)$ in the phase space connecting x',t' with x'',t'', skeletonize it, and write down the skeletonized action (4.17):

$$S(x(t_{(N)}),t_{(N)}|x(t_{(K)}),p(t_{(K)}),t_{(K)}) = \sum_{K=0}^{N-1} S(x(t_{(K+1)}),t_{(K+1)}|x(t_{(K)}),p(t_{(K)}),t_{(K)}). \tag{4.19}$$

We expand each link in the chain (4.19) as in Eq. (4.18) and refine the skeletonization so that $\tau_{\max} = \max_{K=0,...,N-1}\tau_{(K)} \to 0$. In this way we conclude that the canonical action integral (2.26)–(2.27), (4.1) is a continuous limit of the skeletonized action,

$$\lim_{\tau_{\max}\to 0} S(x(t_{(N)}),t_{(N)}|x(t_{(K)}),p(t_{(K)}),t_{(K)}) = \lim_{\tau_{\max}\to 0}\sum_{K=0}^{N-1}\{p_a(t_{(K)})\dot{x}^a(t_{(K)}) - \tfrac{1}{2}g^{ab}(x(t_{(K)}))p_a(t_{(K)})p_b(t_{(K)})\}\tau_{(K)}$$
$$\equiv \int_{t'}^{t''} dt\,(p_a\dot{x}^a - \tfrac{1}{2}g^{ab}(x)p_ap_b) \equiv S[x(t),p(t)]. \tag{4.20}$$

In the limit $\tau_{\max}\to 0$, of course, it does not matter if we use the skeletonized action (4.19) or the action in which the first link $S(x(t_{(1)}),t_{(1)}|x(t_{(0)}),p(t_{(0)}),t_{(0)})$ was replaced by $S(x(t_{(1)}),t_{(1)}|x(t_{(0)}),t_{(0)})$. We have thus satisfied the last requirement III), Eq. (2.27), and completed our program.

When we use the skeletonization (4.17), (4.10) in the phase space path integral and perform the momentum integrations, we recover the original Feynman prescription. In curved space, this leads to an extrea potential term $\frac{1}{6}\hbar^2 R$ in the Hamilton operator. We shall now study alternative phase space principal functions which lead to different scalar curvature terms in the Hamilton operator.

5. GEODESIC MOTION: ALTERNATIVE PHASE SPACE PRINCIPAL FUNCTIONS

The principal function (4.8) can be completed into a square,

$$S(x'',t''|x',p',t') = -\tfrac{1}{2}g^{a'b'}P_{a'}P_{b'}\tau + \tfrac{1}{2}g_{a'b'}\dot{x}^{a'}x^{b'}\tau, \quad (5.1)$$

with

$$P_{a'} \equiv p_{a'} - \dot{x}_{a'}. \quad (5.2)$$

Of course, $\dot{x}^{a'}$ is still given by Eq. (4.9). From Eqs. (3.6) and (3.8) we then discover that

$$S(x'',t''|x',p',t') = -\tfrac{1}{2}g^{a'b'}P_{a'}P_{b'}\tau + S(x'',t''|x',t'). \quad (5.3)$$

The last form of the principal function brings into clear focus our fundamental observations: (1) The extremization with respect to $p_{a'}$ leads to the equation

$$P^{a'} = 0, \quad (5.4)$$

which, by Eqs. (5.2) and (4.9), yields the actual momentum, and (2) when Eq. (5.4) is substituted back into the principal function (5.3), we recover the Hamilton principal function.

At this point, it becomes obvious how to modify the principal function (5.3) while preserving the essential features I–III of our algorithm: We simply replace $g^{a'b'}$ by an arbitrary regular bitensor $G^{a'b'}(x''|x')$ with correct coincidence limit

$$[G^{a'b'}] = g^{a''b''}. \quad (5.5)$$

There is no lack of suitable candidates. First of all, we can multiply $g^{a'b'}$ by an arbitrary power of the determinantal factor (3.43):

$$G^{a'b'}(x''|x') = D^{-\nu/n}(x''|x')g^{a'b'}(x'). \quad (5.6)$$

For future convenience, we chose the exponent $-\nu/n$ as a ratio of an arbitrary real number $\nu\in(-\infty,\infty)$ with the number of degrees of freedom n of the system. The factors (3.43) are all positive in a finite neighborhood of x', and the expansion formula (3.52) ensures the proper continuous limit.

Another candidate for $G^{a'b'}$ is a chain constructed from M bitensors $\Sigma_{|a'b'}$ or from M inverse bitensors $\Sigma^{a'b'}$,

$$\Sigma^{a'c'}\Sigma_{|b'c'} = \delta^{a'}_{b'}. \quad (5.7)$$

Such a chain takes the form

$$G^{a'b'} = \Sigma^{|a'}_{\ \ k'_1}\Sigma^{|k'_1}_{\ \ k'_2}\cdots\Sigma^{|k'_{M-1}}_{\ \ k'_M}g^{k'_Mb'} \equiv (\Sigma^{|a'b'})^M \quad (5.8)$$

or

$$G^{a'b'} = \Sigma^{a'}_{\ \ k'_1}\Sigma^{k'_1}_{\ \ k'_2}\cdots\Sigma^{k'_{M-1}}_{\ \ k'_M}g^{k'_Mb'} \equiv (\Sigma^{a'b'})^M. \quad (5.9)$$

Of course, the indices in $\Sigma_{|a'b'}$ and $\Sigma_{a'b'}$ are raised and lowered by the metric tensor at x'. Equation (3.20) ensures the correct coincidence limit (5.5).

Still another candidate for $G^{a'b'}$ is a chain constructed from an even number $2M$ of bitensors $\Sigma_{,a'b''}$ or an even number of inverse bitensors $\Sigma^{a'b''}$:

$$G^{a'b'} = \Sigma^{,a'}_{\ \ k'_1}\Sigma^{,k''_1}_{\ \ k'_2}\cdots\Sigma^{,k'_{2M-2}}_{\ \ k''_{2M-1}}\Sigma^{,k''_{2M-1}}_{\ \ k'_{2M}}\cdot g^{k'_{2M}b'} \equiv (\Sigma^{,a'b''})^{2M}, \quad (5.10)$$

or

$$G^{a'b'} = \Sigma^{a'}_{\ \ k''_1}\Sigma^{k''_1}_{\ \ k'_2}\cdots\Sigma^{k'_{2M-2}}_{\ \ k''_{2M-1}}\Sigma^{k''_{2M-1}}_{\ \ k'_{2M}}\cdot g^{k'_{2M}b'} \equiv (\Sigma^{a'b''})^{2M}. \quad (5.11)$$

Again, the primed and double primed indices are raised and lowered by appropriate metric tensors at x' or x''. This time, Eq. (3.19) ensures the correct coincidence limit (5.5).

The principal function (5.3) was originally written in the form (4.10). The modified principal functions (5.8)–(5.11) can be cast into a very similar form,

$$S(x'',t''|x',p',t') = (-p_{a'}\Sigma^{,a'}(x''|x')\tau^{-1} - \tfrac{1}{2}G^{a'b'}(x''|x')p_{a'}p_{b'})\tau, \quad (5.12)$$

by a repeated use of Eqs. (3.9)–(3.10). The relationship between the modified function (5.6) and the original function (4.10), which we for a while call $S_{\|}$, is exhibited in the formula

$$S(x'',t''|x',p',t') = S_{\|}(x'',t''|x',p',t')D^{-\nu/n}(x''|x') + S(x'',t''|x',t')[1 - D^{-\nu/n}(x''|x')]. \quad (5.13)$$

The simplest chain (5.9), namely,

$$G^{a'b'} = \Sigma^{a'k''}g_{k''l''}\Sigma^{b'l''} \equiv (\Sigma^{a'b''})^2 \quad (5.14)$$

is particularly interesting because, in the phase space path integral formalism, it leads to the Hamilton operator without curvature term. We can describe the construction of the principal function (5.12), (5.14) in the following manner: We take the initial momentum $p_{a'}$ at $x^{a'}$ and transport it instantaneously to the point $x^{a'}$ by the prescription

$$p_{a''} = p_{b'}\Sigma^{b'}_{\ \ a''}(x''|x'). \quad (5.15)$$

We then start parallel propagating $p_{a''}$ back to $x^{a'}$ along the geodesic, evaluating thereby the canonical action integral (2.26)–(2.27), (4.1). This integral yields the modified function (5.12), (5.14).

Though permissible by our general criteria, the modified principal functions we have just studied do not have the clear geometric appeal of the original function (4.10). There exists, however, a phase space principal function obtained by a geometric construction as natural as the parallel transport, which leads in the end to the Hamilton operator without curvature term. We shall now turn to its discussion.

6. GEODESIC MOTION: PHASE SPACE PRINCIPAL FUNCTION FROM THE GEODESIC DEVIATION TRANSPORT OF MOMENTUM

Equation (4.5) is not the only covariant differential equation satisfied by the actual momentum (4.2):

$$\nabla_t^2 p^a + \lambda R^{\ a}{}_{bcd}\dot{x}^b p^c \dot{x}^d = 0, \quad \lambda = \text{const.}, \tag{6.1}$$

is another such equation. However, unlike the parallel transport equation (4.5), Eq. (6.1) is of the second order. Therefore, if we want to propagate the momentum vector p^a by this equation along the geodesic (4.4) from x',t' to x'',t'', we need to know not only the initial value $p^{a'}$ of the momentum, but also the initial value $[\nabla_t p^a]_{t=t'}$ of its first derivative. We thus adjoin to Eq. (6.1) our old equation (4.5) as the initial condition:

$$[\nabla_t p^a]_{t=t'} = 0. \tag{6.2}$$

Taken together, Eqs. (6.1)–(6.2) define p^a at any point of the fiducial geodesic once $p_{a'}$ is given. Moreover, they are linear and homogeneous in the momentum, which implies that p^a must be a linear homogeneous function of the initial momentum,

$$p^a = G^{ab'}(x|x')p_{b'}. \tag{6.3}$$

For $\lambda = 1$, Eq. (6.1) has a simple geometric meaning: It is the equation of geodesic deviation for the vector $p^a(t)$ transported along the fiducial geodesic. We can describe the geometric content of Eqs. (6.1), (6.2), $\lambda = 1$, in the following terms:

Take a geodesic hypersurface perpendicular to the fiducial geodesic at x'. Parallel-transport the velocity vector $\dot{x}^{a'}$ along the spike geodesics which generate the surface to nearby points on the surface. At all times, keep emitting free systems with these velocities, generating thus a geodesic flow in a thin tube about the fiducial geodesic. The velocity field within the tube has some distribution $v^a(x)$ which, because the metric does not depend on time, does not depend on time either. Take then the base and the tip of the momentum vector $p^{a'}$, attach them to the systems which are just passing by at a time t', and let the systems carry the vector as they move. Formally, let p^a be Lie propagated by the flow $v^a(x)$:

$$\pounds_v p^a \equiv p^a{}_{,b}v^b - p^b v^a{}_{,b} = \nabla_t p^a - p^b v^a{}_{|b} = 0. \tag{6.4}$$

Because the field v^a starts parallel to $\dot{x}^{a'}$ in the vicinity of $x^{a'}$, $v^{a'}{}_{|b'} = 0$. The momentum p^a then satisfies the initial condition (6.2). As the transport continues, it is well known that the Lie transport (6.4) along any geodesic flow leads to the equation of geodesic deviation.

It is worthwhile to note that it is the contravariant rather than the covariant form of the momentum which is Lie-transported. The covariant momentum $p_a = g_{ab}p^b$ in general, and the actual covariant momentum p_a in particular, is *not* Lie-transported along the described flow, because

$$\pounds_v p_a = \pounds_v g_{ab}\cdot p^b + g_{ab}\pounds_v p^b = \pounds_v g^{ab}\cdot p^b \tag{6.5}$$

does not in general vanish, unless the field v^a happens to be a Killing vector field.

Instead of talking in the dynamical language, we could have carried the whole construction in terms of the arc length s along the geodesics, by propagating the unit vector $d_s x^{a'}$. In Eqs. (6.1) and (6.2) it does not matter whether we use an affine parameter t or another affine parameter $s = vt$, $v = \text{const}$. For a system moving under the influence of forces the distinction becomes nontrivial.

For $\lambda = 0$, Eqs. (6.1) and (6.2) are equivalent to Eq. (4.5) of the parallel transport. Which transport, the parallel transport (4.5) or the geodesic deviation transport (6.1)–(6.2), with $\lambda = 1$, is more natural from the standpoint of Hamiltonian mechanics? In the parallel transport, the tip of the transported vector p^a does not in general span a geodesic, i.e., an actual configuration path of the particle; in the geodesic deviation transport it does. For this reason, we feel that preference should be given to the geodesic deviation transport.

The distinction between the two transports is geometrically significant. The congruence of lines obtained by parallel transporting the tips of different vectors $p^{a'}$ does not have any rotation, shear, or expansion. The cross section of a narrow tube formed by such lines stays the same. Under the geodesic deviation transport (6.1)–(6.2), the geodesics start originally parallel, but they soon begin deviating from each other under the influence of curvature. Physically, when we replace the geodesic deviation transport by parallel transport, we must use force on neighboring systems to enforce the rigidity of the flow.

In the geodesic flow, the disordering of the originally parallel geodesics, due to the second order nature of Eq. (6.1), becomes apparent only in the second order terms in σ (or τ) along the fiducial geodesic. In the phase space principal function which we are going to construct, such terms drop out in the continuous limit of the skeletonized action, and we obtain the old canonical action integral (2.26)–(2.27), (4.1). However, they do not entirely drop out from the phase space path integral, giving rise to a nontrivial measure in the configuration space path integral. This nontrivial measure leads to the Hamilton operator without curvature term.

Contrast this situation with that obtained under parallel transport. Parallel transport forcibly corrects for any disorder created in the flow by curvature. By maintaining order and rigidity by force, it produces the phase space principal function which, in the configuration space path integral, gives rise to a tame and seemingly trivial measure. This measure, however, ultimately leads to a curvature term in the Hamilton operator. Imposing rigid rules on the classical flow curves the backbone of the Hamilton operator. The moral: Resisting curvature by force does not pay off in the last judgement.

Anticipating this outcome, we have mainly in mind Eq. (6.1) with $\lambda = 1$ even when carrying out our calculations for an arbitrary λ. In this spirit, we shall losely talk about Eq. (6.1) with an arbitrary λ as the equation of geodesic deviation.

Proceed now with the construction of the appropriate phase space principal function. We connect the points x',t' and x'',t'' by the actual configuration path $x(t)$ of the system (the geodesic) and propagate the initial momentum $p^{a'}$ along this geodesic by the equation of geodesic deviation (6.1) with the initial condition (6.2). We define the principal function $S(x'',t''|x',p',t')$ as the canonical action integral (2.26)–(2.27),

(4.1) evaluated along the phase space path $x(t), p(t)$ which we have just described.

We split the canonical action integral into two pieces along the lines of the canonical Lagrangian (2.27). We investigate first the piece obtained from the Cartan form $p_a\dot{x}^a$. From Eqs. (4.4) and (6.1),

$$d_t^2(p_a\dot{x}^a) = \nabla_t^2(p_a\dot{x}^a) = \nabla_t^2 p_a\cdot\dot{x}^a = 0 \tag{6.6}$$

at all times. Further, from Eqs. (4.4) and (6.2),

$$[d_t(p_a\dot{x}^a)]_{t=t'} = [\nabla_t(p_a\dot{x}^a)]_{t=t'} = [\nabla_t p_a]_{t=t'}\dot{x}^{a'} = 0 \tag{6.7}$$

at least at the initial moment t'. Therefore,

$$p_a\dot{x}^{a'} = p_{a'}\dot{x}^{a'} = -p_{a'}\Sigma^{,a'}(x''|x')\tau^{-1} \tag{6.8}$$

and the first piece of the canonical action integral is

$$\int_{t'}^{t''} dt\, p_a\dot{x}^a = -p_{a'}\Sigma^{,a'}(x''|x'). \tag{6.9}$$

The second piece of the canonical action integral

$$S_H \equiv \int_{t'}^{t''} dt\, \tfrac{1}{2}g^{ab}(x)p_a p_b \tag{6.10}$$

cannot be written in such an explicit from, but we can infer a number of its properties.

When we reparametrize the path $x(t)$ by the arc length s, we see that $S_H(x'',t''|x',p',t')$ is proportional to $\tau = t''-t'$:

$$S_H = \tfrac{1}{2}\tau G(x''|x',p'), \tag{6.11}$$

where

$$G(x''|x',p') = \sigma^{-1}\int_{s'}^{s''} ds\, g^{ab}p_a p_b \tag{6.12}$$

does not contain any reference to time.

From Eq. (6.3) we learn that $G(x''|x',p')$ is a quadratic form in p':

$$G(x''|x',p') = G^{a'b'}(x''|x')p_{a'}p_{b'}, \tag{6.13}$$

with

$$G^{a'b'}(x''|x') = \sigma^{-1}\int_{s'}^{s''} ds\, g^{cd}G_c^{\ a'}G_d^{\ b'}. \tag{6.14}$$

We replace the quadratic form (6.13) by the bilinear form

$$G^{a'b'}(x''|x')p_{a'}\bar{p}_{b'} = \sigma^1\int_{s'}^{s''} ds\, g^{ab}p_a\bar{p}_b, \tag{6.15}$$

in which both $p_a = G_a^{\ b'}p_{b'}$ and $\bar{p}_a = G_a^{\ b'}\bar{p}_{b'}$ are propagated by the geodesic deviation equation (6.1)–(6.2) in the s parametrization. Choose now a special vector propagated in this way: $\bar{p}^a = d_s x^a$. The integral in (6.15) can be calculated in the same manner as the integral (6.9), and we thus get

$$G^{a'b'}p_{a'}\Sigma_{,b'} = p_{a'}\Sigma^{,a'}. \tag{6.16}$$

Because $p_{a'}$ is arbitrary,

$$G^{a'b'}\Sigma_{,b'} = \Sigma^{,a'}. \tag{6.17}$$

The tensor–scalar $G^{a'b'}(x''|x')$ acts thus on $\Sigma_{,b'}$ in the same way as the metric tensor $g^{a'b'}$. From Eq. (3.6) is also follows that

$$\tfrac{1}{2}G^{a'b'}\Sigma_{,a'}\Sigma_{,b'} = \Sigma. \tag{6.18}$$

We summarize our results by saying that the phase space principal function is a quadratic function of the initial momentum,

$$S(x'',t''|x',p',t') = -p_{a'}\Sigma^{,a'} - \tfrac{1}{2}G^{a'b'}(x''|x')p_{a'}p_{b'}\cdot\tau, \tag{6.19}$$

in which the coefficient $G^{a'b'}$ has the property (6.17). From these facts alone it is easy to show that the funciton (6.19) can be used for interpolation in the same manner as our old function (4.10). The extremization of the principal function (6.19) is p' gives the equation

$$\frac{\partial S}{\partial p_{a'}}(x'',t''|x',p',t') = -\Sigma^{,a'} - G^{a'b'}p_{b'}\cdot\tau = 0, \tag{6.20}$$

which, by the property (6.17), has the correct solution (4.11) for the actual momentum. Further, when we substitute this solution back into Eq. (6.19), the principal function reduces to the Hamilton principal function (3.3) by virtue of Eqs. (3.6) and (6.18). Again, these results could have been perceived immediately, without any calculation, from the constructive definition of our new principal function.

To prove that the continuous limit of the skeletonized action (4.19) based on the new principal function (6.19) again leads to the canonical action integral, we need to expand the coefficient $G^{a'b'}(x''|x')$ in powers of the geodesic separation $\sigma(x''|x')$. When the Hamiltonian $H(s) = \frac{1}{2}g^{ab}(x(s))p_a(s)p_b(s)$ entering into the integral (6.12) is expanded in powers of $(s-s')$,

$$H(s) = H(s') + [d_s H(s)]_{s'}(s-s') + \tfrac{1}{2}[d_s^2 H(s)]_{s'}(s-s')^2 + o((s-s')^3), \tag{6.21}$$

we can write

$$G(x''|x',p') = 2H(s') + [d_s H(s)]_{s'}\sigma + \tfrac{1}{3}[d_s^2 H(s)]_{s'}\sigma^2 + o(\sigma^3). \tag{6.22}$$

By differentiating $H(s)$ along the phase space path (4.2), (6.1),

$$\begin{aligned} d_s H &= \nabla_s H = p^a\nabla_s p_a, \\ d_s^2 H &= \nabla_s^2 H = \nabla_s p^a\nabla_s p_a + p^a\nabla_s^2 p_a \\ &= \nabla_s p^a\nabla_s p_a - \lambda R_{abcd}p^a u^b p^c u^d, \\ u^a &\equiv d_s x^a \end{aligned} \tag{6.23}$$

and using Eqs. (6.2), (3.12), we obtain the coefficients

$$\begin{aligned} &[d_s H(s)]_{s'} = 0, \\ &[d_s^2 H(s)]_{s'} = -\lambda R^{a'c'b'd'}\Sigma_{,c'}\Sigma_{,d'}\sigma^{-2}p_{a'}p_{b'}. \end{aligned} \tag{6.24}$$

As a result,

$$G(x''|x',p') = (g^{a'b'} - \tfrac{1}{3}\lambda R^{a'c'b'd'}\Sigma_{,c'}\Sigma_{,d'})p_{a'}p_{b'} + o(\sigma^3) \tag{6.25}$$

and

$$G^{a'b'} = g^{a'b'} - \tfrac{1}{3}\lambda R^{a'c'b'd'}\Sigma_{,c'}\Sigma_{,d'} + o(\sigma^3). \tag{6.26}$$

The coefficient $G^{a'b'}$ differs from the metric $g^{a'b'}$ only in terms which are of the second and higher order in σ; this ensures the correct continuous limit of the skeletonized action. The exact form of the second order term in Eq. (6.26) plays a vital role in phase space path integration. It is entirely responsible for the appearance of a nontrivial measure in the configuration space path integral.

To summarize, the principal function obtained from the geodesic deviation transport has the form (6.19) in which the bitensor $G^{a'b'}$ has the expansion (6.26). Due to Eqs. (6.17) and (6.18), we can again complete it into a square and write it thus in the form

$$S(x'',t''|x',p',t') = -\tfrac{1}{2}G^{a'b'}P_{a'}P_{b'}\tau + S(x'',t''|x',t'), \tag{6.27}$$

with

$$P_{a'} = p_{a'} - \dot{x}_{a'} = p_{a'} - \tau^{-1}\Sigma_{,a'}. \tag{6.28}$$

7. SYSTEM WITH POTENTIALS: PHASE SPACE PRINCIPAL FUNCTION FROM PARALLEL TRANSPORT WITH A FORCE TERM

Having learned the rules of the game on geodesic motion, we turn to a system moving under the influence of a generalized potential $V - A_a\dot{x}^a$. From the Lagrangian (2.35) we get the covariant acceleration

$$\nabla_t\dot{x}_a = F_a. \tag{7.1}$$

The system is driven away from a geodesic by the force

$$F_a = B_{ab}\dot{x}^b - V_{,a}, \tag{7.2}$$

which has the "electric" component $-V_{,a}$ and the "magnetic" component

$$B_{ba} = A_{a,b} - A_{b,a}. \tag{7.3}$$

Note that the metric tensor is not needed to obtain the force, but it enters into the covariant derivative $\nabla_t\dot{x}_a$.

The first Hamilton equation

$$\dot{x}^a = p^a - A^a \equiv \pi^a \tag{7.4}$$

is covariant under point transformations. The second Hamilton equation is not, and we better replace it by a covariant equation

$$\nabla_t\pi_a = \nabla_t\dot{x}_a = F_a. \tag{7.5}$$

We have introduced the abbreviation π^a for the mechanical momentum of the system. This momentum is driven away from the parallel transport by the force (7.2).

Equations (7.1)–(7.3) define the actual configuration path of the system from x',t' to x'',t''. Equation (7.5) propagates any initial momentum $p_{a'}$ along this path. We define the phase space principal function as the canonical action integral (2.26)–(2.27), (2.36) evaluated along this path $x(t), p(t)$ in the phase space.

Find what happens to the canonical Lagrangian (2.27), (2.36) under the transport (7.1)–(7.3), (7.5). Because

$$\begin{aligned}\nabla_t(\tfrac{1}{2}g^{ab}\pi_a\pi_b) &= \pi_a\nabla_t\pi^a = \pi^a\nabla_t\dot{x}^a \\ &= \nabla_t(\pi_a\dot{x}^a) - \dot{x}^a\nabla_t\pi_a \\ &= \nabla_t(\pi_a\dot{x}^a) - \dot{x}^a\nabla_t\dot{x}_a = \nabla_t(\pi_a\dot{x}^a - \tfrac{1}{2}g^{ab}\dot{x}^a\dot{x}^b),\end{aligned} \tag{7.6}$$

we see that the rate of change of the canonical Lagrangian L is the same as the rate of change of the configuration space Lagrangian l:

$$\begin{aligned}d_tL &= \nabla_tL = \nabla_t((\pi_a + A_a)\dot{x}^a - \tfrac{1}{2}g^{ab}\pi_a\pi_b - V) \\ &= \nabla_t(\tfrac{1}{2}g^{ab}\dot{x}^a\dot{x}^b + A_a\dot{x}^a - V) = V_tl = d_tl.\end{aligned} \tag{7.7}$$

From here,

$$L(t) = L(t') + l(t) - l(t').$$

Integrating,

$$S(x'',t''|x',p',t') = (p_{a'}\dot{x}^{a'} - H')\tau - l'\tau + \int_{t'}^{t''} dt\, l(t).$$

However, the last integral is nothing else but the Hamilton principal function $S(x'',t''|x',t')$. We write the final result in the form

$$\begin{aligned}S(x'',t''|x',p',t') = (p_{a'}\dot{x}^{a'} - H')\tau \\ + S(x'',t''|x',t') - l'\tau.\end{aligned} \tag{7.8}$$

Here, H' is the initial Hamiltonian (2.36), and l' is the initial Lagrangian(2.35). The initial velocity $\dot{x}^{a'}$ is to be expressed in terms of the configuration boundary data as

$$\dot{x}^{a'} = -(S^{,a'}(x'',t''|x',t') + A^{a'}(x')). \tag{7.9}$$

We see that only the first term in the phase space principal function (7.8) depends on the initial momentum $p_{a'}$; the remaining two terms depend solely on the configuration variables $x^{a'}$ and $x^{a''}$. The principal function (7.8) is thus a quadratic function of the initial momentum $p_{a'}$, like the Hamiltonian of the system. When we extremize it in the momentum $p_{a'}$,

$$0 = \frac{\partial S}{\partial p_{a'}} = \frac{\partial S}{\partial \pi_{a'}} = \dot{x}^{a'} - \pi^{a'}, \tag{7.10}$$

we obtain the actual momentum (7.4). When we substitute this momentum back into Eq. (7.8), $L' \to l'$ and the phase space principal function reduces to the Hamilton principal function:

$$S(x'',t''|x',p'(x'',t''|x',t'),t') = S(x'',t''|x',t'). \tag{7.11}$$

From our previous experience, we immediately conclude that the function (7.8) inevitably leads to the correct algorithm for interpolation, skeletonization and taking the continuous limit.

For a free system, $S(x'',t''|x',t') = l'\tau$ and only the first term is left in the expression (7.8). We thus recover our old result (4.10).

The most remarkable feature of the principal function (7.8) is that it works not only for the special system (2.35), (2.36), but for an arbitrary dynamical system. Let $H(x,p)$ be the Hamiltonian of this system, $l(x,\dot{x})$ its Lagrangian, and define the phase space principal function by the prescription (7.8), in which $\dot{x}'$ is interpreted as the function

$$\dot{x}^{a'} = \left.\frac{\partial H(x',p')}{\partial p_{a'}}\right|_{p_{a'} = -S_{,a'}(x'',t''|x',t')} \tag{7.12}$$

of the configuration boundary data x',t' and x'',t''. By extremizing the expression (7.8) in $p_{a'}$,

$$\frac{\partial S}{\partial p_{a'}} = \left[\dot{x}^{a'} - \frac{\partial H(x',p')}{\partial p_{a'}}\right]\cdot\tau = 0, \tag{7.13}$$

we get an equation which is obviously solved by the actual momentum $p_{a'} = -S_{,a'}(x'',t''|x',t')$. As we substitute the actual initial momentum (7.13) into the expression (7.8), the first and the last terms cancel by virtue of the Legendre transformation (2.28), and we are left with the Hamilton principal function $S(x'',t''|x',t')$. We have then only to follow

the familiar steps of our old algorithm dealing with interpolation, skeletonization, and continuous limit.

For special systems (2.35)–(2.36), the principal function (7.8) can again be completed into a square (5.3) with

$$P_{a'} = \pi_{a'} - \dot{x}_{a'} = p_{a'} + S_{,a'}(x'',t''|x',t'). \tag{7.14}$$

We can modify the principal function (7.14) in the same way we have modified the principal function (5.3), without disturbing the properties I–III of our algorithm. This device enables us to introduce an arbitrary measure into the configuration space path integral. In particular, we can choose the modified function such that the curvature term drops out from the Hamilton operator. However, the geodesic deviation transport with forces which we shall discuss in the next section achieves the same aim in a geometrically natural way.

8. SYSTEM WITH POTENTIALS: PHASE SPACE PRINCIPAL FUNCTION FROM GEODESIC DEVIATION TRANSPORT WITH A FORCE TERM

We replace now the propagation equation (7.5) of momentum by the second-order equation

$$\nabla_t^2 \pi_a + \lambda R_{abcd}\dot{x}^b \pi^c \dot{x}^d = \nabla_t^2 \dot{x}_a = \nabla_t F_a \tag{8.1}$$

and adjoin Eq. (7.5) to it as an initial condition,

$$[\nabla_t(\pi_a - \dot{x}_a)]_{t=t'} = 0. \tag{8.2}$$

The momentum is still propagated along the actual configuration path (7.1)–(7.3) of the system. The actual momentum (7.4) satisfies the transport equations (8.1), (8.2). For $\lambda = 1$, Eq. (8.1) tells us that the mechanical momentum π^a is diverted from the geodesic deviation transport by the differentiated force.

We define the phase space principal function as the canonical action integral evaluated along the phase space path determined by conditions (7.10), (7.3), and (8.1)–(8.2), which matches the boundary data x',p',t' and x'',t''. Because the actual momentum

$$p_a(t) = \dot{x}_a + A_a = S_{,a}(x,t|x',t') \tag{8.3}$$

satisfies, on one hand, the transport equations (7.1)–(7.3) and (8.1)–(8.2) and, on the other hand, extremizes the canonical action functional (2.26)–(2.27), (2.36), we can conclude that the initial momentum

$$p_{a'} = \dot{x}_{a'} + A_{a'} = -S_{,a'}(x'',t''|x',t') \tag{8.4}$$

extremizes the principal function $S(x'',t''|x',p',t')$. Further, when we substitute the actual momentum (8.3) back into the canonical action integral, the canonical Lagrangian $L(x^a, \dot{x}^a, p_a = \dot{x}_a + A_a)$ goes by the Legendre dual transformation (2.28) into the ordinary Lagrangian $l(x^a, \dot{x}^a)$ evaluated along the actual configuration path, and the phase space principal function goes over into the Hamilton principal function $S(x'',t''|x',t')$. Without any calculation, we can conclude that the interpolation and chain interpolation algorithm is correctly posed.

The new phase space principal function is again a quadratic function of the initial momenta. To see that, we rearrange the canonical Lagrangian

$$\begin{aligned} L &= (\pi_a + A_a)\dot{x}^a - \tfrac{1}{2}\pi^a\pi_a - V \\ &= -\tfrac{1}{2}(\pi^a - \dot{x}^a)(\pi_a - \dot{x}_a) + \tfrac{1}{2}\dot{x}^a\dot{x}_a + A_a\dot{x}^a - V \\ &= -\tfrac{1}{2}(\pi^a - \dot{x}^a)(\pi_a - \dot{x}_a) + l(x,\dot{x}) \end{aligned} \tag{8.5}$$

and introduce the abbreviation

$$P_a \equiv \pi_a - \dot{x}_a = p_a - A_a - \dot{x}_a. \tag{8.6}$$

The quantities p_a satisfy the transport equation

$$\nabla_t^2 P_a + \lambda R_{abcd}\dot{x}^b P^c \dot{x}^d = 0 \tag{8.7}$$

with the initial condition

$$[\nabla_t P_a]_{t=t'} = 0. \tag{8.8}$$

Because these equations are linear in P_a,

$$P_a(t) = G_a{}^{b'}(x,t|x',t')P_{b'}. \tag{8.9}$$

The phase space principal function is obtained by integrating (8.5) along the path (7.1), (8.7)–(8.9):

$$S(x'',t''|x',p',t') = -\frac{1}{2}\int_{t'}^{t''} dt\, g^{ab}P_aP_b + S(x'',t''|x',t'). \tag{8.10}$$

The first integral is a quadratic form of $P_{a'}$, namely,

$$\begin{aligned} &\frac{1}{2}\int_{t'}^{t''} dt\, g^{ab}P_aP_b \\ &\quad = \frac{1}{2}\int_{t'}^{t''} dt\, g^{cd}(x(t))G_c{}^{a'}(x(t),t|x',t') \\ &\qquad \cdot G_d{}^{b'}(x(t),t|x',t')\cdot P_{a'}P_{b'} \\ &\quad \equiv \tfrac{1}{2}\tau^{-1}G^{a'b'}(x'',t''|x',t')P_{a'}P_{b'}, \end{aligned} \tag{8.11}$$

and thus a quadratic function of the initial momentum $p_{a'}$. We can thus write

$$\begin{aligned} &S(x'',t''|x',p',t') \\ &\quad = S(x'',t''|x',t') - \tfrac{1}{2}\tau G^{a'b'}(x'',t''|x',t')P_{a'}P_{b'}. \end{aligned} \tag{8.12}$$

By extremizing (8.12) in $p_{a'}$, we get the equation

$$0 = \frac{\partial S}{\partial p_{a'}} = G^{a'b'}P_{b'}. \tag{8.13}$$

Around x',t', $G^{a'b'}$ is nondegenerate and thus $P_{b'} = 0$. Due to the meaning of P_a, Eq. (8.6), this yields the correct actual value (8.4) of the initial momentum $p_{b'}$. When we substitute $P_{b'} = 0$ back into Eq. (8.10), we see that $S(x'',t''|x',p',t')$ reduces back to the Hamilton principal function. This gives a formal check of our intuitive result.

To express the coefficient $G^{a'b'}(x'',t''|x',t')$ as a power series in $\tau = t'' - t'$ and $\sigma = \sigma(x''|x')$, we expand $\frac{1}{2}g^{ab}P_aP_b$ in powers of $t - t'$ and integrate. As in Eq. (6.26), we get

$$G^{a'b'} = g^{a'b'} - \tfrac{1}{3}\lambda R^{a'}{}_{c'}{}^{b'}{}_{d'}(\tau\dot{x}^{c'})(\tau\dot{x}^{d'}) + \cdots. \tag{8.14}$$

In a moment, we shall show formally that the quantity $\tau\dot{x}^a(x'',t''|x',t')$ is of the order 1 in the variables τ and σ. Inspecting the propagation equations (7.1)–(7.3) and (8.6)–(8.8), it is not difficult to conclude that all further terms in the expression (8.14) must be at least of the order 3.

As $t'' \to t'$ and x'' approaches x' along a smooth path $x(t)$, $\dot{x}^a(x'',t''|x',t')$ approaches the velocity $\dot{x}^{a'}(t')$ on this path.

Neglecting the higher order terms in τ, we write

$$S(x''(t''),t''|x',t',p') = \{l(x(t''),\dot{x}(t'')) - \tfrac{1}{2}g^{a'b'}P_{a'}P_{b'}\}\tau + o(\tau^2)$$
$$= (p_{a'}\dot{x}^{a'} - \tfrac{1}{2}g^{a'b'}\pi_{a'}\pi_{b'} - V')\tau + o(\tau^2). \quad (8.15)$$

This is sufficient for the skeletonized action to yield the canonical action integral in the continuous limit. Again, the detailed form of the expression (8.14) is needed to derive the configuration space measure in the path integral formulation.

We shall now study the short time form of the Hamilton principal function. For a free system, Eq. (3.6), time and position clearly separate in the Hamilton principal function. This is no longer true when potentials are present. Even in the simplest problems, the Hamilton principal function is a fairly complicated function of time. Two examples may help to illustrate this point. For a linear harmonic oscillator

$$H = \tfrac{1}{2}p^2 + \tfrac{1}{2}x^2, \quad g_{11} = 1, \quad A_a = 0, \quad V = \tfrac{1}{2}x^2 \quad (8.16)$$

we have

$$S(x'',t''|x',t') = \tfrac{1}{2}(x'^2 + x''^2)\cot\tau - x'x''\sin\tau \quad (8.17)$$

and for a charged paticle ($e = 1$) moving in a homogeneous magnetic field $B_a = (0,0,1)$, the uniform motion in the x^3 direction suppressed,

$$H = \tfrac{1}{2}\{(p^1 + \tfrac{1}{2}x^2)^2 + (p^2 - \tfrac{1}{2}x^1)^2\}, \quad (8.18)$$
$$g_{ab} = \delta_{ab}, \quad A_a = \tfrac{1}{2}(-x^2,x^1), \quad V = 0, \quad a = 1,2,$$

we have

$$S(x'',t''|x',t') = \tfrac{1}{4}\delta_{ab}(x^{a''} - x^{a'})(x^{b''} - x^{b'})\cot\tfrac{1}{2}\tau - \tfrac{1}{2}(x^{1''}x^{2'} - x^{1'}x^{2''}). \quad (8.19)$$

For $\tau \to 0$, the Hamilton principal function diverges, as we already see from the expression (3.6) for a free system. We must start our expansion with a term of order τ^{-1}:

$$S(x'',t''|x',t') = \Sigma(x''|x')\tau^{-1} + T(x''|x') + \Phi(x''|x')\tau + o(\tau^2). \quad (8.20)$$

To satisfy the Hamilton–Jacobi equation (2.12), (2.36) in the three lowest powers of τ, namely, τ^{-2}, τ^{-1}, and τ^0, we must put

$$\tfrac{1}{2}g^{a''b''}\Sigma_{,a''}\Sigma_{,b''} - \Sigma = 0, \quad (8.21)$$
$$\Sigma^{,a''}(T_{,a''} - A_{a''}) = 0, \quad (8.22)$$
$$\tfrac{1}{2}g^{a''b''}(T_{,a''} - A_{a''})(T_{,b''} - A_{b''}) + \Phi + \Sigma^{,a''}\Phi_{,a''} + V' = 0. \quad (8.23)$$

Similar equations follow from the second Hamilton–Jacobi equation (2.13)

$$\tfrac{1}{2}g^{a'b'}\Sigma_{,a'}\Sigma_{,b'} - \Sigma = 0, \quad (8.24)$$
$$\Sigma^{,a'}(T_{,a'} + A_{a'}) = 0, \quad (8.25)$$
$$\tfrac{1}{2}g^{a'b'}(T_{,a'} + A_{a'})(T_{,b'} + A_{b'}) + \Phi + \Sigma^{,a'}\Phi_{,a'} + V' = 0. \quad (8.26)$$

Equations (8.21) an (8.22) guarantee that $\Sigma(x''|x')$ is the world function. The expansion (8.20) of the Hamilton principal function thus always starts with the Hamilton principal function of a free system.

We know what we need about the coincidence limits of Σ and its derivatives. To learn something about the coincidence limits of T and Φ, we differentiate Eq. (8.22) twice,

$$\Sigma^{|a''}{}_{b''}(T_{,a''} - A_{a''}) + \Sigma^{,a''}(T_{|a''b''} - A_{a''|b''}) = 0, \quad (8.27)$$
$$\Sigma^{|a''}{}_{b''c''}(T_{|a''} - A_{a''}) + \Sigma^{|a''}{}_{(b''}(T_{|a''c'')} - A_{a''|c'')}) + \Sigma^{|a''}(T_{|a''b''c''} - A_{a''|b''c''}) = 0, \quad (8.28)$$

and take the coincidence limits. We get[14]

$$[T_{,a''} - A_{a''}] = 0, \quad (8.29)$$
$$[T_{|a''b''} - \tfrac{1}{2}A_{(a''|b'')}] = 0. \quad (8.30)$$

Similar equations follow from Eq. (8.25):

$$[T_{,a'} + A_{a'}] = 0, \quad (8.31)$$
$$[T_{|a'b'} + \tfrac{1}{2}A_{(a'|b')}] = 0. \quad (8.32)$$

Equations (8.30) and (8.32) imply that

$$[\Delta''T] = A^{a''}{}_{|a''}, \quad [\Delta'T] = -A^{a'}{}_{|a'}. \quad (8.33)$$

From the coincidence limit of Eq. (8.23) or (8.26) it follows that

$$[\Phi] = -V''. \quad (8.34)$$

When we take the coincidence limit of Eq. (8.20) and follow it by the time limit $\tau \to 0$, we get

$$\lim_{t' \to t''} S(x'',t''|x',t') = [T]. \quad (8.35)$$

From Eq. (2.15) we conclude that

$$[T] = 0. \quad (8.36)$$

This is all we shall need to know about coincidence limits of the biscalars T and Φ.

One can illustrate the derived relations on the two simple examples (8.16) and (8.18). For the linear harmonic oscillator,

$$\Sigma = \tfrac{1}{2}(x'' - x')^2, \quad T = 0, \quad \Phi = -\tfrac{1}{2}(x'^2 + x''^2 + x'x''). \quad (8.37)$$

For a charged particle moving in a homogeneous magnetic field,

$$\Sigma = \tfrac{1}{2}\delta_{ab}(x^{a''} - x^{a'})(x^{b''} - x^{b'}), \quad T = \tfrac{1}{2}(x^{1'}x^{2''} - x^{2'}x^{1''}),$$
$$\Phi = -\tfrac{1}{12}\Sigma. \quad (8.38)$$

One can easily verify that these expression satisfy Eqs. (8.21)–(8.26) and have the coincidence limits (8.29)–(8.36).

From Eq. (8.4) we see that

$$\tau\dot{x}_{a'} = -\Sigma_{,a'} - (T_{,a'} + A_{a'})\tau - \phi_{,a'}\tau^2 + \cdots. \quad (8.39)$$

From the coincidence limits (8.31)–(8.32) we see that the second term, $T_{,a'} + A_{a'}$, is at least of the second order in σ. Therefore, neglecting terms of order 2 or higher in τ and σ, we can write

$$\tau\dot{x}^{a'} \approx -\Sigma^{,a'}. \quad (8.40)$$

This proves our assertion that $\tau\dot{x}^{a'}$ is of the order 1 and enables us also to replace the coefficient (8.14) by the coefficient (6.26) of a free particle. We have thereby all our tools ready for the passage to path integrals.

9. QUANTUM PROPAGATOR. PHASE SPACE AND CONFIGURATION SPACE PATH INTEGRALS

In quantum mechanics, positions x^a and momenta p_a are turned into operators $\mathbf{x}^a$ and $\mathbf{p}_a$ satisfying the commutation relations

$$[\mathbf{x}^a,\mathbf{x}^b]=0, \quad [\mathbf{p}_a,\mathbf{p}_b]=0, \quad [\mathbf{x}^a,\mathbf{p}_b]=i\hbar\delta^a_b. \tag{9.1}$$

We take our state functions $\psi(x,t)$ to be scalars normalized over position variables by the prescription

$$\langle\psi|\psi\rangle=\int d^n x\, g^{1/2}(x)\psi^*(x,t)\psi(x,t). \tag{9.2}$$

We represent $\mathbf{x}^a$ and $\mathbf{p}_a$ by the operators

$$\mathbf{x}^a=x^a, \quad \mathbf{p}_a=-i\hbar g^{-1/4}\partial_a g^{1/4} \tag{9.3}$$

acting on the state functions $\psi(x,t)$. The operators (3.3) satisfy the commutation relations (9.1) and are Hermitian under the inner product (9.2).

The classical Hamiltonian (2.36) must be represented by a covariant operator, which is again Hermitian under the inner product. This requires an appropriate ordering of the noncommuting operators $\mathbf{x}^a$ and $\mathbf{p}_a$. The standard choice

$$\begin{aligned}\mathbf{H}&=\tfrac{1}{2}g^{-1/4}(\mathbf{x})[\mathbf{p}_a-A_a(\mathbf{x})]g^{1/2}g^{ab}(\mathbf{x})\\&\quad\times[\mathbf{p}_b-A_b(\mathbf{x})]g^{-1/4}(\mathbf{x})+V(\mathbf{x})\\&=-\tfrac{1}{2}\hbar^2\Delta+i\hbar(A^a\partial_a+\tfrac{1}{2}A^a{}_{|a})+V+\tfrac{1}{2}A^aA_a\end{aligned} \tag{9.4}$$

leads to the covariant Laplacian $\Delta\equiv g^{ab}\nabla_a\nabla_b$ and to the covariant symmetric ordering of the classical linear term $-A_ap^a$.

The choice (9.4) is, unfortunately, not the only one which is covariant and Hermitian. In particular, one can add to the operator (9.4) an arbitrary multiple of the scalar curvature R,

$$\mathbf{H}_\lambda=\mathbf{H}+\tfrac{1}{8}(1-\lambda)\hbar^2R, \tag{9.5}$$

which effects neither covariance not hermiticity of the Hamilton operator. The transition to quantum theory is thereby ambiguous. We shall see that different skeletonizations of the canonical action lead to Hamilton operators (9.5) with different values of $\lambda\in(-\infty,\infty)$.

Leaving undecided which Hamiltonian, (9.4) or (9.5), we are using, we subject the state function to the Schrödinger equation

$$i\hbar\partial_t\psi=\mathbf{H}\psi. \tag{9.6}$$

The general solution of Eq. (9.6) is provided by the quantum propagator $\langle x'',t''|x',t'\rangle$. This propagator evolves the state function $\psi(x',t')$ at t' into the state function $\psi(x'',t'')$ at t'',

$$\psi(x'',t'')=\int d^n x'\,\langle x'',t''|x',t'\rangle\psi(x',t'). \tag{9.7}$$

For scalar state functions, the propagator must be a scalar in x'' and a scalar density in x'. Because the Schrödinger equation (9.6) is linear, the propagator satisfies the composition law

$$\langle x'',t''|x',t'\rangle=\int\langle x'',t''|x,t\rangle\, d^n x\,\langle x,t|x',t'\rangle. \tag{9.8}$$

The measure $d^n x$ in Eq. (9.8) is consistent with our assignment of weights. The composition law (9.8) can be generalized to an arbitrary number of steps,

$$\langle x'',t''|x',t'\rangle=\int\prod_{I=1}^{N-1}d^n x_{(I)}\cdot\prod_{K=0}^{N-1}\langle x_{(K+1)},t_{(K+1)}|x_{(K)},t_{(K)}\rangle,$$
$$x'=x_{(0)}, \quad t'=t_{(0)}, \quad x''=x_{(N)}, \quad t''=t_{(N)}. \tag{9.9}$$

In this form, it serves as a point of departure for introducing configuration space path integrals.

In addition to the composition law, the propagator must satisfy the Schrödinger equation

$$i\hbar\partial_{t''}\langle x'',t''|x',t'\rangle=\mathbf{H}''\langle x'',t''|x',t'\rangle \tag{9.10}$$

with the boundary condition

$$\langle x'',t'|x',t'\rangle=\delta(x''|x'). \tag{9.11}$$

In fact, the propagator is uniquely determined by the conditions (9.10)–(9.11). Note that our δ functions are scalars in the first argument and scalar densities in the second argument.

At this stage, we have finally approached the main task of this paper. We want to represent the quantum propagator by an integral over all phase space paths $x(t),p(t)$ which start in the configuration x' at t' and end in the configuration x'' at t'',

$$\langle x'',t''|x',t'\rangle\, d^n x'=\int Dx\,Dp\, e^{i\hbar^{-1}S[x(t),p(t)]}. \tag{9.12}$$

Here, $S[x(t),p(t)]$ is the canonical action integral and $Dx\,Dp$ is a measure in the space of phase space paths.

The phase space principal functions which we have introduced enable us to interpret the formal expression (9.7) in a manifestly covariant way. We skeletonize the canonical action by the chain (4.17) of phase space principal functions $S(x_{(K+1)},t_{(K+1)}|x_{(K)},p_{(K)},t_{(K)})$ and take the measure $Dx\,Dp$ to be the product

$$Dx\,Dp=\prod_{K=0}^{N-1}(2\pi\hbar)^{-n}d^n x_{(K)}\,d^n p_{(K)} \tag{9.13}$$

of the invariant Liouville measures $d^n x\,d^n p$ on phase space, each measure divided by the quantum cell $(2\pi\hbar)^n$ of the phase space. There is one such measure at each time $t_{(K)}$ $K=0,1,\ldots,N-1$, with the exception of the final time $t_{(N)}$. The integration is performed over all of the momenta $p_{(K)}$, $K=0,1,\ldots,N-1$, but only over the interpolated coordinates $x_{(I)}$, $I=1,\ldots,N-1$. The differential $d^n x'$ thus remains unused in the integral (9.12), and it appears on both sides of the equation. An asymmetric way in which the x integrations and p integrations are performed reflects the fact that the paths have fixed boundary configurations, but free boundary momenta. Of course, the integration over the initial position x' is performed when the quantum propagator is used to evolve the state function by Eq. (9.7).

The phase space principal function can be any of the functions which we have discussed in previous sections. Note that the Planck constant occurs only at two places: It divides each principal function to turn it into a dimensionless phase, and it divides each Liouville measure $d^n x\,d^n p$ to turn it into a dimensionless measure $(2\pi\hbar)^{-n}d^n x\,d^n p$. There is no $\hbar$ in the canonical action integral itself, such as

that which occurs in "compensating quantum potential" terms $\sim\hbar^2 R$ introduced by DeWitt.[2]

We define the path integral (9.12) as a limit of the described $Nn\cdot(N-1)n$-fold integral (x' integration omitted) as $N\to\infty$ while the skeletonization is infinitely refined:

$$\int Dx\,Dp\,e^{i\hbar^{-1}S[x(t),p(t)]}$$

$$\equiv \lim_{\tau_{\max}\to 0}\int\prod_{K=0}^{N-1}(2\pi\hbar)^{-n}\,d^{n}x_{(K)}\,d^{n}p_{(K)}$$
$$\times\exp\{i\hbar^{-1}S(x_{(K+1)},t_{(K+1)}|x_{(K)},p_{(K)},t_{(K)})\}. \quad (9.14)$$

During the integrations (9.14), $x_{(K+1)}, p_{(K+1)}, t_{(K+1)}$ may differ from $x_{(K)}, p_{(K)}$ at a neighboring $t_{(K)}$ by an arbitrary amount. This reflects an intuitive notion that "paths" which we use in the path integral (9.14) do not need to be smooth. In fact, it is well known that such "nondifferentiable paths" dominate the path integral. It is exactly this circumstance which brings into prominence the second-order terms in the coefficient $G^{a'b'}$, which could have been neglected in the continuous limit when passing to the canonical action integral along a smooth path.

The prescription (9.14) for the propagator is manifestly covariant under point transformations (2.32). This is obvious, because the integrand of the skeletonized expression (9.14) is a multiscalar under point transformations and the Liouville measure $d^{n}x\,d^{n}p$ is an invariant measure under all canonical transformations. Of course, only "half the measure" is used at the initial p' integration at $t_{(0)}=t'$. Consequently, at x', the integral (9.14) transforms so that it gets multiplied by $\det|\partial\bar{x}^{b'}/\partial x^{a'}|$ under the point transformation (2.32); i.e., it behaves as a scalar density at x'. This is consistent, because the propagator which we are calculating is also a scalar density at x'. Of course, with the coordinate cell $d^{n}x'$ included on both sides of Eq. (9.12), each side again transforms as a scalar at x'.

For special systems (2.36), the momentum integrals can be explicitly evaluated. The phase space path integral (9.14) reduces thereby to a configuration space path integral

$$\langle x'',t''|x',t'\rangle\,d^{n}x'$$
$$=\lim_{\tau_{\max}\to 0}\int\prod_{K=0}^{N-1}d^{n}x_{(K)}\,C(x_{(K+1)},t_{(K+1)}|x_{(K)},t_{(K)}), \quad (9.15)$$

in which

$$C(x'',t''|x',t')\equiv\int d^{n}p'\,(2\pi\hbar)^{-n}e^{i\hbar^{-1}S(x'',t''|x',p',t')}. \quad (9.16)$$

Equation (9.15) looks like the limit of the multiple composition law (9.9), in which $C(x'',t''|x',t')$ replaced the quantum propagator. We shall call $C(x'',t''|x',t')$ "the classical propagator." To show that the expression (9.15) is indeed the quantum propagator, we need only to prove that each classical propagator in the chain (9.15) approximates the corresponding quantum propagator in the chain (9.9) so fast that the discrepancy between the two chains disappears in the limit $\tau_{\max}\to 0$. This happens when

$$\lim_{\tau\to 0}\tau^{-1}(\langle x'',t''|x',t'\rangle - C(x'',t''|x',t'))=0. \quad (9.17)$$

Because the quantum propagator is uniquely determined by the differential equation (9.10) with the boundary condition (9.11), Eq. (9.17) is established once we prove that

$$\lim_{t''\to t'}C(x'',t''|x',t')=\delta(x''|x') \quad (9.18)$$

and that $C(x'',t''|x',t')$ satisfies the Schrödinger equation in the limit $t''\to t'$.

$$\lim_{t''\to t'}(ih\,\partial_{t''}-\mathbf{H}'')C(x'',t''|x',t')=0. \quad (9.19)$$

This we shall do in the last section.

10. MEASURE FOR MEASURE

All phase space principal functions which we have considered are ultimately reducible to the same basic form (8.12), with P_a given by Eq. (8.6). They differ only by the coefficients $G^{a'b'}$. For those principal functions which were generated by parallel transport of momentum, $G^{a'b'}$ is simply the metric tensor $g^{a'b'}$ [cf. Eq. (5.3)]. We modified such functions by two methods: By multiplicative factors (5.6), (3.43), or by replacing $g^{a'b'}$ by the chains (5.8)–(5.11), (5.14). For those principal functions which were generated by the geodesic deviation transport of momentum [Eqs. (6.19) and (6.26) for the geodesic motion, Eq. (8.12) for motion under forces], $G^{a'b'}$ differs from $g^{a'b'}$ in the second order by a Riemann tensor term, Eq. (6.26). We enter the coefficients $G^{a'b'}$ into the first column of the Table I. The second column refers back to the place where the coefficient $G^{a'b'}$ was first introduced. The meaning of the following columns will unravel as we proceed.

The form (8.12) of the phase space principal function implies that the classical propagator (9.16) is a product of a Gaussian integral in the momenta with a phase factor given by the Hamilton principal function,

$$C(x'',t''|x',t')$$
$$=(2\pi\hbar)^{-n}\int d^{n}p'\exp(-\tfrac{1}{2}i\hbar^{-1}\tau G^{a'b'}p_{a'}p_{b'})$$
$$\times\exp\{i\hbar^{-1}S(x'',t''|x',t')\}. \quad (10.1)$$

The Gaussian integral with any positive definite matrix H^{ab} can be evaluated, yielding

$$\int d^{n}p\exp(-H^{ab}p_a p_b)=\pi^{n/2}(\det H_{ab})^{1/2}, \quad (10.2)$$

where H_{ab} is the inverse matrix to H^{ab}. Modulo the standard difficulty associated with the presence of i in the Gaussian integral of Eq. (10.1), we get

$$C(x'',t''|x',t')=(2\pi i\hbar\tau)^{-n/2}(\det G_{a'b'})^{1/2}$$
$$\times\exp\{i\hbar^{-1}S(x'',t''|x',t')\}. \quad (10.3)$$

The quantum propagator (9.15) thus takes the form

$$\langle x'',t''|x',t'\rangle\,d^{n}x'$$
$$=\lim_{\tau_{\max}\to 0}\int\prod_{K=0}^{N-1}d^{n}x_{(K)}\,(2\pi i\hbar\tau_{(K)})^{-n/2}$$
$$\times(\det G_{a_{(K)}b_{(K)}}(x_{(K+1)}|x_{(K)}))^{1/2}$$
$$\times\exp\{i\hbar^{-1}S(x_{(K+1)},t_{(K+1)}|x_{(K)},t_{(K)})\}, \quad (10.4)$$

TABLE I.

Phase space principal function		$G^{a'b'}(x''\|x')$	Eq.	$\Lambda(x''\|x')$	Measure discussed by	λ	$\frac{1}{6}(1-\lambda)$
Parallel transport				1	Feynman	0	$\frac{1}{6}$
		$D^{-\nu/n}(x''\|x')\,g^{a'b'}(x')$, with	(5.6)	D^{ν}, with			
		$D=D_{\cdot\cdot}\approx 1-\frac{1}{3}R^{a''b''}\Sigma_{\cdot a''}\Sigma_{\cdot b''}$		$D^{\nu}_{\cdot\cdot}\approx 1-\frac{1}{3}\nu R^{a''b''}\Sigma_{\cdot a''}\Sigma_{\cdot b''}$		$-\nu$	$\frac{1}{6}(1+\nu)$
Factor-modified			(3.41)				
		$D=D_{\cdot\cdot}\approx 1-\frac{1}{3}R^{a'b''}\Sigma_{\cdot a''}\Sigma_{\cdot b''}$	(3.43) (3.52)	$D^{\nu}_{\cdot\cdot}\approx 1-\frac{1}{3}\nu R^{a'b''}\Sigma_{\cdot a''}\Sigma_{\cdot b''}$		$-\nu$	$\frac{1}{6}(1+\nu)$
					Parker: ν	$\frac{1}{2}\nu$	$\frac{1}{6}(1-\nu)$
		$D=D_{\cdot\cdot}\approx 1+\frac{1}{6}R^{a''b''}\Sigma_{\cdot a''}\Sigma_{\cdot b''}$		$D^{\nu}_{\cdot\cdot}\approx 1+\frac{1}{6}\nu R^{a''b''}\Sigma_{\cdot a''}\Sigma_{\cdot b''}$	Parker: $\nu=2$ C. DeWitt	1	0
					Parker: $\nu=1$ Pauli–B. DeWitt	$\frac{1}{2}$	$\frac{1}{12}$
		$(\Sigma^{(a'b')})^M$	(5.8)	$D^{-M}\approx 1+\frac{1}{3}MR^{a''b''}\Sigma_{\cdot a''}\Sigma_{\cdot b''}$		M	$\frac{1}{6}(1-M)$
		$(\Sigma^{a'b'})^M$	(5.9)	$D^{M}\approx 1-\frac{1}{3}MR^{a''b''}\Sigma_{\cdot a''}\Sigma_{\cdot b''}$		$-M$	$\frac{1}{6}(1+M)$
Chain-modified							
		$(\Sigma^{\cdot a'b''})^{2M}$	(5.10)	$D^{-2M}\approx 1-\frac{1}{3}MR^{a''b''}\Sigma_{\cdot a''}\Sigma_{\cdot b''}$		$-M$	$\frac{1}{6}(1+M)$
		$(\Sigma^{a'b''})^{2M}$	(5.11)	$D^{2M}\approx 1+\frac{1}{3}MR^{a''b''}\Sigma_{\cdot a''}\Sigma_{\cdot b''}$		M	$\frac{1}{6}(1-M)$
		$(\Sigma^{a'b''})^2$	(5.14)	$D_{\cdot\cdot}\approx 1+\frac{1}{3}R^{a''b''}\Sigma_{\cdot a''}\Sigma_{\cdot b''}$	Parker	1	0
Geodesic deviation transport	True $\lambda=1$	$\approx g^{a'b'}-\frac{1}{3}R^{a'b'}\Sigma_{\cdot a'}\Sigma_{\cdot b'}$	(6.25)	$\approx 1+\frac{1}{3}R^{a''b''}\Sigma_{\cdot a''}\Sigma_{\cdot b''}$	Parker C. DeWitt	1	0
	Modified λ	$\approx g^{a'b'}-\frac{1}{3}\lambda R^{a'b'}\Sigma_{\cdot a'}\Sigma_{\cdot b'}$		$\approx 1+\frac{1}{3}\lambda R^{a''b''}\Sigma_{\cdot a''}\Sigma_{\cdot b''}$		λ	$\frac{1}{6}(1-\lambda)$

which we can interpret as the skeletonized configuration space path integral

$$\langle x'',t''|x',t'\rangle\, d^{n}x' = \int \bar{D}x\, e^{i\hbar^{-1}S[x(t)]}, \tag{10.5}$$

whose phase is given by the Lagrangian action functional (2.1) and whose measure in the space of configuration paths connecting the fixed boundaries x' at t' with x'' at t'' has the skeletonized form

$$\bar{D}x = \prod_{K=0}^{N-1} d^{n}x_{(K)}\,(2\pi i\hbar\tau_{(K)})^{-n/2} \times(\det G_{a_{(K)}b_{(K)}}(x_{(K+1)}|x_{(K)}))^{1/2}. \tag{10.6}$$

The measure (10.6) is determined entirely by the phase space principal function, which we used to skeletonize the canonical action integral. For the principal functions generated by parallel transport,

$$\bar{D}x = \prod_{K=0}^{N-1} d^{n}x_{(K)}\, g^{1/2}(x_{(K)})\cdot(2\pi i\hbar\tau_{(K)})^{-n/2}. \tag{10.7}$$

In other words, the measure

$$\bar{D}_{(K)}x = d^{n}x_{(K)}\, g^{1/2}(x_{(K)})\cdot(2\pi i\hbar\tau_{(K)})^{-n/2} \tag{10.8}$$

associated with a short stretch of the beam of configuration space paths starting at the gate $d^{n}x_{(K)}$ and propagating for a short time $\tau_{(K)}=t_{(K+1)}-t_{(K)}$ to the next gate at $x_{(K+1)}$ is proportional to the proper area $d^{n}x_{(K)}\,g^{1/2}(x_{(K)})$ of the initial gate, the factor of proportionality $(2\pi i\hbar\tau_{(K+1)})^{-n/2}$ being the same whatever place $x_{(K+1)}$ is the next gate located. In flat space treated in Cartesian coordinates, $g^{1/2}=1$ and the measure (10.6) is the original Feynman measure. The expression (10.6) coincides with DeWitt's generalization of the Feynman measure to curved spaces.[2]

The Feynman measure is trivial, because it does not depend on the position $x_{(K+1)}$ of the next gate to which the beam of paths is propagating. This is no longer true if we base the phase space path integral on one of the modified phase space principal functions. Each element (10.7) of the Feynman measure gets multiplied by a factor $\Lambda^{1/2}(x_{(K+1)}|x_{(K)})$ coming from the determinant in Eq. (10.5):

$$\bar{D}_{(K)}x = d^{n}x_{(K)}\, g^{1/2}(x_{(K)})\cdot(2\pi i\hbar\tau_{(K)})^{-n/2}\cdot\Lambda^{1/2}(x_{(K+1)}|x_{(K)}). \tag{10.9}$$

We list such factors in the third column of Table I. In the fourth column, the people who discussed various configuration space path measures are mentioned by name. If the factor is known only in the expanded form, we enter $\approx$ into the third column. In fact, the approximate form of all factors is

always the same, namely

$$\Lambda(x''|x') \approx 1 + \tfrac{1}{3}\lambda R^{a''b''}\Sigma_{,a''}\Sigma_{,b''}. \tag{10.10}$$

We thus enter into the fifth column simply the values of the parameter λ. For the chain modified measures, this parameter is a whole number.

Let us finally discuss the measure which follows from the geodesic deviation transport. From Eq. (6.25),

$$G_{a'b'} \approx g_{a'b'} + \tfrac{1}{3}\lambda R_{a'}{}^{c'}{}_{b'}{}^{d'}\Sigma_{,c'}\Sigma_{,d'}, \tag{10.11}$$

and by the trace rule,

$$\begin{aligned}\det G_{a'b'} &\approx g'(1 + \tfrac{1}{3}\lambda R^{a'b'}\Sigma_{,a'}\Sigma_{,b'}) \\ &\approx g'(1 + \tfrac{1}{3}\lambda R^{a''b''}\Sigma_{,a''}\Sigma_{,b''}).\end{aligned} \tag{10.12}$$

This leads exactly to the factor (10.10).

The conclusion at which we have just arrived in worth emphasizing over and over again. We started from phase integrals with the natural Liouville measure, but with skeletonizations of the canonical action integral ruled by different phase space principal functions. Each choice of the phase space principal function induces a different measure in the space of configuration paths. The configuration path measure obtained from the rigid parallel propagation of momentum is trivial, but other measures are not. In particular, the measure imposed by the geodesic deviation transport whose sovreignity comes directly from geometry, the ultimate source of authority for a relativist, is nontrivial. The Liouville measure is unique and unambiguous, but the configuration space path measure is at the mercy of the phase space principal function. The law is unambiguous, but the result of a trial depends on the judge. We must know the character of the phase space principal function to forsee what configuration space path measure will result from the phase space path measure.

Having made our point sufficiently memorable, we return back to more mundane affairs. For all choices of the phase space principal function, the classical propagator has the form

$$\begin{aligned}C(x'',t''|x',t') = \Lambda^{1/2}(x''|x')\cdot(2\pi i\hbar\tau)^{-n/2}g^{1/2}(x') \\ \times e^{i\hbar^{-1}S(x'',t''|x',t')},\end{aligned} \tag{10.13}$$

where Λ differs from the expression (10.10) at most by terms which are of the order 3 and higher in the variables τ and σ. The Hamilton principal function itself can be expanded in powers of τ, Eq. (8.20), and the classical propagator (10.13) rearranged into

$$C(x'',t''|x',t') = \Lambda^{1/2}e^{i\hbar^{-1}(T+\Phi\tau)}\delta_\tau(x''|x'), \tag{10.14}$$

where

$$\delta_\tau(x''|x') \equiv (2\pi i\hbar\tau)^{-n/2}g^{1/2}(x')e^{i\Sigma(x''|x')/\hbar\tau} \tag{10.15}$$

is a family of scalar–scalar density functions labeled by the continuous parameter τ. We can interpret $\delta_\tau(x''|x')$ as the classical propagator obtained under parallel transport of momentum along the geodesic motion.

To complete our argument that the phase space path integral (9.14) correctly represents the quantum propagator (9.12), we must show that the classical propagator (10.14) has the properties (9.18), (9.19). The first step in this direction is the proof that, in the limit $\tau\to 0$, the one-parameter family of functions (10.15) defines a δ function. Because of its close association with the concept of a geodesic, we shall call this particular function the geodesic δ funtion.

11. GEODESIC δ FUNCTION

Let $F(x''|x')$ be any biscalar with a regular coincidence limit $[F] = F(x''|x'')$ and

$$I(\tau) \equiv \int d^n x'\, F(x''|x')\delta_\tau(x''|x'). \tag{11.1}$$

We want to show that

$$\lim_{\tau\to 0} I(\tau) = F(x''|x''). \tag{11.2}$$

To do that, we replace the integration variables x' by the Riemann normal coordinates (3.26) of the point x' based on the fixed origin x''. The Jacobian of this transformation was introduced in Eq. (3.42). In the new variables $y_{a''}$,

$$\begin{aligned}I(\tau) = \int d^n y''\, \Sigma_{,-}^{-1}(x''|x'(x'',y''))F(x''|x'(x'',y'')) \\ \times g^{1/2}(x'(x'',y''))(2\pi i\hbar\tau)^{-n/2} \\ \times \exp\{(2i\hbar\tau)^{-1}g^{a''b''}y_{a''}y_{b''}\}.\end{aligned} \tag{11.3}$$

We change the variables once more, scaling them by $(\hbar\tau)^{-1/2}$:

$$y_{a''} = (\hbar\tau)^{1/2}z_{a''}, \quad d^n y'' = (\hbar\tau)^{n/2}\, d^n z''. \tag{11.4}$$

As $\tau\to 0$ at a fixed $z_{a''}$, $y_{a''}\to 0$. This corresponds to taking the coincidence limit: $x'(x'',y''=0) = x''$. Because $[\Sigma_{,-}] = g(x'')$,

$$\begin{aligned}\lim_{\tau\to 0} I(\tau) = F(x''|x'')g^{-1/2}(x'')\cdot(2\pi i)^{-n/2} \\ \times \int d^n z''\, \exp\{-(2i)^{-1}g^{a''b''}z_{a''}z_{b''}\}.\end{aligned} \tag{11.5}$$

By Eq. (10.2),

$$\int d^n z''\, \exp[-(2i)^{-1}g^{a''b''}z_{a''}z_{b''}] = (2\pi i)^{n/2}g^{1/2}(x''). \tag{11.6}$$

This shows that for $\tau\to 0$ the family $\delta_\tau(x''|x')$ defines the δ function, Eqs. (11.1)–(11.2).

Let now $F(x''|x')$ be a biscalar whose expansion in the powers of σ starts with a second-order term,

$$F(x''|x') = F^{a'b'}(x')\Sigma_{,a'}\Sigma_{,b'} + o(\sigma^3). \tag{11.7}$$

We want to prove that the integral

$$J(\tau) \equiv \tau^{-1}\int d^n x'\, F(x''|x')\delta_\tau(x''|x') \tag{11.8}$$

has a finite limit J as $\tau\to 0$. By the same substitutions which brought us to Eq. (11.5), we get

$$\begin{aligned}J = \lim_{\tau\to 0} J(\tau) = \hbar F^{a''b''}(x'')g^{-1/2}(x'')(2\pi i)^{-n/2} \\ \times \int d^n z''\, z_{a''}z_{b''}\, \exp\{-(2i)^{-1}g^{a''b''}z_{a''}z_{b''}\}.\end{aligned} \tag{11.9}$$

When we differentiate Eq. (11.6) with respect to $g^{a''b''}$, we

learn that

$$(2\pi i)^{-n/2}\int d^{n}z''\, z_{a''}z_{b''} \times\exp\{-(2i)^{-1}g^{a''b''}z_{a''}z_{b''}\} = ig^{1/2}(x'')g_{a''b''}. \quad (11.10)$$

Therefore,

$$J = i\hbar g_{a''b''}F^{a''b''}. \quad (11.11)$$

To summarize[15]: Let $F(x''|x')$ be a biscalar with the expansion (11.6). Then

$$\lim_{\tau\to 0}\tau^{-1}\int d^{n}x'\, F(x''|x')\delta_{\tau}(x''|x') = i\hbar g_{a''b''}F^{a''b''}. \quad (11.12)$$

12. SCHRÖDINGER'S EQUATION FOR THE PROPAGATOR

The proof that the classical propagator (10.14) has the desired properties (9.18), (9.19) is entirely based on the two properties, Eqs. (11.1)–(11.2) and (11.12), of the geodesic δ function.

It is obvious that the classical propagator reduces to the δ function as $t''\to t'$. Indeed,

$$\lim_{\tau\to 0}\Delta^{1/2}e^{i(T+\Phi\tau)}\delta_{\tau} = [\Delta]^{1/2}e^{i[T]}\delta(x''|x') = \delta(x''|x'), \quad (12.1)$$

because $[T]=0$ by Eq. (8.36) and $[\Delta]=1$ by Eq. (10.10). The only remaining task thus is to prove that $C(x'',t''|x',t')$ satisfies the Schrödinger equation (9.19) as $\tau\to 0$.

We must apply the operator $i\hbar\partial_{t''}-\mathbf{H}''$ to the function (10.14)–(10.15) which is a product of the factor $(2\pi i\hbar\tau)^{-n/2}\Delta^{1/2}(x''|x')g^{1/2}(x')$ with the exponential $\exp\{i\hbar^{-1}(\tau^{-1}\Sigma+T+\tau\Phi)\}$. We can neglect the terms proportional to τ and τ^2 and collect the remaining powers of τ, namely, τ^{-2}, τ^{-1}, and τ^{0}. Using Eq. (9.4) for the Hamilton operator H, we get

$$\begin{aligned}(i\hbar\partial_{t''}-\mathbf{H}'')&C(x'',t''|x',t') \\ &= \{\tau^{-1}(\Sigma-\tfrac{1}{2}\Sigma_{,a''}\Sigma^{,a''}) \\ &\quad+\tau^{-1}(-\Sigma_{,a''}(T^{,a''}-A^{a''}) \\ &\quad+\tfrac{1}{2}i\hbar(\Delta''\Sigma-n+\Delta^{-1}\Delta_{,a''}\Sigma^{,a''})) \\ &\quad+\tau^{0}((-\tfrac{1}{2}(T_{,a''}-A_{a''})(T^{,a''}-A_{a''}) \\ &\quad+\Sigma_{,a''}\Phi^{,a''}+\Phi+V'') \\ &\quad+\tfrac{1}{2}i\hbar(\Delta''T-A^{a''}{}_{|a''})+i\hbar\Delta^{-1}\Delta_{,a''}(T^{,a''}-A^{a''}) \\ &\quad+\tfrac{1}{4}\hbar^{2}\Delta^{-1}(\Delta''\Delta-\Delta^{-1}\Delta_{,a''}\Delta^{,a''}))+o(\tau)\} \\ &\quad\times C(x'',t''|x',t'). \end{aligned} \quad (12.2)$$

This equation vastly simplifies when we take into account the Hamilton–Jacobi equations (8.21)–(8.23) for the expanded Hamilton's principal function. We get

$$\begin{aligned}(i\hbar\partial_{t''}-\mathbf{H}'')&C(x'',t''|x',t') \\ &= \{\tau^{-1}\tfrac{1}{2}i\hbar(\Delta''\Sigma-n+\Delta^{-1}\Delta_{,a''}\Sigma^{,a''}) \\ &\quad+\tfrac{1}{2}i\hbar(\Delta''T-A^{a''}{}_{|a''})+i\hbar\Delta^{-1}\Delta_{,a''}(T^{,a''}-A^{a''}) \\ &\quad+\tfrac{1}{4}\hbar^{2}\Delta^{-1}(\Delta''\Delta-\Delta^{-1}\Delta_{,a''}\Delta^{,a''})+o(\tau)\} \\ &\quad\times C(x'',t''|x',t'). \end{aligned} \quad (12.3)$$

Let us note in passing an alternative form of the coefficients:

$$\Delta''\Sigma-n+\Delta^{-1}\Delta_{,a''}\Sigma^{,a''} = \Delta^{-1}((\Delta\Sigma^{,a''})_{|a''}-n\Delta), \quad (12.4)$$

$$\tfrac{1}{4}\Delta^{-1}(\Delta''\Delta-\Delta^{-1}\Delta_{,a''}\Delta^{,a''}) = \tfrac{1}{2}\Delta^{-1/2}\Delta''\Delta^{1/2}.$$

We shall now take the limit $\tau\to 0$. This is easy in the absolute term τ^{0}: Due to Eqs. (11.1)–(11.2), we have only to replace the coefficient of $C(x'',t''|x',t')$ by its coincidence limit. Taking into account the coincidence limits (8.29) and (8.33), we are left with the expression

$$\tfrac{1}{4}\hbar^{2}[\Delta]^{-1}([\Delta''\Delta]-[\Delta]^{-1}[\Delta_{,a''}][\Delta^{,a''}])\delta(x''|x'). \quad (12.5)$$

From Eq. (10.10) and the coincidence limits (3.17), (3.20),

$$[\Delta]=1, \quad [\Delta_{|a''}]=0, \quad [\Delta''\Delta]=\tfrac{2}{3}\lambda R''. \quad (12.6)$$

The expression (12.5) thus reduces to

$$\tfrac{1}{6}\lambda\hbar^{2}R''\delta(x''|x'). \quad (12.7)$$

The term proportional to τ^{-1} is handled through Eq. (11.11). The biscalar $\Delta''\Sigma-n$ was expanded in Eq. (3.40). From the expanded form of Δ, Eq. (10.10), we conclude that

$$\Delta^{-1}\Delta_{,a''}\Sigma^{,a''}\approx\tfrac{2}{3}\lambda R^{a''b''}\Sigma_{,a''}\Sigma_{,b''}. \quad (12.8)$$

As a consequence, the τ^{-1} term yields

$$-\tfrac{1}{6}(-1+2\lambda)\hbar^{2}R''\delta(x''|x') \quad (12.9)$$

in the $\tau\to 0$ limit.

Putting the two contributions, (12.7) and (12.9), together, we conclude that the classical propagator $C(x'',t''|x',t')$ satisfies the Schrödinger equation (9.19) corresponding to the Hamilton operator (9.5) in the limit $\tau\to 0$. It has thus the vital properties (9.18)–(9.19) which guarantee that the path integral (9.14) leads to the quantum propagator (9.12) which satisfies the Schrödinger equation (9.10) with the Hamilton operator (9.5), (9.4).

The curvature term $\hbar^{2}R''$ occurs in the Schrödinger equation with the coefficient $\frac{1}{6}(1-\lambda)$. In the last column of our Table I, we give the values of this coefficient for different phase space principal functions. For $\lambda=1$, the scalar curvature drops out from the Schrödinger equation. In particular, this happens for the principal function modified by the factors $D^{1/n}$, $D^{1/n}$ and $D^{2/n}$ and by the chain measures $(\Sigma^{|a'b'})$, $(\Sigma^{a'b'})^{-1}$, $(\Sigma^{,a'b''})^{-2}$, and $(\Sigma^{a'b''})^{2}$. From these contrived principal functions, the last one seems the least contrived. All these cases, however, are of little interest in comparison with the main result of this paper:

Theorem: The phase space principal function generated by the geodesic deviation transport leads to the quantum propagator which satisfies the Schrödinger equation without curvature term.

Two other cases deserve special attention. For $\lambda=0$, we get the trivial measure ($\Delta=1$) corresponding to Feynman's procedure. The absolute term (12.5) then strictly vanishes, but the τ^{-1} term gives a nonvanishing contribution (12.9) leading to the $\frac{1}{6}\hbar^{2}R''$ potential in the Schrödinger equation. On the other hand, for $\lambda=\frac{1}{2}$ we get the Pauli–DeWitt measure. The τ^{-1} term then strictly vanishes, but the absolute term gives a nonvanishing contribution (12.7) leading to

the $\frac{1}{12}\hbar^2 R''$ potential in the Schrödinger equation. For $\lambda = 1$, both terms are present, but they mutually cancel. For $\lambda \in (-\infty, \infty)$, we get an arbitrary potential term in the Schrödinger equation.

ACKNOWLEDGMENT

This paper is an offshot of collaboration with James Hartle on path integrals in parametrized and gauge theories.

[1]R. P. Feynman, Rev. Mod. Phys. **20**, 327 (1978).

[2]B. S. DeWitt, Rev. Mod. Phys. **29**, 377 (1957).

[3]W. Pauli, "Feldquantisierung," lecture notes, ETH, Zürich, 1950–1951. English translation in: *Pauli Lectures on Physics: Volume 6. Selected Topics in Field Quantization* (MIT Press, Cambridge, MA, 1973).

[4]J. H. Van Vleck, Proc. Natl. Acad. Sci. USA **14**, 178 (1928).

[5]L. Parker, Phys. Rev. D **19**, 438 (1979). Parker's paper discusses only a free relativistic particle in the proper time formalism, but the modification to Newtonian systems with potentials is straightforward.

[6]C. DeWitt-Morette, K. D. Elworthy, B. L. Nelson, and G. S. Sammelman, Ann. Inst. H. Poincaré **32**, 327 (1980).

[7]For a summary, see, e.g., L. D. Fadeev, lectures notes, Les Houches, 1976, or L. D. Fadeev and A. A. Slavnov, *Gauge Fields: Introduction to Quantum Theory* (Benjamin, Reading, MA, 1980).

[8]A nice discussion of this point may be found in J. R. Klauder, Path Integrals, lecture given at XIX Internationale Universitätswochen für Kernphysik, Schladming, Austria, 1980.

[9]The term "point transformations" has many meanings. We are using it for transformations in the phase space induced by coordinate transformations in the configuration space. Cf. Eq. (2.32).

[10]See, e.g., the discussion in Refs. 7.

[11]M. S. Marinov, Phys. Rep. **60**, 2 (1980).

[12]J. Hartle and K. Kuchař, work in progress.

[13]This term comes from Synge. The biscalar itself was introduced by H. S. Ruse, Proc. London Math. Soc. **32**, 87 (1931). An excellent discussion of the biscalar and its properties can be found in J. L. Synge, *Relativity: The General Theory* (North-Holland, Amsterdam, 1960).

[14]Our convention for the symmetrization bracket is $A...(a...b)... \equiv A...a...b... + A...b...a...$.

[15]Cf. B. DeWitt, Ref. 2, Eq. (7.21).

A description of virtual scattering states in the rigged Hilbert space formulation of quantum mechanics

M. Gadella
Facultad de Ciencias, Av. los Castros s/n, Santander, Spain

(Received 27 July 1982; accepted for publication 24 September 1982)

We define a pair of Gel'fand triplets $\Phi_i \subset \mathscr{H}_i \subset \Phi_i'$, $i = 1,2$, where the linear spaces Φ_i are formed by suitably chosen entire analytic functions on the Riemann surface associated to the transformation $\omega = z^2$. Virtual scattering states are well defined as pairs of functionals, respectively, in Φ_1' and Φ_2'.

PACS numbers: 03.65.Db, 24.90. + d

I. INTRODUCTION

At low energy, proton-neutron scattering presents an anomalously large cross section corresponding to the presence of a single state.[1-4] The scattering length of this state is -23.7×10^{-13} cm. The minus sign reveals that it cannot be produced by a bound state.[1-2]

In the S-matrix formalism such a state is being characterized by a pole in the negative imaginary axis of $S(p)$ in the momentum representation. This corresponds to a pole on the negative real axis in the second sheet in the energy representation. Conversely, every pole of the S-matrix on the negative real axis in the second sheet in the energy representation produces a large cross section. The closer the pole is to zero, the bigger is the perturbation. We say that the perturbation is due to the presence of a "virtual scattering state," one example of it being the proton-neutron singlet. The aim of the present paper is to present a description of virtual scattering states in the framework of rigged Hilbert spaces (RHS).

As the reader already knows, a rigged Hilbert space is a triplet of vector spaces:

$$\Phi \subset \mathscr{H} \subset \Phi', \tag{1}$$

where $\mathscr{H}$ is a Hilbert space. Φ is a dense subspace of $\mathscr{H}$ endowed with a nuclear topology which is finer than that one inherited from $\mathscr{H}$ and makes Φ complete. Φ' is the topological antidual of Φ, i.e., the vector space of all continuous antilinear functionals on Φ.[5-9] Rigged Hilbert spaces have been used as an alternative to Hilbert spaces for the mathematical apparatus of quantum mechanics.[1,6-9] Moreover, it offers some advantages: it rigorizes the Dirac formulation of quantum mechanics[6-8] and allows us to disregard domain problems of operators provided that the basic observables for every actual physical problem have a common invariant dense domain on which they are essentially self-adjoint.[8,10] Furthermore, the RHS formulation appears specially suitable for the description of resonance phenomena.[11-15]

Our method to describe virtual scattering states consists in a refinement of that one used for the description of resonances.[11,12,15] In Ref. 15 the space Φ in (1) was a nuclear space of Hardy class functions in a half-plane. Now we want the functions in Φ to be analytically continued at least to a strip of the other half-plane (here we work in the second sheet of the Riemann surface associated to the transformation $E = p^2/2m$) and still have their restrictions to R^+ fulfill the imposed conditions to Φ. In Sec. II, we shall see that the functions can be chosen to be entire. In Sec. III, we shall apply the mathematical construction to describe virtual scattering states and resonances.

II. A RIGGED HILBERT SPACE OF ENTIRE HARDY CLASS FUNCTIONS

Let Φ be the space of $L^2(R^+)$ functions such that

(i) If $\varphi(E) \in \Phi$, $\varphi(E)$ is the restriction to R^+ of a Hardy class function in the lower half-plane.

(ii) If $\varphi(E) \in \Phi$, $E^n \varphi(E) \in \Phi$ for all $n \in N$.

(iii) The space Φ can be endowed with a nuclear (non-necessarily metrizable) topology.

(iv) The space Φ is dense in $L^2(R^+)$.

(v) Every $\varphi(E)$ is an entire analytic function.

Φ can be constructed in the following way: Consider $\mathscr{D}(R^+)$ the space of the Schwartz functions with compact support in R^+.

A Paley–Wiener theorem asserts that the Fourier transforms of Schwartz functions with compact support are entire analytic.[16] A second Paley–Wiener theorem says that the Fourier transform of a $L^2(R^+)$ function is Hardy class on the lower half-plane.[17] With these ideas in mind, we conclude: we can construct the required space Φ with the restrictions to the positive real axis of Fourier transforms of the functions in $\mathscr{D}(R^+)$.

The proof comes as follows: (i) and (v) are immediate consequences of the Paley–Wiener theorems. (ii) is due to the fact that the Fourier transform of the nth derivative of a Schwartz function $f(x)$ is the Fourier transform $\hat{f}(E)$ of $f(x)$ times $(-iE)^n$.

We carry out the proof of (iv) as in the analogous density requirement in Ref. 15. If we call Ψ the space $\mathscr{F}(\mathscr{D}(R^+))$ of all Fourier transforms of functions in $\mathscr{D}(R^+)$, we claim that Ψ is dense in H^2_-. This is simple: $\mathscr{D}(R^+)$ is dense in $L^2(R^+)$ and the Fourier transform is unitary and hence it preserves dense sets. Therefore, Ψ is dense in H^2_- and hence Φ is dense in H^2_{-+}, the space of the restrictions to R^+ of the functions in H^2_-. On the other hand H^2_{-+} is dense in $L^2(R^+)$.[18] Consequently Φ is dense in $L^2(R^+)$.

We now only have to endow Φ with a nuclear topology. The space $\Psi = \mathscr{F}(\mathscr{D}(R^+))$ has its own topology as a subspace of $\mathscr{S}(R)$. This topology is nuclear; however, Ψ is not a closed subspace of $\mathscr{S}(R)$. Let $\mathscr{L}$ be equal to $\mathscr{F}(\mathscr{S}(R))$, the

 0022-2488/83/082142-04$02.50

space of Fourier transforms of all Schwartz functions on R having a compact support. As we have noted earlier, all the functions in $\mathscr{Z}$ are entire analytic. Furthermore, $\mathscr{Z}$ has a natural nuclear topology and the Fourier transform is an homeomorphism between $\mathscr{Z}$ and $\mathscr{D}(R)$.[19] We can therefore endow Ψ with the subspace topology inherited from $\mathscr{Z}$. Since $\mathscr{D}(R^+)$ is a closed subspace of $\mathscr{D}(R)$ and the Fourier transform is an homeomorphism, Ψ is closed in $\mathscr{Z}$ and hence complete. Since any subspace of a nuclear space is also nuclear,[20] so is Ψ.

It only remains to fix the topology on Φ. The Heaviside function $\theta(x)$ is an one-to-one mapping from Ψ onto Φ because Hardy class functions are defined by their boundary values on a half-line. The topology on Φ is going to be the final one with respect to Ψ and the Heaviside function $\theta(x)$. Thus Φ is a nuclear complete space because of the properties of initial and final topologies.[8]

As we shall see later, the functions in Φ are to represent physical state vectors of some system in the energy representation. In this representation, the Hamiltonian becomes the multiplication operator on Φ. A well behaving observable should be an essentially self-adjoint operator, i.e., a symmetric operator having a unique self-adjoint extension. Let us check whether the multiplication operator H on Φ is essentially self-adjoint. $[H\varphi(E) = E\varphi(E)\ \forall(E)\in\Phi]$.

We shall use the following theorem[21]: A symmetric operator B defined on a dense subspace $\mathscr{D}$ of a Hilbert space $\mathscr{H}$ is essentially self-adjoint if and only if the equation

$$Af = \pm if \tag{2}$$

has no solution in $\mathscr{D}$.

Consider now the equation

$$E\varphi(E) = \pm i\varphi(E), \tag{3}$$

where $\varphi(E)\in\Phi$. φ is continuable to an entire function. Thus, we can expand it in Taylor series around the origin:

$$\varphi(E) = \sum_{n=0}^{\infty} a_m E^n. \tag{4}$$

However, Eqs. (3) and (4) give $a_n = 0$, for all n. This means that $\varphi(E) = 0$ which proves our assertion.

In summary we have constructed two rigged Hilbert spaces

$$\Phi\subset L^2(R^+)\subset\Phi' \tag{5}$$

and

$$\Psi\subset H^2_-\subset\Psi', \tag{6}$$

where Φ' and Ψ' are the dual spaces, respectively, of Φ and Ψ. Every vector in Φ' (resp. Ψ') is a continuous antilinear functional on Φ (resp. Ψ). These two triplets are equivalent in the sense that every term of one of them is homeomorphic as well as isomorphic to the corresponding term in the other. However, the functions which give this correspondence are not the same. The correspondence between Ψ and Φ is given by means of the Heaviside function which has nothing to do with the unitary transformation that provides the equivalence betwen H^2_- and $L^2(R^+)$. The Heaviside function transforms H^2_- onto H^2_{-+}. However, H^2_{-+} with the inherited topology from $L^2(R^+)$ is not a Hilbert space but a dense subspace.

A change in the notation is now necessary. Henceforth, we shall call the triplet (5)

$$\Phi_-\subset L^2(R^+)\subset\Phi'_- \tag{7}$$

and the triplet (6)

$$\Psi_-\subset H^2_-\subset\Psi'_-. \tag{8}$$

The reason for such a change is the following: We could have constructed Φ as the space $\mathscr{F}(\mathscr{D}(R^-))$ of all Fourier transforms of Schwartz functions with compact support in R^-. This new space Φ (henceforth Φ_+) fulfills all the properties (i)–(v) except that any $\varphi\in\Phi_+$ and be extended to a Hardy class function of the upper half-plane. We now obtain two new triplets in the same way that we have done with (7) and (8),

$$\Phi_+\subset L^2(R^+)\subset\Phi'_+, \tag{9}$$

$$\Psi_+\subset H^2_+\subset\Psi'_+. \tag{10}$$

The sign plus or minus means that the respective functions are Hardy class on the upper or lower half-plane.

Finally, note that every function in Φ_+ or Φ_- can be extended to an analytic function on the whole Riemann surface associated to the transformation $\omega = z^2$. In fact, on every sheet we shall have a function which is the exact copy of the value of the function on the other sheet at the same point.

III. THE VIRTUAL STATE VECTORS AND THEIR ASSOCIATION WITH THE CORRESPONDING POLE OF THE S-MATRIX

In this section we shall proceed as in the second part of Ref. 15. We may suppose that the interacting Hamiltonian H does not have bound states. The corresponding S-matrix in the energy representation has a simple pole on the negative part of the real axis in the second sheet plus pairs of conjugate poles near the positive real axis in this second sheet and no more singularities. As we have already mentioned, the pole on the negative real axis in the second sheet is to be associated with the appearance of a virtual state. The pairs of conjugate poles are linked to resonance phenomena. The S-matrix is assumed not to grow faster than a polynomial in the second sheet.

As in Ref. 15, we consider the integral

$$(\psi^{\text{out}}(t), S\phi^{\text{in}}(t)) = (\psi^-(0),\phi^+(0)) = \int_{R^+}\psi^-(E)S(E+i0)\phi^+(E)\,dE, \tag{11}$$

where $\psi^-(E)\in\Phi_+$ and $\phi^+(E)\in\Phi_-$. $\psi^-(E)$ and $\phi^+(E)$ are the vectors in the energy representation which, respectively, correspond to $\psi^-(0) = \Omega^-\psi^{\text{out}}(0)$ and $\phi^+(0) = \Omega^+\psi^{\text{in}}(0)$ in the Schrödinger representation. According to Ref. 5, we may find antilinear functionals $|E\pm\rangle\in\Phi'_{\mp}$ such that for almost every $E\in R^+$,

$$[\phi^+(E)]^* = \langle{}^+\phi\,|E^+\rangle = \langle{}^+E\,|\phi^+\rangle^*,$$
$$[\psi^-(E)]^* = \langle{}^-\psi|E^-\rangle = \langle{}^-E\,|\psi^-\rangle^*. \tag{12}$$

Since the functions we are dealing with are continuous, it results that the functions $|E^\pm\rangle$ are defined for every $E>0$.

These vectors are generalized eigenvectors of the total Hamiltonian. Therefore, Eq. (11) can also be written

$$\langle \psi^- | \phi^+ \rangle = \int_{R^-} \langle {}^-\psi | E^- \rangle S(E + i0) \langle {}^+E | \phi^+ \rangle \, dE. \tag{13}$$

Now we assume that the S-matrix has a simple pole on the lower half-plane in the second sheet at $z_R = E_R - i\Gamma/2$. Also we find a simple pole at $z_v = -E_v$ on the second sheet. The functions in $\Phi_\pm$ can be extended to be functions of $\Psi_\pm$ in the second sheet. There

$$\begin{aligned}\langle \psi^-, \phi^+ \rangle = &\int_{R_\epsilon^-} d\omega \langle {}^-\psi | \omega^- \rangle S_{II}(\omega) \langle {}^+\omega | \phi^+ \rangle \\ &+ \int_C d\omega \langle {}^-\psi | \omega^- \rangle S_{II}(\omega) \langle {}^+\omega | \phi^+ \rangle \\ &+ \int_{C_1} d\omega \langle {}^-\psi | \omega^- \rangle S_{II}(\omega) \langle {}^+\omega | \phi^+ \rangle. \end{aligned} \tag{14}$$

R_ϵ^- is the path over R^- plus the small semicircle of radius ϵ around z_v. C_1 is the inner path which embraces z_v and z_R (see Fig. 1). C is the lowest circular path. $S_{II}(\omega)$ is the S-matrix in the second sheet in the energy representation and $\langle {}^-\psi | \omega^- \rangle$ and $\langle {}^+\omega | \phi^+ \rangle$ are the corresponding extensions of $\langle {}^-\psi | E^- \rangle$ and $\langle {}^+E | \phi^+ \rangle$. We may deform the integration paths to reach in the limit a situation like this: C tends to the infinite semicircle in the lower half-plane and

$$\begin{aligned}&\int_{R_\epsilon^-} d\omega \langle {}^-\psi | \omega^- \rangle S_{II}(\omega) \langle {}^+\omega | \phi^+ \rangle \\ &= \int_0^{-E_v+\epsilon} dE \langle {}^-\psi | E^- \rangle S_{II}(E) \langle {}^+E | \phi^+ \rangle \\ &\quad + \int_{C_\epsilon} d\omega \langle {}^-\psi | \omega^- \rangle S_{II}(\omega) \langle {}^+\omega | \phi^+ \rangle \\ &\quad + \int_{-E_v-\epsilon}^{-\infty} d\omega \langle {}^-\psi | E^- \rangle S_{II}(E) \langle {}^+E | \phi^+ \rangle, \end{aligned} \tag{15}$$

where C_ϵ is the small semicircle above of $-E_v$ (see Fig. 1). The integral over C has been proven to be zero in Ref. 15. When ϵ goes to zero (15) tends to

$$\begin{aligned}&P\int_0^{-\infty} dE \langle {}^-\psi | E^- \rangle S_{II}(E) \langle {}^+E | \phi^+ \rangle \\ &\quad - \pi i \langle {}^-\psi | E_v^- \rangle s_v \langle {}^+E_v | \phi^+ \rangle, \end{aligned} \tag{16}$$

s_v being the residuum of $S_{II}(\omega)$ at $-E_v$. P stands for the principal value. The integral over C_1 becomes

$$\begin{aligned}&-2\pi i \langle {}^-\psi | z_R^- \rangle s_R \langle {}^+z_R | \phi^+ \rangle \\ &-2\pi i \langle {}^-\psi | -E_v^- \rangle s_v \langle {}^+(-E_v) | \phi^+ \rangle. \end{aligned} \tag{17}$$

Therefore we finally obtain

$$\begin{aligned}\langle \psi^-, \phi^+ \rangle = &\text{ background} - 2\pi i \langle {}^-\psi | z_R^- | \phi^+ \rangle s_R \langle {}^+z_R | \phi^+ \rangle \\ &- \frac{3}{2} \pi i \langle {}^-\psi | -E_v^- \rangle s_v \langle {}^+(-E_v) | \phi^+ \rangle, \end{aligned} \tag{18}$$

where

$$\text{background} = P\int_0^{-\infty} dE \langle {}^-\psi | E^- \rangle S_{II}(E) \langle {}^+E | \phi^+ \rangle. \tag{19}$$

The analysis of resonance behavior does not differ from that we have already made in Ref. 15. If we omit the arbitrary vector $\psi^- \in \Psi_+$ we have

$$\begin{aligned}\phi^+ = &\int_0^{-\infty} dE \, | E^- \rangle S_{II}(E) \langle {}^+E | \phi^+ \rangle \\ &- 2\pi i s_R | z_R^- \rangle \langle {}^+z_R | \phi^+ \rangle \\ &- \frac{3}{2} \pi i s_v | -E_v^- \rangle \langle {}^+(-E_v) | \phi^+ \rangle. \end{aligned} \tag{20}$$

This equation has to be understood as an equation in Ψ'_+. The integral exists in the weak sense only. If ϕ^+ is to represent a decaying state vector, it is the sum of three parts: $\alpha | z_R^- \rangle$ decaying exponentially,[12,15] the background, and $\beta | -E_v^- \rangle$ which corresponds to the virtual state.

IV. BEHAVIOR OF THE VECTOR $| -E_v^- \rangle$

The evolution $U(t) = e^{-itE}$ $(E > 0)$ is a well-defined operator on $L^2(R^+)$. On the other hand, if $t > 0$, $e^{itE} \Phi_+ \subset \Phi_+$, as can be easily proven. Hence, if $f(E) \in \Phi_+$, $U^*(t) f(E) = e^{itE} f(E) \in \Phi_+$ and, therefore, can be extended to a function in Ψ_+ which we also denote by $e^{itE} f(E)$ $(E \in R$ now). Corresponding to $U^*(t)$ $(t < 0)$ on Φ_+, we have an operator $U_1^*(t) = e^{itE}$ $(t > 0$ and $-\infty < E < \infty)$ so that $U_1^*(t) f(E) = e^{itE} f(E) \forall f(E) \Psi_+$. Since $U_1^*(t)$ is the counterpart on Ψ_+ of $U^*(t)$ and it is also unitary, we can look at its adjoint $U_1(t)$ as the evolution operator on the extended space Ψ_+. We are interested in obtaining the action of $U_1(t)$ on $| -E_v^- \rangle$ and the result will be the time evolution of the "state" $| -E_v^- \rangle$ for $t > 0$.[22]

Let $[\phi^-(E)]^* = \langle {}^-\phi | E^- \rangle \in \Psi_+$. Since $\phi^-(E)$ is continuous everywhere the functional $| -E_v^- \rangle$ when applied to $\phi^-(E)$ gives the value of $[\phi^-(E)]^*$ on $-E_v$: $\langle {}^-\phi | -E_v^- \rangle$. If we now apply $| -E_v^- \rangle$ to the function $U_1^*(t) \phi^-(E)$, we have

$$\begin{aligned}&\langle {}^-U_1^*(t)\phi | -E_v^- \rangle = \langle {}^-\phi | U_1(t) | -E_v^- \rangle, \\ &\langle {}^-e^{itE}\phi | -E_v^- \rangle = [e^{-itE}\phi^-(-E_v)]^* \\ &\quad = e^{itE_v}[\phi^-(-E_v)]^* = e^{itE_v} \langle {}^-\phi | -E_v^- \rangle. \end{aligned} \tag{21}$$

The first equality defines the extension of $U_1(t)$ to Ψ'_+. By omitting the arbitrary vector $\phi^- \in \Psi_+$ we finally obtain

$$U_1(t) | -E_v^- \rangle = e^{itE_v} | -E_v^- \rangle \quad (t > 0) \tag{22}$$

which gives us the evolution of the vector state $| -E_v^- \rangle$ for positive times.

The virtual state is equally well described by the func-

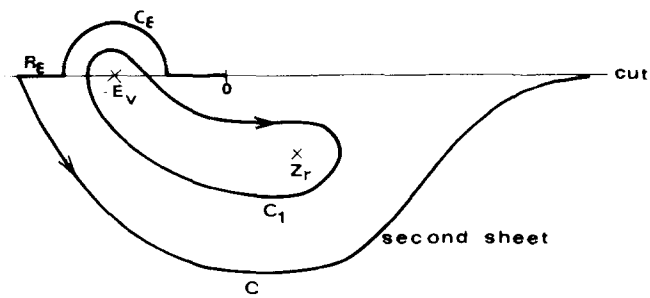


FIG. 1.

tional $|-E_v^+\rangle$ acting on Ψ_-. We can repeat the preceding procedure in the upper half-plane in the second sheet to obtain for ψ^- a similar formula to (20). Time evolution for $|-E_v^+\rangle$ is now only defined when $t<0$ and its explicit form is

$$U_1(t)|-E_v^+\rangle = e^{-itE_v}|-E_v^+\rangle \quad (t<0). \tag{23}$$

V. CONCLUDING REMARKS

Virtual scattering states can be described by a pair of functionals $|-E_v^-\rangle$ and $|-E_v^+\rangle$, respectively, acting on the spaces Ψ_+ and Ψ_- defined as above. Both functionals are continuous on the corresponding spaces. They associate every function on Ψ_+ or Ψ_- with its value at the point $-E_v$. Using a similar argument to the one showing the non-normalizability of the Dirac delta, we can see that $|-E_v^-\rangle$ and $|-E_v^+\rangle$ are non-normalizable as well.

In Ref. 15 we have described resonances with the help of a less refined model. We can here translate this description practically word by word. The difference is that the state spaces $\Phi_\pm$ are now made up by functions which are analytic in the whole Riemann surface, allowing the study of possible virtual states along with resonances.

Summarizing, we could compare virtual states and resonances within the present framework:

1. Virtual states are described by a pair of non-normalizable continuous functionals $|-E_v^+\rangle$ and $|-E_v^-\rangle$. They, respectively, exist and are stationary when $t<0$ and $t>0$.

2. Resonances are described by a pair of normalizable vectors in $L^2(R)$. They are, respectively, called the growing and decaying Gamow vector. However, their time evolution should be studied as functionals on $\Psi_\pm$ and not as vectors in $L^2(R)$.[12,15] Then, the decaying vector exists only when $t>0$ and its norm vanishes exponentially. The growing vector exists when $t<0$ only and its norm grows exponentially in the positive direction of time. Norm conservation is assured with the existence of a background term.

[1] A. Böhm, *Quantum Mechanics* (Springer, New York, 1979).
[2] J. M. Blatt and V. F. Weisskopf, *Theoretical Nuclear Physics* (Springer, New York, 1979).
[3] H. M. Nussenzveig, *Causality and Dispersion Relations* (Academic, New York, 1972).
[4] J. R. Taylor, *Scattering Theory* (Wiley, New York, 1972).
[5] I. M. Gel'fand and N. Y. Vilenkin, *Generalized Functions: Applications of Harmonic Analysis* (Academic, New York, 1964), Vol. 4.
[6] A. Böhm, *The Rigged Hilbert Space and Quantum Mechanics, Lecture Notes in Physics*, Vol. 78 (Springer, New York, 1978).
[7] J. P. Antoine, J. Math. Phys. **10**, 53 (1969).
[8] J. E. Roberts, Comm. Math. Phys. **3**, 98 (1966).
[9] O. Melsheimer, J. Math. Phys. **15**, 902 (1974).
[10] K. Maurin, Bull. Acad. Pol. Sci. Ser. Sci. Math. **VIII**, 471 (1959).
[11] A. Böhm, Lett. Math. Phys. **3**, 455 (1979).
[12] A. Böhm, J. Math. Phys. **22**, 2813 (1981).
[13] H. Baumgärtel, Math. Nachr. **75**, 133 (1976).
[14] G. Parravicini, V. Gorini, and E. C. G. Sudarshan, J. Math. Phys. **21**, 2208 (1980).
[15] M. Gadella, "A rigged Hilbert space of Hardy-class functions: Applications to resonances," J. Math. Phys. **24**, 1462 (1983).
[16] W. Rudin, *Functional Analysis* (McGraw-Hill, New York, 1973), p. 183.
[17] See for instance, P. Koosis, *Introduction to Hp Spaces*, London Math. Soc. Lecture Notes Series No. 40 (Cambridge U. P., New York, 1980), p. 176.
[18] C. Van Winter, Trans. Am. Math. Soc. **162**, 103 (1971); C. Van Winter, J. Math. An. **47**, 633 (1974).
[19] I. M. Gel'fand and G. E. Shilov, *Generalized Functions* (Academic, New York, 1968), Vol. 2, p. 126.
[20] A. Pietsch, *Nuclear Locally Convex Spaces* (Springer, Berlin, 1972).
[21] M. Reed and B. Simon, *Functional Analysis* (Academic, New York, 1972), p. 257.

[22] When $t<0$, $U_1(t)|-E_v^-\rangle$ is not well defined because $U_1^\dagger(t)\Psi_+ \not\subset \Psi'_+$. See Ref. 15 for a proof of this statement.

The Gel'fand–Levitan equation on a finite interval

Harry E. Moses[a)]
University of Lowell, Center for Atmospheric Research, Lowell, Massachusetts 01854
Reese T. Prosser[b)]
Dartmouth College, Department of Mathematics, Hanover, New Hampshire 03755

(Received 1 June 1982; accepted for publication 25 March 1983)

It is shown here, by using the methods of Gel'fand and Levitan, that one may alter in certain ways the spectral data (eigenvalues and normalizations) of the Schrödinger equation for a finite interval, by suitably altering the potential. These alterations include the deletion and insertion of a finite number of eigenvalues and/or renormalization of a finite number of eigenfunctions. The deletion of an eigenvalue requires the addition of a singular potential, while the insertion of an eigenvalue requires an adjustment of the boundary conditions resulting in a non-self-adjoint operator. Similarly the insertion of a continuous spectrum leads to non-self-adjoint operators. The non-self-adjoint operators have, nevertheless, complete sets of eigenfunctions and a corresponding spectral decomposition in a suitable subspace.

PACS numbers: 03.65.Nk, 03.65.Ge

1. INTRODUCTION

The Gel'fand-Levitan equation has long been used to obtain potentials from spectral data for the one-dimensional Schrödinger equation on the infinite interval $[0,\infty]$. In their original paper[1] Gel'fand and Levitan also considered the same equation on the finite interval $[0,l]$. See also Crum.[2] In this case the Schrödinger equation with appropriate boundary conditions becomes a Sturm–Liouville system with a necessarily discrete spectrum. Gel'fand and Levitan show that this spectrum may be replaced by any other satisfying the same asymptotic conditions by a suitable change in the potential.

In this paper we reconsider the problem on the finite interval, to see whether it is possible to *insert* or *delete* an eigenvalue in the spectrum—cases not covered by the work of Gel'fand and Levitan. We consider here only *finite rearrangements* of the spectrum, i.e., modification of only a finite number of eignevalues and/or normalizations, to avoid dealing with convergence questions. We show that it is always possible to *delete* an eigenvalue, at the cost of adding a singular potential, but that it is possible to insert an eigenvalue only in a generalized sense, and only at the cost of sacrificing the self-adjointness of the Schrödinger operator.

2. THE GEL'FAND–LEVITAN THEOREM

We consider here only the interval $[0,\pi]$, and only Dirichlet boundary conditions on $[0,\pi]$. Let $\mathscr{H}$ denote the Hilbert space of all square-integrable functions on the interval $[0,\pi]$ and $\mathscr{D}_0$ the subspace of twice-differentiable functions which vanish at 0 and π:

$$\mathscr{H} = L^2[0,\pi],$$
$$\mathscr{D}_0 = \{f\in C^2:\ f(0) = f(\pi) = 0\}. \tag{2.1}$$

Set $H_0 = -d^2/dx^2$ on $\mathscr{D}_0$. Then H_0 is essentially self-adjoint on $\mathscr{H}$, with domain $\mathscr{D}(H_0) = \mathscr{D}_0$, with spectrum

$$\{\lambda_n = n^2,\ n = 1,2,3,...\}, \tag{2.2}$$

and with eigenfunctions

$$\{\psi_n = \sin nx/n,\ n = 1,2,3,...\}, \tag{2.3}$$

normalized so that

$$\psi_n(0) = 0,$$
$$\psi_n'(0) = 1. \tag{2.4}$$

These eigenfunctions satisfy the orthogonality conditions

$$\int_0^\pi \frac{\psi_i(x)\psi_j(x)dx}{c_i} = \delta_{ij} \tag{2.5}$$

and the completeness relations

$$\sum_{i=1}^{\infty} \frac{\psi_i(x)\psi_i(y)}{c_i} = \delta(x-y), \tag{2.6}$$

where the normalization constants c_i are given by

$$c_i = \pi/2i^2. \tag{2.6'}$$

Next we define a family of functions $\zeta(x,\lambda)$, parameterized by the real variable λ:

$$\zeta(x,\lambda) = \begin{cases} \dfrac{\sin\sqrt{\lambda}\,x}{\sqrt{\lambda}}, & \lambda > 0, \\ x, & \lambda = 0, \\ \dfrac{\sinh\sqrt{-\lambda}\,x}{\sqrt{-\lambda}}, & \lambda < 0. \end{cases} \tag{2.7}$$

Then $\zeta(0,\lambda) = 0, \zeta'(0,\lambda) = 1$, and

$$-\frac{d^2}{dx^2}\zeta(x,\lambda) = \lambda\zeta(x,\lambda); \tag{2.8}$$

note that $\zeta(x,i^2) = \psi_i(x)$.

Now we introduce a sequence $\{\lambda_i, C_i\}_{i=1}^{\infty}$ of *eigenvalues* λ_i (numbered in increasing order) and *normalization constants* C_i, comprising the new spectral data, and define

$$\Omega(x,y) = \left(\sum_{i=1}^{\infty} \frac{\zeta(x,\lambda_i)\zeta(y,\lambda_i)}{C_i}\right) - \left(\sum_{i=1}^{\infty} \frac{\psi_i(x)\psi_i(y)}{c_i}\right). \tag{2.9}$$

[a)] This research was sponsored by the Air Force Office of Scientific Research, Air Force Systems Command, United States Air Force, under Grant No. AFOSR-81-0253.
[b)] This work was supported in part by National Science Foundation Grant No. MCS-8103409.

 0022-2488/83/082146-06$02.50

We assume that $\Omega(x,y)$ is well defined and well behaved [cf. Ref. 2; this is the case in all finite rearrangements, since then (2.9) reduces to a finite sum] and construct the Gel'fand–Levitan equation[1]:

$$K(x,y)+\Omega(x,y)+\int_0^x K(x,z)\Omega(z,y)dz=0. \tag{2.10}$$

Theorem 1: Assume that $\Omega(x,y)$ is a twice-differentiable function of x and y for $0\leqslant y\leqslant x\leqslant\pi$. Then (a) The Gel'fand–Levitan equation (2.10) admits a unique twice-differentiable solution $K(x,y)$ valid for $0\leqslant y\leqslant x\leqslant\pi$. (b) The functions $\chi(x,\lambda_i)$ defined by

$$\chi(x,\lambda_i)=\zeta(x,\lambda_i)+\int_0^x K(x,y)\zeta(y,\lambda_i)dy \tag{2.11}$$

satisfy the differential equations

$$\left(-\frac{d^2}{dx^2}+V(x)\right)\chi(x,\lambda_i)=\lambda_i\chi(x,\lambda_i), \tag{2.12}$$

where

$$V(x)=2\frac{d}{dx}K(x,x). \tag{2.13}$$

(c) The functions $\chi(x,\lambda_i)$ satisfy the boundary conditions

$$\chi(0,\lambda_i)=0;\quad \chi'(0,\lambda_i)=1. \tag{2.14}$$

(d) Finally, the functions $\chi(x,\lambda_i)$ satisfy the completeness relation

$$\sum_{i=1}^{\infty}\frac{\chi(x,\lambda_i)\chi(y,\lambda_i)}{C_i}=\delta(x-y) \tag{2.15}$$

with the normalization constants C_i.

The proof of Theorem 1 is similar to proofs in Refs. 1, 3, 4 and 5 and will not be given here.

This theorem and its proof do not depend in any way on the boundary condition at $x=\pi$. Indeed, the theorem says nothing about whether or not the functions $\chi(x,\lambda_i)$ are orthogonal on $[0,\pi]$ or whether or not they vanish at $x=\pi$ and so lie in $\mathscr{D}(H_0)$. It is precisely these points which we wish to investigate here: Under what conditions do the $\chi(x,\lambda_i)$ form an orthogonal basis in $\mathscr{D}(H_0)$?

Gel'fand and Levitan essentially show in Ref. 2 that if $\lambda_i=i^2+O(1/i)$ and if $C_i=\pi/2i^2+O(1/i^3)$, then indeed the functions $\chi(x,\lambda_i)$ do vanish at $x=\pi$. We verify this directly for finite rearrangements below. If, however, an eigenvalue is *inserted* or *deleted*, then for large i, $\lambda_i=(i\pm1)^2$, and these results do not apply. We show directly that if an eigenvalue is deleted, then the $\chi(x,\lambda_i)$ are orthogonal on $[0,\pi]$ and do vanish at $x=\pi$, while if an eigenvalue is inserted, then they are not and do not. More precisely, we establish for any finite rearrangement, i.e., any modification of a finite number of eigenvalues and/or normalizations, the following.

Theorem 2: In any finite rearrangement of the spectral data for H_0, the resulting eigenfunctions $\chi(x,\lambda_i)$ are orthogonal on $[0,\pi]$ and vanish at $x=0$ and $x=\pi$ if and only if no extra eigenvalues are inserted.

The proof consists essentially of examining three cases, which are treated separately below.

3. DELETION OF AN EIGENVALUE

We calculate explicitly the case where the lowest eigenvalue $\lambda_n=n^2=1$ is deleted from the spectrum of H_0. In this case $\{\lambda_i,C_i\}=\{(i+1)^2,\pi/2(i+1)^2\}$, and the Gel'fand–Levitan kernel (2.9) becomes

$$\Omega(x,y)=-(2/\pi)\psi_1(x)\psi_1(y), \tag{3.1}$$

where $\psi_1(x)=\sin x$.

The Gel'fand–Levitan equation (2.10) becomes

$$K(x,y)=\frac{2}{\pi}\psi_1(x)\psi_1(y)+\frac{2}{\pi}\int_0^x K(x,z)\psi_1(z)\psi_1(y)dz. \tag{3.2}$$

Since the kernel $\Omega(x,y)$ is separable, we may assume a solution of the form

$$K(x,y)=(2/\pi)\psi_1(x)b(x)\psi_1(y) \tag{3.3}$$

with $b(x)$ to be determined. Substituting (3.3) into (3.2) yields

$$b(x)=1+b(x)d(x), \tag{3.4}$$

where

$$d(x)=\frac{2}{\pi}\int_0^x\psi_1(z)^2dz. \tag{3.5}$$

Hence

$$b(x)=1/[1-d(x)] \tag{3.6}$$

and

$$K(x,y)=\frac{2}{\pi}\frac{\psi_1(x)\psi_1(y)}{1-(2/\pi)\int_0^x\psi_1(z)^2dz}. \tag{3.7}$$

Now for $\lambda_i=(i+1)^2$ and $\chi_i(x)=\chi(x,\lambda_i)$,

$$\chi_i(x)=\psi_{i+1}(x)+\frac{2}{\pi}\frac{\psi_1(x)\int_0^x\psi_1(y)\psi_{i+1}(y)dy}{1-(2/\pi)\int_0^x\psi_1(z)^2dz}. \tag{3.8}$$

As $x\to\pi,\psi_{i+1}(x)\to0$, and one verifies easily that in the remaining term the numerator is of fourth order in $(\pi-x)$ while the denominator is of third order, so that this term $\to0$ also. Thus

$$\chi_i(x)\to0 \text{ as } x\to\pi. \tag{3.9}$$

We also note that

$$\begin{aligned}(\chi_i,\chi_j)=\int_0^\pi\chi_i(x)\chi_j(x)dx=\lim_{x\to\pi}\Bigg\{&\int_0^x\psi_{i+1}(y)\psi_{j+1}(y)dy\\&+\frac{2}{\pi}\int_0^x\left[\frac{\psi_{i+1}(y)\psi_1(y)\int_0^y\psi_1(z)\psi_{j+1}(z)dz}{1-(2/\pi)\int_0^y\psi_1(z)^2dz}\right]dy\\&+\frac{2}{\pi}\int_0^x\left[\frac{\int_0^y\psi_{i+1}(z)\psi_1(z)dz\psi_1(y)\psi_{j+1}(y)}{1-(2/\pi)\int_0^y\psi_1(z)^2dz}\right]dy\\&+\frac{4}{\pi^2}\int_0^x\left[\frac{\int_0^y\psi_{i+1}(z)\psi_1(z)dz\psi_1(y)^2\int_0^y\psi_1(z)\psi_{j+1}(z)dz}{(1-(2/\pi)\int_0^y\psi_1(z)^2dz)^2}\right]\Bigg\}dy.\end{aligned} \tag{3.10}$$

Since

$$\frac{d}{dy}\left(1-\frac{2}{\pi}\int_0^y \psi_1(z)^2dz\right)^{-1} = +\frac{2}{\pi}\psi_1(y)^2\left(1-\frac{2}{\pi}\int_0^y \psi_1(z)^2dz\right)^{-2} \tag{3.11}$$

we may integrate the last term of (3.10) by parts to obtain

$$(\chi_i,\chi_j) = \lim_{x\to\pi}\left[\int_0^x \psi_{i+1}(y)\psi_{j+1}(y)dy + \frac{\int_0^x \psi_{i+1}(y)\psi_1(y)dy\int_0^x \psi_1(z)\psi_{j+1}(z)dz}{1-(2/\pi)\int_0^x \psi_1(z)^2dz}\right]. \tag{3.12}$$

Here the first term $\to(\psi_{i+1},\psi_{j+1})$ as $x\to\pi$. The second term vanishes to sixth order in the numerator and to third order in the denominator, and so vanishes as $x\to\pi$. Hence

$$(\chi_i,\chi_j) = (\psi_{i+1},\psi_{j+1}) = \frac{2(i+1)^2}{\pi}\delta_{ij} \tag{3.13}$$

so that the χ_i satisfy the same orthogonality conditions as do the ψ_{i+1}. They are complete by (2.15)

What happens to the missing eigenfunction? Formally, we may define

$$\chi_1(x) = \psi_1(x) + \frac{2}{\pi}\int_0^x \frac{\psi_1(x)\psi_1(y)^2dy}{1-(2/\pi)\int_0^x \psi_1(z)^2dz} = \frac{\psi_1(x)}{1-(2/\pi)\int_0^x \psi_1(z)^2dz}. \tag{3.14}$$

As $x\to\pi$, the numerator vanishes to first order, but the denominator vanishes to third order, so that $\chi_1(x)$ has a pole of second order at $x=\pi$. It follows that $\chi_1\notin\mathscr{H}$; i.e., the missing eigenfunction χ_1 is not normalizable.

What about $V(x)$? We have

$$V(x) = 2\frac{d}{dx}K(x,x) = 2\frac{d}{dx}\frac{2}{\pi}\frac{\psi_1(x)^2}{1-(2/\pi)\int_0^x \psi_1(z)^2dz} = -2\frac{d^2}{dx^2}\ln\left(1-\frac{2}{\pi}\int_0^x \psi_1(z)^2dz\right). \tag{3.15}$$

Now as $x\to\pi$, $K(x,x)$ vanishes to second order in the numerator, and to third order in the denominator. Hence $V(x)$ also has a pole of second order at $x=\pi$, and hence is not bounded. Nevertheless, $H=H_0+V$ is essentially self-adjoint on $\mathscr{D}(H_0)$, with eigenvalues $\{\lambda_i=(i+1)^2\}_{i=1}^{\infty}$ and eigenfunctions $\{\chi_i\}_{i=1}^{\infty}$.

4. INSERTION OF AN EIGENVALUE

Now we calculate explicitly the case where an extra eigenvalue is inserted into the spectral data of H_0. We take $\lambda_1=0$ for the extra eigenvalue, with normalization constant $C_1=1$. In this case $\{\lambda_i,C_i\}=\{0,1\}\{(i-1)^2,\pi/2(i-1)^2\}$, and $\Omega(x,y)$ becomes

$$\Omega(x,y) = \psi_0(x)\psi_0(y),$$

where

$$\psi_0(x) = x\equiv\zeta(x,0). \tag{4.1}$$

Now we assume that $K(x,y)$ has the form

$$K(x,y) = \psi_0(x)b(x)\psi_0(y). \tag{4.2}$$

Substituting (4.1) and (4.2) into the Gel'fand–Levitan equation (2.10) and solving for $b(x)$ as in the previous section, we find

$$K(x,y) = -\frac{\psi_0(x)\psi_0(y)}{1+\int_0^x \psi_0(z)^2dz}. \tag{4.3}$$

Hence if $\chi(x,\lambda_i)=\chi_i$, then

$$\chi_1(x) = \psi_0(x) - \frac{\int_0^x \psi_0(x)\psi_0(y)^2dy}{1+\int_0^x \psi_0(z)^2dz} = \frac{\psi_0(x)}{1+\int_0^x \psi_0(z)^2dz} \tag{4.4}$$

and for $i>1$,

$$\chi_i(x) = \psi_{i-1}(x) - \frac{\int_0^x \psi_0(x)\psi_0(y)\psi_{i-1}(y)dy}{1+\int_0^x \psi_0(z)^2dz}. \tag{4.5}$$

We first consider the orthogonality relations:

$$(\chi_i,\chi_j) = \int_0^\pi \chi_i(x)\chi_j(x)dx. \tag{4.6}$$

A calculation entirely similar to that in the previous section shows that

$$(\chi_i,\chi_j) = (\psi_{i-1},\psi_{j-1}) - \frac{(\psi_{i-1},\psi_0)(\psi_0,\psi_{j-1})}{1+(\psi_0,\psi_0)}. \tag{4.7}$$

Since $\psi_0(x)=x$ and $\psi_n(x)=\sin nx/n$, we find

$$(\psi_0,\psi_n) = 2[(-1)^{n+1}/n^2],\qquad (\psi_0,\psi_0) = \pi^3/3, \tag{4.8}$$

so that for $i,j>1$

$$(\chi_i,\chi_j) = (\psi_{i-1},\psi_{j-1}) - 12\frac{(-1)^{i+j}}{(3+\pi^3)(i-1)^2(j-1)^2}, \tag{4.9}$$

while for $i>1, j=1$

$$(\chi_i,\chi_1) = \frac{(\psi_{i-1},\psi_0)}{1+(\psi_0,\psi_0)} = 6\frac{(-1)^i}{(3+\pi^3)(i-1)^2},\qquad (\chi_1,\chi_1) = \frac{(\psi_0,\psi_0)}{1+(\psi_0,\psi_0)} = \frac{\pi^3}{3+\pi^3}. \tag{4.10}$$

We see here that the eigenfunctions χ_i are not orthogonal at all, and that $(\chi_1,\chi_1)\neq C_1, (\chi_i,\chi_i)\neq(\psi_{i-1},\psi_{i-1})$ for $i>1$, so that χ_i are not properly normalized, either. Nevertheless, the completeness relation (2.15) still holds.

At the boundary $x=\pi$, we find

$$\chi_i(\pi) = \psi_{i-1}(\pi) - \frac{\psi_0(\pi)\int_0^\pi \psi_0(y)\psi_{i-1}(y)dy}{1+\int_0^\pi \psi_0(z)^2dz} = \psi_{i-1}(\pi) - \frac{(\psi_{i-1},\psi_0)\psi_0(\pi)}{1+(\psi_0,\psi_0)} = \psi_{i-1}(\pi) - (\chi_i,\chi_1)\psi_0(\pi). \tag{4.11}$$

Since $\psi_0(\pi)$, and $\psi_n(\pi)=0$, $n>0$, we have

$$\chi_i(\pi) = 6\pi\frac{(-1)^{i+1}}{(3+\pi^3)(i-1)^2},\ i>1,\qquad \chi_1(\pi) = 3\pi/(3+\pi^3). \tag{4.12}$$

Thus $\chi_i(\pi)\neq 0$, and $\chi_i \notin \mathscr{D}_0 = \mathscr{D}(H_0)$, although χ_i is bounded.

If we define the subspace $\mathscr{D}_1$ of $\mathscr{H}$ by [cf. (2.1)]

$$\mathscr{D}_1 = \{f \in C^2 : f(0) = 0, f(\pi) = -(f,\chi_1)\psi_0(\pi)\}, \tag{4.13}$$

then $\chi_i \in \mathscr{D}_1$ for $i > 1$. It follows that the operator $U = I + K$ carries $\mathscr{D}_0$ into $\mathscr{D}_1$, and since the ψ_i span $\mathscr{H}$, the $\chi_{i+1} = U\psi_i$ must also span $\mathscr{H}$. In fact, if $f \in \mathscr{D}_0, f = \Sigma_{i=1}^{\infty}(2i^2/\pi)(f,\psi_i)\psi_i$, then $Uf = \Sigma_{i=1}^{\infty}(2i^2/\pi)(f,\psi_i)\chi_{i+1}$, and since U is invertible, every function in $\mathscr{D}_1$ is of the form Uf. Moreover, since $HU = UH_0$, we see that $\mathscr{D}_1 = \mathscr{D}(H)$ is the domain of $H = H_0 + V$, and not $\mathscr{D}_0 = \mathscr{D}(H_0)$. Since the boundary condition defining $\mathscr{D}_1$ at $x = \pi$ is not of self-adjoint form, we see that H cannot be self-adjoint; this is why the eigenfunctions of H are not orthogonal.

What about χ_1? From (4.12) we see that

$$\chi_1(\pi) = +\pi\frac{\psi_0(\pi)}{1 + (\psi_0,\psi_0)} \tag{4.14}$$

so that χ_1 does *not* lie in $\mathscr{D}_1$. This is no real surprise, since ψ_0 does not lie in $\mathscr{D}_0$. Nevertheless the completeness relations (2.15) still hold! In fact, for any $f \in \mathscr{H}$, and $g = Uf$,

$$g = \sum_{i=1}^{\infty}\frac{2i^2}{\pi}(f,\psi_i)\chi_{i+1}; \tag{4.15}$$

in particular, if $f = \psi_0$. Then $g = \chi_1$, and

$$\chi_1 = \sum_{i=1}^{\infty}\frac{2i^2}{\pi}(\psi_0,\psi_i)\chi_{i+1}. \tag{4.16}$$

Since by (4.10)

$$(g,\chi_1) = \frac{(f,\psi_0)}{1 + (\psi_0,\psi_0)} \tag{4.17}$$

and

$$(g,\chi_{i+1}) = (f,\psi_i) - \frac{(f,\psi_0)(\psi_0,\psi_i)}{1 + (\psi_0,\psi_0)}, \tag{4.18}$$

we have

$$\begin{aligned} g - (g,\chi_1)\chi_1 &= \sum_{i=1}^{\infty}\frac{2i^2}{\pi}((f,\psi_i) - (g,\chi_1)(\psi_0,\psi_i))\chi_{i+1} \\ &= \sum_{i=1}^{\infty}\frac{2i^2}{\pi}(g,\chi_{i+1})\chi_{i+1}, \end{aligned} \tag{4.19}$$

which leads to the completeness relation (2.15).

If $g \in \mathscr{D}_1$ then $Hg \in L^2$ and formally we have

$$\begin{aligned} Hg &= \sum_{i=1}^{\infty}\frac{2i^2}{\pi}(g,\chi_{i+1})H\chi_{i+1} + (g,\chi_1)H\chi_1 \\ &= \sum_{i=1}^{\infty}\frac{2i^4}{\pi}(g,\chi_{i+1})\chi_{i+1}. \end{aligned} \tag{4.20}$$

The trouble here is that $\chi_1 \notin \mathscr{D}_1$, so $H\chi_1$ is not defined except formally. However, if $g \in \mathscr{D}_1$ and $g(\pi) = 0$, then $g = Uf$ with $f \in \mathscr{D}_0$, and

$$0 = g(\pi) = f(\pi) - (g,\chi_1)\psi_0(\pi). \tag{4.21}$$

Hence if $g \in \mathscr{D}_1$ and $g(\pi) = 0$, then $(g,\chi_1) = 0$, and it follows that (4.20) holds for all such g. In fact, for all such g we obtain from (4.19) that

$$\begin{aligned} g &= \sum_{i=1}^{\infty}\frac{2i^2}{\pi}(f,\psi_i)\chi_{i+1}, \\ Hg &= HUf = UH_0f = \sum_{i=1}^{\infty}\frac{2i^4}{\pi}(f,\psi_i)\chi_{i+1} \\ &= \sum_{i=1}^{\infty}\frac{2i^4}{\pi}(g,\chi_{i+1})\chi_{i+1} \end{aligned} \tag{4.22}$$

so that (4.20) holds for the smaller domain $\mathscr{D}_1 \cap \mathscr{D}_0$. In this way we see that the true spectrum of H is the same as that of H_0, and that we have inserted the spectral data $\{\lambda = 0, C = 1\}$, *not into the spectrum of* H, *but rather into the completeness relation* (2.15).

In the same way we find that if $g \in \mathscr{D}_1 \cap \mathscr{D}_0$ then

$$\exp(\sqrt{-1}Ht)g = \sum_{i=1}^{\infty}\exp(\sqrt{-1}i^2t)\frac{2i^2}{\pi}(g,\chi_{i+1})\chi_{i+1}. \tag{4.23}$$

The sum here actually converges in norm for all g, and represents the one-parameter group $W(t) = \exp(\sqrt{-1}Ht)$ of (nonunitary) operators generated by H, related to $W_0(t) = \exp(\sqrt{-1}H_0t)$ by the similarity transformation $W(t)U = UW_0(t)$.

Finally, for $V(x)$ we have

$$V(x) = 2\frac{d}{dx}K(x,x) = -2\frac{d}{dx}\ln\left(1 + \int_0^x \psi_0^2(z)dz\right). \tag{4.24}$$

It should be understood that the operator $H = -d^2/dx^2 + V(x)$ with domain $\mathscr{D}_1$ discussed here is *not* the same as the Sturm–Liouville operator $H_2 = -d^2/dx^2 + V(x)$ with domain $\mathscr{D}_0$. The operator H_2 is self-adjoint and has for its eigenvalues and normalization constants a set of data completely different from that of our operator H. These data can be calculated by standard Sturm–Liouville methods.

More generally, if $\xi(x,\lambda)$ is any eigenfunction of $-d^2/dx^2$ [cf. (2.7)] with eigenvalue λ, normalized so that $\xi(0,\lambda) = 0, \xi'(0,\lambda) = 1$, and if $K(x,y)$ is given by (4.3), then $\chi(x,\lambda)$, defined by

$$\chi(x,\lambda) = \xi(x,\lambda) + \int_0^x K(x,z)\xi(z,\lambda)dz, \tag{4.25}$$

is an eigenfunction of $-d^2/dx^2 + V(x)$ with the same eigenvalue and the same normalization, if $V(x)$ is given by (4.24). Hence the spectrum of the operator H_2 consists precisely of those λ for which $\chi(x,\lambda)$ vanishes at $x = \pi$, while the spectrum of H consists of precisely those λ for which $\chi(\pi,\lambda) = (\chi,\chi_1)\psi_0(\pi)$. In fact, one can determine in this way the spectrum of any operator of the form $-d^2/dx^2 + V(x)$ on any interval $[0,l]$ with boundary conditions $\chi(0) = 0$ and $\chi(l)$ specified.

5. REPOSITIONING AN EIGENVALUE

Finally, we calculate explicitly the case where the lowest eigenvalue is moved from $\lambda_1 = 1$ to $\lambda_1 = 0$. This case is already covered by the work of Gel'fand and Levitan, and we know that the eigenfunctions $\chi_i(x)$ must vanish at $x = \pi$. On the other hand, this case can be considered as a deletion of $\lambda_1 = 1$ followed by an insertion of $\lambda_1 = 0$ with the same nor-

malization and in view of the results of the preceding sections, it is not clear how this vanishing comes about.

In this case we take $\Omega(x,y)$ to be

$$\Omega(x,y) = (2/\pi)(\psi_0(x)\psi_0(y) - \psi_1(x)\psi_1(y)) \times \begin{cases}\psi_1(x) = \sin x,\\ \psi_0(x) = x(=\zeta(x,0)),\end{cases}$$
$$= \frac{2}{\pi}\sum \psi_i(x)C_{ij}\psi_j(y) \quad \begin{cases} i,j = 0 \text{ or } 1,\\ C_{ij} = (-1)^i\delta_{ij}\end{cases} \tag{5.1}$$

and assume that $K(x,y)$ has the form

$$K(x,y) = \frac{2}{\pi}\sum_{ij}\psi_i(x)b_{ij}(x)\psi_j(y), \quad i,j = 0 \text{ or } 1. \tag{5.2}$$

Substituting (5.1) and (5.2) into the Gel'fand–Levitan equation (2.10) and solving for $b_{ij}(x)$ as in the previous sections, we find

$$b_{ij}(x) = -C_{ij} - \sum_{kl} b_{ik}(x)d_{kl}(x)C_{lj}, \tag{5.3}$$

where

$$C_{ij} = (-1)^i\delta_{ik}$$

and

$$d_{ij}(x) = \frac{2}{\pi}\int_0^x \psi_i(y)\psi_j(y)dy. \tag{5.4}$$

In short, we have

$$B = -C - BDC \tag{5.5}$$

with solution

$$B = -(C + D)^{-1} \tag{5.6}$$

or

$$b_{ij}(x) = \begin{pmatrix} \dfrac{1 - d_{11}(x)}{\delta(x)} & \dfrac{d_{01}(x)}{\delta(x)} \\ \dfrac{d_{10}(x)}{\delta(x)} & \dfrac{-1 - d_{00}(x)}{\delta(x)} \end{pmatrix},$$
$$\delta(x) = \det(C + D) = (d_{11}(x) - 1)(d_{00}(x) + 1) - d_{01}(x)\,d_{10}(x). \tag{5.7}$$

We note here for later reference that

$$B^{-1}(x)\frac{d}{dx}B(x) = D'(x)B(x),$$

where

$$D'(x) = \frac{d}{dx}D(x),$$
$$\delta^{-1}(x)\frac{d}{dx}\delta(x) = -\operatorname{tr}(D'(x)B(x)) = -K(x,x). \tag{5.8}$$

From (5.2) and (5.7) we find

$$\chi_0(x) = (I + K)\psi_0(x) = \psi_0(x)\frac{(1 - d_{11}(x))}{\delta(x)} + \psi_1(x)\frac{d_{10}(x)}{\delta(x)},$$
$$\chi_1(x) = (I + K)\psi_1(x) = -\psi_0(x)\frac{d_{01}(x)}{\delta(x)} + \psi_1(x)\frac{(1 + d_{00}(x))}{\delta(x)},$$

and for $n > 1$

$$\chi_n(x) = (I + K)\psi_n(x) = \psi_n(x) + \frac{2}{\pi}\sum_{ij}\psi_i(x)b_{ij}(x)\int_0^x \psi_j(y)\psi_n(y)dy. \tag{5.9}$$

For the orthogonality relations we calculate

$$(\chi_m,\chi_n) = (\psi_m,\psi_n) + \frac{2}{\pi}\sum_{ij}\int_0^\pi \psi_m(x)\psi_i(x)b_{ij}(x)\int_0^x \psi_j(y)\psi_n(y)dy\,dx + \frac{2}{\pi}\sum_{ij}\int_0^\pi\int_0^x \psi_m(y)\psi_i(y)dy b_{ij}(x)\psi_j(x)\psi_n(x)dx + \frac{4}{\pi^2}\sum_{ijkl}\int_0^\pi\int_0^x \psi_m(y)\psi_j(y)dy b_{ik}(x)\psi_k(x)\psi_l(x)b_{lj}(x) \times \int_0^x \psi_j(y)\psi_n(y)dy\,dx. \tag{5.10}$$

Integrating the last term by parts and using (5.8), we find [cf. (3.12) and (4.7)]

$$(\chi_m,\chi_n) = (\psi_m,\psi_n) + \frac{2}{\pi}\sum_{ij}(\psi_m,\psi_i)b_{ij}(\pi)(\psi_j,\psi_n). \tag{5.11}$$

But from (5.7)

$$b_{ij}(\pi) = \begin{pmatrix} \dfrac{1 - (2/\pi)(\psi_1,\psi_1)}{-(4/\pi^2)(\psi_0,\psi_1)^2} & \dfrac{(2/\pi)(\psi_0,\psi_1)}{-(4/\pi^2)(\psi_0,\psi_1)^2} \\ \dfrac{(2/\pi)(\psi_1,\psi_0)}{-(4/\pi^2)(\psi_0,\psi_1)^2} & \dfrac{1 + (2/\pi)(\psi_0,\psi_0)}{(4/\pi^2)(\psi_0,\psi_1)^2} \end{pmatrix}. \tag{5.12}$$

Hence $b_{00}(\pi) = 0$, and since $(\psi_n,\psi_1) = 0$ if $n > 1$, we see that for $m, n > 1$,

$$(\chi_m,\chi_n) = (\psi_m,\psi_n). \tag{5.13}$$

Moreover, if $m > 1$,

$$(\chi_m,\chi_0) = (\psi_m,\psi_0) + \frac{2}{\pi}(\psi_m,\psi_0)b_{01}(\pi)(\psi_1,\psi_0) = (\psi_m,\psi_0)\left(1 - \frac{1}{(2/\pi)(\psi_1,\psi_0)}\frac{2}{\pi}(\psi_1,\psi_0)\right) = 0 \tag{5.14}$$

and

$$(\chi_0,\chi_0) = (\psi_0,\psi_0) - \frac{(\psi_0,\psi_0)(\psi_0,\psi_1)(\psi_1,\psi_0)}{(\psi_0,\psi_1)^2} - \frac{(\psi_0,\psi_1)(\psi_1,\psi_0)(\psi_0,\psi_0)}{(\psi_1,\psi_0)^2} + \frac{(\pi/2)(\psi_0,\psi_1)(1 + (2/\pi)(\psi_0,\psi_0))(\psi_1,\psi_0)}{(\psi_0,\psi_1)^2} = \pi/2. \tag{5.15}$$

In summary:

$$(\chi_m,\chi_n) = (\psi_m,\psi_n), \quad m,n = 2,3,4\ldots$$
$$(\chi_0,\chi_0) = (\psi_1,\psi_1) = \pi/2. \tag{5.16}$$

As $x \to \pi$, we see from (5.9) that for $n > 1$

$$\chi_n(\pi) = \psi_n(\pi) + \sum_{ij}\psi_i(\pi)b_{ij}(\pi)(\psi_j,\psi_n) = 0 + \psi_0(\pi)b_{00}(\pi)(\psi_0,\psi_n) = 0. \tag{5.17}$$

If $n = 0$,

$$\chi_0(\pi) = \psi_0(\pi)b_{00}(\pi) + \psi_1(\pi)b_{10}(\pi) = 0. \tag{5.18}$$

But if $n = 1$

$$\chi_1(\pi) = -\psi_0(\pi)b_{01}(\pi) + \psi_1(\pi)b_{11}(\pi)$$
$$= \frac{\psi_0(\pi)}{(\psi_0,\psi_1)}$$
$$\neq 0. \tag{5.19}$$

Thus the eigenfunctions χ_i all vanish at $x = \pi$ except the deleted one, χ_1. The other eigenfunctions all lie in $\mathscr{D}(H_0)$ and satisfy the orthogonality relations (5.16) and the completeness relations (2.15).

Finally, we see from (5.8) that in this case the potential $V(x)$ has the form [cf. (3.15), (4.24)]

$$V(x) = 2\frac{d}{dx}K(x,x)$$
$$= 2\frac{d^2}{dx^2}\operatorname{tr}\ln B$$
$$= -2\frac{d^2}{dx^2}\ln\delta(x), \tag{5.20}$$

where

$$\delta(x) = \det(C + D(x)) = \det\begin{pmatrix} d_{00}+1 & d_{01} \\ d_{10} & d_{11}-1 \end{pmatrix}, \tag{5.21}$$
$$d_{ij}(x) = \frac{2}{\pi}\int_0^x \psi_i(y)\psi_j(y)dy.$$

In particular, $V(x)$ is smooth and bounded on $[0,\pi]$.

Note that we could move the eigenvalue λ_1 to λ_0, where λ_0 is any real number different from the λ_n. If λ_0 is positive (negative) then $\psi_0(x)$ must be replaced by

$$\zeta(x,\lambda_0) = \frac{\sin\sqrt{\lambda_0}x}{\sqrt{\lambda_0}} \quad \frac{\sinh\sqrt{-\lambda_0}x}{\sqrt{-\lambda_0}}$$

and everything else goes through as before. If we move λ_1 above 2, however, then we must reorder the eigenvalues, so that we actually move more than one. Note that we cannot move λ_1 exactly to λ_2, since then the spectrum of H would have a double eigenvalue. In that case the determinant $\delta(x)$ in (5.7) vanishes to sixth order at $x = P$, so that $b_{ij}(x)$ has a pole of order at least 3, and so both $\psi_0(x)$ and $\psi_1(x)$ have poles of order 2 at $x = \pi$.

6. DISCUSSION

It is clear that by iterating the arguments of the previous sections, we can treat any finite rearrrangement of the spectral data for H_0. We have not discussed explicitly the case of a simple change of normalization constant, say C_1, since that is covered by the results of Gel'fand–Levitan and in any case it follows a similar pattern. With some care, the general case of a finite rearrangement can be treated by matrix methods as in Sec. 5, giving a complete proof of our Theorem 2, but the results are not so transparent.

It is formally possible to add a continuous spectrum to the spectral data for H_0, in the same manner that we add a single eigenvalue. One finds that in this case none of the eigenfunctions of H again will vanish at $x = \pi$. We consider this case in the last section.

7. INSERTION OF A CONTINUOUS SPECTRUM

In this last section we consider the possibility of inserting an absolutely continuous segment into the spectrum of H_0. Explicitly, we propose for the spectrum of H the set

$$\left\{i^2, \frac{2i^2}{\pi}\right\}\{(0,\infty),\pi\}.$$

We take for $\Omega(x,y)$ the following:

$$\Omega(x,y) = \frac{2}{\pi}\int_0^\infty \frac{\sin kx \sin ky}{k^2}dk$$
$$= \frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{\sin kx \sin ky}{k^2}dk. \tag{7.1}$$

Residue calculus gives

$$\Omega(x,y) = y\eta(x-y) + x\eta(y-x), \tag{7.2}$$

where $\eta(x)$ is the Heaviside function:

$$\eta(x) = \begin{cases} 1, & \text{if } x > 0, \\ \frac{1}{2}, & \text{if } x = 0, \\ 0, & \text{if } x < 0. \end{cases} \tag{7.3}$$

For $x \geqslant y$, then, the Gel'fand–Levitan equation becomes

$$K(x,y) = -y - \left(\int_y^x K(x,z)dz\right)y - \int_0^y K(x,z)z\,dz. \tag{7.4}$$

One readily obtains from (7.4) the differential equation

$$\frac{\partial^2}{\partial y^2}K(x,y) = K(x,y) \tag{7.5}$$

with general solution

$$K(x,y) = A(x)e^y + B(x)e^{-y}. \tag{7.6}$$

Substituting (7.6) into (7.4), we find

$$A(x) = -B(x) = -\tfrac{1}{2}\operatorname{sech} x. \tag{7.7}$$

Hence

$$K(x,y) = -\operatorname{sech}x \sinh y, \tag{7.8}$$

$$V(x) = 2\frac{d}{dx}K(x,x) = -2\operatorname{sech}^2 x, \tag{7.9}$$

$$\chi_i(x) = (i^2+1)^{-1}[(i^2+1)\psi_i(x) + \tanh x\psi_i'(x)],$$
$$\chi(x,k^2) = (k^2+1)^{-1}[(k^2+1)\zeta(x,k^2) + \tanh x\zeta'(x,k^2)]. \tag{7.10}$$

The boundary condition at $x = \pi$ becomes

$$\chi_i(\pi) = -(\psi_i,\zeta_{-1})\operatorname{sech}(\pi), \tag{7.11}$$

where $\zeta_{-1} = \zeta(x,-1) = \sinh x$ [cf. (2.7)]. Again this condition is satisfied by the eigenfunctions χ_i corresponding to the original spectrum, and not by the eigenvalues $\chi(x,k^2)$, corresponding to the inserted segment. Nevertheless, the completeness relation holds:

$$\sum_{i=1}^\infty \frac{2i^2}{\pi}\chi_i(x)\chi_i(y) + \frac{2}{\pi}\int_0^\infty \chi(x,k^2)\chi(y,k^2)dk = \delta(x-y). \tag{7.12}$$

[1] I. M. Gel'fand and B. M. Levitan, Amer. Math. Soc. Transl. **1**, 253–304 (1955).
[2] M. M. Crum, Q. J. Math. XX, 121–127 (1955).
[3] I. Kay and H. E. Moses, Nuovo Cimento **3**, 276–304 (1956). Reprinted in I. Kay and H. E. Moses, *Inverse Scattering Papers 1955–1963* (Math. Sci. Press, Brookline, MA, 1982).
[4] H. E. Moses, J. Math. Phys. **20**, 2047–2053 (1979).
[5] P. B. Abraham and H. E. Moses, Phys. Rev. A **22**, 1333–1340 (1980).

Inverse scattering by a local impurity in a periodic potential in one dimension

Roger G. Newton
Physics Department, Indiana University, Bloomington, Indiana 47405

(Received 12 October 1982; accepted for publication 25 March 1983)

Hill's equation, modified by a potential that vanishes as $x \to \pm \infty$, is considered. The direct scattering problem is studied; analytic and asymptotic properties of solutions of Hill's equation as well as of solutions of the modified equation are established. A new version of Levinson's theorem is proved. The inverse scattering problem is solved by means of a Marchenko-like equation.

PACS numbers: 03.65.Nk, 72.10.Fk

1. INTRODUCTION

This paper treats the direct and inverse scattering problems in one dimension for Hill's equation modified by a decreasing function that is added to the periodic one. In physical terms we are dealing with the scattering of Bloch waves by an impurity in a crystal. Such scattering problems have been discussed in the literature,[1] but little is known about the analyticity and asymptotic properties of the solutions, which form the basis on which the solution of the inverse problem in the Marchenko technique rests. Even for Hill's equation, i.e., a purely periodic potential, not enough information about the needed analytic and asymptotic properties exists in the literature.[2]

Somewhat surprisingly, it is necessary first to consider the scattering, on the full line $-\infty < x < \infty$, by the potential of one cell, that is, without its periodic repetition. The well-known solution[3,4] to this problem is given in Sec. 2, mostly to establish notation and some needed facts. Section 3 sets up the Bloch solutions of Hill's equation. Rather than expressing these in terms of the conventional solutions[2] defined by boundary conditions at a point, it turns out to be more useful to express them in terms of the scattering solutions of Sec. 2. Analytic and asymptotic properties of the Bloch functions as functions either of the square root of the energy, or of the wave number, are proved and collected in Lemma 1. The conventional solutions are discussed in an appendix.

In Sec. 4 Hill's equation is modified by adding a potential that tends to zero as $x \to \pm \infty$ to the periodic potential. The direct scattering problem for such an "impurity" is solved, a "regular solution" is defined, and a Jost function is introduced. Lemma 2 establishes the needed analytic and asymptotic properties. Bound states of the modified Hill's equation are treated in Sec. 5, and a Levinson theorem[5] (Theorem 1) is proved.

Section 6 solves the inverse scattering problem. The technique is based on the solution of a Riemann–Hilbert problem by means of a Marchenko equation, as in Refs. 4 and 6–9. Theorem 2 states the reconstruction method if an underlying potential is known to exist, and Theorem 3 solves the construction problem if the existence of such a potential is not known. In both cases the periodic potential is assumed known, and so are the bound states and their characters,[4] as well as the scattering data. It turns out to be necessary to assume that all the bound states lie below the first allowed band of the spectrum. If they do, the solution of the inverse problem is unique if it exists. The existence of a solution depends on the occurrence of a "miracle" of the same kind as in Refs. 4 and 6–8. No solution technique with nonsingular integral equations is known if there are bound states in any of the regions between allowed spectral bands.

2. THE SCHRÖDINGER EQUATION IN ONE CELL

A. The scattering solution

It is useful first to consider the Schrödinger equation on $\mathbb{R}$

$$y'' + Ey = \hat{V}y, \tag{2.1}$$

where $\hat{V}$ vanishes for $x < 0$ and $x > 1$. We shall write

$$\hat{V}(x) = V(x)\Theta(x)\Theta(1-x),$$

in which $\Theta(x)$ is the Heaviside function. The results and precedures in this case are, of course, well known.[4] We write them down here for later convenience.

Let us define a pair of standard scattering solutions $\gamma_1(x)$ and $\gamma_2(x)$ and combine them in a two-component column vector[10] $\gamma(x)$, $\tilde{\gamma} = (\gamma_1, \gamma_2)$, by the integral equation

$$\gamma(x) = X(x)\,\hat{1} + \frac{1}{2i\lambda}\int_0^1 dy\, e^{i\lambda|x-y|}V(y)\gamma(y), \tag{2.2}$$

where

$$X = \begin{pmatrix} e^{i\lambda x} & 0 \\ 0 & e^{-i\lambda x} \end{pmatrix},$$

$\lambda = \sqrt{E}$ and $\hat{1}$ is the column vector each of whose entries are unity, $\tilde{\hat{1}} = (1,1)$. If $V \in L^2(0,1)$, then Eq. (2.2) defines $\gamma(x)$ in $0 \leqslant x \leqslant 1$ uniquely unless its Fredholm determinant vanishes. The vanishing of this Fredholm determinant at $E = E^0$ implies a bound state on the full line, $-\infty < x < \infty$, for $\hat{V}$, and hence $E^0 < 0$. Thus $\gamma(x)$ is uniquely defined for all real λ. It has the symmetry[11] $\gamma(-\lambda, x) = \gamma(\lambda, x)^*$.

We obtain from (2.2)

$$\begin{aligned} &\gamma_1(0) = 1 + \eta_1, \quad \gamma_1'(0) = i\lambda(1-\eta_1), \\ &\gamma_2(0) = 1 + \eta_3, \quad \gamma_2'(0) = -i\lambda(1+\eta_3), \end{aligned} \tag{2.3}$$

where

$$\eta_1 = \frac{1}{2i\lambda}\int_0^1 dx\, e^{i\lambda x}V(x)\gamma_1(x),$$

$$\eta_3 = \frac{1}{2i\lambda}\int_0^1 dx\, e^{i\lambda x}V(x)\gamma_2(x).$$

Similarly,

 0022-2488/83/082152-11$02.50

$$\gamma_1(1) = e^{i\lambda}(1+\eta_4), \quad \gamma_1'(1) = i\lambda e^{i\lambda}(1+\eta_4),$$
$$\gamma_2(1) = e^{i\lambda}\eta_2 + e^{-i\lambda}, \quad \gamma_2'(1) = i\lambda\,(e^{i\lambda}\eta_2 - e^{-i\lambda}), \tag{2.4}$$

where

$$\eta_2 = \frac{1}{2i\lambda}\int_0^1 dx\, e^{-i\lambda x} V(x)\gamma_2(x),$$
$$\eta_4 = \frac{1}{2i\lambda}\int_0^1 dx\, e^{-i\lambda x} V(x)\gamma_1(x).$$

Evaluating the Wronskian $W(\gamma_1,\gamma_2) = \gamma_1\gamma_2' - \gamma_1'\gamma_2$ at $x = 0$ and at $x = 1$, we find

$$W(\gamma_1,\gamma_2) = -2i\lambda\,(1+\eta_3) = -2i\lambda\,(1+\eta_4), \tag{2.5}$$

and therefore $\eta_3 = \eta_4$. One also easily obtains expressions of the complex conjugates (for real λ) of γ_1 and γ_2 in terms of γ_1 and γ_2,

$$\gamma_1^* = \eta_1^*\gamma_1 + (1+\eta_3^*)\gamma_2,$$
$$\gamma_2^* = (1+\eta_3^*)\gamma_1 + \eta_2^*\gamma_2, \tag{2.6}$$

i.e.,

$$\gamma^* = QS_0^*\gamma, \tag{2.6'}$$

where

$$Q = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

and

$$S_0 = \begin{pmatrix} 1+\eta_3 & \eta_2 \\ \eta_1 & 1+\eta_3 \end{pmatrix}.$$

S_0 is the S matrix of the scattering problem defined by $\hat{V}$. Therefore,

$$|\eta_1|^2 + |1+\eta_3|^2 = |\eta_2|^2 + |1+\eta_3|^2 = 1,$$
$$\eta_1(1+\eta_3^*) + \eta_2^*(1+\eta_3) = 0, \tag{2.7}$$

i.e., S_0 is *unitary*[12]:

$$S_0 S_0^\dagger = S_0^\dagger S_0 = \mathbb{1}. \tag{2.7'}$$

It also satisfies the equation[13]

$$\tilde{S}_0 = QS_0 Q. \tag{2.8}$$

The effect of shifting $\hat{V}(x)$ by t, that is, of replacing $\hat{V}(x)$ by $\hat{V}_t(x) = \hat{V}(x+t)$, is to replace $\gamma(x)$ by

$$\gamma_t(x) = X(t)^{-1}\gamma(x+t), \tag{2.9}$$

as is well known, and to replace S_0 by

$$S_{0t} = X(t)S_0 X(t)^{-1}. \tag{2.10}$$

B. Analyticity and asymptotics

The integral equation (2.2) implies that γ_1 and γ_2, and hence also η_1, η_2, and η_3 are meromorphic functions of λ, except for the possibility of singularities at $\lambda = 0$. In order to show that they are analytic also at $\lambda = 0$, we define a solution of (2.1) by the Volterra equation

$$z(x) = e^{i\lambda x} - \frac{1}{\lambda}\int_x^1 dy \sin\lambda\,(x-y)\,\hat{V}(y)\,z(y). \tag{2.11}$$

From its value and derivative at $x = 1$ we infer that

$$\gamma_1(x) = (1+\eta_3)z(x), \text{ and therefore}$$
$$\frac{1}{1+\eta_3} = 1 - \frac{1}{2i\lambda}\int_0^1 dx\, e^{-i\lambda x} V(x)z(x),$$
$$\frac{\eta_1}{1+\eta_3} = \frac{1}{2i\lambda}\int_0^1 dx\, e^{i\lambda x} V(x)z(x).$$

If $\hat{V} \in L^1$, then $z(x)$ is an entire analytic function of λ. Since (2.7) implies that, as $\lambda \to 0$, η_3 remains bounded, η_1, η_3, and γ_1 are analytic there also. The same holds for η_2 and γ_2, by similar arguments.

Note that there are two possibilities at $\lambda = 0$. In case (a) $\lim_{\lambda\to 0}\int_0^1 dx\, Vz \neq 0$, the generic case, we have $\lim \eta_1 = \lim \eta_2 = \lim \eta_3 = -1$ and $\lim \gamma_1(x) = \lim \gamma_2(x) \equiv 0$. In case (b) $\lim_{\lambda\to 0}\int_0^1 dx\, Vz = 0$. Then $\lim \eta_1 = \lim \eta_2 = \lim \eta_3 = ic/(1-ic)$, c real.

The poles of γ_1, γ_2, η_1, η_2, and η_3 in[14] $\mathbb{C}^+$ are simple, on the imaginary λ axis, and all at the same points, namely, where the Fredholm determinant of (2.2) vanishes, i.e., at the full-line bound states of $\hat{V}$.

The integral equation (2.2) also implies that as $\lambda \to \pm\infty$ or as $|\lambda| \to \infty$ in $\mathbb{C}^+$

$$\gamma_1(x)e^{-i\lambda x}, \quad \gamma_2 e^{i\lambda x} = 1 + O(\lambda^{-1}). \tag{2.12}$$

Consequently,

$$\eta_1, \eta_3 = O(\lambda^{-1}) \tag{2.13}$$

and

$$\eta_2 e^{2i\lambda} = O(\lambda^{-1}). \tag{2.13'}$$

C. The Jost function

The fact that S_0 is the S matrix of the scattering problem for the potential V implies the existence of a unique Jost function J_0 with the following properties[4]: J_0 is an analytic function of λ, holomorphic in $\mathbb{C}^+$ and continuous on the real axis. In $\mathbb{C}^+$ $\det J_0$ has simple zeros at the points where the Fredholm determinant of (2.2) vanishes (i.e., the bound states of $\hat{V}$), and nowhere else. These points lie on the imaginary axis and are finite in number. The null space of J_0 at each of these points is determined by the "character" of the corresponding bound state of the full-line problem with $\hat{V}$. As $\lambda \to \pm\infty$ or as $|\lambda| \to \infty$,

$$J_0 = \mathbb{1} + O(1/\lambda). \tag{2.14}$$

Finally, J_0 satisfies the equation

$$J_0^* = QJ_0 QS_0. \tag{2.15}$$

It should also be noted that a shift of $\hat{V}$ leaves the bound states of (2.1) unchanged.

3. HILL'S EQUATION

A. The Bloch solutions

We next study the Schrödinger equation

$$y'' + Ey = Vy, \tag{3.1}$$

with a periodic potential, $V(x) = V(x+1)$, which is also known as *Hill's equation*. It will be understood that the function $V(x)$ is identical with $\hat{V}$ in (2.1) for $0 \leqslant x \leqslant 1$.

A solution $\beta(x)$ of (3.1) with the property

$$\beta(x+1)=\alpha\beta(x) \tag{3.2}$$

exists by Floquet's theorem.[15] In the interval [0,1] it must be expressible as a linear combination of γ_1 and γ_2. The requirement (3.2), being equivalent to

$$\beta(1)=\alpha\beta(0), \quad \beta'(1)=\alpha\beta'(0),$$

leads to the equation

$$\alpha^2-2\alpha\Delta+1=0, \tag{3.3}$$

where, by (2.7),

$$\Delta=\frac{1}{2}\frac{e^{i\lambda}}{1+\eta_3^*}+\frac{1}{2}\frac{e^{-i\lambda}}{1+\eta_3} \tag{3.4}$$

for real λ; Δ is known as the *discriminant*. Thus there are two solutions of (3.3),

$$\alpha_\pm=\Delta\pm i(1-\Delta^2)^{1/2}. \tag{3.5}$$

The *periodic spectrum* consists of those values of E for which $\Delta=\pm 1$: If $\Delta=1$, the corresponding solution β has period 1; if $\Delta=-1$, it has period 2. The periodic spectrum

$$-\infty<E_0<E_1\leqslant E_2<E_3\cdots$$

separates those intervals $(E_0, E_1),(E_2, E_3), \cdots$ for which $\Delta^2<1$, which are the *allowed bands*, from those where $\Delta^2>1$, the *forbidden bands*. The intervals $(E_1, E_2),(E_3, E_4), \cdots$ will be called the *band gaps*. In the former $\alpha_\pm$ are complex, $\alpha_-=\alpha_+^*$, of modulus 1, and the solutions β_i, $i=1,2$, which correspond to α_+ and α_-, when continued by (3.2) to all of $\mathbb{R}$, are uniformly bounded; in the latter $\alpha_+=1/\alpha_-$ are real and the solutions β_i are unbounded either as $x\to+\infty$ or as $x\to-\infty$. We shall denote the set of points E where $\Delta^2\leqslant 1$ as $\mathbb{R}_a$; the set of points $\lambda=\sqrt{E}$ such that $\lambda^2\in\mathbb{R}_a$ will be denoted by $\mathbb{R}_a'$.

The variable E in (2.1) is defined only to within an additive constant. If we start with one definition, $y''+E^1y=V^1y$, we may also write $y''+Ey=Vy$, where $E=E^1+a$, $V=V^1+a$. Let the first point in the periodic spectrum, in the first parametrization, be E_0^1. Then we may choose $a=-E_0^1$, and, as a result, in the second parametrization the first point in the periodic spectrum is $E_0=0$. We shall from now on assume that this shift has been done so that $E_0=0$.

We define the function $d(E)$ as that branch of $k=\cos^{-1}\Delta(E)$ for which $k=0$ when $E=E_0=0$. The function $k(\lambda)$ has branch points of the square-root type at the periodic spectrum, which we connect by cuts along the band gaps. The inverse function $\lambda(k)$ has branch points in each band gap, and we connect each such branch point in the upper half of the k plane by a cut with the symmetric point in the lower half-plane. The left-hand forbidden band, $-\infty<E<0$, which corresponds to the positive imaginary λ axis on the first sheet of the E plane, and to the negative imaginary λ axis on the second sheet, is mapped into the positive and negative imaginary k axis, respectively. The resulting mapping is indicated in Fig. 1. Except for the cuts, each quadrant of the λ plane corresponds to the same quadrant in the k plane.

We now have $\alpha_\pm=e^{\pm ik}$. The corresponding solutions of (3.1) are the *Bloch functions* $\beta_i(x)$, $i=1,2$. They are such that

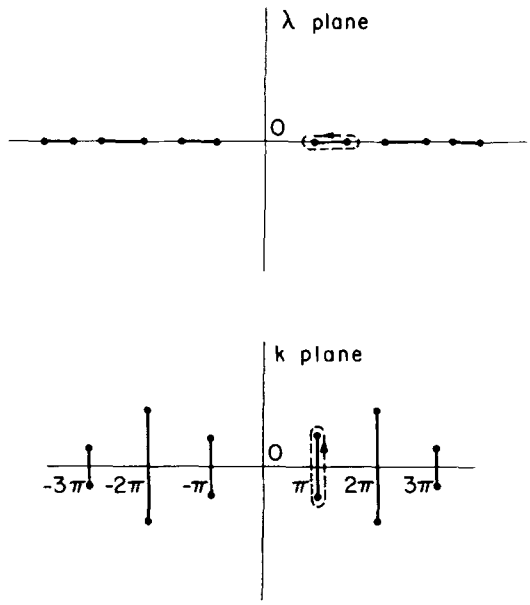


FIG. 1. The mapping $\lambda\leftrightarrow k$. The heavy lines in the λ plane indicate branch cuts of $k(\lambda)$, whose end points are the periodic spectrum. The heavy lines in the k plane are branch cuts of $\lambda(k)$.

$$\beta_1(x+1)=e^{ik}\beta_1(x), \quad \beta_2(x+1)=e^{-ik}\beta_2(x),$$

which implies that

$$\beta_1(x)=e^{ikx}\xi_1(x), \quad \beta_2(x)=e^{-ikx}\xi_2(x), \tag{3.6}$$

where $\xi_i(x)$ is periodic,

$$\xi_i(x+1)=\xi_i(x).$$

Let us combine β_1 and β_2 in a two-component column vector β. Since in the interval [0,1] Eqs. (2.1) and (3.1) are identical, the solutions β and γ are there related by

$$\beta(x)=M\gamma(x), \tag{3.7}$$

where M can be chosen as

$$M=\frac{1}{2i\lambda}\begin{pmatrix} e^{i\lambda}l^*-e^{-ik} & e^{i\lambda}-le^{ik} \\ e^{i\lambda}l^*-e^{ik} & e^{i\lambda}-le^{-ik}\end{pmatrix} \tag{3.8}$$

and its inverse is given by

$$M^{-1}=-\frac{i(1+\eta_3)}{2\epsilon\sin k}\begin{pmatrix} e^{i\lambda}-le^{-ik} & le^{ik}-e^{i\lambda} \\ e^{ik}-l^*e^{i\lambda} & l^*e^{i\lambda}-e^{-ik}\end{pmatrix}, \tag{3.9}$$

in which

$$l=(1+\eta_1)/(1+\eta_3)=(1+\eta_3-l\eta_2)^* \tag{3.10}$$

and

$$\epsilon=(l^*e^{i\lambda}-le^{-i\lambda})/2i\lambda. \tag{3.11}$$

We also define

$$\epsilon_1=\frac{i\lambda}{2}\left(e^{i\lambda}\frac{1-\eta_1^*}{1+\eta_3^*}-e^{-i\lambda}\frac{1-\eta_1}{1+\eta_3}\right) \tag{3.12}$$

and

$$\epsilon_2=\frac{1}{2}\frac{1-\eta_1^*}{1+\eta_3^*}e^{i\lambda}+\frac{1}{2}\frac{1-\eta_1}{1+\eta_3}e^{-i\lambda}. \tag{3.13}$$

The quantities ϵ, ϵ_1, and ϵ_2 are all real for real λ. Equation

(3.7) may now be used to continue $\gamma(x)$ for all $x \in \mathbb{R}$ as a solution of (3.1) rather than of (2.1). *From now on this will be the meaning of* $\gamma(x)$.

It follows from the expressions (A4) in the Appendix that when E is in an allowed band, then β_1 and β_2 are one another's complex conjugates, $\beta_1 = \beta_2^*$. In the band gaps on the other hand, each β_i is real, $\beta_i = \beta_i^*$. Their Wronskian is given by

$$W(\beta_1, \beta_2) = -2i\epsilon \sin k. \tag{3.14}$$

One readily finds that

$$\epsilon\epsilon_1 = (\epsilon_2 - e^{ik})(e^{-ik} - \epsilon_2). \tag{3.15}$$

This equation implies that when either $\epsilon = 0$ or $\epsilon_1 = 0$ then either $\epsilon_2 = e^{ik}$ or $\epsilon_2 = e^{-ik}$, which shows that neither ϵ nor ϵ_1 can vanish in an allowed band. The points E_n' where $\epsilon = 0$ are those values of E for which a solution of (3.1) exists that vanishes at $x = 0$ and $x = 1$ (see the Appendix). These points constitute the first *tied spectrum* (for boundary conditions at $x = 0$) in the terminology of Trubowitz.[16] At the points E_n'' where $\epsilon_1 = 0$, a solution of (3.1) exists whose first derivative vanishes at $x = 0$ and $x = 1$. These points make up the second tied spectrum (for boundary conditions at $x = 0$). The tied spectra are known to interlace the periodic spectrum[17]:

$$-\infty < E_0 < E_1 \leqslant E_1' \leqslant E_2 < E_3 \leqslant E_2' \leqslant \cdots,$$

$$-\infty < E_1'' \leqslant E_0 < E_1 \leqslant E_2'' \leqslant E_2 < E_3 \leqslant E_3'' \leqslant \cdots.$$

We note that the tied spectra depend on the point at which boundary conditions are assigned, whereas the periodic spectrum does not, and neither do the solutions β_i [except for their scale as defined by (3.7)].

The relation between the Bloch solutions and the functions γ_1 and γ_2 may be written as

$$\beta_{\frac{1}{2}} = (1/2i\lambda)(e^{\pm ik} - \epsilon_2)(\gamma_1 - l\gamma_2) + \frac{1}{2}\epsilon\left(\gamma_1 + \frac{1-\eta_1}{1+\eta_3}\gamma_2\right), \tag{3.16}$$

which, together with (3.15), shows that when $\epsilon = 0$, i.e., at the first tied spectrum, one of the Bloch solutions vanishes identically, and the other vanishes at $x = 0 \pmod 1$ (see Appendix). For E in an allowed band, i.e., $\Delta^2 < 1$, β_i cannot vanish for any value of x. This is because if $\beta_1(x) = 0$, then also $\beta_2(x) = \beta_1^*(x) = 0$ and hence *every* solution would have to vanish there. Suppose, on the other hand, that, for $E = E_s$ not in an allowed band, $\beta_1(x_0) = 0$ [or $\beta_2(x_0) = 0$]. Then also $\varphi_1(x_0 + 1) = 0$ [or $\beta_2(x_0 + 1) = 0$], and hence E_s is in the first tied spectrum for boundary conditions at $x = x_0$. Since the tied spectra are known to be real, E_s must be real. Thus β_i can vanish for any point x only if E is in a band gap or in the periodic spectrum. In fact, for every point x_0 there is exactly one value of E in each band gap for which one of the two Bloch solutions vanishes.

For real λ we find from (2.6') and (3.8)

$$\beta^* = M^*\gamma^* = M^*QS_0^* M^{-1}\beta;$$

but for real k we also have $\beta^* = Q\beta$ while, for purely imaginary $k \pmod \pi$, $\beta^* = \beta$. It follows that in the allowed bands

$$S_0 = QM^{-1}QM^*, \tag{3.17}$$

and in the band gaps

$$S_0 = QM^{-1}M^*. \tag{3.17'}$$

Direct calculation also shows that, although M^{-1} is unbounded at the periodic spectrum, $M^{-1}M^*$ remains bounded as $\sin k \to 0$, both as approached from the allowed side and from the gap.

The solutions β_i serve also to define a Green's function $\mathcal{G}^+(x, x')$, such that

$$\left[\frac{\partial^2}{\partial x^2} + E - V(x)\right]\mathcal{G}^+(x, x') = \delta(x - x').$$

Using (3.7), we obtain, for all x and x',

$$\mathcal{G}^+(x, x') = \frac{\beta_1(x_>)\beta_2(x_<)}{2i\epsilon \sin k}, \tag{3.18}$$

where $x_> = \max(x, x')$ and $x_< = \min(x, x')$. Note that (3.15) and (3.16) imply that $\mathcal{G}^+(x, x')$ remains bounded at the points where $\epsilon = 0$. In the band gaps $\mathcal{G}^+$ is real. Another Green's function is given by

$$\mathcal{G}(x, x') = \begin{cases} \dfrac{\beta_1(x)\beta_2(x') - \beta_1(x')\beta_2(x)}{2i\epsilon \sin k}, & x' < x, \\ 0, & x' > x. \end{cases} \tag{3.19}$$

This function is real for real E, $\mathcal{G}^*(x, x') = \mathcal{G}(x, x')$. The alternative form (A7) for (3.19) shows that the singularities of (3.19) at $k = 0 \pmod \pi$ are removable.

Let us consider the effects of shifting $V(x)$ by an amount t, that is, of replacing $V(x)$ by $V_t(x) = V(x + t)$. We shall refer to such a translation as a *comprehensive shift* if at the same time we also shift the reference points 0 and 1, which enter into (2.2) and hence into the definitions of γ and β. For γ as a solution of (2.1) on $\mathbb{R}$ this is an ordinary shift of $\hat{V}$, and as a solution of (3.1), γ still changes as in (2.9), and S_0 changes as in (2.10). The Green's function (3.18) does not change, i.e.,

$$\mathcal{G}_t^+(x, x') = \mathcal{G}^+(x + t, x' + t). \tag{3.20}$$

B. Analyticity and asymptotics

We state the main results in the form of

Lemma 1: If $V \in L^2(0,1)$, then the quantities ϵ, l, l^*, and Δ have analytic continuations off the real axis that are entire analytic functions of λ with the asymptotic forms as $\lambda \to \pm\infty$ or as $|\lambda| \to \infty$ in $\mathbb{C}^+$,

$$\epsilon = (\sin\lambda)/\lambda + O(\lambda^{-2}e^{\nu}), \tag{3.21}$$

$$l = 1 + O(\lambda^{-1}), \tag{3.22}$$

$$\Delta = \cos\lambda + O(\lambda^{-1}e^{\nu}), \tag{3.23}$$

where $\nu = \operatorname{Im}\lambda$. In the same sense

$$k = \lambda + O(\lambda^{-1}) \pmod{2\pi}. \tag{3.24}$$

The matrix M is an analytic functions of λ holomorphic everywhere in the plane cut as in Fig. 1, except for the possibility of a simple pole at $\lambda = 0$; M^{-1} is meromorphic in the cut λ plane with poles at the zeros of the Fredholm determinant of (2.2), i.e., the bound states of $\hat{V}$. The matrices $J_0 M^{-1}$ and MJ_0^{-1} are holomorphic in $\mathbb{C}^+$ and continuous on $\mathbb{R}$. As $\lambda \to \pm\infty$ or as $|\lambda| \to \infty$ in $\mathbb{C}^+$,

$$M = \mathbb{1}(\sin\lambda)/\lambda + O(\lambda^{-2}e^{\nu}). \tag{3.25}$$

The Bloch functions are analytic functions of λ, holomor-

phic in the cut λ-plane. The function $\gamma(x) = M^{-1}\beta(x)$ has the asymptotic behavior, for each fixed $x \in \mathbb{R}$, as $\lambda \to \pm\infty$ or $|\lambda| \to \infty$ in $\mathbb{C}^+$

$$X^{-1}\gamma(x) = \hat{1} + O(\lambda^{-1}). \tag{3.26}$$

The Green's function (3.18) is the boundary value of an analytic function of λ, holomorphic in $\mathbb{C}^+$, continuous on $\mathbb{R}$, and as $\lambda \to \pm\infty$ or $|\lambda| \to \infty$ in $\mathbb{C}^+$,

$$g^+(x,y) = (1/2i\lambda)e^{i\lambda|x-y|} + O(\lambda^{-2}e^{-\nu|x-y|}). \tag{3.27}$$

The Green's function (3.19) is an entire analytic function of λ^2 with the asymptotics on $\mathbb{R}$ and in $\mathbb{C}^+$

$$g(x,y) = \sin\lambda(x-y)/\lambda + O(\lambda^{-2}e^{\nu|x-y|}). \tag{3.28}$$

Proof: One easily sees that $l = z(0)$, where $z(x)$ is the function defined by (2.11). Therefore, l is entire. Particularly, it follows that if $\eta_3 = -1$, then also $\eta_1 = -1$ and $\eta_2 = -1$. This can (and generally will) happen only at $\lambda = 0$. It also follows that l^* has an analytic continuation that is entire analytic. Furthermore, l is real for $\lambda = 0$. Therefore, by (3.11), ϵ is entire. The expression for Δ given in the appendix shows that it is an entire function of λ^2. The asymptotic forms (3.21)–(3.23) follow from (2.13), (2.13′) and (3.10); and (3.24) follows from (3.23). [Note that (3.24) implies the well-known fact[18] that the lengths of the band gaps tend to zero as $E \to \infty$.]

The analyticity of M follows from that of l and k. In case (a) of Sec. 2B, M has a simple pole at $\lambda = 0$, and so does $\det M$, but M^{-1} is holomorphic there. In case (b) $\lim_{\lambda\to 0} l = 1$ and both M and M^{-1} are holomorphic there. That M^{-1} is meromorphic follows from (3.9) and the fact that η_3 is meromorphic. We have $\det M = \epsilon \sin k/\lambda(1+\eta_3)$, which has simple zeros at the bound states of $\hat{V}$. Thus the residue of M^{-1} at such a pole is a singular matrix of rank one. That $J_0 M^{-1}$ is holomorphic there follows because γ has a pole at a bound state of $\hat{V}$ but β does not, and the null space of J_0 equals the range of the residue of γ. As the zeros of $\det J_0$ and $\det M$ are both simple, MJ_0^{-1} is holomorphic too. The asymptotic form (3.25) follows from (3.8) and (2.13).

The analyticity of the Bloch solutions follows most simply from their expression given in the Appendix. The asymptotics (3.26) follow from (2.12). For $0 \leqslant x \leqslant 1$, (2.12) is directly applicable. Therefore, $X(1)^{-1}\gamma(1) = \gamma(0) + O(\lambda^{-1})$ and $X(1)^{-1}\gamma'(1) = \gamma'(0) + O(1)$. As a result, $X(x+1)^{-1}\gamma(x+1) = X(x)\gamma(x) + O(\lambda^{-1})$ for $0 \leqslant x \leqslant 1$. Repetition leads to (3.26) for all fixed $x \in \mathbb{R}$.

The analyticity and asymptotics of the Green's functions follow from those of ϵ, k, and β_i.

4. SCATTERING BY IMPURITIES

A. The solutions

We now consider the Schrödinger equation

$$y'' + Ey = (V+U)y, \tag{4.1}$$

where V is periodic, $V(x+1) = V(x)$, and U tends to zero as $x \to \pm\infty$. A suitable Green's function is given by (3.19), and we define two solutions of (4.1) by the integral equations

$$\chi_1(x) = \beta_1(x) - \int_x^\infty dx'\, g(x',x)U(x')\chi_1(x'),$$
$$\chi_2(x) = \beta_2(x) + \int_{-\infty}^x dx'\, g(x,x')U(x')\chi_2(x'). \tag{4.2}$$

If $u \in L^1(-\infty,\infty)$, these equations can be solved by iteration.

As $x \to \pm\infty$ the form (3.19) of $g(x,x')$ shows that when E is in an allowed band,

$$\chi_1(x) = \beta_1(x) + o(1), \quad \text{as } x \to \infty,$$
$$= \beta_1(x)/T_1 + \beta_2(x)R_1/T_1 + o(1), \quad \text{as } x \to -\infty, \tag{4.3}$$

$$\chi_2(x) = \beta_2(x)/T_2 + \beta_1(x)R_2/T_2 + o(1), \quad \text{as } x \to \infty,$$
$$= \beta_2(x) + o(1), \quad \text{as } x \to -\infty, \tag{4.3′}$$

where

$$\frac{1}{T_1} = 1 - \frac{\csc k}{2i\epsilon}\int_{-\infty}^{\infty} dx\,\beta_2(x)U(x)\chi_1(x),$$
$$\frac{R_1}{T_1} = \frac{\csc k}{2i\epsilon}\int_{-\infty}^{\infty} dx\,\beta_1(x)U(x)\chi_1(x),$$
$$\frac{1}{T_2} = 1 - \frac{\csc k}{2i\epsilon}\int_{-\infty}^{\infty} dx\,\beta_1(x)U(x)\chi_2(x), \tag{4.4}$$
$$\frac{R_2}{T_2} = \frac{\csc k}{2i\epsilon}\int_{-\infty}^{\infty} dx\,\beta_2(x)U(x)\chi_2(x).$$

The Wronskian of χ_1 and χ_2 may be evaluated at $x \to +\infty$ or $x \to -\infty$, which yields

$$W(\chi_1,\chi_2) = W(\beta_1,\beta_2)/T_1 = W(\beta_1,\beta_2)/T_2$$
$$= -(2i\epsilon \sin k)/T_1, \tag{4.5}$$

and hence $T_1 = T_2 \equiv T$. We form a two-component column vector χ out of χ_1 and χ_2, $\tilde{\chi} = (\chi_1,\chi_2)$.

It is also convenient to define another two-component solution ψ by the integral equation

$$\psi(x) = \beta(x) + \int_{-\infty}^{\infty} dy\, g^+(x,y)U(y)\psi(y), \tag{4.6}$$

where g^+ is defined in (3.18). As $x \to \pm\infty$, (4.6) gives

$$\psi_1(x) = \beta_1(x)T_1 + o(1), \quad \text{as } x \to \infty,$$
$$= \beta_1(x) + \beta_2(x)R_1 + o(1), \quad \text{as } x \to -\infty, \tag{4.7}$$

$$\psi_2(x) = \beta_2(x) + \beta_1(x)R_2 + o(1), \quad \text{as } x \to \infty,$$
$$= \beta_2(x)T_2 + o(1), \quad \text{as } x \to -\infty. \tag{4.7′}$$

These asymptotic forms may also be written more compactly as

$$\psi(x) - \beta(x) = \beta_1(x)\tilde{A}\begin{pmatrix}1\\0\end{pmatrix} + o(1), \quad \text{as } x \to \infty,$$
$$\beta_2(x)\tilde{A}\begin{pmatrix}0\\1\end{pmatrix} + o(1), \quad \text{as } x \to -\infty, \tag{4.7″}$$

where the 2×2 matrix A is given by

$$A = \begin{pmatrix} T & R_2 \\ R_1 & T \end{pmatrix} - \mathbb{1} \tag{4.8}$$

and its elements have the representation

$$A_{ij} = \frac{\csc k}{2i\epsilon}\int_{-\infty}^{\infty} dx\,\beta_i^*(x)U(x)\psi_j(x). \tag{4.9}$$

Because U is real, the integral equatins (4.6) lead to the alternative representation

$$A_{lj} = \frac{\csc k}{2i\epsilon} \int_{-\infty}^{\infty} dx\, \psi_n(x) U(x) \beta_j(x), \quad l \neq n. \tag{4.9'}$$

In matrix form (4.9) and (4.9') imply

$$\tilde{A} = QAQ. \tag{4.10}$$

We note that comparison of (4.7) and (4.7') with (4.3) and (4.3') leads to the conclusion

$$\psi = T\chi. \tag{4.11}$$

By the same argument as in Appendix A of Ref. 4 one proves that the Fredholm determinant of (4.6) equals $1/T$.

When E is in an allowed band, the function $\psi_- \equiv Q\psi^*$, $\tilde{\psi}_- = (\psi_2^*, \psi_1^*)$, is also a solution of (4.1). Hence there must exist a 2×2 matrix S such that

$$\psi = \tilde{S}\psi_- = \tilde{S}Q\psi^*. \tag{4.12}$$

Comparing the asymptotic form for $x \to \pm\infty$ of ψ with that of ψ^*, we conclude that

$$S = \mathbb{1} + A, \tag{4.13}$$

and, furthermore, that S is unitary,

$$SS^\dagger = S^\dagger S = \mathbb{1}. \tag{4.14}$$

According to (4.8), S is made up of the transmission and reflection amplitudes,

$$S = \begin{pmatrix} T & R_2 \\ R_1 & T \end{pmatrix}. \tag{4.8'}$$

The unitarity equations (4.14) read more explicitly

$$\begin{aligned} &|R_1|^2 = |R_2|^2 = 1 - |T|^2, \\ &R_1 R_2/|R_1 R_2| = -T^2/|T|^2. \end{aligned} \tag{4.14'}$$

They also imply

$$\det S = T/T^*. \tag{4.15}$$

Equations (4.12) and (4.14) imply that for λ in an allowed band

$$\psi^* = QS^*\psi. \tag{4.12'}$$

Now define

$$\hat{\psi} = M^{-1}\psi, \tag{4.16}$$

$$\hat{\chi} = M^{-1}\chi, \tag{4.16'}$$

where M is defined in (3.8). Equation (4.12') then leads to

$$\hat{\psi}^* = Q\hat{S}^*\hat{\psi}, \tag{4.17}$$

where

$$\hat{S} = QM^{-1}QSM^* \tag{4.18}$$

in allowed bands, and by (3.17')

$$\hat{\psi}^* = QS_0^*\hat{\psi} \tag{4.17'}$$

in the band gaps. We note that if $S = 1$, then $\hat{S} = S_0$, by (3.17).

Consider the effect of a "comprehensive shift" of the lattice and the impurity together. That is, let $V(x)$ and $U(x)$ be replaced by

$$V_t(x) = V(x+t), \quad U_t(x) = U(x+t),$$

and at the same time let the reference points 0 and 1 be shifted to $-t$ and $1-t$. Then $\gamma(x) = M^{-1}\beta(x)$ changes as in (2.9) while the Green's function φ^+ changes as in (3.20). Therefore, by (4.6) the corresponding change in $M^{-1}\psi$ is $(M^{-1}\psi)_t(x) = X(t)^{-1}M^{-1}\psi(x+t)$. Consequently, the matrix $\hat{S}_t$ after the shift is

$$\hat{S}_t = X(t)\hat{S}X(t)^{-1} \tag{4.19}$$

in the allowed bands, and in the band gaps the shifted S_0 is, as in (2.10),

$$S_{0t} = X(t)S_0X(t)^{-1}. \tag{4.19'}$$

For reasons that will become clear in Sec. 4B [see Eq. (4.32)] it is useful to consider the function

$$\Psi(x) = X^{-1}\hat{\psi}(x). \tag{4.20}$$

According to (4.17), (4.17'), (4.19), and (4.19'), this function satisfies the relation

$$\Psi^*(x) = Q\hat{S}_x^*\Psi(x) \tag{4.21}$$

in allowed bands, and

$$\Psi^*(x) = QS_{0x}^*\Psi(x) \tag{4.21'}$$

in band gaps.

The x dependence of Ψ will from now on no longer be explicitly indicated. Instead it will be its dependence on λ that is of importance, and we shall often simply write $\Psi(\lambda)$.

A pair ϕ of "regular" solutions of (4.1), $\tilde{\phi} = (\phi_1, \phi_2)$, is defined for $x \geqslant 0$ by the Volterra equation

$$\phi(x) = J_0\gamma(x) + \int_0^x dy\, \varphi(x,y)U(y)\phi(y) \tag{4.22}$$

and for $x \leqslant 0$ by

$$\phi(x) = J_0\gamma(x) - \int_x^0 dy\, \varphi(y,x)U(y)\phi(y), \tag{4.22'}$$

where J_0 is the "single cell" Jost function with the property (2.15). Because φ is real, Eqs. (2.6'), (2.15), and (2.7') imply that for real λ

$$\phi^*(x) = Q\phi(x). \tag{4.23}$$

The solution ϕ must be expressible as a linear combination of $\hat{\psi}_1$ and $\hat{\psi}_2$. Therefore, there must exist an x-independent 2×2 matrix J such that

$$\phi = J\hat{\psi}. \tag{4.24}$$

Since by (4.5)

$$W(\chi,\tilde{\chi}) = -\frac{2i\epsilon \sin k}{T} P,$$

where $P = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$, and, by (4.11), J is given by

$$J = (\csc k/2i\epsilon)W(\phi,\tilde{\hat{\chi}})\tilde{M}PM. \tag{4.25}$$

We also have $W(\phi,\tilde{\phi}) = -2i\epsilon \sin k\, J_0\, M^{-1}P\tilde{M}^{-1}\tilde{J}_0$, from which it follows that

$$\det J = \det J_0/T. \tag{4.26}$$

Equation (4.18), together with (4.23) and (2.15), then leads to the decomposition

$$\hat{S} = QJ^{-1}QJ^*,$$

which for later purposes we shall write in the form

$$J^{*-1} = Q\hat{S}^*J^{-1}Q \tag{4.27}$$

in the allowed bands, while in the band gaps

$$J^{*-1} = QS_0^* J^{-1} Q. \tag{4.27'}$$

B. Analyticity and asymptotics

We state the principal conclusions in the form of

Lemma 2: Assume that $V \in L^2(0,1)$ and that $U \in L^1(-\infty,\infty) \cap L^2(-\infty,\infty)$, $|x|U \in L^1(-\infty,\infty)$. Then χ is an analytic function of E, holomorphic in $\mathbb{C}\backslash\mathbb{R}_a$. As a function of k it is analytic in $\mathbb{C}^+$ cut as indicated in Fig. 1 with branch points of the square-root type at those purely imaginary (mod π) points where $d\Delta/dE = 0$. The functions ψ and T are analytic in λ, meromorphic in $\mathbb{C}^+$ and in the band gaps and continuous in the allowed bands and the periodic spectrum. The poles of T and ψ in $\mathbb{C}^+$ and in the gaps are simple and occur at the same points, on the positive imaginary axis or in the band gaps only. The asymptotic forms of $\hat{\psi}$ and T as $\lambda \to \pm\infty$ or as $|\lambda| \to \infty$ in $\mathbb{C}^+$ are

$$e^{-ikx}\hat{\psi}_1(x), \quad e^{ikx}\hat{\psi}_2(x) = 1 + O(\lambda^{-1}), \tag{4.28}$$

$$T = 1 + O(\lambda^{-1}), \tag{4.29}$$

while as $\lambda \to \pm\infty$

$$R_1, R_2 = O(\lambda^{-1}). \tag{4.30}$$

The solution ϕ of (4.22) and (4.22') is an analytic function of E, holomorphic in $\mathbb{C}\backslash\mathbb{R}_a$. The Jost function J is an analytic function of λ, holomorphic in $\mathbb{C}^+$ and in the gaps and continuous in the allowed bands and at the points of the periodic spectrum. The asymptotic form of J as $\lambda \to \pm\infty$ or as $|\lambda| \to \infty$ in $\mathbb{C}^+$ is

$$J = \mathbb{1} + O(\lambda^{-1}). \tag{4.31}$$

The function Ψ defined by (4.16) and (4.20) is an analytic function of λ, meromorphic in $\mathbb{C}^+$ and in the band gaps, continuous in $\mathbb{R}_a'$. Its asymptotic form as $\lambda \to \pm\infty$ or as $|\lambda| \to \infty$ in $\mathbb{C}^+$ is

$$\Psi(\lambda, x) = \hat{\mathbb{1}} + O(\lambda^{-1}), \tag{4.32}$$

where $\hat{\mathbb{1}} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. The matrix $\hat{S}$ defined by (4.18) has the asymptotic form as $\lambda \to \pm\infty$

$$\hat{S} = \mathbb{1} + O(\lambda^{-1}). \tag{4.33}$$

(This is to be understood in the sense that as $\lambda \to \pm\infty$ it takes on only values in the allowed bands.) In $\mathbb{R}_a'$ and the periodic spectrum $\hat{S}$ is continuous.

Proof: The Green's function (3.19) in (4.2) may be expressed in the form (for $x' < x$)

$$g(x,x') = \frac{\csc k}{2i\epsilon}\left[e^{ik(x-x')}\xi_1(x)\xi_2(x') - e^{ik(x'-x)}\xi_1(x')\xi_2(x)\right],$$

and thus the integral equation for χ_1 may be written as

$$\chi_1(x)e^{-ikx} = \xi_1(x) - \frac{\csc k}{2i}\int_x^\infty dy\,[\xi_1(x)\xi_2(y) - e^{2ik(y-x)}\xi_1(y)\xi_2(x)]\,U(y)\chi_1(y)e^{-iky}. \tag{4.2'}$$

The functions ξ_1 and ξ_2 are periodic and uniformly bounded, and $g(x,x')$ is an entire analytic function of E. Therefore, $\chi_1 e^{ikx}$ is well defined by the iteration of (4.2') for all $k \in \mathbb{C}^+$ if $U \in L^1$. If furthermore, $|x|U \in L^1$, then χ_1 is an analytic function of E holomorphic in $\mathbb{C}\backslash\mathbb{R}_a$. As a function of k its analyticity follows from Sec. 3A. The same holds for χ_2.

In contrast to (4.2), (4.6) cannot in general be solved by iteration. If $U \in L^1 \cap L^2$, then its kernel is in the trace class for all $k \in \mathbb{C}^+$, and for all real k except where $\sin k = 0$. As $\sin k \to 0$, $\beta_1 \to \beta_2$, and according to (4.4), $1/T \sim 1 - a\csc k$, $R_1/T \sim a\csc k$, where $a = \int\beta_1 U\chi_1\,dx$. Therefore, if $a \neq 0$, then $T \sim -(\sin k)/a$ and $R_1 \sim Ta\csc k \to -1$; similarly for R_2. Therefore, *in general* (i.e., unless $a = 0$) at the periodic spectrum, i.e., at the end points of the allowed bands, $R_1 = R_2 = -1$ and $T = 0$. If $a = 0$, then, at $\sin k = 0$, $R_1 = R_2 = 0$ and $T = 1$.

It follows from this analysis and (4.11), since χ is bounded at $\sin k = 0$, that at the periodic spectrum $\psi = 0$ in general. In exceptional cases $T = 1$ at such points and ψ is bounded.

Since (4.6) is a Fredholm equation, its solution ψ will not exist at the exceptional points $E = E_e$, where the Fredholm determinant vanishes. It will be seen in Sec. 5 that these points E_e must lie on the real axis, in fact in band gaps or below E_0. In $\mathbb{C}^+$ and the band gaps (except at the points E_e) ψ is a holomorphic functions of λ, and on $\mathbb{R}_a'$ it is continuous if $U \in L^2$. The Fredholm determinant of (4.6), and hence also $1/T$, is analytic in $\mathbb{C}^+$ and in the gaps and continuous in $\mathbb{R}_a'$. The poles of T at the exceptional points will be seen below (in Sec. 5) to be simple.

It follows from (3.28) that as $\nu = \operatorname{Im}\lambda \to \infty$ the Hilbert–Schmidt norm of $|U|^{1/2}g^+U|U|^{-1/2}$ tends to naught. Therefore, the Fredholm determinant of (4.6) approaches unity as $\nu \to \infty$. If $U \in L^1$, the same holds as $\lambda \to \pm\infty$. For the same reason we find (4.28) from (4.6) and (3.26). The asymptotic form (4.30) follows from (4.28) and (4.9).

If $U \in L^1$, the solution ϕ of the Volterra equations (4.22) and (4.22') is easily seen to be an analytic function of E holomorphic in $\mathbb{C}\backslash\mathbb{R}_a$, because $J_0\gamma(x)$ is entire analytic. The analyticity and asymptotic form (4.32) of J follow from (4.25). The only point at which J appears to have a singularity is $\lambda = 0$. However, it was shown in Sec. 2 that in the generic case $\gamma(x) = 0$ at $\lambda = 0$, and therefore $\phi(x) = \hat{\chi}(x) = 0$; furthermore, it is then easily seen from (3.8) that $\lim_{\lambda\to 0}\lambda\tilde{M}PM$ exists. In the special case (b) when M and M^{-1} are holomorphic at $\lambda = 0$, $W(M\phi,\tilde{\chi})$ tends to zero as $\lambda \to 0$ and hence J again is holomorphic. Thus J is holomorphic as a function of λ in $\mathbb{C}^+\backslash\mathbb{R}_a'$.

The analyticity and asymptotics of $\Psi(x)$ follow from those of $\psi(x)$ and M. The results (4.29) and (4.30) imply that as $\lambda \to \pm\infty$ (always in allowed bands)

$$S = \mathbb{1} + O(\lambda^{-1}).$$

The definition (4.18) together with (3.25) therefore lead to (4.33).

We also note that as E approaches the end of an allowed band, i.e., $\sin k \to 0$, in general, as we saw, $S \to -Q$; in exceptional cases $S \to \mathbb{1}$. It then follows from (3.17) and the remark below (3.17') that in all cases $\hat{S}$ remains bounded even at the periodic spectrum.

As $\lambda \to 0$, one easily shows by means of (4.18), (3.9), and

(3.10) that $\hat{S}$ remains finite both in the generic and in the exceptional cases. Q.E.D.

5. BOUND STATES

A. Eigenvalues and characters

Suppose that $E = E_b$ is not in an allowed band or in the periodic spectrum, and that the imaginary part of $k_b = \cos^{-1}\Delta(E_b)$ is positive. Then, as $x \to \infty$,

$$\chi_1(x)e^{-ikx} = \xi_1(x) + o(1)$$

and, as $x \to -\infty$,

$$\chi_2(x)e^{ikx} = \xi_2(x) + o(1).$$

Therefore, if $W(\chi_1,\chi_2) = 0$, then $\chi_1 = c\chi_2$ decreases exponentially as $x \to \pm\infty$ and hence E_b is an eigenvalue of (4.1); there is a *bound state*. It follows from the self-adjointness of (4.1) that E_b must be real, i.e., k_b purely imaginary (mod π). By (4.5) at such a point $1/T$ vanishes and hence E_b must be a zero of the Fredholm determinant of (4.6) and thus a zero of $\det J$. It follows that the points E_b are identical with the exceptional points E_e defined in Sec. 4B.

When λ is real and $k(\lambda)$ is purely imaginary (mod π), then $k(-\lambda) = -k^*(\lambda)$. Therefore, $\beta(-\lambda) = \beta(\lambda)$ and $\chi(-\lambda) = \chi(\lambda)$. Consequently, if T has a pole at $\lambda = \lambda_b$, then it also has one at $\lambda = -\lambda_b$; similarly for ψ. Thus the bound-state poles in the band gaps occur in pairs symmetric with respect to the imaginary axis. (This is true both in the λ plane and in the k plane.)

From the Schrödinger equation (4.1) we also find that

$$\frac{d}{d\lambda}W(\chi_1,\chi_2)\Big|_{E=E_b} = -2\lambda_b c\int_{-\infty}^{\infty}\chi_2^2(x)\,dx,$$

and since for purely (mod π) imaginary k and real E χ_2 is real, this cannot vanish. Therefore, T has *simple* poles at the bound states. We note that bound states may occur at positive or negative E; the only requirement is that E_b not be in an allowed band.[19]

Suppose then that $E = E_b$, in a band gap or $E_b < 0$, is a bound-state eigenvalue. Then ψ has a simple pole at $E = E_b$ and its residue is of the form

$$\psi_b^{\rm res}(x) = u_b(x)s,$$

where $u_b(x)$ is the eigenfunction. [Apart from normalization this function is unique because otherwise *every* solution of (4.1) at E_b would have to vanish as $x \to \pm\infty$; it may be chosen to be real.] From the fact that the residue of the full resolvent of (4.1) at $E = E_b$ is proportional to $u_b(x)u_b(x')$ it follows that the normalization of u_b may be chosen so that

$$s = \int_{-\infty}^{\infty} dx\, u_b(x)U(x)\beta(x).$$

On the other hand, it follows from (3.18) and (4.6) that

$$(2i\epsilon\sin k_b)u_b(x) = \begin{cases}\beta_1(x)s_2 + o(e^{ik_bx}) & \text{as } x \to +\infty,\\ \beta_2(x)s_1 + o(e^{-ik_bx}) & \text{as } x \to -\infty.\end{cases}$$

Thus the relative values of the components of s measure the relative asymptotic values of the bound-state eigenfunction as $x \to \pm\infty$. We refer to the ray s as the *character* of the bound state. Define $\hat{s} = M^{-1}s$, with M evaluated at E_b.

Now at $E = E_b$ we must also have $\det J = \det J_0/T = 0$. So there must exist a vector s' such that $Js' = 0$ at $E = E_b$. Since $\phi = J\hat{\psi} = (TJ)(\hat{\psi}/T)$ is holomorphic at $E = E_b$ and $\hat{\psi}$ and T have simple poles there, the residue $\hat{\psi}_b^{\rm res}(x) = u_b(x)\hat{s}$ of $\hat{\psi}(x)$ must be such that $J\hat{s} = 0$. Apart from normalization therefore $\hat{s} = s'$. (It is easy to see that if $J = 0$ at $E = E_b$ then there are two linearly independent bound-state eigenfunctions. Since this is impossible, two linearly independent vectors s' cannot exist.) We conclude that *the character of the bound state is the null space of* JM^{-1} *at* $E = E_b$. As a function of λ, J has equal nontrivial null spaces at $\lambda = \lambda_b$ and at $\lambda = -\lambda_b$ if $\lambda_b \in \mathbb{R}\setminus\mathbb{R}_a'$.

It follows from (4.26) that J is a singular matrix also at those points at which J_0 is a singular matrix, i.e., at the bound states of $\hat{V}$. Equations (4.22), (4.22'), and (4.25) show that the range of J there equals the range of J_0, but their null spaces generally differ.

Suppose there are N bound states with the eigenvalues $E_b = \lambda_b^2$, $b = 1,\ldots,N$, including both the eigenvalues of $U + V$ and those of $\hat{V}$. We then define N real projections $B_b = B_b^* = B_b^2$ successively as follows, for $b = 1,\ldots,N$:

If $E_b < 0$, so that $k_b = \cos^{-1}\Delta(E_b)$ is positive imaginary, then

$$\Gamma_b(k) = \mathbb{1} - B_b + B_b\frac{k + k_b}{k - k_b};$$

if $E_b > 0$, so that k_b is positive imaginary (mod π), then

$$\Gamma_b(k) = \mathbb{1} - B_b + B_b\frac{k - k_b^*}{k - k_b}\frac{k + k_b}{k + k_b^*};$$

$$C_b = \Gamma_1(k_b)\cdots\Gamma_{b-1}(k_b), \quad C_1 = \mathbb{1},$$

$$(\mathbb{1} - B_b)C_b^{-1}\Psi_b^{\rm res} = 0,$$

where $\Psi_b^{\rm res}$ is the residue of $\Psi(x) = X^{-1}\hat{\Psi}(x)$ at $E = E_b$. Also define

$$\Pi = \Gamma_1\cdots\Gamma_N$$

and

$$\Psi^{\rm red} = \Pi^{-1}\Psi. \tag{5.1}$$

Then $\Psi^{\rm red}$ is holomorphic at $\lambda = \lambda_b$ and otherwise has the same analyticity and asymptotic properties as Ψ.

Next we define, for $E \in \mathbb{R}_a$

$$\hat{S}_x^{\rm red} = Q\Pi^{-1}Q\hat{S}_x\Pi^*. \tag{5.2}$$

Then by (5.1) and (4.21) for $E \in \mathbb{R}_a$

$$\Psi^{\rm red*} = Q\hat{S}_x^{\rm red*}\Psi^{\rm red}. \tag{5.3}$$

In the band gaps (4.21') leads to

$$\Psi^{\rm red*} = QS_{0x}^{\rm red*}\Psi^{\rm red}, \tag{5.3'}$$

where

$$S_{0x}^{\rm red} = Q\Pi^{-1}QS_{0x}\Pi^*. \tag{5.4}$$

We note that $|k| \to \infty$

$$\Pi = \mathbb{1} + O(k^{-1}). \tag{5.5}$$

Therefore, by (4.34), (2.13), and (2.13') as $\lambda \to \pm\infty$ in allowed bands

$$\hat{S}_x^{\text{red}} = 1 + O(\lambda^{-1}) \tag{5.6}$$

and in the band gaps

$$S_{0x}^{\text{red}} = 1 + O(\lambda^{-1}). \tag{5.6'}$$

A similar reduction is performed on the Jost function. If we denote by J_x the Jost function for a potential comprehensively shifted by x, then we define

$$J_x^{\text{red}} = J_x \Pi \tag{5.7}$$

and the decompositions (4.27) and (4.27') lead to

$$(J_x^{\text{red}\,-1})^* = Q\hat{S}_x^{\text{red}} * J_x^{\text{red}\,-1} Q \tag{5.8}$$

in the allowed bands, and

$$(J_x^{\text{red}\,-1})^* = QS_{0x}^{\text{red}} * J_x^{\text{red}\,-1} Q \tag{5.8'}$$

in the band gaps. It follows from the fact that the null space of J_x at λ_b equals the range of the residue of Ψ there that J_x^{red} has the same analytic properties as J_x, and $\det J_x$ has no zeros in $\mathbb{C}^+$ or in the gaps. Its asymptotic behavior is the same as (4.31).

B. Levinson's theorem

Let us define $\bar{\delta}$ as the argument of T, $T = |T|e^{i\bar{\delta}}$. Then we have by (4.15) in the allowed bands

$$\det S = e^{2i\delta}, \tag{5.9}$$

where $\delta = \bar{\delta} \pmod{\pi}$. We shall require that $\delta(-\lambda) = -\delta(\lambda)$, which we may since $S(-\lambda) = S^*(\lambda)$ [see Sec. 6].

In the band gaps, on the other hand, T is real and hence $\bar{\delta} = 0 \pmod{\pi}$. Since S is continuous, δ will be defined to be continuous in the allowed bands for $E > 0$. At the points of the periodic spectrum, T has (in the generic case) simple zeros, and, as we circumscribe a simple zero in the clockwise sense in the upper half-plane, the phase decreases by π. The phase δ, which is defined only in the allowed bands, will therefore be defined so that at each band gap the difference between the left-hand limit of δ at the left gap end and its right-hand limit at the right gap end is equal to 2π. As $E \to \infty$, Eq. (4.29) allows us to define δ to be zero. This defines δ uniquely in all the allowed regions. Its value at $E = 0+$ is then given by

Theorem 1 (Levinson theorem): Let n be the total number of bound states (i.e., those at negative energies and those in band gaps). If at each point of the periodic spectrum (including $E = 0$) $T = 0$ (which is the generic case), then the phase δ defined above is such that its value $\delta(0) = \lim \delta(E)$ as $E \to 0+$ is $\delta(0) = \pi(n - \frac{1}{2})$.[20]

Proof: We apply the argument principle to the analytic (by Lemma 2) function T and a contour that runs along the real axis from $-\infty$ to $+\infty$, avoiding each point of the periodic spectrum, including the origin, and each pole of T in a band gap, by a small semicircle in $\mathbb{C}^+$, and that is closed by a large semicircle in $\mathbb{C}^+$.

The result is that $\bar{\delta}(-\infty) - \bar{\delta}(\infty) = 2\pi n'$ if n' is the number of poles of T in $\mathbb{C}^+$, i.e., the number of negative-energy bound states. We define $\delta(\infty) = \bar{\delta}(\infty) = 0$. Since each bound-state pole in a gap appears once for $\lambda > 0$ and once for $\lambda < 0$, and at each such pole $\bar{\delta}$ increases from left to right by π while at each gap end it decreases by π, whereas δ decreases by 2π across each gap, we have $\delta(\epsilon) = \bar{\delta}(\epsilon) + \pi n''$ and $-\delta(\epsilon) = \delta(-\epsilon) = \bar{\delta}(-\epsilon) - \pi n'' - \bar{\delta}(-\infty)$ if ϵ is a small positive number and n'' is the number of positive-energy bound states. Furthermore, $\bar{\delta}(\epsilon) - \bar{\delta}(-\epsilon) = \pi$ because T has a simple zero at the origin. The solution of these equations is $\delta(\epsilon) = \pi(n' + n'' - \frac{1}{2})$. We then allow ϵ to approach zero and set $n = n' + n''$. Q.E.D.

6. THE INVERSE SCATTERING PROBLEM

We now have all the ingredients needed for the solution of the inverse scattering problem associated with Eq. (4.1). It will be given by the solution of the Riemann–Hilbert problem suggested by Eq. (4.12') and the analyticity of ψ. More specifically, it is the function Ψ defined by (4.20) and (4.16) rather than ψ that has to be considered, because of its simple asymptotic form (4.32).

The first need is to recognize a symmetry possessed by all the functions we have dealt with. The function γ defined by (2.2) satisfies the relation $\gamma^*(-\lambda, x) = \gamma(\lambda, x)$ for real λ. As a result of this and of the definition of k adopted the Bloch solution, β has the same symmetry property. This is true even in the band gaps because it is the limit on the real axis of $\beta^*(-\lambda^*, x) = \beta(\lambda, x)$ and $-\lambda^*$ is mapped into $-k^*$. As a result, all the functions defined in Sec. 3 obey this relation, and those defined in Secs. 4 and 5 inherit it. For any function f of λ, we shall use the notation f_-, defined by $f_-(\lambda) = f(-\lambda)$.

Combining Eqs. (4.21) and (4.21') into a single equation for all real λ we may now write them in the form

$$\Psi_- = \omega Q \Psi, \tag{6.1}$$

where

$$\omega = \begin{cases} Q\hat{S}_x^* Q, & \text{when } \lambda \in \mathbb{R}_a^*, \\ QS_{0x}^* Q, & \text{when } \lambda \notin \mathbb{R}_a'. \end{cases}$$

Similarly, (5.3) and (5.3') become

$$\Psi_-^{\text{red}} = \omega^{\text{red}} Q \Psi^{\text{red}}, \tag{6.2}$$

where

$$\omega^{\text{red}} = \Pi^{*-1} \omega Q \Pi Q \tag{6.3}$$

with Π as defined in Sec. 5. Similarly, Eqs. (4.27) and (4.27') may be combined to read

$$(J_x^{-1})_- = \omega Q J_x^{-1} Q, \tag{6.4}$$

and (5.8) and (5.8')

$$(J_x^{\text{red}\,-1})_- = \omega^{\text{red}} Q J_x^{\text{red}\,-1} Q. \tag{6.5}$$

The relevant Riemann–Hilbert problem $\mathscr{H}$ to be solved now is this: Given ω^{red} on the real λ axis, we wish to find a (2×2)-matrix-valued function $J_x^{\text{red}\,-1}$ that is holomorphic and zero-free (i.e., its determinant nowhere vanishes) in $\mathbb{C}^+$, continuous on the real axis, satisfies the asymptotic relation (4.32), and that obeys Eq. (6.5). If $J_x^{\text{red}\,-1}$ solves this problem, then

$$\Psi^{\text{red}} = J_x^{\text{red}\,-1} \hat{1}$$

satisfies (6.2) and all other requirements for Ψ^{red}. The advantage of solving (6.5) instead of (6.2) directly is that if an under-

lying potential is known to exist, then we know that the problem $\mathscr{H}$ has a solution. By Lemma 3 of Ref. 9 this solution is *unique*. It then follows that the solution of the problems for Ψ^{red} is also unique. No such uniqueness could be directly guaranteed for a Hilbert problem with vector-values solutions.

To find the solution of $\mathscr{H}$ we first define the Fourier transform of $\omega^{\text{red}} - \mathbb{1}$:

$$\mathscr{G}(p) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\lambda\, e^{i\lambda p}[\omega^{\text{red}}(\lambda) - \mathbb{1}]. \tag{6.6}$$

If there are no positive-energy bound states, i.e., no bound states in the band gaps, the function ω^{red} is continuous for all $\chi \in \mathbb{R}$, and by (5.6) and (5.6'),

$$\omega^{\text{red}} = \mathbb{1} + O(\lambda^{-1}) \tag{6.7}$$

as $\lambda \to \pm\infty$, both in allowed bands and in the band gaps. If there are positive-energy bound states, then Π has poles and hence ω^{red} has poles in the band gaps. We shall from now on assume that *there are no positive-energy bound states*. If we regard Ψ^{red}, J^{red}, and ω^{red} as functions of λ on $\mathbb{R}$, then $(\Psi^{\text{red}} - \hat{1}) \in L^2$, $(J^{\text{red}} - \mathbb{1}) \in L^2$, and $(\omega^{\text{red}} - \mathbb{1}) \in L^2$. The Fourier transform (6.6) therefore exists in the L^2 sense. We then have the following two theorems, the first dealing with the reconstruction of an underlying potential that is known to exist for the given data and the other dealing with the existence and construction of an underlying potential from arbitrarily given data:

Theorem 2: Suppose that $\hat{V} \in L^2(0,1)$, $U \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, S_0 is the S matrix of $\hat{V}$ [defined below (2.1)], S is the S matrix of $V + U$ as defined in Sec. 4, and ω^{red} is defined as in (6.3) and Sec. 5 in terms of the bound state eigenvalues and characters of $\hat{V}$ and $V + U$. Assume that there are no bound states in the band gaps. Define $\mathscr{G}$ as in (6.6) and let $\mathscr{g}$ be the self-adjoint operator $L^2(\mathbb{R}_+) \to L^2(\mathbb{R}_+)$ whose kernel is $\mathscr{G}(p+q)$. Then $\mathscr{g}$ is Hilbert–Schmidt; assume that 1 is not in its spectrum. Then the linear integral equation

$$\sigma(p) = \mathscr{G}(p)\hat{1} + \int_0^{\infty} dq\, \mathscr{G}(p+q)Q\sigma(q) \tag{6.8}$$

for $p \geqslant 0$ has a unique solution in $L^2(\mathbb{R}_+)$. If $\mathscr{g}$ does not have the eigenvalue -1 either, then the Neumann series of (6.8) converges. The Fourier transform of the solution of (6.8)

$$\Psi^{\text{red}}(\lambda) = \hat{1} + \frac{1}{2\pi}\int_0^{\infty} dp\, e^{i\lambda p}\sigma(p) \tag{6.9}$$

is such that $\hat{\psi} = X\Pi\Psi^{\text{red}}$ satisfies the Schrödinger equation (4.1) with the potential

$$V + U = -2\frac{d}{dx}\bar{\sigma}_1(x) = 2\frac{d}{dx}\bar{\sigma}_2(x), \tag{6.10}$$

where

$$\bar{\sigma}(x) = \sigma(0+) - \lim_{p\to 0+}\frac{1}{2\pi}\int_{-\infty}^{\infty} d\lambda e^{i\lambda p}[\Pi(\lambda) - \mathbb{1}]\hat{1}. \tag{6.11}$$

[Recall that Π and ω depend on x, and hence, so do $\mathscr{G}(p)$ and $\sigma(p)$.]

Theorem 3: Suppose that a real $V(x)$, $x \in \mathbb{R}$, and a unitary (2×2)-matrix-valued function $S(\lambda) = S^*(-\lambda)$, $\lambda \in \mathbb{R}'_a$, are given such that $V(x+1) = V(x)$ and $\hat{V}$ [as defined below (2.1)] $\in L^2(0,1)$, S is continuous, $(S - \mathbb{1}) \in L^2(\mathbb{R}'_a)$ and S satisfied the Levinson theorem (Theorem 1) for n bound states. Also given are n negative eigenvalues and corresponding bound-state characters. Define the self-adjoint operator $\mathscr{g}$ on $L^2(\mathbb{R}_+)$ by its integral kernel $\mathscr{G}(p+q)$ in terms of the Fourier transform of ω^{red} as in (6.6), where ω^{red} is defined by (6.3), (4.18), and (4.19), and Π as in Sec. 5; also define $\mathscr{g}_0$ by the kernel $-\mathscr{G}(-p-q)$. Suppose that $\mathscr{g}$ is Hilbert–Schmidt and neither $\mathscr{g}$ nor $\mathscr{g}_0$ has the eigenvalue 1, and that $\mathscr{G}(p)$ is continuous at $p = 0$. Then Eq. (6.8) has a unique solution $\sigma(p)$ in $L^2(\mathbb{R}_+)$. Define $\sigma(x)$ as in (6.11). Suppose further that the solution σ is "miraculous" in the sense that $\bar{\sigma}_1(x) = -\bar{\sigma}_2(x)$. Then the Fourier transform of σ as in (6.9) is such that $\hat{\psi} = X\Pi\Psi^{\text{red}}$ satisfies the Schrödinger equation (4.1) with the potential given by (6.10), S is the corresponding S matrix, the chosen eigenvalues and characters are those of (4.1) with the potential (6.10), and the residues of $\hat{\psi}$ at the poles of Π are the corresponding bound-state eigenfunctions.

The proofs of these two theorems are identical to those of Theorem 2.1 of Ref. 7 and of Theorem 3.1 of Ref. 8, and they need not be repeated here. It should be noted that an important ingredient of the proof of the latter theorem is the fact that (6.1) may be obtained from the Riemann–Hilbert problem $\mathscr{H}$ whose solution is an operator-valued (or in the present case, square-matrix-valued) function by operating on $\hat{1}$. This was the main purpose of the introduction of the Jost function here.

ACKNOWLEDGMENTS

This work was supported in part by Grant PHY-80-20457 from the National Science Foundation. I am also happy to acknowledge useful conversations with Mr. S. H. Yoon.

APPENDIX

It is customary to define two solutions by boundary conditions at $x = 0$:

$$y_1(0) = y_2'(0) = 1, \quad y_1'(0) = y_2(0) = 0. \tag{A1}$$

For $0 \leqslant x \leqslant 1$ these functions satisfy the Volterra equations (where $\lambda = \sqrt{E}$)

$$\begin{aligned} y_1(x) &= \cos\lambda x \\ &\quad + \frac{1}{\lambda}\int_0^x dt \sin\lambda(x-t)V(t)y_1(t), \\ y_2(x) &= (\sin\lambda x)/\lambda \\ &\quad + \frac{1}{\lambda}\int_0^x dt \sin\lambda(x-t)V(t)y_2(t), \end{aligned} \tag{A2}$$

which have unique solutions obtainable by iteration if $V \in L^1(0,1)$. The two functions y_1 and y_2 are linearly independent and their Wronskian is

$$W(y_1,y_2) = y_1y_2' - y_1'y_2 = 1. \tag{A3}$$

The Bloch solutions β_i, $i = 1,2$, are expressible as linear combinations of y_1 and y_2 as follows:

$$\beta_1(x) = (e^{ik} - \epsilon_2)y_2(x) + \epsilon y_1(x), \tag{A4}$$

and the quantities ϵ, ϵ_1, and ϵ_2 are given by $\epsilon = y_2(1)$, $\epsilon_1 = y_1'(1)$, $\epsilon_2 = y_1(1)$. The discriminant is given by $\Delta = \frac{1}{2}[\, y_2'(1) + \epsilon_2\,]$, and the relation between γ_i and y_i is

$$\begin{aligned} \gamma_1(x) &= (1 + \eta_1)y_1(x) + i\lambda\,(1 - \eta_1)y_2(x), \\ \gamma_2(x) &= (1 + \eta_3)y_1(x) - i\lambda\,(1 + \eta_3)y_2(x), \end{aligned} \tag{A5}$$

and, conversely,

$$\begin{aligned} y_1(x) &= \frac{1}{2}\gamma_1(x) + \frac{1}{2}\frac{1 - \eta_1}{1 + \eta_3}\gamma_2(x), \\ y_2(x) &= \frac{1}{2i\lambda}\,[\gamma_1(x) - l\gamma_2(x)]. \end{aligned} \tag{A6}$$

The Green's function (3.19) may, by (A4), be expressed in the alternative form

$$g(x, x') = y_1(x)y_2(x') - y_1(x')y_2(x), \quad x' < x. \tag{A7}$$

The Volterra equations for y_1 and y_2 imply that if $\hat{V} \in L^1(0,1)$, then both of these solutions are entire analytic functions of E. Furthermore, they easily lead to the conclusion that as $|E| \to \infty$ with $\nu = \operatorname{Im}\lambda$,

$$\begin{aligned} y_1(x) &= \cos\lambda x + O(\lambda^{-1}e^{|\nu|\,|x|}), \\ y_2(x) &= (\sin\lambda x)/\lambda + O(\lambda^{-2}e^{|\nu|\,|x|}). \end{aligned} \tag{A8}$$

Therefore, Δ is an entire analytic function of E of order $\frac{1}{2}$, as is well known.

[1] T. H. Dupree, Ann. Phys. (N.Y.) **15**, 63 (1961); G. J. Morgan, Proc. Phys. Soc. **89**, 365 (1966); N. A. W. Holzwarth, Phys. Rev. B **11**, 3718 (1975).

[2] P. Ungar, Comm. Pure Appl. Math. **14**, 707 (1961); H. Hochstadt, Arch. Rat. Mech. Anal. **19**, 353 (1965); W. Magnus and S. Winkler, *Hill's Equation* (Interscience, New York, 1966); H. P. McKean and P. van Moerbeke, Invenciones Math. **30**, 217 (1975); E. Trubowitz, Comm. Pure Appl. Math. **30**, 321 (1977).

[3] L. D. Faddeev, T. Mat. Inst. Steklov **73**, 314 (1964) [Am. Math. Soc. Transl. **2**, 139 (1964)]; P. Deift and E. Trubowitz, Comm. Pure Appl. Math. **32**, 121 (1979).

[4] R. G. Newton, J. Math. Phys. **21**, 493 (1980).

[5] N. Levinson, K. Dan. Vidensk. Selsk. Mat.-Phys. Medd. **25**(9) (1949).

[6] R. G. Newton, J. Math. Phys. **21**, 1698 (1980); **22**, 631 (1981); **23**, 693 (1982).

[7] R. G. Newton, J. Math. Phys. **22**, 2191 (1981); **23**, 693 (1982).

[8] R. G. Newton, J. Math. Phys. **23**, 594 (1982).

[9] R. G. Newton, J. Math. Phys. **23**, 2257 (1982).

[10] A tilde denotes the transpose.

[11] An asterisk denotes the complex conjugate.

[12] A dagger denotes the adjoint.

[13] This is known as the *reciprocity theorem*.

[14] $C^{\pm}$ mean the upper and lower half-planes, not including the real axis.

[15] See, for example, Magnus and Winkler, Ref. 2.

[16] E. Trubowitz, Ref. 2.

[17] B. M. Levitan and I. S. Sargsjan, *Introduction to Spectral Theory* (Am. Math. Soc., Providence, RI, 1975), p. 21.

[18] H. Hochstadt, Ref. 2.

[19] For results on the distribution of bound states among the band gaps, see F. S. Rofe-Beketov, Dokl. Akad. Nauk SSSR **156**, 1029 (1964) [Sov. Math. **5**, 772 (1964)]; V. A. Zheludev, in *Topics in Mathematical Physics*, edited by M. Sh. Birman (Consultants Bureau, New York, 1968, 1971), Vol. 2, p. 87 and Vol. 4, p. 55; M. Klaus, Helv. Phys. Acta **55**, 49 (1982).

[20] The statement of the Levinson theorem in Ref. 4, Sec. 3, is in error. When $T(0) = 0$, which is the generic case, then $\delta(0) - \delta(\infty) = \pi(n - \frac{1}{2})$; when there is a half-bound state, i.e., $T(0) = 1$, then $\delta(0) - \delta(\infty) = \pi n$.

Transmission through a system of potential barriers. I. Transmission coefficient

Örjan Dammert
Institute of Theoretical Physics, University of Uppsala, Thunbergsvägen 3, S-752 38 Uppsala, Sweden

(Received 25 June 1981; accepted for publication 28 January 1983)

The transmission coefficient for a one-dimensional system of N potential barriers of general shapes (unequal or equal) is expressed exactly, by means of the phase-integral method of N. Fröman and P. O. Fröman, in terms of quantities characterizing the separate barriers and wells of the system. The exact formula, being of interest for further rigorous studies of transmission properties, can for evaluation readily be converted into a useful approximate formula by insertion of available phase-integral expressions (of an arbitrary order) for the characteristic quantities mentioned above. Error estimates are given. The sub- and superbarrier cases are treated in a unified way. Some key facts about the phase-integral method are given in Appendix A.

PACS numbers: 03.65.Sq

1. INTRODUCTION

Tunneling phenomena, widely discussed in physical literature ever since the early days of quantum mechanics, play an important role in several fields of physics such as, e.g., solid state[1,2] and surface physics,[3] soliton theory,[4,5] molecular theory, and the theory of nuclear fission.[6] Some general results concerning tunneling[7] have recently been obtained in connection with the study of black holes.

In the present paper and some papers to follow,[8–10] transmission properties of smooth one-dimensional multibarrier potentials of general shapes will be studied by means of a powerful phase-integral method, due to N. Fröman and P. O. Fröman, which is developed in five publications.[11–15] In Ref. 11 a rigorous method for mastering connection problems was jointly constructed by N. Fröman and P. O. Fröman. Subsequently, N. Fröman introduced the "symmetric" higher-order phase-integral approximations[12–13] into this method which were later generalized by both authors, in Ref. 14 and on pp. 126–131 in Ref. 15, yielding a versatile tool in semiclassical physics, sometimes useful even outside that domain.[16] The advantages, as documented in Ref. 17, of using higher-order phase-integral approximations instead of the related higher-order JWKB approximations are primarily due to their simpler form, the relation between phase and amplitude being, in a classically allowed region, the same as for the first-order JWKB approximation and, more important, the same as for the exact solution of the Schrödinger equation (see p. 14 in Ref. 11). It is this formal structure of the phase-integral functions that renders them useful in the phase-integral method. In a series of papers during the past 15 years the Fröman method has been successfully applied to various physical problems. Simple formulae admitting accurate evaluations have been given for a number of physical quantities such as energy eigenvalues, level densities, normalization factors, quantal expectation values, quantal matrix elements, dispersion relations, and transmission and reflection coefficients.[16,18–28]

Some important features of the treatment in the present and subsequent papers[8–10] are comprised in the following points: (i) the considered multibarrier potentials are of general shapes; (ii) the subbarrier and superbarrier cases are treated in a unified way; (iii) the calculations are performed without approximations resulting in exact final formulae; (iv) phase-integral expressions (of an arbitrary order) for characteristic quantities appearing in the final formulae are given with rigorous error estimates. Particularly in higher orders, these expressions are often extremely accurate. Hence, the exact final formulae can readily be converted into useful approximate ones of high accuracy.

In Secs. 2–8 below we shall consider a one-dimensional potential function $V(x)$ forming a system of N potential barriers of general shapes (unequal or equal). Certain quantities characterizing the individual barriers and wells in the system will be introduced, by means of which convenient matrices called **P** matrices are defined, one for each barrier interval. The transmission and reflection coefficients are finally expressed in terms of these **P** matrices, by Eqs. (43a), (43b) and (44a), (44b), which are the end results of this paper and which, further, form the starting point for the investigations in Refs. 8–10. For the benefit of the reader who is not already acquainted with the phase-integral method, some key facts about this method have been gathered in Appendix A. In order to make the present paper more easy to read, we shall sometimes refer to this appendix instead of to the original papers. Further references are then given in Appendix A. In Appendix B we give a list of phase-integral expressions (of an arbitrary order) for the characteristic quantities appearing in the exact final formulae. Error estimates are given, and ranges of validity are discussed.

Specializing (43a), (43b) and (44a), (44b) to the subbarrier case and first-order approximation, using phase-integral expressions given in Appendix B (with omission of correction terms) and putting all σ quantities equal to zero, we arrive at approximate formulae which are equivalent with Eqs. (109) and (110) in a paper by Ponomarev[29] (if misprints in his equations are corrected).

 0022-2488/83/082163-13$02.50

2. EXACT FORMULAE FOR THE TRANSMISSION COEFFICIENT

The phase-integral method, developed by N. Fröman and P. O. Fröman in Refs. 11–14 and on pp. 126–131 in Ref. 15, forms the basis for the treatment in the present paper of particle transmission through a system of an arbitrary number of potential barriers of various shapes. The particular cases of a single potential barrier and a system of two potential barriers have previously been treated by means of the phase-integral method, in Refs. 11, 26 and Ref. 27, respectively. By approaching the general case of N potential barriers along similar lines we shall be able to utilize some of the considerations and arguments in Refs. 27 and 11. Certain knowledge about the principal features of the phase-integral method will be presupposed below; this refers, for instance, to the function $q(z)$, the phase-integral functions $f_1(z)$ and $f_2(z)$, the Riemann surface for defining $w(z)$, the choice of lower limit in the $w(z)$ integral, and the matrix $\mathbf{F}(z_1, z_2)$. The reader is referred to Appendix A, where some key facts about the phase-integral method extracted from the works mentioned above, have been collected. For further details we can refer, besides to the original papers, to a recent review article by N. Fröman in a monograph[30] on semiclassical methods in molecular scattering and spectroscopy.

We study the one-dimensional Schrödinger equation

$$\frac{d^2\psi}{dz^2} + Q^2(z)\psi = 0, \tag{1}$$

where

$$Q^2(z) = 2m[E - V(z)]/\hbar^2, \tag{2}$$

z being a complex variable (the real values of which will be denoted by x) and E being the energy of a particle with mass m moving in the potential field $V(x)$ which, having N humps, forms a general system of N potential barriers. We assume that $Q^2(z)$ is real on the real axis, i.e., $V(x)$ real. Furthermore, we assume that $Q^2(z)$ is an analytical function of z in the complex plane.

We shall restrict the possible choices of the function $q(z)$ to the unmodified phase-integral expressions given by (A36) in Appendix A.

Figures 1(a) and 1(b) show the leftmost barrier, $B(1)$, of the considered system in the cases of subbarrier and superbarrier transmission, respectively. Figures 2(a)–(d) show (for $n \geqslant 2$) the nth barrier from the left, $B(n)$, together with the neighboring barrier to the right, $B(n+1)$, in the different cases of subbarrier and superbarrier transmission that are possible. With the exception of part β, Figs. 2(a)–(d) are applicable also for $n = 1$. The phase of $q^{1/2}(z)$ on the real axis of the first Riemann sheet as well as the lower limit of integration in the integral defining $w(z)$ have been chosen, as in Ref. 27 (see there pp. 629–35), in the way described in the last fourth part of Appendix A.

We shall make the following assumptions concerning the behavior of $V(x)$ for large values of $|x|$. We assume that $Q^2(x)$ and hence also $q^2(x)$ is positive for sufficiently large values of $|x|$ and that the integral $\mu(x_1, x_{2N+1})$, defined by (4.2) in Ref. 11, with the integration performed along some convenient path on the upper half of the first Riemann sheet,

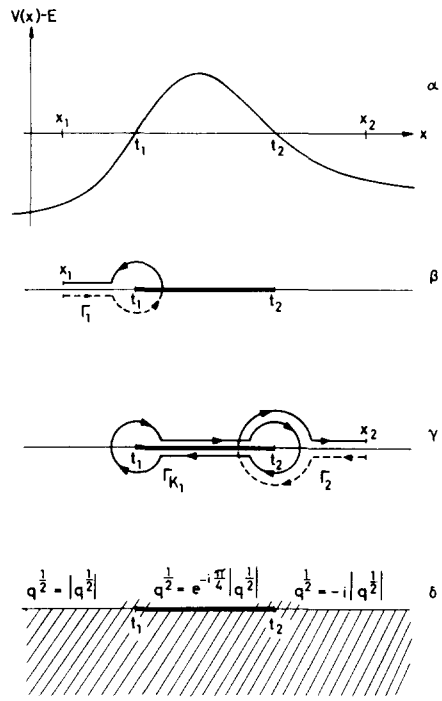


FIG. 1(a). Barrier $B(1)$ in the subbarrier case: (α) Qualitative behavior of $V(x) - E$. ($\beta - \gamma$) Contours of integration for obtaining $w(z)$. Those parts of the contours which lie on the second Riemann sheet are indicated by dashed lines. The heavy line indicates a cut. (δ) Phase of $q^{1/2}(z)$ on the real axis of the first Riemann sheet.

converges as x_1 tends to $-\infty$ and as x_{2N+1}, lying to the right of the barrier system, tends to $+\infty$. From this it follows, similarly as on p. 92 in Ref. 11, that if z does not coincide with a singularity or zero of $q^2(z)$, the limits $\mathbf{F}(-\infty, z)$, $\mathbf{F}(z, +\infty)$, and $\mathbf{F}(-\infty, +\infty)$ exist and are finite, which in

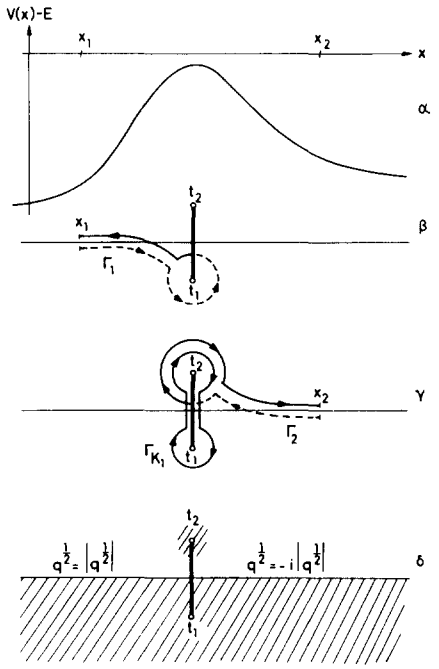
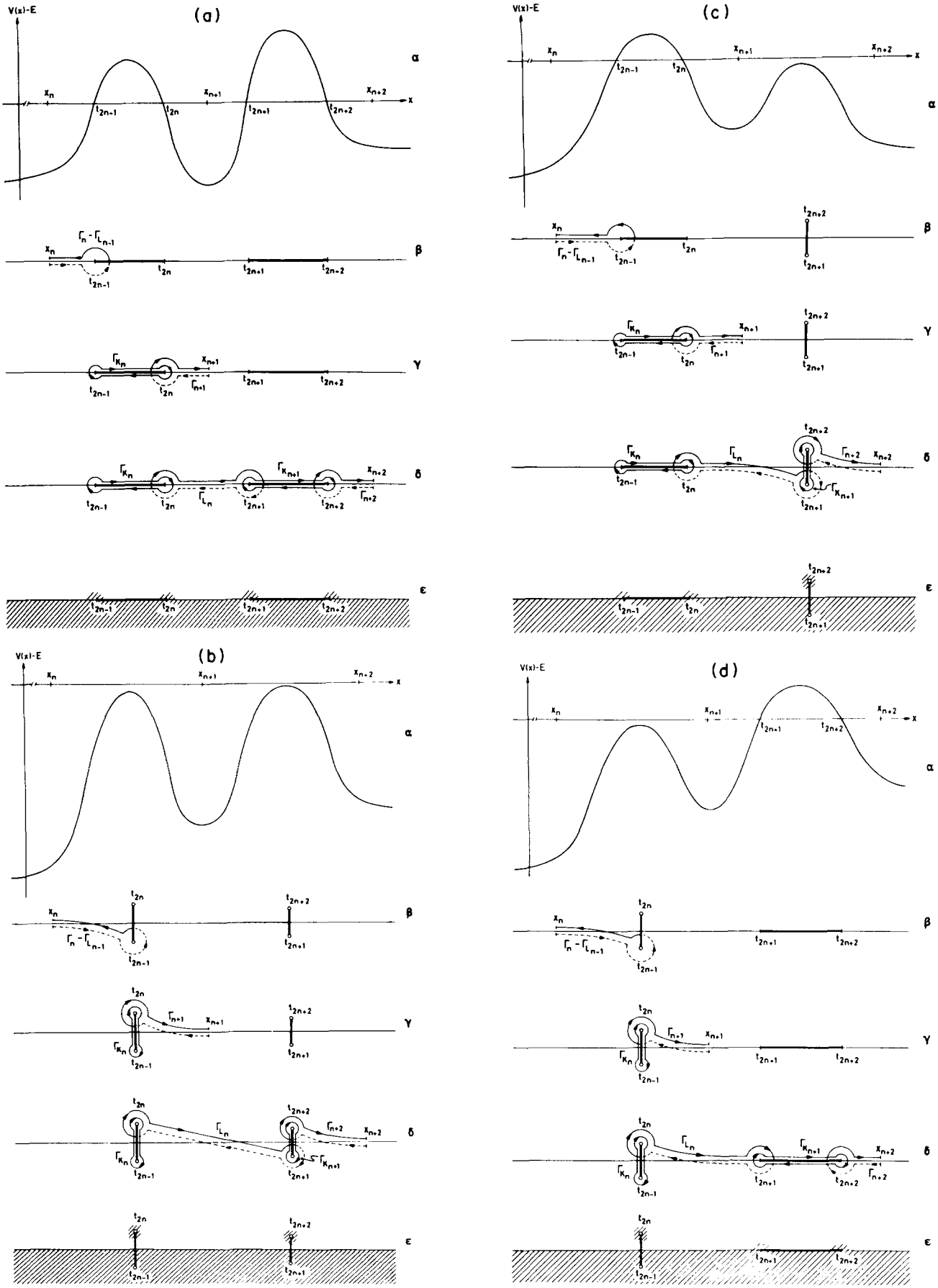


FIG. 1(b). Barrier $B(1)$ in the superbarrier case. The figure should be interpreted in the same manner as Fig. 1(a).



FIGS. 2(a)–(d). The barriers $B(n)$ and $B(n+1)$ in the four possible cases of subbarrier and superbarrier transmission: (α) Qualitative behavior of $V(x) - E$. ($\beta - \delta$) Contours of integration for obtaining $w(x)$. Dashed lines indicate those parts of the contours which lie on the second Riemann sheet. The heavy lines indicate cuts between zeros of $Q^2(z)$. (ϵ) The phase of $q^{1/2}(z)$ on the real axis of the first Riemann sheet is specified in Figs. 1(a) and 1(b).

turn implies that the limits $a_1(\pm\infty)$ and $a_2(\pm\infty)$ exist and are finite, and that $f_1(x)$ and $f_2(x)$ for increasing values of $|x|$ approach two linearly independent solutions of the Schrödinger equation. These solutions may be interpreted as waves traveling in opposite directions.

Thus, in the expression

$$\psi(x) = a_1(x)f_1(x) + a_2(x)f_2(x) \tag{3}$$

for the solution of the Schrödinger equation on the real axis obtained from (A3) in Appendix A, the two terms may, as $|x|\to\infty$, be interpreted as waves moving in opposite directions. According to (9.5) in Ref. 11, the corresponding probability currents are given by

$$(\hbar/m)q|a_1f_1|^2, \tag{4a}$$

and

$$-(\hbar/m)q|a_2f_2|^2 \tag{4b}$$

respectively, where f_1 and f_2 are defined by (A4a) and (A4b) in Appendix A. We observe that the sign of q decides the directions of the two waves. In consequence of the choice of phase of $q^{1/2}(z)$, shown in Figs. 1(a) and 1(b), the sign of $q(z)$ on the real axis of the first Riemann sheet is positive to the left of the barrier system, while to the right of the system it is negative in the case of an odd number of barriers but positive for an even number.

Assuming that there is only an outgoing wave far to the right of the barrier system, we conclude from (4a) and (4b) that the transmission coefficient is exactly given by

$$T_{\text{odd}} = \left| \frac{a_2(+\infty)e^{-iw(+\infty)}}{a_1(-\infty)e^{iw(-\infty)}} \right|^2 \tag{5a}$$

when the system contains an odd number of barriers, and by

$$T_{\text{even}} = \left| \frac{a_1(+\infty)e^{iw(+\infty)}}{a_1(-\infty)e^{iw(-\infty)}} \right|^2 \tag{5b}$$

in the case of an even number of barriers. It is clear from the remarks above, that the difference in appearance between (5a) and (5b) is due to the fact that the outgoing wave, far to the right of the barrier system, is represented by $a_2(z)f_2(z)$ in the case of an odd number of barriers but by $a_1(z)f_1(z)$ in the case of an even number.

3. EXPRESSIONS FOR $w(x)$

It is convenient to introduce for $n = 1,2,3,\cdots$ the quantities

$$K_n = (-1)^{n+1}\frac{i}{2}\int_{\Gamma_{K_n}} q(z)\,dz, \tag{6}$$

$$L_n = (-1)^n \text{Re}\left(\frac{1}{2}\int_{\Gamma_{L_n}} q(z)\,dz\right) \quad \text{(positive)}, \tag{7}$$

which both are real, in consequence of the reality of $Q^2(z)$ on the real axis. The integration contours are shown in Figs. 2(a)–(d). When the first-order expression for $q(z)$ is used [i.e., $q(z) = Q(z)$ according to (A36) in Appendix A], the quantity K_n is positive in the case of subbarrier transmission for the barrier $B(n)$ but negative in the superbarrier case while the quantity L_n is always positive. The same is true also when a higher-order expression for $q(z)$ is used, except for energy values lying close to the top of the barrier $B(n)$ or, alternatively (as concerns L_n), the barrier $B(n+1)$. The behavior of K_n and L_n as the energy approaches a proper barrier top is elucidated in the discussion on pp. 316–18 in Ref. 19. We remark that K_1, K_2, and L_1 are the same quantities as are defined by Eqs. (6a), (6b), and (7) in Ref. 27.

We further define for $n = 1,2,3,\cdots$

$$w_1(x_1) = \text{Re}\left(\frac{1}{2}\int_{\Gamma_1} q(z)\,dz\right), \tag{8a}$$

$$w_{2n+1}(x_{n+1}) = \text{Re}\left(\frac{1}{2}\int_{\Gamma_{n+1}-\Gamma_{L_n}} q(z)\,dz\right), \tag{8b}$$

$$w_{2n}(x_{n+1}) = \text{Re}\left(\frac{1}{2}\int_{\Gamma_{n+1}} q(z)\,dz\right). \tag{8c}$$

The integration contours are shown in Figs. 1(a), 1(b), and 2(a)–(d). The subscript $2n$ in the notation $w_{2n}(x_{n+1})$ indicates that the proper integration contour encloses the turning point t_{2n} with the same subscript. Analogously for the subscript $2n+1$. The quantity $w_{2n}(x_{n+1})$ is positive if n is an even number, and negative otherwise. The quantity $w_{2n+1}(x_{n+1})$ is instead negative if n is an even number, and positive otherwise.

From (7), (8b) and (8c) it follows that

$$w_{2n}(x_{n+1}) = w_{2n+1}(x_{n+1}) + (-1)^n L_n, \quad n = 1,2,3,\cdots. \tag{9}$$

Having chosen the lower limit in the integral defining $w(z)$ as described in Appendix A, we obtain with the aid of (6)–(9) the following expressions for $w(z)$ in the different regions of the real axis

$$w(x_1) = w_1(x), \tag{10a}$$

$$w(x_{n+1}) = w_{2n+1}(x_{n+1}) + \sum_{\nu=1}^{n}(-1)^\nu(L_\nu + iK_\nu),$$

$$n = 1,2,3,\cdots. \tag{10b}$$

We note that

$$\text{Im}\, w(x_1) = 0, \tag{11a}$$

$$\text{Im}\, w(x_{n+1}) = \sum_{\nu=1}^{n}(-1)^\nu K_\nu, \quad n = 1,2,3,\cdots. \tag{11b}$$

4. ALTERNATIVE FORMULAE FOR T

Since there is only a transmitted wave far to the right of the barrier system, we have $a_1(+\infty) = 0$ in the case of an odd number of barriers but $a_2(+\infty) = 0$ in the case of an even number [see the text in connection with formulae (3)–(5)]. From (3.16) in Ref. 11 it then follows that

$$a_1(-\infty) = F_{12}(-\infty,+\infty)a_2(+\infty) \tag{12a}$$

when the system contains an odd number of barriers and that

$$a_1(-\infty) = F_{11}(-\infty,+\infty)a_1(+\infty) \tag{12b}$$

in the case of an even number of barriers.

Inserting (11a), (11b) and (12a), (12b) into the exact expressions (5a) and (5b) for the transmission coefficient T, we get the exact formulae

$$\frac{1}{T_{\text{odd}}} = |F_{12}(-\infty,+\infty)|^2 \exp\left[-2\sum_{\nu=1}^{N}(-1)^\nu K_\nu\right],$$

$$\text{for } N \text{ odd}, \tag{13a}$$

$$\frac{1}{T_{\text{even}}} = |F_{11}(-\infty,+\infty)|^2 \exp\left[2\sum_{\nu=1}^{N}(-1)^{\nu}K_{\nu}\right],$$

for N even, (13b)

where N is the number of potential barriers in the system. Formula (13a) should be used when N is odd and (13b) when N is even.

5. BARRIER CHARACTERISTICS

The transmission properties of the leftmost barrier $B(1)$ are contained in the matrix $\mathbf{F}(x_1,x_2)$, where x_1 and x_2 are two points on the real axis on opposite sides of the barrier, defining the barrier interval.

The following notations are introduced

$$A_1 = |F_{12}(x_1,x_2)|, \tag{14a}$$

$$B_1 = |F_{22}(x_1,x_2)|, \tag{14b}$$

$$\sigma_1 = \tfrac{1}{2}[\arg F_{12}(x_1,x_2) - \tfrac{1}{2}\pi], \tag{14c}$$

$$\tau_1 = \tfrac{1}{2}\arg F_{22}(x_1,x_2). \tag{14d}$$

From Eqs. (5.7a) and (5.7b) in Ref. 11, which are valid when the unmodified phase-integral expressions are used for specifying $q(z)$ and when the lower limit in the integral $w(z)$ is chosen as in the present paper, we obtain with the aid of (11a) and (11b) the symmetry relations

$$F_{11}(x_1,x_2) = -F_{22}^{*}(x_1,x_2)\exp(2K_1), \tag{15a}$$

$$F_{21}(x_1,x_2) = -F_{12}^{*}(x_1,x_2)\exp(2K_1). \tag{15b}$$

By virtue of (14) and (15) the matrix $\mathbf{F}(x_1,x_2)$ can be written

$$\mathbf{F}(x_1,x_2) = \begin{pmatrix} -B_1\exp(2K_1 - i2\tau_1) & A_1\exp[i(2\sigma_1 + \frac{1}{2}\pi)] \\ -A_1\exp[2K_1 - i(2\sigma_1 + \frac{1}{2}\pi)] & B_1\exp(i2\tau_1) \end{pmatrix}. \tag{16}$$

Inserting (16) into the exact equation $\det \mathbf{F}(x_1,x_2) = 1$, which is valid according to (3.19) in Ref. 11, we obtain

$$A_1^2 - B_1^2 = \exp(-2K_1). \tag{17}$$

We shall now define corresponding quantities for the other barriers in the system. Consider $B(n)$, the nth barrier from the left. In order to obtain quantities which properly correspond to A_1, B_1, σ_1, and τ_1 for the barrier $B(1)$, we must, when defining A_n, B_n, σ_n, and τ_n for the barrier $B(n)$ in accordance with (14), take care that the phase of $q(z)$ on the real axis as well as the lower limit in the $w(z)$ integral are now specified so as to be related to the barrier $B(n)$ in the same way as they were related to the barrier $B(1)$ in the definitions of A_1, B_1, σ_1, and τ_1.

We introduce the notation $\mathbf{F}(t_{2m-1},n;z_1,z_2)$ for that $\mathbf{F}(z_1,z_2)$ matrix which is obtained by choosing the lower limit in the $w(z)$ integral in the neighborhood of the turning point t_{2m-1} instead of the turning point t_1, but otherwise in precisely the same way as it was chosen in the neighborhood of t_1 earlier in the present paper, and by choosing the phase of $q(z)$ on the real axis in the way shown in Fig. 3, where the situation at the barrier $B(n)$ is depicted. The matrix $\mathbf{F}(z_1,z_2)$ appearing in (14) is thus identical with $\mathbf{F}(t_1,1;z_1,z_2)$.

We now define, for $n = 1,2,\ldots,N$,

$$A_n = |F_{12}(t_{2n-1},n;x_n,x_{n+1})|, \tag{18a}$$

$$B_n = |F_{22}(t_{2n-1},n;x_n,x_{n+1})|, \tag{18b}$$

$$\sigma_n = \tfrac{1}{2}[\arg F_{12}(t_{2n-1},n;x_n,x_{n+1}) - \tfrac{1}{2}\pi], \tag{18c}$$

$$\tau_n = \tfrac{1}{2}\arg F_{22}(t_{2n-1},n;x_n,x_{n+1}). \tag{18d}$$

The definition of σ_n is consistent with the definitions of σ_1 and σ_2 in Ref. 27. A comparison between the contributions from the individual barriers to the transmission coefficient is facilitated by these uniform definitions of the barrier characteristics. Phase-integral expressions (of an arbitrary order) for the quantities A_n, B_n, σ_n, and τ_n are given by (B1), (B2), (B8), and (B9) in Appendix B.

With the aid of Eqs. (18) and the symmetry relations for the matrix $\mathbf{F}(t_{2n-1},n;x_n,x_{n+1})$, which are the same as those for the matrix $\mathbf{F}(x_1,x_2)$ displayed in (15) if only K_1 is replaced by K_n, we can write

$$\mathbf{F}(t_{2n-1},n;x_n,x_{n+1}) = \begin{pmatrix} -B_n\exp(2K_n - i2\tau_n) & A_n\exp[i(2\sigma_n + \frac{1}{2}\pi)] \\ -A_n\exp[2K_n - i(2\sigma_n + \frac{1}{2}\pi)] & B_n\exp(i2\tau_n) \end{pmatrix}, \quad n = 1,2,\ldots,N. \tag{19}$$

Insertion of (19) into the relation $\det \mathbf{F}(t_{2n-1},n;x_n,x_{n+1}) = 1$, which follows from (3.19) in Ref. 11, yields

$$A_n^2 - B_n^2 = \exp(-2K_n). \tag{20}$$

6. DEPENDENCE OF THE F-MATRIX ELEMENTS ON THE LOWER LIMIT IN THE $w(z)$ INTEGRAL AND ON THE PHASE OF $q(z)$

It follows from formulae (3.22a)–(3.22d) in Ref. 11 that the quantities $F_{12}(z_1,z_2)\exp[2iw(z_1)]$, $F_{21}(z_1,z_2)\exp[-2iw(z_1)]$ and the diagonal elements $F_{11}(z_1,z_2)$ and $F_{22}(z_1,z_2)$ are independent of the lower limit of integration in the integral defining $w(z)$. This implies that

$$F_{12}(t_{2n-1},1;z_1,z_2) = F_{12}(t_1,1;z_1,z_2)\exp(2i\Theta_{n-1}), \tag{21a}$$

$$F_{21}(t_{2n-1},1;z_1,z_2) = F_{21}(t_1,1;z_1,z_2)\exp(-2i\Theta_{n-1}), \tag{21b}$$

where

$$\Theta_n = w(t_1;z_1) - w(t_{2n+1};z_1), \tag{22}$$

the quantity $w(t_1;z)$ being identical with $w(z)$ as it is defined in the present paper (as well as in Ref. 27), while the quantity $w(t_{2n+1};z)$ is obtained from $w(t_1;z)$ by replacing, in the definition of $w(t_1;z)$, the turning point t_1 by t_{2n+1}, which amounts to moving the lower limit of integration in the $w(z)$ integral from the neighborhood of t_1 to the neighborhood of t_{2n+1}.

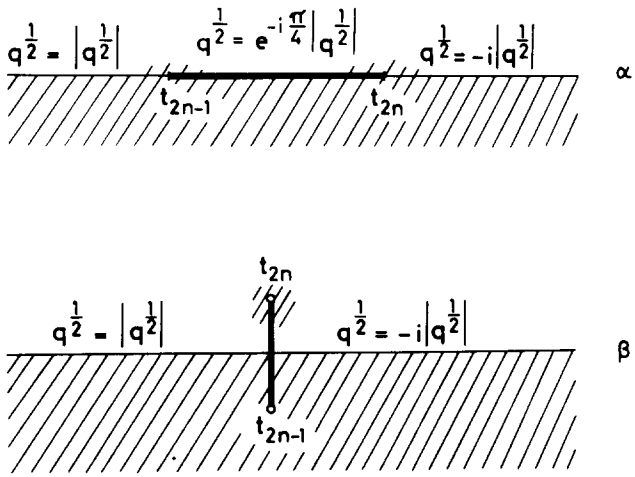


FIG. 3. Choice of phase of $q^{1/2}(z)$ on the real axis of the first Riemann sheet for the definition of the matrix $F(t_{2m-1},n;z_1,z_2)$. The picture shows the region in the neighborhood of the barrier $B(n)$ in the (α) subbarrier and (β) superbarrier case.

Formulae (21) are valid for two arbitrary points z_1 and z_2 situated in the region where the **F** matrix is uniquely defined. It is evident that the value of Θ_n is independent of the position of the point z_1. For $n = 0$, Eq. (22) simply reads

$$\Theta_0 = 0. \tag{23a}$$

Putting first $z_1 = x_{n+1}$ in (22) and then using the identities

$$w(t_1;x_{n+1}) = w(x_{n+1}),$$
$$w(t_{2n+1};x_{n+1}) = w_{2n+1}(x_{n+1}),$$

where the latter identity can easily be verified with the aid of (8b) together with Figs. 2(a)–(d), we find from (22) and (10b) that

$$\Theta_n = \sum_{\nu=1}^{n} (-1)^{\nu}(L_\nu + iK_\nu), \quad n = 1,2,...,N-1. \tag{23b}$$

Choosing $z_1 = x_n$ and $z_2 = x_{n+1}$ in Eqs. (21), we obtain

$$F_{12}(t_{2n-1},1;x_n,x_{n+1}) = F_{12}(t_1,1;x_n,x_{n+1})\exp(2i\Theta_{n-1}), \tag{24a}$$

$$F_{21}(t_{2n-1},1;x_n,x_{n+1}) = F_{21}(t_1,1;x_n,x_{n+1})\exp(-2i\Theta_{n-1}), \quad n = 1,...,N. \tag{24b}$$

For the sake of completeness we shall also write down the corresponding relations for the diagonal elements [see the text above (21)]

$$F_{11}(t_{2n-1},1;x_n,x_{n+1}) = F_{11}(t_1,1;x_n,x_{n+1}), \tag{24c}$$

$$F_{22}(t_{2n-1},1;x_n,x_{n+1}) = F_{22}(t_1,1;x_n,x_{n+1}), \quad n = 1,...,N. \tag{24d}$$

Examining the structure of formulae (3.22a)–(3.22d) in Ref. 11, we find that if the phase of $q(z)$ is altered by an odd multiple of π, i.e., if q is replaced by $-q$, the elements of the matrix $\mathbf{F}(z_1,z_2)$ change their values in such a way that the new matrix elements are obtained from the original ones according to the scheme

$$F_{11}^{\text{new}}(z_1,z_2) = F_{22}^{\text{orig}}(z_1,z_2), \tag{25a}$$

$$F_{12}^{\text{new}}(z_1,z_2) = F_{21}^{\text{orig}}(z_1,z_2), \tag{25b}$$

$$F_{21}^{\text{new}}(z_1,z_2) = F_{12}^{\text{orig}}(z_1,z_2), \tag{25c}$$

$$F_{22}^{\text{new}}(z_1,z_2) = F_{11}^{\text{orig}}(z_1,z_2). \tag{25d}$$

A change in the phase of $q(z)$ by an even multiple of π, however, leaves the matrix $\mathbf{F}(z_1,z_2)$ unaffected [cf. the remark below (3.22d) in Ref. 11].

We can now express the matrix $\mathbf{F}(x_n,x_{n+1})$, for $n = 1,2,...,N$, in terms of elements of $\mathbf{F}(t_{2n-1},n;x_n,x_{n+1})$. From (24) and (25), with comments, we obtain, if n is an odd integer,

$$F_{11}(x_n,x_{n+1}) = F_{11}(t_{2n-1},n;x_n,x_{n+1}), \tag{26a}$$

$$F_{12}(x_n,x_{n+1}) = F_{12}(t_{2n-1},n;x_n,x_{n+1})\exp(-2i\Theta_{n-1}), \tag{26b}$$

$$F_{21}(x_n,x_{n+1}) = F_{21}(t_{2n-1},n;x_n,x_{n+1})\exp(2i\Theta_{n-1}), \tag{26c}$$

$$F_{22}(x_n,x_{n+1}) = F_{22}(t_{2n-1},n;x_n,x_{n+1}), \tag{26d}$$

and, if n is even, we find

$$F_{11}(x_n,x_{n+1}) = F_{22}(t_{2n-1},n;x_n,x_{n+1}), \tag{27a}$$

$$F_{12}(x_n,x_{n+1}) = F_{21}(t_{2n-1},n;x_n,x_{n+1})\exp(-2i\Theta_{n-1}), \tag{27b}$$

$$F_{21}(x_n,x_{n+1}) = F_{12}(t_{2n-1},n;x_n,x_{n+1})\exp(2i\Theta_{n-1}), \tag{27c}$$

$$F_{22}(x_n,x_{n+1}) = F_{11}(t_{2n-1},n;x_n,x_{n+1}). \tag{27d}$$

7. THE P MATRIX

In the present section we shall introduce a matrix $\mathbf{P}(x_n,x_{n+1})$, for $n = 1,2,...,N$, the elements of which are simple functions of the quantities A_n and B_n and (if $n \geqslant 2$) also of the linear combination $L_{n-1} - \sigma_{n-1} - \sigma_n - \tau_{n-1} + \tau_n$. For $n \geqslant 2$, the matrix $\mathbf{P}(x_n,x_{n+1})$ is thus associated with the barriers $B(n)$ and $B(n-1)$ and the potential well between these barriers, but for $n = 1$ it is associated only with the barrier $B(1)$.

The transmission coefficient, besides being a function of the quantities A_n and B_n, $n = 1, 2, ..., N$, is further a function of the quantities L_n, σ_n, and τ_n, but in fact only through the linear combinations $L_n - \sigma_n - \sigma_{n+1} - \tau_n + \tau_{n+1}$, where $n = 1,2,...,N-1$. It will later be shown, in Eqs. (43), that the transmission coefficient for a system of N barriers can be conveniently expressed in terms of the matrices $\mathbf{P}(x_n,x_{n+1})$, $n = 1,2,...,N$, which are now defined by

$$\mathbf{P}(x_n,x_{n+1}) = \begin{pmatrix} B_n & A_n \\ A_n \exp(2i\beta_{n-1}) & B_n \exp(2i\beta_{n-1}) \end{pmatrix}, \quad n = 1,2,...,N, \tag{28}$$

where

$$\beta_0 = 0, \tag{29a}$$

$$\beta_n = (-1)^n\alpha_n = (-1)^n[L_n - (\sigma_n + \sigma_{n+1}) - (\tau_n - \tau_{n+1})] \quad \text{for } n = 1,2,...,N-1. \tag{29b}$$

The quantity α_1 is identical with α defined by (27) in Ref. 27.

From (28) we obtain the symmetry relations

$$P_{11}(x_n,x_{n+1}) = P_{22}^{*}(x_n,x_{n+1})\exp(2i\beta_{n-1}), \tag{30a}$$

$$P_{12}(x_n,x_{n+1}) = P_{21}^*(x_n,x_{n+1})\exp(2i\beta_{n-1}), \tag{30b}$$

and from (20) and (28) we see that

$$\det \mathbf{P}(x_n,x_{n+1}) = -\exp(-2K_n + 2i\beta_{n-1}) \tag{31}$$

and

$$|P_{12}(x_n,x_{n+1})|^2 - |P_{11}(x_n,x_{n+1})|^2 = \exp(-2K_n). \tag{32}$$

We also introduce a matrix $\mathbf{P}(x_1,x_{n+1})$, associated with all the barriers $B(1)$, $B(2)$, ..., $B(n)$, by the definition

$$\mathbf{P}(x_1,x_{n+1}) = \mathbf{P}(x_1,x_2)\mathbf{P}(x_2,x_3)\cdots\mathbf{P}(x_n,x_{n+1})$$
$$\text{for } n = 1,2,\ldots,N. \tag{33}$$

From (33) and (31) it follows that

$$\det \mathbf{P}(x_1,x_{n+1}) = \det \mathbf{P}(x_1,x_2)\det \mathbf{P}(x_2,x_3)\cdots\det \mathbf{P}(x_n,x_{n+1})$$
$$= (-1)^n\exp\left(-2\sum_{\nu=1}^{n}K_\nu + 2i\sum_{\nu=1}^{n-1}\beta_\nu\right). \tag{34}$$

Using repeatedly the symmetry relations (30) in (33), we find the following symmetry relations for the matrix $\mathbf{P}(x_1,x_{n+1})$:

$$P_{11}(x_1,x_{n+1}) = P_{22}^*(x_1,x_{n+1})\exp\left(2i\sum_{\nu=1}^{n-1}\beta_\nu\right), \tag{35a}$$

$$P_{12}(x_1,x_{n+1}) = P_{21}^*(x_1,x_{n+1})\exp\left(2i\sum_{\nu=1}^{n-1}\beta_\nu\right). \tag{35b}$$

From (34) and (35) it follows that

$$|P_{12}(x_1,x_{n+1})|^2 - |P_{11}(x_1,x_{n+1})|^2 = (-1)^{n+1}\exp\left(-2\sum_{\nu=1}^{n}K_\nu\right) \tag{36}$$

for $n = 1,2,\ldots,N$.

8. THE TRANSMISSION COEFFICIENT IN TERMS OF THE MATRIX $\mathbf{P}(x_1,x_{N+1})$

From (19) and (28) we obtain, for $n = 1,2,\ldots,N$,

$$F_{11}(t_{2n-1},n;x_n,x_{n+1}) = -P_{11}(x_n,x_{n+1})\times\exp[2K_n - i2\tau_n], \tag{37a}$$

$$F_{12}(t_{2n-1},n;x_n,x_{n+1}) = P_{12}(x_n,x_{n+1})\times\exp[i(2\sigma_n + \tfrac{1}{2}\pi)], \tag{37b}$$

$$F_{21}(t_{2n-1},n;x_n,x_{n+1}) = -P_{21}(x_n,x_{n+1})\times\exp[2K_n - i(2\sigma_n + 2\beta_{n-1} + \tfrac{1}{2}\pi)], \tag{37c}$$

$$F_{22}(t_{2n-1},n;x_n,x_{n+1}) = P_{22}(x_n,x_{n+1})\times\exp[i(2\tau_n - 2\beta_{n-1})]. \tag{37d}$$

Equations (26) and (37) yield

$$F_{11}(x_n,x_{n+1}) = P_{11}(x_n,x_{n+1})\exp[2K_n + i(\pi - 2\tau_n)], \tag{38a}$$

$$F_{12}(x_n,x_{n+1}) = P_{12}(x_n,x_{n+1})\times\exp[i(\tfrac{1}{2}\pi - 2\Theta_{n-1} + 2\sigma_n)], \tag{38b}$$

$$F_{21}(x_n,x_{n+1}) = P_{21}(x_n,x_{n+1})\times\exp[2K_n + i(\tfrac{1}{2}\pi + 2\Theta_{n-1} - 2\beta_{n-1} - 2\sigma_n)], \tag{38c}$$

$$F_{22}(x_n,x_{n+1}) = P_{22}(x_n,x_{n+1})\exp[i(-2\beta_{n-1} + 2\tau_n)], \tag{38d}$$

which are valid when n is an odd integer. Equations (27), (28), and (37) give

$$F_{11}(x_n,x_{n+1}) = P_{11}(x_n,x_{n+1})\exp[i2\tau_n], \tag{39a}$$

$$F_{12}(x_n,x_{n+1}) = P_{12}(x_n,x_{n+1})\times\exp[2K_n + i(\tfrac{1}{2}\pi - 2\Theta_{n-1} - 2\sigma_n)], \tag{39b}$$

$$F_{21}(x_n,x_{n+1}) = P_{21}(x_n,x_{n+1})\times\exp[i(\tfrac{1}{2}\pi + 2\Theta_{n-1} - 2\beta_{n-1} + 2\sigma_n)], \tag{39c}$$

$$F_{22}(x_n,x_{n+1}) = P_{22}(x_n,x_{n+1})\times\exp[2K_n + i(\pi - 2\beta_{n-1} - 2\tau_n)], \tag{39d}$$

valid when n is an even integer.

With the aid of (33), (38), and (39), and the formula

$$\mathbf{F}(x_1,x_{n+1}) = \mathbf{F}(x_1,x_2)\mathbf{F}(x_2,x_3)\cdots\mathbf{F}(x_n,x_{n+1}), \tag{40}$$

which follows from (3.27) in Ref. 11, it can be proved by complete induction in a straightforward way that

$$F_{11}(x_1,x_{2n}) = P_{11}(x_1,x_{2n})\exp\left[2\sum_{\nu=1}^{n}K_{2\nu-1} + i\left(n\pi + 2\sum_{\nu=1}^{2n-1}(-1)^\nu\tau_\nu\right)\right], \tag{41a}$$

$$F_{12}(x_1,x_{2n}) = P_{12}(x_1,x_{2n})\exp\left[2\sum_{\nu=1}^{n-1}K_{2\nu} + i\left((n-\tfrac{1}{2})\pi + 2\sigma_{2n-1} + 2\sum_{\nu=1}^{2n-2}(-1)^\nu(\tau_\nu - L_\nu)\right)\right], \tag{41b}$$

and

$$F_{11}(x_1,x_{2n+1}) = P_{11}(x_1,x_{2n+1})\exp\left[2\sum_{\nu=1}^{n}K_{2\nu-1} + i\left(n\pi + 2\sum_{\nu=1}^{2n}(-1)^\nu\tau_\nu\right)\right], \tag{42a}$$

$$F_{12}(x_1,x_{2n+1}) = P_{12}(x_1,x_{2n+1})\exp\left[2\sum_{\nu=1}^{n}K_{2\nu} + i\left((n+\tfrac{1}{2})\pi - 2\sigma_{2n} + 2\sum_{\nu=1}^{2n-1}(-1)^\nu(\tau_\nu - L_\nu)\right)\right], \tag{42b}$$

for $n = 1,2,3,\ldots$.

Inserting finally (41b) into (13a) and (42a) into (13b), we obtain with the aid of (36) the following exact expressions for the transmission coefficient pertaining to a system of N potential barriers:

$$\frac{1}{T_{\text{odd}}} = |P_{12}(-\infty,+\infty)|^2\exp\left(2\sum_{\nu=1}^{N}K_\nu\right) = 1 + |P_{11}(-\infty,+\infty)|^2\exp\left(2\sum_{\nu=1}^{N}K_\nu\right) = \frac{|P_{12}(-\infty,+\infty)|^2}{|P_{12}(-\infty,+\infty)|^2 - |P_{11}(-\infty,+\infty)|^2}, \quad N \text{ odd}, \tag{43a}$$

$$\frac{1}{T_{\text{even}}} = |P_{11}(-\infty,+\infty)|^2 \exp\left(2\sum_{\nu=1}^{N} K_\nu\right)$$
$$= 1 + |P_{12}(-\infty,+\infty)|^2 \exp\left(2\sum_{\nu=1}^{N} K_\nu\right)$$
$$= \frac{|P_{11}(-\infty,+\infty)|^2}{|P_{11}(-\infty,+\infty)|^2 - |P_{12}(-\infty,+\infty)|^2}, \quad N \text{ even}, \tag{43b}$$

For the reflection coefficient R we obtain, since $R = 1 - T$,

$$R_{\text{odd}} = \left| \frac{P_{11}(-\infty,+\infty)}{P_{12}(-\infty,+\infty)} \right|^2, \quad N \text{ odd}, \tag{44a}$$

$$R_{\text{even}} = \left| \frac{|P_{12}(-\infty,+\infty)}{P_{11}(-\infty,+\infty)} \right|^2, \quad N \text{ even}. \tag{44b}$$

The elements of the matrix $\mathbf{P}(-\infty,+\infty)$ appearing in (43) and (44) are the limiting values of the elements of the matrix $\mathbf{P}(x_1,x_{N+1})$ which are obtained by letting $x_1 \to -\infty$ and $x_{N+1} \to +\infty$. For the definitions of the $\mathbf{P}$ matrices, we refer to (33), (28), (29) and also to (7) and (18). By substituting for A_n, B_n, σ_n, and τ_n the phase-integral expressions given in Appendix B (omitting correction terms) we can immediately convert (43) and (44) into useful approximate formulae of high accuracy.

Formulae (43) and (44) form the starting point for some further investigations.[8–10] In Ref. 8 a necessary condition for complete transparency is derived and a maximum transmission problem of a general kind is solved. In Ref. 9 a system of N identical barriers is considered. Concise formulae for T are obtained and, also, equations determining the resonance energies which give T equal to unity. In Ref. 10 an equation for calculating the bound state energies of a general potential well containing N humps of arbitrary shapes is derived from a condition for total transmission through a system of potential barriers.

ACKNOWLEDGMENTS

I wish to thank Professor N. Fröman and Professor P. O. Fröman for suggesting the multibarrier transmission problem, for introducing me to the phase-integral method, and for numerous stimulating discussions and constructive criticism of the manuscript.

APPENDIX A: SOME KEY FACTS ABOUT THE PHASE-INTEGRAL METHOD

We consider the one-dimensional Schrödinger equation

$$\frac{d^2\psi}{dz^2} + Q^2(z)\psi = 0 \tag{A1}$$

with

$$Q^2(z) = 2m[E - V(z)]/\hbar^2, \tag{A2}$$

z being a complex variable (the real values of which will be denoted by x) and E being the energy of a particle with mass m moving in the potential field $V(x)$. We assume that $Q^2(z)$ is real on the real axis, i.e., $V(x)$ real. Furthermore, we assume that $Q^2(z)$ is an analytical function of z in the complex plane.

We write the solution of Eq. (A1) as follows:

$$\psi(z) = a_1(z) f_1(z) + a_2(z) f_2(z), \tag{A3}$$

where $f_1(z)$ and $f_2(z)$, the phase-integral functions, are defined by

$$f_1(z) = q^{-1/2}(z)\exp[iw(z)], \tag{A4a}$$

$$f_2(z) = q^{-1/2}(z)\exp[-iw(z)], \tag{A4b}$$

with

$$w(z) = \int^z q(z)\,dz, \tag{A5}$$

the function $q(z)$ being a so-far unspecified function of z, the choice of which will be described further on, and the lower limit of integration for $w(z)$ being any arbitrary point where the integral for $w(z)$ converges [cf. Eqs. (3.1), (3.25a), (3.24a), (3.24b), and (3.3) in Ref. 11].

By imposing on $a_1(z)$ and $a_2(z)$ the condition

$$a_1'(z) f_1(z) + a_2'(z) f_2(z) = 0, \tag{A6}$$

we can write the derivative of the solution (A3) as follows:

$$\psi'(z) = a_1(z) f_1'(z) + a_2(z) f_2'(z). \tag{A7}$$

This expression is formally obtained from (A3) by treating a_1 and a_2 as constants in the differentiation [cf. (3.10) and (3.25b) in Ref. 11].

Introducing the row vector $\mathbf{f}(z)$

$$\mathbf{f}(z) = (f_1(z), f_2(z)) \tag{A8}$$

and the column vector $\mathbf{a}(z)$

$$\mathbf{a}(z) = \begin{pmatrix} a_1(z) \\ a_2(z) \end{pmatrix}, \tag{A9}$$

we can write (A3) and (A7) in the form

$$\psi(z) = \mathbf{f}(z)\mathbf{a}(z), \tag{A10a}$$

$$\psi'(z) = \mathbf{f}'(z)\mathbf{a}(z). \tag{A10b}$$

By specifying the values of $a_1(z)$ and $a_2(z)$ at some point z_0, we specify a unique solution $\psi(z)$ of Eq. (A1), since, according to Eqs. (A10), the values of $\psi(z_0)$ and $\psi'(z_0)$ thereby become determined. We are able to trace this solution $\psi(z)$ from the point z_0 to an arbitrary point z by using the formula [cf. (3.23), (3.12), and (3.26) in Ref. 11]

$$\mathbf{a}(z) = \mathbf{F}(z,z_0)\mathbf{a}(z_0), \tag{A11}$$

where $\mathbf{F}(z,z_0)$ is a 2×2 matrix, the elements of which are given by the convergent series (3.22a)–(3.22d) in Ref. 11. Thus, formulae (A3) and (A11) together give a representation of the solution $\psi(z)$.

According to (3.27) and (3.19)–(3.20) in Ref. 11, the matrix $\mathbf{F}(z,z_0)$ has the following properties:

$$\mathbf{F}(z,z_0) = \mathbf{F}(z,z_1)\mathbf{F}(z_1,z_0), \tag{A12}$$

$$\det \mathbf{F}(z,z_0) = 1, \tag{A13}$$

$$\mathbf{F}(z,z_0) = [\mathbf{F}(z_0,z)]^{-1}, \tag{A14}$$

where z_1 is an arbitrary point. The last equation can also be written

$$\begin{pmatrix} F_{11}(z,z_0) & F_{12}(z,z_0) \\ F_{21}(z,z_0) & F_{22}(z,z_0) \end{pmatrix} = \begin{pmatrix} F_{22}(z_0,z) & -F_{12}(z_0,z) \\ -F_{21}(z_0,z) & F_{11}(z_0,z) \end{pmatrix}. \tag{A15}$$

Useful estimates of the elements of $\mathbf{F}(z,z_0)$ have been derived in Ref. 11 on the assumption that z_0 and z can be connected by a path Λ in the complex plane a long which $|\exp[iw(z)]|$ increases monotonically from z_0 to z. These estimates, given by (4.3a)–(4.3d) in Ref. 11, can be written as follows [cf. (3.7a)–(3.7d) in the article by N. Fröman in Ref. 30]

$$|F_{11}(z,z_0) - 1| \leqslant \tfrac{1}{2}\mu + \cdots, \tag{A16a}$$

$$|F_{12}(z,z_0)| \leqslant |\exp[-2iw(z_0)]|(\tfrac{1}{2}\mu + \cdots), \tag{A16b}$$

$$|F_{21}(z,z_0)| \leqslant |\exp[2iw(z_0)]|(\tfrac{1}{2}\mu + \cdots), \tag{A16c}$$

$$|F_{22}(z,z_0) - 1| \leqslant \tfrac{1}{2}\mu + |\exp\{2i[w(z) - w(z_0)]\}|(\tfrac{1}{4}\mu^2 + \cdots). \tag{A16d}$$

The quantity μ appearing in Eqs. (A16) is defined by

$$\mu = \mu(z,z_0) = \int_{z_0}^{z} \left| \left(\frac{Q^2 - q^2}{q} + q^{-1/2} \frac{d^2}{dz^2} q^{-1/2} \right) dz \right| \tag{A17}$$

with the integration performed along the path of monotonicity [cf. (4.2), (3.3), and (3.5a) in Ref. 11]. Introducing the quantity $\epsilon(z)$ by the definition

$$\epsilon(z) = \frac{Q^2 - q^2}{q^2} + q^{-3/2} \frac{d^2}{dz^2} q^{-1/2}, \tag{A18}$$

we can write

$$\mu(z,z_0) = \int_{z_0}^{z} |\epsilon(z) q(z)\, dz|. \tag{A19}$$

By means of the relations

$$q^{1/2} \frac{d^2}{dz^2} q^{-1/2} = \frac{2n+1}{4n^2} \left(q^{-n} \frac{d}{dz} q^n \right)^2 - \frac{1}{2n} \left(q^{-n} \frac{d^2}{dz^2} q^n \right) \tag{A20a}$$

and

$$q^{-3/2} \frac{d^2}{dz^2} q^{-1/2} = -\left(\frac{1}{2n} q^{-n-1} \frac{d}{dz} q^n \right)^2 - \frac{d}{dw} \left(\frac{1}{2n} q^{-n-1} \frac{d}{dz} q^n \right), \tag{A20b}$$

which hold for n being an arbitrary real number ($\neq 0$), we can write $\epsilon(z)$ in different useful ways. The successive choices $n = -\frac{1}{2}, 1, -1, 2$ in (A20a) yield the expressions (3.5a), (3.5b), (3.5c), and (3.5d) in Ref. 11.

We recognize that in order that the estimates (A16) will be useful, we must choose the function $q(z)$ in such a way that $\mu(z,z_0)$ becomes small compared to unity [see (A17)]. In the extreme case that $q(z)$ is chosen such that $\epsilon(z)$ is everywhere equal to zero, implying that $\mu(z,z_0) = 0$, we find from (A16) and the multiplication rule (A12) that the matrix $\mathbf{F}(z,z_0)$ is equal to the unit matrix, not only for two points which can be joined by a path of monotonicity, but in fact for arbitrary points z_0 and z. From (A3) and (A11) we conclude that $f_1(z)$ and $f_2(z)$ are in this case two linearly independent exact solutions of the Schrödinger equation (A1).

General properties of the function $Q^2(z)$ in Eq. (A1) are reflected in properties of the matrix $\mathbf{F}$. Thus, if the functions $Q^2(z)$ and $q^2(z)$ are real on the real axis, there exist certain relations between the elements of the matrix $\mathbf{F}(x_1,x_2)$ for points x_1 and x_2 lying on the real axis. These symmetry relations are given by Eqs. (5.7a), (5.7b), (5.8a)–(5.8d), (5.9a), and (5.9b) in Ref. 11. Similarily, in the special case that $Q^2(z) = Q^2(-z)$ and $q^2(z) = q^2(-z)$, or if $Q^2(z)$ and $q^2(z)$ are periodic with respect to z, there exist other symmetry relations between the elements of $\mathbf{F}$ [see, e.g., Eqs. (A.10) and (A.11) in Appendix A in Ref. 18.]

When applying the phase-integral method to physical problems, the general procedure is to derive an exact expression for the physical quantity or physical relation of interest and then omit "small" quantities, expressed in terms of elements of the matrix $\mathbf{F}$. In general situations when the points z_0 and z cannot be joined by a path along which $|\exp[iw(z)]|$ is monotonic, one has to divide the path chosen into monotonic parts, use the relations (A12)–(A15), and possibly other relations existing, to bring the expression for the quantity of interest into a convenient form, and then use the basic estimates (A16) for obtaining upper bounds for the terms neglected (cf. p. 10 in Ref. 30). It is sometimes preferable not to resolve the exact expression into its smallest components, i.e., $\mathbf{F}$-matrix elements for the separate monotonic parts, but rather to stop at an intermediate stage, retaining certain $\mathbf{F}(z_1,z_2)$-matrix elements also for some points z_1 and z_2 which cannot be joined by a monotonic path. The needed estimates of these $\mathbf{F}$-matrix elements can be derived from the basic estimates (A16) for the separate monotonic parts by means of the multiplication rule (A12) together with other relations existing. They can suitably be written in the form of exact expressions for those matrix elements [see, e.g., Eqs. (B1)–(B5) in Appendix B], expressions which contain a main part plus a "small" additional part [estimated in terms of the quantity $\mu(z_1,z_2)$ evaluated for the path chosen between z_1 and z_2], which is to be omitted in the final step, when these expressions are substituted for the proper $\mathbf{F}$-matrix elements in the exact expression for the physical quantity of interest.

Let us now describe how $q(z)$ is specified in the phase-integral method. Writing $\psi(z)$ in the form

$$\psi(z) = q^{-1/2}(z) \exp\left[\pm i \int^{z} q(z)\, dz \right], \tag{A21}$$

and inserting this expression into Eq. (A1), we obtain for $q(z)$ the differential equation [cf. (3.6) in Ref. 11]

$$Q^2(z) - q^2(z) + q^{+1/2}(z) \frac{d^2}{dz^2} q^{-1/2}(z) = 0, \tag{A22}$$

which is rigorously equivalent to the initial Schrödinger equation. Inserting, into (A22), the expression

$$q(z) = Q_{\mathrm{mod}}(z) g(z), \tag{A23}$$

where $Q_{\mathrm{mod}}(z)$ is a so-far unspecified function, we obtain a differential equation for the function $g(z)$ instead, which can be written as follows:

$$1 + \epsilon_0 - g^2(z) + g^{+1/2}(z) \frac{d^2}{d\zeta^2} g^{-1/2}(z) = 0, \tag{A24}$$

where

$$\epsilon_0 = \frac{Q^2(z) - Q^2_{\mathrm{mod}}(z)}{Q^2_{\mathrm{mod}}(z)} + Q^{-3/2}_{\mathrm{mod}}(z) \frac{d^2}{dz^2} Q^{-1/2}_{\mathrm{mod}}(z) \tag{A25}$$

and

$$\zeta = \int^z Q_{\mathrm{mod}}(z)\,dz \tag{A26}$$

(cf. pp. 105–6 in Ref. 14). By means of the relations (A20), used for the special case $q = Q_{\mathrm{mod}}$, we can write ϵ_0 in different useful ways. The choice $n = 2$ in (A20b) gives expression (11″) in Ref. 15.

Substituting (A23) into (A18), we obtain the following relation between ϵ_0 and ϵ:

$$\epsilon = \frac{(1+\epsilon_0) - g^2}{g^2} + g^{-3/2}\frac{d^2}{d\zeta^2}g^{-1/2} \tag{A27}$$

[cf. (A18) and (A25)]. We note that if we could determine a function $Q^2_{\mathrm{mod}}(z)$ which makes ϵ_0, given by (A25), identically equal to zero, Eq. (A24) would have the solution $g(z) = 1$, which according to (A23) means that $q(z) = Q_{\mathrm{mod}}(z)$ would be a solution of Eq. (A22). Hence, on substitution of Q_{mod} for $q(z)$, formula (A21) would yield two linearly independent exact solutions of the Schrödinger equation (A1); cf. the discussion below (A20).

Suppose that somehow we have found a function $Q^2_{\mathrm{mod}}(z)$ which makes ϵ_0 [(A25)] small compared to unity (even though not exactly equal to zero) in a region of the complex plane. We shall then seek to find a solution $g(z)$ of Eq. (A24) which deviates only very little from unity, by trying a series expansion for $g(z)$ (in powers of a small parameter), the first and dominant term of which is equal to unity. By truncating this series, to be determined below, after some fixed number of terms and then substituting the truncated series for $g(z)$ in formula (A27), we obtain an expression for $\epsilon(z)$ which can be expected to be smaller in magnitude, the smaller we have made ϵ_0 [(A25)] by means of a clever choice of $Q^2_{\mathrm{mod}}(z)$. As is seen from (A19), a small ϵ is of great importance for obtaining a small value of μ, and a small μ is what guarantees the goodness of our final approximate formulae, which are obtained from the exact formulae by omission of "small" parts, for which upper bounds are given in terms of μ.

Let us now obtain the formal series solution for $g(z)$; cf. Ref. 14. We note that the function $Q^2(z)$, given by (A2), is inversely proportional to the small constant $\hbar^2$. This in general makes $Q^2(z)$ large at points in the complex plane lying well away from any zero of $Q^2(z)$. In order to take the largeness of $Q^2(z)$ explicitly into account, we introduce a small parameter p^2, and consider $Q^2(z)$ to be inversely proportional to p^2. We shall represent this way of looking at $Q^2(z)$ by writing $Q^2(z)/\lambda^2$ instead of only $Q^2(z)$. The notation is meant to indicate that we consider $Q^2(z)$, in itself, to be inversely proportional to p^2. The symbol λ^2, attached to $Q^2(z)$, is not supposed to add anything to the value of the expression, but should be regarded merely as a sign, carrying information about $Q^2(z)$. For convenience, we shall sometimes permit ourselves to say that we represent the dependence of $Q^2(z)$ on p^2 by means of $Q^2(z)/\lambda^2$, or that we represent the largeness of $Q^2(z)$ by means of $Q^2(z)/\lambda^2$, or even that we represent $Q^2(z)$ by $Q^2(z)/\lambda^2$, when we mean to say, precisely, that we shall consider $Q^2(z)$ inversely proportional to p^2. Having utilized λ for obtaining the series expansion for $g(z)$, we shall then delete λ from all expressions, which we technically achieve by putting $\lambda = 1$.

It is not necessary to let the value of p^2 be equal to the value of $\hbar^2$ although this is a possible specification. In any case, we shall not regard p^2 as firmly tied to the actual occurrence of the symbol $\hbar^2$. The advantage of using a parameter, not tied to $\hbar^2$, becomes apparent when we want to treat cases where $Q^2(z)$ is not given by expression (A2) but instead by a sum of two parts, where one part is inversely proportional to $\hbar^2$ while the other does not contain $\hbar$ at all, as in the radial Schrödinger equation

$$\frac{d^2\chi}{dr^2} + Q^2(r)\chi = 0, \tag{A28}$$

where (with obvious notations) $Q^2(r)$ is given by

$$Q^2(r) = 2m[E - V(r)]/\hbar^2 - l(l+1)/r^2 \tag{A29}$$

[cf. Eqs. (11.1) and (11.2) in Ref. 11]. Also the largeness of the second term in (A29) can be taken into account by means of the parameter p^2, and indeed in different ways, which may be more or less favorable within different ranges of values for r and l. For instance, if the expression $[l(l+1)/r^2]\lambda^0$ satisfactorily serves to represent the second term in (A29) for a certain range of r values and for small values of l, it seems reasonable to expect that the expression $[l(l+1)/r^2]/\lambda^2$ will be more suitable when we (keeping to the same range of r values) turn to consider values of l which are sufficiently large for us to regard the second term in (A29) as being on a level with the first term, which is represented by $[2m[E - V(r)]/\hbar^2]/\lambda^2$ in both cases.

The following rules are now introduced. The function $Q^2_{\mathrm{mod}}(z)$ shall always be considered inversely proportional to p^2, whereas the difference $Q^2(z) - Q^2_{\mathrm{mod}}(z)$ shall be considered proportional to p^0. Hence, when using the technique to exhibit the assumed dependence on p^2, we shall replace $Q_{\mathrm{mod}}(z)$ by $Q_{\mathrm{mod}}(z)/\lambda$ and $Q^2(z) - Q^2_{\mathrm{mod}}(z)$ by $[Q^2(z) - Q^2_{\mathrm{mod}}(z)]\lambda^0$. According to (A26) and (A25) this leads to the replacements of ζ by ζ/λ and of ϵ_0 by $\epsilon_0\lambda^2$. With these replacements, Eq. (A24) becomes

$$1 + \epsilon_0\lambda^2 - g^2(z) + g^{1/2}(z)\frac{d^2}{d\zeta^2}g^{-1/2}(z)\lambda^2 = 0. \tag{A30}$$

Inserting the formal expansion

$$g(z) = \sum_{n=0}^{\infty} Y_{2n}\lambda^{2n} \tag{A31}$$

into (A30), and putting the coefficients of successive powers of λ^2 equal to zero, we obtain $Y_0 = \pm 1$ and a recursion formula for obtaining the functions Y_{2n} for $n \geqslant 1$ in terms of ϵ_0 and derivatives of ϵ_0 with respect to ζ. This formula can be written as follows [cf. Eq. (7) in Ref. 14]

$$\begin{aligned} Y_{2n} = {} & \frac{1}{2}\sum_{\substack{\alpha+\beta=n \\ 0\leqslant\alpha,\beta,\gamma,\delta\leqslant n-1}} Y_{2\alpha}Y_{2\beta} \\ & - \frac{1}{2}\sum_{\substack{\alpha+\beta+\gamma+\delta=n \\ 0\leqslant\alpha,\beta,\gamma,\delta\leqslant n-1}} Y_{2\alpha}Y_{2\beta}Y_{2\gamma}Y_{2\delta} \\ & + \frac{1}{2}\sum_{\substack{\alpha+\beta=n-1 \\ 0\leqslant\alpha,\beta\leqslant n-1}}\left[\epsilon_0 Y_{2\alpha}Y_{2\beta} + \frac{3}{4}\frac{dY_{2\alpha}}{d\zeta}\frac{dY_{2\beta}}{d\zeta}\right. \\ & \left. - \frac{1}{4}\left(Y_{2\alpha}\frac{d^2Y_{2\beta}}{d\zeta^2} + \frac{d^2Y_{2\alpha}}{d\zeta^2}Y_{2\beta}\right)\right], \quad n \geqslant 1. \end{aligned} \tag{A32}$$

Since we have already the double signs $\pm$ in the expo-

nent of (A21), we see that it is no restriction to choose $Y_0 = +1$. Starting from $Y_0 = +1$ we can successively determine $Y_2, Y_4, \cdots$ by means of (A32). Explicit expressions for Y_{2n} for $n \leqslant 10$ can be found in Ref. 32 (cf. also Ref. 26 for $n \leqslant 4$). The first few of Y_{2n} are

$$Y_0 = 1, \tag{A33a}$$

$$Y_2 = \tfrac{1}{2}\epsilon_0, \tag{A33b}$$

$$Y_4 = -\frac{1}{8}\left(\epsilon_0^2 + \frac{d^2\epsilon_0}{d\zeta^2}\right). \tag{A33c}$$

Having utilized λ for deriving the series expansion for $g(z)$ given by (A31) and (A32), we now put $\lambda = 1$ in the formulae above. Inserting (A31) into (A23) and truncating the series after $n = M$, we obtain

$$q(z) = Q_{\mathrm{mod}}(z)\sum_{n=0}^{M} Y_{2n}, \quad M = 0,1,2,\cdots, \tag{A34}$$

which is the phase-integral expression of order $2M + 1$ for the function $q(z)$. (We use the letter M so as to avoid possible confusion with N, equal to the number of barriers in the potential function considered in the present paper). Inserting this expression into (A4a) and (A4b), we get

$$f_1(z) = q^{-1/2}(z)\exp\left[+i\int^z q(z)\,dz\right], \tag{A35a}$$

$$f_2(z) = q^{-1/2}(z)\exp\left[-i\int^z q(z)\,dz\right], \tag{A35b}$$

with $q(z)$ given by (A34) and the functions Y_{2n} given by (A32) with $Y_0 = 1$. Being approximate solutions of the Schrödinger equation, the functions (A35a) and (A35b) are called phase-integral approximations of order $2M + 1$.

The function $Q_{\mathrm{mod}}(z)$ can be chosen conveniently to suit the particular problem to be treated. For the special choice $Q^2_{\mathrm{mod}}(z) = Q^2(z)$, formula (A34) turns to

$$q(z) = Q(z)\sum_{n=0}^{M} Y_{2n}, \quad M = 0,1,2,\cdots, \tag{A36}$$

which is the unmodified phase-integral expression of order $2M + 1$ for the function $q(z)$. Formulae (A35) yield in this special case so-called unmodified phase-integral approximations of order $2M + 1$, which for $M = 0$ are identical with the first-order JWKB approximations. The choice $Q^2_{\mathrm{mod}} = Q^2$ is satisfactory for a wide class of physical problems. However, in cases where the unmodified approximations break down at certain singular points of $Q^2(z)$, we have the possibility of finding instead, by means of a judicious choice of $Q^2_{\mathrm{mod}}(z)$, modified phase-integral approximations which are good also at the above-mentioned critical points. The way how to choose $Q^2_{\mathrm{mod}}(z)$ successfully is discussed on pp. 106–7 in Ref. 14 and on pp. 129–31 in Ref. 15.

In the present paper we specify the function $Q^2_{\mathrm{mod}}(z)$ to be equal to $Q^2(z)$, which means that we are using the unmodified phase-integral expressions (A36) for $q(z)$. This choice will be presupposed in the following discussion in Appendix A.

For the definition of $w(z)$ by (A5), one uses a Riemann surface consisting of two superposed sheets which are appropriately joined together along suitable cuts emerging from zeros and poles of $Q^2(z)$, drawn such that $Q(z)$ is single-valued on the surface. Examples of such cuts are shown in Figs. 1–3, where they are indicated by heavy lines. On the introduced Riemann surface, the function $q(z) = Q(z)\Sigma_{n=0}^{M} Y_{2n}$ is in fact also single-valued since ϵ_0 and all the functions $Y_{2n}(z)$ are single-valued functions of z (cf. p. 544 in Ref. 12).

When using the first-order phase-integral expression for $q(z)$ [i.e., $q(z) = Q(z)$], we can choose a simple zero of $Q^2(z)$ as lower limit in the integral defining $w(z)$. Unfortunately, this is not possible when a higher-order expression (A36) for $q(z)$ is used, since in that case $q(z)$ is infinite at the zero, the function $\Sigma_{n=0}^{M} Y_{2n}(z)$ having at that point a pole of order $3M$ (see p. 456 in Ref. 13). However, by a convenient choice of the lower limit of integration in (A5) it is possible to express the function $w(z)$ by means of certain contours of integration on the Riemann surface avoiding the zeros of $Q^2(z)$; cf. pp. 459–60 in Ref. 13. This mode of expressing $w(z)$, first introduced on pp. 545–6 in Ref. 12 and further discussed on pp. 84–5 in Ref. 18, was used in Ref. 26 (see pp. 608–11) and in Ref. 27 (see pp. 629–35), and is also adopted in the present paper.

Thus, in the present paper we have for $w(x_1)$ the formula [see (8a) and (10a)]

$$w(x_1) = \mathrm{Re}\,\frac{1}{2}\int_{\Gamma_1} q(z)\,dz, \tag{A37}$$

where x_1 is any point on the real axis to the left of the barrier system. The contour Γ_1 is shown in Figs. 1(a) and 1(b) in the two possible cases of subbarrier penetration and superbarrier transmission, respectively. Some remarks should be made in this context. From the expressions (A33) and the recursion formula (A32), it is realized that the functions $Y_{2n}(z)$ are real on the real axis. This means that $q(z)$, as given by (A36), is real in classically allowed regions on the real axis, i.e., in regions where $Q^2(x) > 0$, but purely imaginary in classically forbidden regions on the real axis, i.e., in regions where $Q^2(x) < 0$. One should further observe, that at each position along a cut, the values of $q(z)$ on opposite edges of the cut (but on the same Riemann sheet) differ only in sign. Moreover, the value of $q(z)$ at any point on the upper Riemann sheet differs only in sign from the value at the corresponding point lying on the lower sheet.

With these remarks in mind, one realizes that when the first-order unmodified expression for $q(z)$ [i.e., $q = Q$] is used, the way (A37) of expressing $w(x_1)$ corresponds to choosing the point t_1 in Fig. 1(a) as lower limit of integration in the subbarrier case. In the superbarrier case shown in Fig. 1(b), however, it corresponds instead to choosing $x_s - 0$ as lower limit of integration, where by x_s we mean the point of intersection between the real axis and the cut from t_1 to t_2 when conveniently drawn along Stokes' line [i.e., the line along which $q(z)dz$ is purely imaginary] joining these two points. To see that this is true, it is helpful to deform the contour Γ_1 in Fig. 1(b) so as to run close to the cut along Stokes' line up to the real axis and from there along this axis on to x_1.

With formula (A37) for $w(x_1)$ at hand, the problem of finding the value of $w(z)$ at an arbitrary point z is reduced to evaluating the integral (A5) from x_1 to z along a convenient path. Some reflection convinces us that the integration path

from x_1 to z may be replaced, equivalently, by contours on the Riemann surface of the kind shown in Figs. 1(a), 1(b), and 2(a)–(d). It is then easily realized that $w(z)$, in different regions of the real axis, can be expressed in terms of the contour integrals (6), (7), (8b), and (8c) in the way presented by formula (10b).

A point z', which is a simple zero of $Q^2(z)$, is not a zero of the function $q(z)$ defined by (A36), when $M>0$, but instead a point of singularity, as was remarked above. However, there exist for $q(z)$, if $M>0$, a certain number of zeros lying in the neighborhood of z', which are due to the function $\Sigma_{n=0}^{M} Y_{2n}(z)$ in (A36); cf. pp. 456–60 in Ref. 13. These zeros, which are not branch points for $q(z)$, cause no complication for the evaluation of $w(z)$ since the contours of integration may pass through the regions where the zeros are located. But these zeros are branch points for $q^{1/2}(z)$ and hence, according to (A35a) and (A35b), they are singular branch points for $f_1(z)$ and $f_2(z)$.

Therefore, when dealing with the unmodified phase-integral approximations $f_1(z)$ and $f_2(z)$, defined by (A35) and (A36), we cut the Riemann surface along the real axis and consider the function $q^{1/2}(z)$, and hence also $f_1(z)$ and $f_2(z)$, on the upper half of the first Riemann sheet from which we exclude a certain region around each zero of $Q^2(z)$ where the zeros of $q(z)$ are located, in order to get $q^{1/2}(z)$ single-valued. The shaded areas in Figs. 1–3 indicate the excluded regions on the first Riemann sheet.

We assign the same phase to the functions $q^{1/2}(z)$ and $Q^{1/2}(z)$ on the real axis at some point well to the left of the barrier system. As explained on p. 461 in Ref. 13, this means that the phase of $q^{1/2}(z)$ will be the same as the phase of $Q^{1/2}(z)$ at any other point on the real axis, outside the regions where the zeros of $q(z)$ are located. Figures 1(a) and 1(b) show the phase of $q^{1/2}(z)$ on the real axis around the leftmost barrier, as determined by the choice $q^{1/2}(x)=|q^{1/2}(x)|$ for $x<x_1$.

APPENDIX B: PHASE-INTEGRAL EXPRESSIONS FOR A_n, B_n, σ_n, AND τ_n

Phase-integral expressions (with rigorous error estimates) are given for the matrix elements $F_{22}(x_1,x_2)$ and $F_{12}(x_1,x_2)$ by formulae (43a) and (43b) in Ref. 26, and further, provided $\mu\exp[\frac{1}{2}(|K|-K)]\ll 1$, for the combinations $|F_{12}|+|F_{22}|$, $|F_{12}|-|F_{22}|$, and $|F_{12}/F_{22}|$ by formulae (53a), (53b), and (52b) in Ref. 26.

However, in the energy region where K is negative and large, i.e., for energies lying well above the top of the barrier situated between x_1 and x_2, the given condition for the validity of (53a) and (53b) in Ref. 26 can be considerably relaxed. In fact, it is shown in Ref. 31 that the conditions $\mu<0.4$ and $\mu<0.1$ are safely sufficient for the validity of (53a) and (53b), respectively, and that (43a) and (43b) in the same paper[26] are certainly valid if $\mu<0.4$.

By applying above mentioned phase-integral formulae to the matrix $\mathbf{F}(t_{2n-1},n;x_n,x_{n+1})$ defined in Sec. 5, and using the definitions (18a) and (18b) of A_n and B_n we obtain, for $n=1,2,\ldots,N$,

$$A_n=[1+\exp(-2K_n)]^{1/2}\times(1+\exp[-\tfrac{1}{2}(|K_n|-K_n)]O(\mu_n)+O(\mu_n^2)), \quad \text{(B1)}$$

$$B_n=1+\exp[\tfrac{1}{2}(|K_n|-K_n)]O(\mu_n), \quad \text{(B2)}$$

$$A_n+B_n=([1+\exp(-2K_n)]^{1/2}+1)[1+O(\mu_n)], \quad \text{(B3)}$$

$$A_n-B_n=([1+\exp(-2K_n)]^{1/2}-1)[1+O(\mu_n)], \quad \text{(B4)}$$

$$A_n/B_n=[1+\exp(-2K_n)]^{1/2}+\exp(-2K_n)O(\mu_n)=[1+\exp(-2K_n)]^{1/2}(1+\exp[-2K_n-\tfrac{1}{2}(|K_n|-K_n)]O(\mu_n)), \quad \text{(B5)}$$

where μ_n denotes the integral

$$\mu_n=\mu(x_n,x_{n+1})=\int_{x_n}^{x_{n+1}}\left|\left(\frac{Q^2-q^2}{q}+q^{-1/2}\frac{d^2}{dz^2}q^{-1/2}\right)dz\right| \quad \text{(B6)}$$

along the appropriate one of the paths of integration depicted in Figs. 6.4, 6.5, and 6.6 in Ref. 11. The quantity μ_n is assumed to be small compared to unity. The symbol $O(\mu_n)$ denotes a quantity at the most of the order of magnitude μ_n. Some comments concerning the significance of μ are given in Appendix A. The formulae (B1), (B2), and (B3) are safely valid if $\mu_n<0.4$, formula (B4) if $\mu_n<0.1$, and formula (B5) if $\mu_n\exp[\frac{1}{2}(|K_n|-K_n)]<0.1$. See Ref. 31.

The absolute value of the error term in (B2) is supposed to be less than unity for every n. Hence, from (B2) and (20) it follows that

$$A_n>B_n>0, \quad n=1,2,\ldots,N. \quad \text{(B7)}$$

From (51) in Ref. 26 we get the estimate

$$\sigma_n=O(\mu_n), \quad n=1,2,\ldots,N, \quad \text{(B8)}$$

which is valid provided $|K_n|$ is sufficiently large and the μ integral is performed along a path such as that in Fig. 6.4 of Ref. 11 when K_n is positive but along the real axis from x_n to x_{n+1} when K_n is negative. However, for energies in a region enclosing the top of the barrier $B(n)$, where $|K_n|$ is small and σ_n may be large (even infinite), the estimate (B8) is not sufficient. Being particularly important for this critical region, there exist, given by (10) together with (10a)–(10c) in Ref. 19, approximate formulae for σ_n pertaining to the phase-integral appoximation of order $2M+1$ in the cases $M=0,1$, and 2, which are valid also for energies near the top of the barrier. Furthermore, a general approximate formula for σ_n of order $2M+1$ where M can take arbitrary integer values (>0), which reproduces above-mentioned formulae as special cases, has recently been proposed[33].

From (B2) and the definitions (18b) and (18d), if follows that

$$\tau_n=\exp[\tfrac{1}{2}(|K_n|-K_n)]O(\mu_n), \quad \text{(B9)}$$

provided

$$\mu_n\exp[\tfrac{1}{2}(|K_n|-K_n)]\ll 1. \quad \text{(B10)}$$

The condition (B10) is assumed to be fulfilled for $K_n\geqslant 0$ and also for small negative values of K_n. However, by application to concrete cases, it has been found that (B10) may hold also for rather large negative values of K_n.

[1]C. B. Duke, *Tunneling in Solids* (Academic, New York and London, 1969).

[2]E. Burstein and S. Lundqvist, Eds., *Tunneling Phenomena in Solids* (Plenum, New York, 1969).
[3]B. Feuerbacher, B. Fitton, and R. F. Willis, Eds., *Photoemission and the Electronic Properties of Surfaces* (Wiley, New York, 1978).
[4]C. Rebbi, "Solitons," Sci. Am. **240** (2), 76 (1979).
[5]R. K. Bullough and P. J. Caudrey, Eds., *Solitons* (Springer, New York, 1980).
[6]S. Bjørnholm and J. E. Lynn, Rev. Mod. Phys. **52**, 725 (1980).
[7]S. Chandrasekhar, Proc. R. Soc. Lond. A **369**, 425 (1980).
[8]Ö. Dammert, "Transmission through a system of potential barriers II. Necessary condition for complete transparency. A maximum transmission problem" (to be published).
[9]Ö. Dammert, "Transmission through a system of N identical potential barriers of a general shape" (to be published).
[10]Ö. Dammert, "Bound state energies of a general potential well containing N humps of arbitrary shapes" (to be published).
[11]N. Fröman and P. O. Fröman, *JWKB Approximation, Contributions to the Theory* (North-Holland, Amsterdam, 1965), Russian transl (MIR, Moscow, 1967).
[12]N. Fröman, Ark. Fys. **32**, 541 (1966).
[13]N. Fröman, Ann. Phys. (N.Y.) **61**, 451 (1970).
[14]N. Fröman and P. O. Fröman, Ann. Phys. (N.Y.) **83**, 103 (1974).
[15]N. Fröman and P. O. Fröman, Nuovo Cimento B **20**, 121 (1974).
[16]N. Fröman, J. Phys. A: Math. Gen. **12**, 2355 (1979).
[17]Ö. Dammert and P. O. Fröman, J. Math. Phys. **21**, 1683 (1980).
[18]N. Fröman, Ark. Fys. **32**, 79 (1966).
[19]N. Fröman, P. O. Fröman, U. Myhrman, and R. Paulsson, Ann. Phys. (N.Y.) **74**, 314 (1972).
[20]N. Fröman, Phys. Rev. A **17**, 493 (1978).
[21]S. Yngve, J. Math. Phys. **13**, 324 (1972).
[22]P. O. Fröman, Ann. Phys. (N.Y.) **88**, 621 (1974).
[23]N. Fröman, Phys. Lett. A **48**, 137 (1974).
[24]N. Fröman and P. O. Fröman, J. Math. Phys. **18**, 903 (1977).
[25]N. Fröman, P. O. Fröman, and F. Karlsson, Molec. Phys. **38**, 749 (1979).
[26]N. Fröman and P. O. Fröman, Nucl. Phys. A **147**, 606 (1970).
[27]N. Fröman and Ö. Dammert, Nucl. Phys. A **147**, 627 (1970).
[28]B. Lundborg, Math. Proc. Camb. Phil. Soc. **85**, 493 (1979).
[29]L. I. Ponomarev, Lectures on quasiclassics ITF-67-53, Institute for Theoretical Physics, Acad. Sci. Ukr. SSR (Kiev, 1968).
[30]M. S. Child, Ed., *Semiclassical Methods in Molecular Scattering and Spectroscopy* (D. Reidel, Dortrecht, Holland, 1980).
[31]Ö. Dammert, Transmission through a real potential barrier treated by means of certain phase-integral approximations. II, UUITP, Report No. 27, April 1980.
[32]J. A. Campbell, J. Comp. Phys. **10**, 308 (1972).
[33]F. Karlsson (to be published).

Static conformally invariant scalar field

A. J. Accioly, A. N. Vaidya, and M. M. Som
Universidade Federal do Rio de Janeiro, Instituto de Física, Cidade Universitária-Ilha do Fundão, Rio de Janeiro, CEP. 21.944, Brazil

(Received 5 January 1983; accepted for publication 8 April 1983)

For a massless conformally invariant scalar field, a class of solutions is obtained to the Einstein equations for which the geometry of the space-time admits one timelike and two spacelike Killing vectors. The class of solutions admit plane-symmetric and conformally flat solutions as special cases. The metrics can be interpreted as static cylindrically symmetric solutions if one of the spacelike Killing vectors is associated with the rotational motion.

PACS numbers: 04.20.Jb

INTRODUCTION

The conformally invariant action integral for a system of scalar field ϕ coupled to the gravitational field is[1]

$$I=\int d^4x\sqrt{-g}\left[\frac{1}{2k}\left(1-\frac{k\phi^2}{6}\right)R+\partial_\mu\phi\partial^\mu\phi\right], \quad (1.1)$$

where R is the scalar curvature. The action integral (1.1) is invariant under Weyl's scale transformation

$$g_{\mu\nu}(x)\to\psi(x)g_{\mu\nu}(x), \quad (1.2a)$$

$$g^{\mu\nu}(x)\to\psi^{-1}(x)g^{\mu\nu}(x), \quad (1.2b)$$

$$\phi(x)\to\psi^{-1/2}(x)\phi(x). \quad (1.2c)$$

There is also an additional discrete symmetry $\phi\to-\phi$. The equations of motion are

$$(R_{\mu\nu}-\tfrac{1}{2}g_{\mu\nu}R)f(\phi)=\kappa[-\partial_\mu\phi\partial_\nu\phi+\tfrac{1}{2}g_{\mu\nu}\partial_\alpha\phi\partial^\alpha\phi]+g_{\mu\nu}\nabla^\alpha\nabla_\alpha f-\nabla_\mu\nabla_\nu f, \quad (1.3)$$

$$(1/\sqrt{-g})\partial_\mu(\sqrt{-g}\,g^{\mu\nu}\partial_\mu\phi)+\tfrac{1}{6}R\phi=0, \quad (1.4)$$

where ∇ denotes covariant derivative, and

$$f=1-\tfrac{1}{6}\kappa\phi^2. \quad (1.5)$$

The trace of (1.3) leads, with the help of (1.4), to

$$R=0. \quad (1.6)$$

Thus the energy-momentum tensor of the conformally invariant scalar field is trace-free provided there is no other material distribution. Defining

$$u=\sqrt{\kappa/6}\phi, \quad (1.7)$$

we can rewrite the field equations (1.3) and (1.4) in the forms

$$f(u)R^\mu_{\ \nu}=\delta^\mu_{\ \nu}\partial_\alpha u\partial^\alpha u-4\partial^\mu u\partial_\nu u+2u\nabla^\mu\nabla_\nu u, \quad (1.8)$$

$$(1/\sqrt{-g})\partial_\mu(\sqrt{-g}\,g^{\mu\nu}\partial_\nu\mu)=0. \quad (1.9)$$

Frøyland[2] first studied the solutions of (1.3) for the static spherically symmetric case. Among the solutions obtained is one with a conformally flat and asymptotically flat metric with total energy equal to the Schwarzschild mass. Recently, Vaidya and Som[3] treated the same problem in the static plane-symmetric case. Unlike the spherically symmetric case, their solutions are expressible in a simple form. The disklike singularity of the static plane-symmetric vacuum solution[4] disappears in some cases of their solutions.

In this paper, we treat the problem in a more general space-time. We consider the solutions of (1.8) and (1.9) in a space-time admitting a timelike and two spacelike Killing vectors. These solutions include the plane-symmetric solutions and conformally flat solutions as special cases. By choosing one of the spatial coordinates as azimuthal angle, the solutions can be interpreted as static cylindrically symmetric solutions of the conformally invariant scalar field.

II. SOLUTIONS

We consider the metrics

$$ds^2=e^{2a}(dt^2-dx^2)-e^{2b}dy^2-e^{2c}dz^2, \quad (2.1)$$

where a, b, and c depend on the variable x only. These metrics admit one timelike Killing vector with contravariant components

$$\xi_t=(1,0,0,0) \quad (2.2)$$

and two spacelike vectors defined as

$$\xi_y=(0,0,1,0) \text{ and } \xi_z(0,0,0,1). \quad (2.3)$$

Equation (1.9) for the metrics (2.1) is equivalent to

$$u_1=k_1e^{-(b+c)}, \quad (2.4)$$

where on the left-hand side the suffix 1 denotes $\partial/\partial x$ and k_1 is an integration constant. Equations (1.8) for the metric (2.1) are given by

$$a_{11}+a_1(b_1+c_1)=u_1^2/f-(f_1/f)a_1, \quad (2.5)$$

$$a_{11}+b_{11}c_{11}+b_1^2+c_1^2-a_1(b_1+c_1)=-3u_1^2/f+(f_1/f)(a_1+b_1+c_1), \quad (2.6)$$

$$b_{11}+b_1(b_1+c_1)=u_1^2/f-(f_1/f)b_1, \quad (2.7)$$

$$c_{11}+c_1(b_1+c_1)=u_1^2/f-(f_1/f)c_1. \quad (2.8)$$

Equations (2.5) – (2.8) are not all independent. They are related algebraically by Eq. (1.6). We shall consider (2.5), (2.7), and (2.8) to determine a, b, and c uniquely. We make the assumption

$$a=a(u), \quad b=b(u), \quad \text{and } c=c(u).$$

From (2.4) we have

$$u_{11}/u_1=-(b_1+c_1), \text{ i.e.,}$$

$$u_{11}=-(b'+c')u_1^2, \quad (2.9)$$

where prime denotes differentiation with respect to u. Using (2.9) one obtains, from (2.5),

$$a''=1/f-(f'/f)a' \quad (2.10)$$

Equation (2.10) yields, on integration,

$$a=\tfrac{1}{2}\mu\ln(1+u)/(1-u)-\tfrac{1}{2}\ln f+\tfrac{1}{2}\ln k_2, \quad (2.11)$$

where μ and k_2 are integration constants. Similarly one obtains, from (2.7) and (2.8),

$$b=-\tfrac{1}{2}\nu\ln(1+u)/(1-u)-\tfrac{1}{2}\ln f+\tfrac{1}{2}\ln k_3, \quad (2.12)$$

$$c=-\tfrac{1}{2}\lambda\ln(1+u)/(1-u)-\tfrac{1}{2}\ln f+\tfrac{1}{2}\ln k_4, \quad (2.13)$$

 0022-2488/83/082176-02$02.50

where k_3, k_4 are arbitrary constants of integrations. The constants of integration ν, λ, and μ are constrained through (1.6) to the relation

$$\mu\nu + \nu\lambda + \lambda\mu = 3. \tag{2.14}$$

The metrics (2.1), then, are given by

$$ds^2 = \frac{k_2}{f}\left(\frac{1-u}{1+u}\right)^{\mu}[dt^2 - dx^2] - \left[\frac{k_3}{f}\left(\frac{1-u}{1+u}\right)^{\nu} dy^2 + \frac{k_4}{f}\left(\frac{1-u}{1+u}\right)^{\lambda} dz^2\right]. \tag{2.15}$$

The expression for the scalar field u may be obtained from (2.4) and (2.15) as

$$u = k(1+u)^{1+\rho}(1-u)^{1-\rho}, \tag{2.16}$$

where

$$k = k_1/\sqrt{k_3 k_4} \text{ and } \rho = (\nu + \lambda)/2. \tag{2.17}$$

Equation (2.16) yields, on integration,

$$((1-u)/(-1+u))^{\rho} = -2\rho(k_5 + kx), \tag{2.18}$$

where k_5 is the constant of integration. k_5 can be fixed by the requirement that at $x = 0$, $u = 0$. Then from (2.18), we have

$$k_5 = 1/2\rho. \tag{2.19}$$

Now adjusting constant k suitably, we can write (2.18), using (2.19), as

$$((1-u)/(1+u))^{\rho} = 1 + \beta x. \tag{2.20}$$

When $\beta = 0$, $u = 0$, the metric (2.15) reduces to the Minkowski metric. However, another vacuum solution can be recovered by the following limiting process: Consider the case where $u < 1$. In this case one can take $\lim_{\mu \to 0} f \sim 1$. The metric takes the form, using (2.20),

$$ds^2 \to k_2(1+\lambda x)^{2\mu/\nu+\lambda}[dt^2 - dx^2] - [k_3(1+\lambda x)^{2\nu/\nu+\lambda} dy^2 + k_4(1+\lambda x)^{2\lambda/\nu+\lambda} dz^2]. \tag{2.21}$$

The metric (2.21) is identical with the Kasner metric.[5]

As the theory is invariant under $u \to -u$, the metric (2.15) is singular both at $u = \pm 1$. From (2.20), it is evident when $u = +1$, that the plane $x = -1/\beta$ is singular if any of the parameters μ, ν, or λ is less than 1. For $u = -1$, the plane $x = \infty$ is then singular.

III. CONCLUSION

The solutions (2.15) are obtained on the assumption that there exists a functional relationship between metric elements and the conformally invariant scalar field. The solution (2.15) can be interpreted as static cylindrically symmetric solutions of the Einstein equations when the source of the gravitational field is a conformally invariant scalar field if one of the coordinates y or z represents the azimuthal coordinate. One can obtain the plane-symmetric solution first obtained by Vaidya and Som simply by choosing $\nu = \lambda$. Then from (2.14), one gets $\nu(\nu + 2\mu) = 3$.

The conformally flat metric is found to be a very special case when $\nu = \mu = \lambda = 1$. The metric (2.15) then takes the very simple form:

$$ds^2 = (1+u)^{-2}[dt^2 - dx^2 - dy^2 - dz^2]. \tag{3.1}$$

It is singular at $u = -1$. It is worth noticing that for $u = \text{constant} \neq -1$, the induced three-dimensional metric is flat again. If $\nu = \mu = \lambda = -1$, then the metric is singular at $u = +1$ and flat at $u = -1$.

Another interesting feature of the solution is that one can write u explicitly as a simple function of x. From (2.14), it is evident that $\nu \neq -\lambda$. Then from (2.20), it follows that

$$u = (1 - (1+\beta x)^{1/\rho})/(1 + (1+\beta x)^{1/\rho}) \tag{3.2}$$

when $x \to \infty$, $u \to 1$, and the metric (2.15) is singular if u, ν, and λ are positive. If the parameters are all negative, then $u \to 1$ as $x \to \infty$.

Substituting (3.2) into (2.15), one finds that

$$ds^2 = [1 + (1+\beta x)^{1/\rho}]^2\{(1+\beta x)^{(\mu-1)/\rho}(dt^2 - dx^2) - (1+\beta x)^{(\nu-1)/\rho} dy^2 - (1+\beta x)^{(\lambda-1)/\rho} dz^2\}. \tag{3.3}$$

ACKNOWLEDGMENTS

The authors gratefully acknowledge financial support from CNPq and FINEP and the CEPEG of Universidade Federal do Rio de Janeiro.

[1] C. G. Callan, S. Coleman, and R. Jackiw, Ann. Phys. (NY) **59**, 42 (1970).
[2] J. Frøyland, Phys. Rev. D **25**, 1470 (1982).
[3] A. N. Vaidya and M. M. Som, Phys. Rev. D **27**, 1728 (1983).
[4] A. H. Taub, Ann. Math. **53**, 472 (1951).
[5] E. Kasner, Am. J. Math. **43**, 217 (1921).

Unification of Ernst-equation Bäcklund transformations using a modified Wahlquist–Estabrook technique[a)]

B. Kent Harrison
Department of Physics and Astronomy, Brigham Young University, Provo, Utah 84602

(Received 30 December 1982; accepted for publication 11 February 1983)

The three known Bäcklund transformations for the Ernst equation are derived using a modification of the Wahlquist–Estabrook prolongation procedure. The modification requires that the equation to be studied be cast into a set of differential forms and their exterior derivatives, such that all coefficients are constant (a "CC ideal"). Analysis of the resulting equations produces 16 solutions composed of the three basic transformations combined with identity and other essentially trivial transformations. The group structure of the transformations is discussed. A Bäcklund transformation (already known) for the Ernst–Maxwell equations can be found by the same method. Promising generalizations are mentioned.

PACS numbers: 04.20.Jb

1. INTRODUCTION

In 1978–1979 three Bäcklund transformations (BT) for the Ernst equation,[1]

$$(\mathrm{Re}\,E)\nabla^2 E = (\nabla E)^2, \tag{1}$$

which is the fundamental equation for vacuum stationary axially symmetric space times and which also occurs in nonlinear field theories,[2] were found—one by the present author[3] and two by G. Neugebauer.[4] (These references will be denoted by I and II, respectively.) The present paper shows how to derive all three BT's in a unified manner, using a modified Wahlquist–Estabrook (MWE) approach, and it is noted, as has been done before, that these BT's are elements of a group. In addition to providing this information about Ernst-equation BT's, this paper demonstrates the general use of the MWE method, which is suitable for systems of equations which can be cast into an ideal of differential forms with constant coefficients (CC ideal). The Ernst–Maxwell equations are also explored. (The term "Bäcklund transformation" is used in this paper to mean a Bäcklund transformation from solutions of an equation to solutions of the same equation, or "auto-Bäcklund transformation.")

2. FORMULATION OF THE EQUATIONS

We write the metric as in I, but with T replaced by f and Q by ω in order to conform to more common usage:

$$ds^2 = \lambda f(dx^1 + \omega\,dx^2)^2 + S^2 f^{-1}(dx^2)^2 + e^{2\gamma} f^{-1}((dx^3)^2 - \lambda\,(dx^4)^2), \tag{2}$$

where $\lambda = \pm 1$ and S, f, ω, and γ are functions of x^3 and x^4 only. $\lambda = 1$ corresponds to cylindrical waves, $\lambda = -1$ to axially symmetric fields. We write $k = \sqrt{\lambda}$ ($= 1$ or i), $x = \frac{1}{2}(x^3 + kx^4)$, and $y = \frac{1}{2}(x^3 - kx^4)$. We define a linear Hodge star operator by $*dx^3 = dx^4$, $*dx^4 = \lambda dx^3$, yielding $*dx = k^{-1}\,dx$, $*dy = -k^{-1}\,dy$. Then $** = \lambda$. Note that $\lambda = -1$ in II.

The field equation for ω may be formally satisfied by defining a potential ϕ such that

$$*d\phi = S^{-1} f^2\,d\omega. \tag{3}$$

$d\phi$ is exact; closure ($dd\phi = 0$) yields the ω field equation. (Alternately: the ω field equation shows that $*S^{-1}f^2\,d\omega$ is closed, so we write it as an exact form $\lambda\,d\phi$.) (Note: our E is Neugebauer's f, and our S is his V.) We write $E = f + i\phi$. Then the field equations for E and S may be written, where subscripts denote differentiation, as

$$S_{xy} = 0, \tag{4}$$

$$E_{xy} + \tfrac{1}{2} S^{-1}(S_x E_y + S_y E_x) = f^{-1} E_x E_y \tag{5}$$

($f = \mathrm{Re}\,E$). Equation (5) is the Ernst equation, in which no specialization of S has yet taken place. An alternate way of writing Eqs. (4) and (5) is

$$d(*dS) = 0, \tag{4'}$$

$$d(Sf^{-1} *df) + Sf^{-2} d\phi \wedge *d\phi = 0, \tag{5'}$$

$$d(Sf^{-2} *d\phi) = 0. \tag{3'}$$

It is usual to satisfy Eq. (4) by choosing $S = x + y = x^3$. In fact, this can always be done for the axially symmetric case, and this choice was made in I. However, this limits the possible Bäcklund transformations to that one which exhibits $S' = S$, i.e., the one found in I. Paper II did not make this restriction, apparently requiring $S' \neq S$, and found the other two BT's. For the purposes of the present paper, therefore, we make no restriction on S beyond Eq. (4).

We now redefine the variables t, u, v, and w of I:

$$t = f^{-1}E_x - S^{-1}S_x, \quad u = f^{-1}E_y - S^{-1}S_y,$$
$$v = f^{-1}\overline{E}_x - S^{-1}S_x, \quad w = f^{-1}\overline{E}_y - S^{-1}S_y. \tag{6}$$

To compare with Neugebauer (II), we first note that his x^1 and x^2 are given in terms of our variables x and y by

[a)]This material is based upon work supported by the National Science Foundation under Grant PHY-8008345.

 0022-2488/83/082178-10$02.50

$x^1 = 2x$ and $x^2 = 2y$. Then his M_i and N_i, in our notation, are

$$\begin{aligned} M_1 &= \tfrac{1}{4} f^{-1} E_x, \quad N_1 = \tfrac{1}{4} f^{-1} \bar{E}_y, \\ M_2 &= \tfrac{1}{4} f^{-1} \bar{E}_x, \quad N_2 = \tfrac{1}{4} f^{-1} E_y, \\ M_3 &= \tfrac{1}{2} S^{-1} S_x, \quad N_3 = \tfrac{1}{2} S^{-1} S_y, \end{aligned} \tag{7}$$

so that we have

$$\begin{aligned} t &= 4M_1 - 2M_3, \quad u = 4N_2 - 2N_3, \\ v &= 4M_2 - 2M_3, \quad w = 4N_1 - 2N_3. \end{aligned} \tag{8}$$

Cosgrove[5] also made this observation; his paper will be denoted by III.

Equation (5) and its complex conjugate may now be written

$$\begin{aligned} t_y &= \tfrac{1}{2} t(u - w) - \tfrac{1}{2} S^{-1}(wS_x + tS_y), \\ u_x &= \tfrac{1}{2} u(t - v) - \tfrac{1}{2} S^{-1}(uS_x + vS_y), \\ v_y &= \tfrac{1}{2} v(w - u) - \tfrac{1}{2} S^{-1}(uS_x + vS_y), \\ w_x &= \tfrac{1}{2} w(v - t) - \tfrac{1}{2} S^{-1}(wS_x + tS_y). \end{aligned} \tag{9}$$

We define potentials η and R by

$$*df = S^{-1} f(d\eta + \omega\, d\phi), \tag{10}$$

$$*dS = dR \tag{11}$$

whose existence is guaranteed by Eqs. (4) and (5), and we define 1-forms as in I:

$$\begin{aligned} \xi_1 &= f^{-1}\, d\phi, \quad \xi_2 = S^{-1} f\, d\omega, \\ \xi_3 &= S^{-1}(d\eta + \omega\, d\phi), \quad \xi_4 = f^{-1}\, df, \\ \xi_5 &= S^{-1}\, dS, \quad \xi_6 = S^{-1}\, dR. \end{aligned} \tag{12}$$

These then satisfy a set of 2-form equations [the differential form versions of Eqs. (4) and (9)], which are given in I and elsewhere.[6–8] [Strictly speaking, the exterior derivatives of the ξ_i may be given in terms of wedge products of the ξ_i themselves, with constant coefficients; and there exists a set of 2-forms, with constant coefficients, in the space of variables ϕ, ω, η, f, S, and R, which—when annulled—represent two-dimensional solution manifolds of Eqs. (4) and (9).]

It is convenient to define new 1-forms η_i, $i = 1$–6, which are self-dual or anti-self-dual up to a factor k, as follows, where the subscript in parentheses goes with the lower sign:

$$\begin{aligned} \eta_{1(2)} &= \xi_4 \pm i\xi_1 + k\xi_3 \pm ik\xi_2, \\ \eta_{3(4)} &= \xi_4 \mp i\xi_1 - k\xi_3 \pm ik\xi_2, \\ \eta_{5(6)} &= \xi_5 \pm k\xi_6. \end{aligned} \tag{13}$$

These satisfy

$$\begin{aligned} *\eta_i &= k^{-1}\eta_i, \quad i = 1,2,5, \\ *\eta_i &= -k^{-1}\eta_i, \quad i = 3,4,6. \end{aligned} \tag{14}$$

We also have, in the notation of I,

$$\begin{aligned} \eta_1 - \eta_5 &= 2t\, dx, \quad \eta_4 - \eta_6 = 2u\, dy, \\ \eta_2 - \eta_5 &= 2v\, dx, \quad \eta_3 - \eta_6 = 2w\, dy, \\ \eta_5 &= 2S^{-1} S_x\, dx, \quad \eta_6 = 2S^{-1} S_y\, dy, \end{aligned} \tag{15}$$

and, in the notation of II,

$$\begin{aligned} \eta_1 &= 8M_1\, dx, \quad \eta_3 = 8N_1\, dx, \\ \eta_2 &= 8M_2\, dx, \quad \eta_4 = 8N_2\, dy, \\ \eta_5 &= 4M_3\, dx, \quad \eta_6 = 4N_3\, dy. \end{aligned} \tag{16}$$

Finally, we note that the exterior derivatives of the η_i are identically:

$$\begin{aligned} 4d\eta_1 &= \eta_1 \wedge (\eta_3 + \eta_6 - \eta_4) - \eta_4 \wedge \eta_5, \\ 4d\eta_2 &= \eta_2 \wedge (\eta_4 + \eta_6 - \eta_3) - \eta_3 \wedge \eta_5, \\ 4d\eta_3 &= \eta_3 \wedge (\eta_1 + \eta_5 - \eta_2) - \eta_2 \wedge \eta_6, \\ 4d\eta_4 &= \eta_4 \wedge (\eta_2 + \eta_5 - \eta_1) - \eta_1 \wedge \eta_6, \\ 2d\eta_5 &= -2d\eta_6 = \eta_5 \wedge \eta_6, \end{aligned} \tag{17}$$

and the following (monomial!) 2-forms vanish:

$$\begin{aligned} &\eta_1 \wedge \eta_2, \quad \eta_1 \wedge \eta_5, \quad \eta_2 \wedge \eta_5 \quad (=0), \\ &\eta_3 \wedge \eta_4, \quad \eta_3 \wedge \eta_6, \quad \eta_4 \wedge \eta_6 \quad (=0). \end{aligned} \tag{18}$$

The content of the field equations appears in the possibility of the definition of the potentials ϕ, η, and R.

3. MODIFIED WAHLQUIST–ESTABROOK APPROACH

We pursue the search for BT's by a modified Wahlquist–Estabrook (MWE) method. The original Wahlquist–Estabrook method[9] consists of (1) the search for a pseudopotential and (2) the search for BT's. The modification given here is suitable for use with a constant coefficient, or CC, ideal, here denoted by C. The basic essentials of the MWE method are given in Eqs. (19) and (20) [or Eqs. (22) and (23)] for the pseudopotential and Eq. (45) for the BT.

A CC ideal, by (limited) definition here, consists of two sets of 2-forms, built from a number of 1-forms τ_i. The first set expresses the exterior derivatives $d\tau_i$, in terms of sums of wedge products of the τ_i. The second set is composed of sums of wedge products of the τ_i, which are to be annulled to obtain the solution manifold. We see that Eqs. (17) and (18) are a CC ideal. CC ideals have been discussed recently by Estabrook.[10] Estabrook has suggested a useful alternate title for a CC ideal: an "invariant Pfaffian system," or IPS.[11] By putting certain ("invariant") 1-forms equal to zero in a set of Cartan–Maurer equations it may be possible to produce an IPS; see Ref. 10.

4. SEARCH FOR A PSEUDOPOTENTIAL

We require that there exist a "pseudopotential" q and a 1-form θ

$$\theta = -dq + F^i(q)\tau_i \quad (\text{sum on } i), \tag{19}$$

where the functions F^i depend only on q. We require that

$$d\theta = 0 \pmod{\theta, C}; \tag{20}$$

i.e., we replace dq, where it occurs, by $F^i\tau_i$, obtained from $\theta = 0$; replace the $d\tau_i$ by their values as given in C; and use the remaining 2-forms in C to further simplify the equation $d\theta = 0$ by treating them as equations among the various

monomial 2-forms. The coefficients are then set equal to zero. We obtain in this way a set of ordinary (nonlinear) differential equations for the F^i.

It is often convenient to use an alternate approach. We consider a column vector Q of pseudopotentials q^α and a column vector Θ of 1-forms θ^α, assumed to be linear in the q^α. We write a linear representation:

$$\theta^\alpha = -dq^\alpha + B^{i\alpha}{}_\beta q^\beta \tau_i \quad (\text{sum on } i,\beta), \tag{21}$$

where the $B^{i\alpha}{}_\beta$ are constant. The index i is summed over the number of 1-forms τ_i; the range of α and β is the dimension of the representation, as yet unspecified (but possibly even infinite). In matrix form we have, where the matrices $B^i = [B^{i\alpha}{}_\beta]$,

$$\Theta = -dQ + (B^i \tau_i)Q. \tag{22}$$

The equation

$$d\Theta = 0 \pmod{\Theta, C} \tag{23}$$

becomes, after substitution for dQ and dropping the vector Q,

$$B^i d\tau_i - B^i B^k \tau_i \wedge \tau_k \pmod{C} = 0. \tag{24}$$

Setting the coefficients of this equation equal to zero yields (generally) an incomplete Lie algebra for the matrices B^i, the "prolongation structure" of the problem. If a two-dimensional representation for the B^i is found, a single pseudopotential q may be defined as q^2/q^1.

The latter approach is suitable for the Ernst equation CC ideal, Eqs. (17)–(18), if one makes one generalization: The matrices B^i must be taken to be functions of the variable

$$\zeta = \{[k(R+l)-S][k(R+l)+S]^{-1}\}^{1/2}, \tag{25}$$

where l is a parameter. If we take $S = x^3$ and $R = x^4$, we see that ζ may be written as

$$\zeta = [(kl-y)(kl+x)^{-1}]^{1/2}, \tag{26}$$

an invariant or similarity variable for the Ernst equation (5). [The scale transformation $x^{3'} = \alpha x^3$, $x^{4'} = \alpha x^4$, and the translation, $x^{4'} = x^4 + \beta$, α and β constant, leave both ζ and Eq. (5) invariant.] We note, from Eqs. (25), (11), and (15), that

$$d\zeta = g\eta_5 + h\eta_6 \tag{27}$$

where $g = \frac{1}{4}\zeta(\zeta^2 - 1)$ and $h = \frac{1}{4}\zeta^{-1}(\zeta^2 - 1)$. (Thus ζ itself is actually a pseudopotential; this point is clear in paper II.)

We now formulate our pseudopotential equations. We write, with the η_i for τ_i,

$$\Theta = -dQ + B^i(\zeta)\eta_i Q \tag{28}$$

and

$$d\Theta = 0 \text{ [mod Eqs. (28), (17), (18)]}. \tag{29}$$

Equation (29) becomes, after expansion and use of Eqs. (27) and (28),

$$0 = B^{i\prime}(g\eta_5 + h\eta_6) \wedge \eta_i - B^i B^k \eta_i \wedge \eta_k + B^i d\eta_i \quad [\text{mod (17), (18)}], \tag{30}$$

where $' = d/d\zeta$. (We choose $B^5 = B^6 = 0$ without loss of generality, since we may reasonably expect $B^5\eta_5 + B^6\eta_6$ to be proportional to $d\zeta$, and then we may eliminate these terms entirely by definition of a new pseudopotential as a function of ζ and the old one.)

Expansion of Eq. (30) now gives

$$\begin{aligned} &B^4 - B^1 = 4gB^{4\prime} = -4hB^{1\prime}, \\ &B^3 - B^2 = 4gB^{3\prime} = -4hB^{2\prime}, \end{aligned} \tag{31}$$

$$\begin{aligned} &B^3 - B^1 = 4[B^3, B^1], \quad B^4 - B^1 = 4[B^1, B^4], \\ &B^3 - B^2 = 4[B^2, B^3], \quad B^4 - B^2 = 4[B^4, B^2]. \end{aligned} \tag{32}$$

Solution of Eq. (31) yields

$$\begin{aligned} &B^1 = a\zeta + b, \quad B^2 = c\zeta + d, \\ &B^3 = c\zeta^{-1} + d, \quad B^4 = a\zeta^{-1} + b, \end{aligned} \tag{33}$$

where a, b, c, and d are constant matrices.

Equation (32) now gives

$$4[a,d] = 4[b,a] = a, \tag{34}$$

$$4[c,b] = 4[d,c] = c, \tag{35}$$

$$4[a,c] + 4[b,d] = b - d, \tag{36}$$

the prolongation structure. If we define

$$e = 4[b,d] \tag{37}$$

and use the Jacobi identity on all commutators, we get, in addition to Eqs. (34), (35), and (37),

$$\begin{aligned} &b - d - e = 4[a,c], \\ &[a,e] = [c,e] = 0, \\ &4[e,b] = 4[e,d] = e. \end{aligned} \tag{38}$$

We can specialize to a homomorphic image of the algebra of Eqs. (34)–(38) by taking $e = 0$, $4a = \tau - i\nu$, $4c = \tau + i\nu$, and $4d = -4b = i\theta$. We get

$$[\tau,\theta] = -\nu, \quad [\nu,\theta] = \tau, \quad \text{and} \quad [\nu,\tau] = \theta, \tag{39a}$$

the sl(2, R) algebra, as obtained before.[3,7]

A particular representation of Eq. (39a) is

$$\tau = -\sigma_z, \quad \nu = \sigma_x, \quad \theta = i\sigma_y, \tag{39b}$$

where the σ_i are the Pauli spin matrices. With this representation, we define $u = q_2/q_1$; u is related to the pseudopotential q defined in I by[7]

$$u = -i(\zeta - 1)(\zeta + 1)^{-1}(q-1)(q+1)^{-1}. \tag{40}$$

We now write Eq. (19) $\theta = 0$, the definition of q, in terms of the η_i [this is merely Eq. (8) of I].

$$4dq = q(1+q\zeta)\eta_1 - (q+\zeta)\eta_2 + \zeta^{-1}q(q+\zeta)\eta_3 - \zeta^{-1}(1+q\zeta)\eta_4 + (1-q^2)(\zeta\eta_5 + \zeta^{-1}\eta_6). \tag{41}$$

We will write further equations in terms of q.

We note the following relations of ζ and q to the variables γ and α in II:

$$\gamma = \zeta^2, \tag{42}$$

$$\alpha = \zeta(1+\zeta q)(q+\zeta)^{-1}. \tag{43}$$

We also note that the pseudopotential q ($= q_c$, say) used in III, Eq. (4.22), is related to q in this paper by

$$q_c = -(1+\zeta q)(q+\zeta)^{-1} = -\zeta^{-1}\alpha. \tag{44}$$

5. SEARCH FOR A BÄCKLUND TRANSFORMATION (BT)

Once a single pseudopotential q is found from Eqs. (19) and (20) [or (22) and (23)], we may assume that a BT exists with the following form:

$$\tau_i' = A_i^{\ j}(q)\tau_j \text{ (sum on } j). \tag{45}$$

This relates the exterior derivatives of the new (primed) variables to those of the old (unprimed) variables, providing the relation between first derivatives of the variables which is typical of BT's. (Estabrook has already published an example of this type of transformation of the KdV equation; see Eq. (23) in Ref. 10.) For the Ernst equation problem, we must generalize by letting the $A_i^{\ j}$ be functions of ζ as well.

[The reader may ask what happens if *both* q_1 and q_2 from Eq. (21), when the B^i are given by Eq. (33) and the 2×2 representation (39b), are used as arguments in Eq. (45) instead of just q. This calculation has been performed; the results are the same.]

Since there are six η_i, we would expect there to be $36\ A_i^{\ j}$—a nearly unmanageable number. However, we defined the η_i to be self- or anti-self-dual [Eq. (14)]—indeed, the current need for simplification was precisely the reason for making the definitions (13). This enables us to treat the two sets of η_i separately, so that we have only $2\times 9 = 18$ coefficients $A_i^{\ j}$. Indeed, we can go further; we assume that the fields S and R—or the independent variables—transform only among themselves (as is true in Refs. I and II), so that η_5' and η_6' are proportional to η_5 and η_6, respectively. The final assumed form is now the following, where as before we must assume the coefficients $a,b,\ldots,v$ to be functions of q and ζ:

$$\begin{aligned}
&\eta_1' = a\eta_1 + b\eta_2 + c\eta_5, \quad \eta_3' = h\eta_3 + m\eta_4 + n\eta_6,\\
&\eta_2' = e\eta_1 + f\eta_2 + g\eta_5, \quad \eta_4' = p\eta_3 + r\eta_4 + t\eta_6,\\
&\eta_5' = u\eta_5, \quad \eta_6' = v\eta_6.
\end{aligned} \tag{46}$$

(*Note*: t, u, v are not the earlier t, u, v.) We require for nondegeneracy that $af - be \neq 0$ and $hr - mp \neq 0$.

The η_i' are to satisfy Eqs. (17) and (18) (primed), as did the η_i. We see that the primed Eqs. (18) are satisfied automatically by virtue of the old Eqs. (18).

We remark parenthetically that we can express Eq. (46) in Neugebauer's matrix notation (II). We write column vector 1-forms as follows:

$$\psi = \begin{bmatrix}\eta_1\\ \eta_2\\ 2\eta_5\end{bmatrix} = 8\begin{bmatrix}M_1\\ M_2\\ M_3\end{bmatrix} dx, \tag{47}$$

$$\sigma = \begin{bmatrix}\eta_3\\ \eta_4\\ 2\eta_6\end{bmatrix} = 8\begin{bmatrix}N_1\\ N_2\\ N_3\end{bmatrix} dy, \tag{48}$$

with

$$\psi' = M\psi, \quad \sigma' = N\sigma. \tag{49}$$

Then

$$M = \begin{bmatrix}a & b & \frac{1}{2}c\\ e & f & \frac{1}{2}g\\ 0 & 0 & u\end{bmatrix} \tag{50}$$

and

$$N = \begin{bmatrix}h & m & \frac{1}{2}n\\ p & r & \frac{1}{2}t\\ 0 & 0 & v\end{bmatrix}. \tag{51}$$

The solution for the coefficients is rather complicated, but a few details will be provided here to show the procedure. We first note that we expect [from (17)']

$$2d\eta_5' = \eta_5' \wedge \eta_6'. \tag{52}$$

Equation (46) now gives

$$2du \wedge \eta_5 + 2u\, d\eta_5 = uv\eta_5 \wedge \eta_6. \tag{53}$$

But $du = u_q\, dq + u_\zeta\, d\zeta$, where dq and $d\zeta$ are supplied from Eqs. (41) and (27) and subscripts indicate derivatives, so that

$$\begin{aligned}
4du \wedge \eta_5 &= u_\zeta \zeta^{-1}(\zeta^2 - 1)\eta_6 \wedge \eta_5\\
&\quad + u_q\zeta^{-1}[q(q+\zeta)\eta_3 - (1+q\zeta)\eta_4\\
&\quad + (1-q^2)\eta_6] \wedge \eta_5\\
&= 2uv\eta_5 \wedge \eta_6 - 2u\eta_5 \wedge \eta_6,
\end{aligned} \tag{54}$$

where we have used Eqs. (17) and (18). Equating coefficients, we see that $u_q = 0$, so that u is a function of ζ only, and ($' = d/d\zeta$)

$$\zeta^{-1}(\zeta^2 - 1)u' = 2u(1-v). \tag{55}$$

In a similar manner, $v = v(\zeta)$ and

$$\zeta(\zeta^2 - 1)v' = 2v(1-u). \tag{56}$$

It may be shown from the other equations that

$$uv = 1. \tag{57}$$

These three equations result in two cases:

$$u = v = 1 \tag{58}$$

or

$$u = \zeta^2, \quad v = \zeta^{-2}. \tag{59}$$

Equation (58) is the choice used in I, equivalent to choosing $S' = S$ and $R' = R$. Equation (59) is that used in II, equivalent to choosing

$$S' \pm kR' = A\,[S \pm k(R + 2l)]^{-1} \pm B, \tag{60}$$

where A and B are parameters.

The $\eta_1 \wedge \eta_3$ coefficient from the $d\eta_1'$ equation and the $\eta_2 \wedge \eta_3$ coefficient from the $d\eta_2'$ equation are, respectively,

$$a(h-p) = -\zeta^{-1}q(q+\zeta)a_q + a, \tag{61}$$

$$f(p-h) = -\zeta^{-1}q(q+\zeta)f_q - f. \tag{62}$$

Elimination of $h - p$ yields $(af)_q = 0$, so that

$$af = A(\zeta). \tag{63}$$

Similar considerations give

TABLE I. Solutions for BT's, Case I. $b = e = m = p = 0$.

·	a	c	f	g	h	n	r	t	u	v
E	1	0	1	0	1	0	1	0	1	1
J	$\zeta\beta$	0	$\zeta\beta^{-1}$	0	$\zeta^{-1}\beta^{-1}$	0	$\zeta^{-1}\beta$	0	ζ^{2}	ζ^{-2}
H	$-q\beta$	μ	$-q^{-1}\beta^{-1}$	$q^{-1}\nu$	$-q\beta^{-1}$	$\zeta^{-1}\nu$	$-q^{-1}\beta$	$\zeta^{-1}q^{-1}\mu$	1	1
K	$-\zeta q$	$\zeta\nu$	$-\zeta q^{-1}$	$\zeta q^{-1}\mu$	$-\zeta^{-1}q$	$\zeta^{-2}\mu$	$-\zeta^{-1}q^{-1}$	$\zeta^{-2}q^{-1}\nu$	ζ^{2}	ζ^{-2}

$$be = B(\zeta), \tag{64}$$

$$hr = C(\zeta), \tag{65}$$

$$mp = D(\zeta). \tag{66}$$

Further combinations show that $A = A_0\,u$, $B = B_0\,u$, $C = C_0 v$, and $D = D_0 v$, where $A_0,\ldots,D_0$ are constants. Other combinations show that if $p \neq 0$, then we must have $r = 0$. Furthermore, $m \neq 0$ implies $h = 0$, $e \neq 0$ implies $f \neq 0$, and $b \neq 0$ implies $a = 0$. The nondegeneracy conditions [following Eq. (46)] play a key role here. Continuing in this vein yields four cases:

$$\begin{aligned} &\text{I. } b = e = m = p = 0,\ afhr \neq 0;\\ &\text{II. } b = e = h = r = 0,\ afmp \neq 0;\\ &\text{III. } a = f = m = p = 0,\ behr \neq 0;\\ &\text{IV. } a = f = h = r = 0,\ bemp \neq 0. \end{aligned} \tag{67}$$

The remaining work is just algebra. It is facilitated by noting that in cases I and II, for example, we get two separate equations for a_q, enabling an algebraic relation among the variables to be found. Equations (63)–(66) are very helpful. In each case the relation promised above, Eq. (57), is found to hold. We find eventually that each of the cases above produces four cases, so that we have 16 final solutions.

6. SOLUTIONS FOR COEFFICIENTS IN EQ. (46)

The solutions in the four cases given in Eq. (67) are listed in Tables I–IV, each table giving the four subcases. We denote each of the 16 solutions by a different English capital letter, listed in the left-hand column in each table. Those coefficients, from Eq. (46), which are nonzero are given across the top of each table, and their expressions, for each case, are given in the body of the table. We abbreviate certain expressions as follows [$\beta = -q_c$, Eq. (44)]:

$$\mu = 1 + \zeta q, \quad \nu = q + \zeta, \quad \beta = \mu\nu^{-1}. \tag{68}$$

Certain of these transformations may be easily identified. E is the identity. J and K are the two Neugebauer transformations, I_1 and I_2, respectively, from Ref. II. H is the Harrison transformation (from I). A is the Neugebauer–Kramer transformation[12]

$$f' = Sf^{-1}, \quad \omega' = -k^{-1}\phi, \quad \phi' = -k\omega, \tag{69}$$

denoted I by Cosgrove in III; change the sign of k for B. (Cosgrove's $I^* = I^{-1}$ and Neugebauer's S.) C is a simple sign change:

$$\omega' = -\omega, \quad \phi' = -\phi. \tag{70}$$

(Various sign conventions may make differences in the statement of some of these transformations.) It will be seen later that the other BT's are combinations of these basic ones.

The matrices M and N, as defined by Eqs. (50) and (51), clearly may be constructed for any of these transformations.

7. GROUP STRUCTURE OF THE SET OF BÄCKLUND TRANSFORMATIONS

The set of 16 Bäcklund transformations, given above, forms a group, with group composition defined as successive transformation. Several obvious subgroups exist. (Some of this material was given by Neugebauer and Cosgrove in II and III; the current treatment presents a more unified, if brief, view.)

The method of successive transformation, to form the group composition, needs to be carefully defined. We first require that the parameter l, occurring in ζ [Eq. (26)], keep the same value from transformation to transformation. This group structure is to be distinguished from the group struc-

TABLE II. Solutions for BT's, Case II. $b = e = h = r = 0$.

·	a	c	f	g	m	n	p	t	u	v
A	-1	1	-1	1	-1	1	-1	1	1	1
M	ζq	$-\zeta q$	ζq^{-1}	$-\zeta q^{-1}$	$\zeta^{-1}q^{-1}$	$-\zeta^{-1}q^{-1}$	$\zeta^{-1}q$	$-\zeta^{-1}q$	ζ^{2}	ζ^{-2}
L	$q\beta$	$-\zeta q$	$q^{-1}\beta^{-1}$	$-\zeta q^{-1}$	$q^{-1}\beta$	$-\zeta^{-1}q^{-1}$	$q\beta^{-1}$	$-\zeta^{-1}q$	1	1
N	$-\zeta\beta$	ζ^{2}	$-\zeta\beta^{-1}$	ζ^{2}	$-\zeta^{-1}\beta$	ζ^{-2}	$-\zeta^{-1}\beta^{-1}$	ζ^{-2}	ζ^{2}	ζ^{-2}

TABLE III. Solutions for BT's, Case III. $a=f=m=p=0$.

	b	c	e	g	h	n	r	t	u	v
B	-1	1	-1	1	-1	1	-1	1	1	1
U	ζq^{-1}	$-\zeta q^{-1}$	ζq	$-\zeta q$	$\zeta^{-1}q$	$-\zeta^{-1}q$	$\zeta^{-1}q^{-1}$	$-\zeta^{-1}q^{-1}$	ζ^2	ζ^{-2}
T	$q^{-1}\beta^{-1}$	$-\zeta q^{-1}$	$q\beta$	$-\zeta q$	$q\beta^{-1}$	$-\zeta^{-1}q$	$q^{-1}\beta$	$-\zeta^{-1}q^{-1}$	1	1
V	$-\zeta\beta^{-1}$	ζ^2	$-\zeta\beta$	ζ^2	$-\zeta^{-1}\beta^{-1}$	ζ^{-2}	$-\zeta^{-1}\beta$	ζ^{-2}	ζ^2	ζ^{-2}

ture discussed in III, Eq. (4.13) and following, in which all of the group elements are the same type of transformation, but with differing parameters. Second, we point out that, after one transformation, with given ζ and q, has been performed, there is no reason to expect ζ and q for the second transformation to be the same; in general, they will be different. (For example, ζ may be replaced by ζ^{-1}.) How do we determine these new quantities, which we denote by ζ' and q'? We outline a method as follows.

From Eqs. (27), (41), (47), and (48), we see that we can write

$$d\zeta = j\psi + s\sigma, \tag{71}$$

$$dq = w\psi + z\sigma, \tag{72}$$

where j, s, w, and z are row vectors:

$$j = \tfrac{1}{8}\zeta(\zeta^2-1)[0 \quad 0 \quad 1], \tag{73}$$

$$S = \tfrac{1}{8}\zeta^{-1}(\zeta^2-1)[0 \quad 0 \quad 1], \tag{74}$$

$$w = \tfrac{1}{4}[q\mu \quad -\nu \quad \tfrac{1}{2}\zeta(1-q^2)], \tag{75}$$

$$z = \tfrac{1}{4}\zeta^{-1}[q\nu \quad -\mu \quad \tfrac{1}{2}(1-q^2], \tag{76}$$

where μ and ν are defined in Eq. (68). We then note that

$$d\zeta' = j'\psi' + s'\sigma' = j'M\psi + s'N\sigma \tag{77a}$$

and

$$dq' = w'\psi' + z'\sigma' = w'M\psi + z'N\sigma, \tag{77b}$$

where we have used Eq. (49) and where primes on j, etc., mean replacement of ζ by ζ' and q by q'. If we assume $\zeta' = F(\zeta,q)$ and $q' = G(\zeta,q)$, substitute into Eq. (77), and equate coefficients of ψ and σ, we get the row vector equations:

$$j'M = F_\zeta j + F_q w, \tag{78a}$$

$$s'N = F_\zeta s + F_q z, \tag{78b}$$

$$w'M = G_\zeta j + G_q w, \tag{79a}$$

$$z'N = G_\zeta s + G_q z. \tag{79b}$$

When solved for F and G, these equations give ζ' and q', which are to be used in the second transformations of any group composition. The matrices M and N of Eq. (49), used above, are those associated with the first transformation.

Equations (78) are easy to solve. We note from Eqs. (50), (51), (73), and (74) that the left-hand sides of Eq. (78) have only a third component. This implies, since w and z have nonzero first and second components, that $F_q = 0$. The remaining equations give ($\zeta' = F$)

$$F(F^2-1)u = \zeta(\zeta^2-1)F_\zeta, \tag{80a}$$

$$F^{-1}(F^2-1)v = \zeta^{-1}(\zeta^2-1)F_\zeta, \tag{80b}$$

where we have used $u = M_{33}$ and $v = N_{33}$. It is easily seen that if $u = v = 1$, then $F = \zeta$; if $u = \zeta^2$ and $v = \zeta^{-2}$, then $F = \zeta^{-1}$.

Equations (79) are

$$\begin{aligned}[G(1+FG) \quad -(F+G) \quad \tfrac{1}{2}F(1-G^2)]M \\ = \tfrac{1}{2}\zeta(\zeta^2-1)G_\zeta[0 \quad 0 \quad 1] \\ + G_q[q(1+q\zeta) \quad -(q+\zeta) \quad \tfrac{1}{2}\zeta(1-q^2)]\end{aligned} \tag{81}$$

and

$$\begin{aligned}F^{-1}[G(F+G) \quad -(1+FG) \quad \tfrac{1}{2}(1-G^2)]N \\ = \tfrac{1}{2}\zeta^{-1}(\zeta^2-1)G_\zeta[0 \quad 0 \quad 1] \\ + \zeta^{-1}[q(q+\zeta) \quad -(1+q\zeta) \quad \tfrac{1}{2}(1-q^2)].\end{aligned} \tag{82}$$

We solve them by choosing one of the 16 transformations, finding M and N and substituting for them, and solving the separate equations for G. All equations are found to be self consistent. We summarize the results in Table V.

TABLE IV. Solutions for BT's, Case IV. $a=f=h=r=0$.

	b	c	e	g	m	n	p	t	u	v
C	1	0	1	0	1	0	1	0	1	1
X	$\zeta\beta^{-1}$	0	$\zeta\beta$	0	$\zeta^{-1}\beta$	0	$\zeta^{-1}\beta^{-1}$	0	ζ^2	ζ^{-2}
W	$-q^{-1}\beta^{-1}$	$q^{-1}\nu$	$-q\beta$	μ	$-q^{-1}\beta$	$\zeta^{-1}q^{-1}\mu$	$-q\beta^{-1}$	$\zeta^{-1}\nu$	1	1
Y	$-\zeta q^{-1}$	$\zeta q^{-1}\mu$	$-\zeta q$	$\zeta\nu$	$-\zeta^{-1}q^{-1}$	$\zeta^{-2}q^{-1}\nu$	$-\zeta^{-1}q$	$\zeta^{-2}\mu$	ζ^2	ζ^{-2}

TABLE V. Transformation of q and ζ for various BT's.

Transformation	ζ'	q'
E	ζ	q
J	ζ^{-1}	q
H	ζ	q^{-1}
K	ζ^{-1}	q^{-1}
A	ζ	$-\beta$
M	ζ^{-1}	$-\beta$
L	ζ	$-\beta^{-1}$
N	ζ^{-1}	$-\beta^{-1}$
B	ζ	$-\beta^{-1}$
U	ζ^{-1}	$-\beta^{-1}$
T	ζ	$-\beta$
V	ζ^{-1}	$-\beta$
C	ζ	q^{-1}
X	ζ^{-1}	q^{-1}
W	ζ	q
Y	ζ^{-1}	q

We are now in a position to explore the composition of successive BT's. If we wish to explore the effect of BT P followed by BT Q, we first construct the matrices M_P, M_Q, N_P, and N_Q. If we apply M_P to ψ, we get $\psi' = M_P\psi$. Before applying M_Q, we must recognize that ζ and q will now be different; so we look up ζ' and q' in Table V for BT P and replace ζ and q in M_Q by ζ' and q'. Call the new matrix M'_Q. We then have $\psi'' = M'_Q\psi' = M'_Q M_P\psi$. Thus the matrix for the composition is $M'_Q M_P$. We repeat for N_P and N_Q. In this way we can demonstrate that we always get one of the other sets (M, N), so that composition is closed.

It is now a simple matter to construct the multiplication table, Table VI, for BT's. We see that the four cases I, II, IV, and III (the order listed in Table VI) transform separately; the subcases in each case transform among themselves.

If we attempt to analyze the group by means of basic generators, we note that E is of order 1; M, N, U, V are of order 4; all other elements are of order 2. It is convenient to choose E, A, C, and M as generators. Then we have $A^2 = C^2 = E$, $M^4 = E$, $AM = M^3A$, $AM^3 = MA$, $AM^2 = M^2A$, $CA = AC$, and $CM = MC$, as the defining equations for the group, which may be denoted[13] $G_{16}{}^9 = G_8{}^4 \otimes G_2{}^1$. We identify the other elements in terms of the generators: $J = MA$, $K = M^3A$, $H = M^2$, $N = M^3$, $L = M^2A$, $T = M^2AC$, $U = MC$, $V = M^3C$, $B = AC$, $X = MAC$, $W = M^2C$, and $Y = M^3AC$. We write the 16 transformations in order, in a square array, in Table VII, in terms of the generators. We see that the first column is composed of the identity E, the two Neugebauer transformations (MA, M^3A), and the Harrison transformation (M^2). The succeeding columns are simply the first column postmultiplied by A (the Neugebauer–Kramer transformation), C (the sign change), and AC $(= B)$. Thus it is clear that the basic three BT's and their transformations are the only nontrivial BT's which can be derived with the given assumptions.

TABLE VII. The 16 BT's in terms of generators.

E	A	C	AC
MA	M	MAC	MC
M^2	M^2A	M^2C	M^2AC
M^3A	M^3	M^3AC	M^3C

It is more transparent, however, to write the 16 BT's as in Table VIII,[14] which clearly shows the effect of B, A, and $C = BA$ on the three basic BT's, J, H, and K.

We use, in the following discussion, Neugebauer's symbols γ and α (paper II). These are defined in Eqs. (42) and (43) above.

Neugebauer states that $I_2 = SI_1S$; in our notation, this is $K = BJB$. We see this easily from Table VI ($JB = U$, $BU = K$; or $BJ = V$, $VB = K$.) This procedure is consistent with Cosgrove's Eqs. (4.8)–(4.12). We consider B acting first; then we multiply by J to get JB, but replacing q in J with $-\beta^{-1}$ (see Table V, entry for B). ζ remains the same. If

TABLE VI. Composition (multiplication) of BT's. The BT listed on the top row is the first transformation, followed by that in the left column. (Thus, for example, $AJ = N$.)

	E	J	H	K	A	M	L	N	C	X	W	Y	B	U	T	V
E	E	J	H	K	A	M	L	N	C	X	W	Y	B	U	T	V
J	J	E	K	H	M	A	N	L	X	C	Y	W	U	B	V	T
H	H	K	E	J	L	N	A	M	W	Y	C	X	T	V	B	U
K	K	H	J	E	N	L	M	A	Y	W	X	C	V	T	U	B
A	A	N	L	M	E	K	H	J	B	V	T	U	C	Y	W	X
M	M	L	N	A	J	H	K	E	U	T	V	B	X	W	Y	C
L	L	M	A	N	H	J	E	K	T	U	B	V	W	X	C	Y
N	N	A	M	L	K	E	J	H	V	B	U	T	Y	C	X	W
C	C	X	W	Y	B	U	T	V	E	J	H	K	A	M	L	N
X	X	C	Y	W	U	B	V	T	J	E	K	H	M	A	N	L
W	W	Y	C	X	T	V	B	U	H	K	E	J	L	N	A	M
Y	Y	W	X	C	V	T	U	B	K	H	J	E	N	L	M	A
B	B	V	T	U	C	Y	W	X	A	N	L	M	E	K	H	J
U	U	T	V	B	X	W	Y	C	M	L	N	A	J	H	K	E
T	T	U	B	V	W	X	C	Y	L	M	A	N	H	J	E	K
V	V	B	U	T	Y	C	X	W	N	A	M	L	K	E	J	H

TABLE VIII. The 16 BT's in terms of J, H, K, B, A.

E	B	A	C	E	B	A	BA
J	V	N	X	J	BJ	AJ	BAJ
H	T	L	W	H	BH	AH	BAH
K	U	M	Y	K	BK	AK	BAK

$q' = -\beta^{-1}$, we get

$$\begin{aligned}\alpha' &= \zeta'\beta' = \zeta\beta' = \zeta(1+q'\zeta)(q'+\zeta)^{-1} \\ &= \zeta(1-\zeta\beta^{-1})(\zeta-\beta^{-1})^{-1} \\ &= \zeta(\beta-\zeta)(\zeta\beta-1)^{-1} = -\zeta q^{-1};\end{aligned} \tag{83}$$

α' can be shown to be the α which occurs in Cosgrove's Eqs. (4.11) and (4.12). Since the third factor, another B, has constant entries, there is no further replacement of ζ and q. Matrix multiplication now gives the proper matrices for K.

Cosgrove notes, in his Eq. (4.19), that a Harrison transformation may be given in terms of I_2 and I_1 by

$$H(\alpha,\gamma) = I_2((\alpha-\gamma)/\gamma(\gamma-1),\gamma^{-1})I_1(\alpha,\gamma). \tag{84}$$

To understand this, we note that the first argument in I_2 is the expression to be substituted for α in Eq. (4.12) for I_2 in III. But this α is the $\alpha' = -\zeta q^{-1}$ given in our Eq. (83) above. Thus we see that the proper procedure for proving Eq. (84) is: (1) Replace α in Cosgrove's Eq. (4.12) by $-\zeta q^{-1}$; this gives, in fact, exactly the matrices M and N found for $K(=I_2)$ in Table I above; (2) now make the change $q \to q$, $\zeta \to \zeta^{-1}$ in these matrices [this is the change prescribed in Table V when there is an initial transformation $J(=I_1)$]; (3) perform the M, N matrix multiplications for $I_2 I_1$. One obtains exactly H. In fact, we see this from Table VI: $H = KJ$ $(= JK)$, or $H = BJBJ = JBJB$. As noted above, the transformations E, H, J, K form a subgroup.

8. ERNST–MAXWELL EQUATIONS

The Ernst–Maxwell equations[15]

$$f(\nabla^2 G + S^{-1}\nabla S\cdot\Delta G) = \nabla G\cdot(\nabla G - 2\lambda\bar{H}\nabla H), \tag{85a}$$

$$f(\nabla^2 H + S^{-1}\nabla S\cdot\nabla H) = \nabla H\cdot(\nabla G - 2\lambda\bar{H}\nabla H), \tag{85b}$$

$$\nabla^2 S = 0, \tag{85c}$$

where $G = f + i\phi + \lambda H\bar{H}$ and H is a complex electromagnetic potential, may be written in a manner similar to Eqs. (3'), (4'), and (5'):

$$d(*dS) = 0, \tag{86a}$$

$$\begin{aligned}&d(Sf^{-1}*df) + Sf^{-2}\Omega\wedge *\Omega \\ &\quad + \lambda Sf^{-1}(dH\wedge *d\bar{H} + d\bar{H}\wedge *dH) = 0,\end{aligned} \tag{86b}$$

$$d(S*\Omega) - 2Sf^{-1}\,df\wedge *\Omega = 0, \tag{86c}$$

$$d(S*dH) - Sf^{-1}(df + i\Omega)\wedge *dH = 0, \tag{86d}$$

where

$$\Omega = d\phi + i\lambda(\bar{H}dH - Hd\bar{H}). \tag{86e}$$

We define

$$\begin{aligned}\xi_1 &= f^{-1}\Omega, \\ \xi_2 &= S^{-1}f\,d\omega, \\ \xi_3 &= S^{-1}(d\eta + \omega\,d\phi + L\,d\bar{H} + \bar{L}\,dH), \\ \xi_4 &= f^{-1}\,df, \\ \xi_5 &= S^{-1}\,dS, \\ \xi_6 &= S^{-1}\,dR, \\ \xi_7 &= S^{-1}f^{1/2}(dK - i\omega\,dH), \\ \xi_8 &= f^{-1/2}\,dH, \\ \xi_9 &= \bar{\xi}_7 = S^{-1}f^{1/2}(d\bar{K} + i\omega\,d\bar{H}), \\ \xi_{10} &= \bar{\xi}_8 = f^{-1/2}\,d\bar{H},\end{aligned} \tag{87}$$

where

$$L = \lambda(K - i\omega H) \tag{88}$$

and

$$dK = Sf^{-1}*dH + i\omega\,dH \tag{89}$$

defines a new potential K. Then we can write a CC ideal for the Ernst–Maxwell equations. The exterior derivatives are

$$\begin{aligned}&d\xi_1 = \xi_1\wedge\xi_4 - 2i\lambda\xi_8\wedge\xi_{10}, \quad d\xi_2 = \xi_2\wedge(\xi_5-\xi_4), \\ &d\xi_3 = \xi_3\wedge\xi_5 - \xi_1\wedge\xi_2 + \lambda(\xi_7\wedge\xi_{10} - \xi_8\wedge\xi_9), \\ &d\xi_4 = 0, \quad d\xi_5 = 0, \quad d\xi_6 = -\xi_5\wedge\xi_6, \\ &d\xi_7 = \xi_7\wedge(\xi_5 - \tfrac{1}{2}\xi_4) - i\xi_2\wedge\xi_8, \\ &d\xi_8 = -\tfrac{1}{2}\xi_4\wedge\xi_8, \quad d\xi_9 = \xi_9\wedge(\xi_5 - \tfrac{1}{2}\xi_4) + i\xi_2\wedge\xi_{10}, \\ &d\xi_{10} = \tfrac{1}{2}\xi_4\wedge\xi_{10},\end{aligned} \tag{90}$$

while the remaining annulled 2-forms are

$$\begin{aligned}&\xi_3\wedge\xi_1 - \xi_2\wedge\xi_4 = 0, \quad \xi_3\wedge\xi_8 - \xi_7\wedge\xi_4 = 0, \\ &\xi_3\wedge\xi_2 - \lambda\xi_1\wedge\xi_4 = 0, \quad \xi_3\wedge\xi_7 - \lambda\xi_8\wedge\xi_4 = 0, \\ &\xi_5\wedge\xi_2 - \xi_1\wedge\xi_6 = 0, \quad \xi_3\wedge\xi_{10} - \xi_9\wedge\xi_4 = 0, \\ &\xi_5\wedge\xi_1 - \lambda\xi_2\wedge\xi_6 = 0, \quad \xi_3\wedge\xi_9 - \lambda\xi_{10}\wedge\xi_4 = 0, \\ &\xi_5\wedge\xi_3 - \xi_4\wedge\xi_6 = 0, \quad \xi_5\wedge\xi_8 - \lambda\xi_7\wedge\xi_6 = 0, \\ &\xi_4\wedge\xi_5 + \lambda\xi_3\wedge\xi_6 = 0, \quad \xi_5\wedge\xi_7 - \xi_8\wedge\xi_6 = 0, \\ &\xi_1\wedge\xi_8 - \lambda\xi_7\wedge\xi_2 = 0, \quad \xi_5\wedge\xi_{10} - \lambda\xi_9\wedge\xi_6 = 0, \\ &\xi_1\wedge\xi_7 - \xi_8\wedge\xi_2 = 0, \quad \xi_5\wedge\xi_9 - \xi_{10}\wedge\xi_6 = 0, \\ &\xi_1\wedge\xi_{10} - \lambda\xi_9\wedge\xi_2 = 0, \quad \xi_7\wedge\xi_{10} - \xi_9\wedge\xi_8 = 0, \\ &\xi_1\wedge\xi_9 - \xi_{10}\wedge\xi_2 = 0, \quad \xi_7\wedge\xi_9 - \lambda\xi_{10}\wedge\xi_8 = 0.\end{aligned} \tag{91}$$

We write a prolongation form in terms of the ξ_i, for variety, instead of the η_i as before:

$$\kappa = -dq + (C^i\xi_i)q, \tag{92}$$

where κ and q are column vectors and C^i are matrix functions of ζ. The prolongation equation becomes

$$\begin{aligned}&C^i d\xi_i - [C^i, C^l]\xi_i\wedge\xi_l + \tfrac{1}{4}\zeta^{-1}(\zeta^2-1)C'' \\ &\quad\times[(\zeta^2+1)\xi_5 + k(\zeta^2-1)\xi_6]\wedge\xi_i = 0.\end{aligned} \tag{93}$$

We set $C^5 = C^6 = 0$ as before. Expansion of Eq. (93) yields a set of differential equations for the C^i plus a set of equations for their commutation relations. Solution of the differential equations yields

$$C^1 = \tfrac{1}{2}\zeta^{-1}(\zeta^2+1)\nu + \theta, \quad C^3 = \tfrac{1}{2}k\zeta^{-1}(\zeta^2-1)\tau,$$
$$C^2 = \tfrac{1}{2}k\zeta^{-1}(\zeta^2-1)\nu, \quad C^4 = \tfrac{1}{2}\zeta^{-1}(\zeta^2+1)\tau + \phi,$$
$$C^7 = \tfrac{1}{2}k\zeta^{-1}(\zeta^2-1)\alpha, \quad C^9 = \tfrac{1}{2}k\zeta^{-1}(\zeta^2-1)\beta,$$
$$C^8 = \tfrac{1}{2}\zeta^{-1}(\zeta^2+1)\alpha + \gamma, \quad C^{10} = \tfrac{1}{2}\zeta^{-1}(\zeta^2+1)\beta + \delta, \tag{94}$$

where $\nu, \theta, \tau, \phi, \alpha, \beta, \gamma$, and δ are constant matrices. τ, θ, and ν play the same role as their counterparts, defined above following Eq. (38). These matrices satisfy

$$[\nu,\theta] = \tau, \quad \nu = [\theta,\tau], \quad [\tau,\gamma] = -\tfrac{1}{2}\alpha,$$
$$\theta = [\theta,\phi] + [\nu,\tau], \quad [\tau,\alpha] + [\phi,\gamma] = -\tfrac{1}{2}\gamma,$$
$$[\nu,\phi] = [\tau,\phi] = 0, \quad [\beta,\gamma] = \lambda(\tau + i\nu),$$
$$[\nu,\gamma] = [\alpha,\theta] = -\tfrac{1}{2}i\alpha, \quad [\alpha,\delta] = \lambda(\tau - i\nu),$$
$$0 = [\nu,\alpha] + [\theta,\gamma], \quad [\alpha,\beta] + [\gamma,\delta] = -2i\gamma\theta,$$
$$[\alpha,\gamma] = [\alpha,\phi] = 0. \tag{95}$$

These equations can be satisfied only by a representation of size 3×3 or greater (reminiscent of the 3×3 matrices required for Ernst–Maxwell in the Ernst–Hauser[16] approach to this problem.) If we make the ansatz that the upper left corners of τ, ν, and θ are given by the spin matrices as before, we find *two* representations, as follows (subscripts indicate the representation label; k_1 and k_2 are constants):

$$\phi_1 = \phi_2 = 0, \quad \tau_1 = \tau_2 = -\tfrac{1}{2}\begin{bmatrix}1&0&0\\0&-1&0\\0&0&0\end{bmatrix},$$
$$\nu_1 = \nu_2 = \tfrac{1}{2}\begin{bmatrix}0&1&0\\1&0&0\\0&0&0\end{bmatrix}, \quad \theta_1 = \theta_2 = \tfrac{1}{2}\begin{bmatrix}0&1&0\\-1&0&0\\0&0&0\end{bmatrix},$$
$$\alpha_1 = k_1\begin{bmatrix}0&0&0\\0&0&0\\1&i&0\end{bmatrix}, \quad \beta_1 = \frac{\lambda}{2k_1}\begin{bmatrix}0&0&1\\0&0&-i\\0&0&0\end{bmatrix},$$
$$\gamma_1 = k_1\begin{bmatrix}0&0&0\\0&0&0\\-1&i&0\end{bmatrix}, \quad \delta_1 = \frac{\lambda}{2k_1}\begin{bmatrix}0&0&1\\0&0&i\\0&0&0\end{bmatrix},$$
$$\alpha_2 = k_2\begin{bmatrix}0&0&1\\0&0&i\\0&0&0\end{bmatrix}, \quad \beta_2 = \frac{\lambda}{2k_2}\begin{bmatrix}0&0&0\\0&0&0\\1&-i&0\end{bmatrix},$$
$$\gamma_2 = k_2\begin{bmatrix}0&0&1\\0&0&-i\\0&0&0\end{bmatrix}, \quad \delta_2 = -\frac{\lambda}{2k_2}\begin{bmatrix}0&0&0\\0&0&0\\1&i&0\end{bmatrix}. \tag{96}$$

We now define the η_i, in the same way as before (the number in parentheses corresponds to the lower sign):

$$\eta_{1(2)} = \xi_4 \pm i\xi_1 + k\xi_3 \pm ik\xi_2,$$
$$\eta_{3(4)} = \xi_4 \mp i\xi_1 - k\xi_3 \pm ik\xi_2, \tag{97}$$
$$\eta_{5(6)} = \xi_5 \pm k\xi_6, \quad \eta_{7(9)} = \xi_{10} \pm k\xi_9, \quad \eta_{8(10)} = \xi_8 \pm k\xi_7.$$

The η_i divide into self-dual η_i ($i = 1,2,5,7,8$; $^*\eta_i = k^{-1}\eta_i$) and anti-self-dual η_i ($i = 3,4,6,9,10$; $^*\eta_i = -k^{-1}\eta_i$). We can show that $\eta_k \wedge \eta_l = 0$ if η_k and η_l are either both self-dual or both anti-self-dual. Equations (90) and (97) enable us to derive expressions for the $d\eta_k$.

Kramer and Neugebauer[17] (denoted by IV) define a set of quantities $A_i,\ldots,E_i$, $i = 1,2$. They are related to the η_i by (remember $\lambda = -1$ in IV)

$$\eta_1 = 4A_1\,dx, \quad \eta_6 = 2C_2\,dy,$$
$$\eta_2 = 4B_1\,dx, \quad \eta_7 = -2iD_1\,dx,$$
$$\eta_3 = 4B_2\,dy, \quad \eta_8 = -2iE_1\,dx, \tag{98}$$
$$\eta_4 = 4A_2\,dy, \quad \eta_9 = -2iD_2\,dy,$$
$$\eta_5 = 2C_1\,dx, \quad \eta_{10} = -2iE_2\,dy.$$

Expansion of Eq. (92) ($\kappa = 0$) for the first representation in Eq. (96) yields, for the pseudopotentials $u(v) = q_1 \pm iq_2$, $w = q_3$:

$$du = -\tfrac{1}{8}(\eta_1 - \eta_2 - \eta_3 + \eta_4)u - \tfrac{1}{4}(\zeta\eta_2 + \zeta^{-1}\eta_3)v + \tfrac{1}{2}ik_1^{-1}\lambda(\zeta\eta_7 + \zeta^{-1}\eta_9)w,$$
$$dv = \tfrac{1}{2}(\eta_1 - \eta_2 - \eta_3 + \eta_4)v - \tfrac{1}{4}(\zeta\eta_1 + \zeta^{-1}\eta_4)u + \tfrac{1}{2}k_1^{-1}\lambda(\eta_7 + \eta_9)w, \tag{99}$$
$$dw = \tfrac{1}{2}k_1[(\zeta\eta_8 + \zeta^{-1}\eta_{10})u - (\eta_8 + \eta_{10})v].$$

Similar equations result for the second representation (with analogous pseudopotentials u', v', w'). Comparison with IV shows that λ in that paper is our ζ, and we also have the following relations for the quantities $\psi, \cdots$ in IV:

$$\psi = \tfrac{1}{2}iak_2^{-1}\sqrt{f}\,u', \quad \tilde{\psi} = -ibk_1\sqrt{f}\,v,$$
$$\chi = -\tfrac{1}{2}iak_2^{-1}\sqrt{f}\,v', \quad \tilde{\chi} = ibk_1\sqrt{f}\,u, \tag{100}$$
$$\sigma = a\sqrt{f}\,w', \quad \tilde{\sigma} = b\sqrt{f}\,w,$$

where a and b are arbitrary constants. In general (whether $\lambda = \pm1$, $k = 1$ or i, and whether or not complex conjugation is taken on k if $k = i$), we have

$$\bar{u} = cv', \quad \bar{v} = cu', \quad \bar{w} = 2\lambda ck_2\bar{k}_1 w', \tag{101}$$

where c is an arbitrary constant.

It is convenient to consider five variables made of the ratios of these variables (uw^{-1}, vw^{-1}, $\bar{u}\bar{w}^{-1}$, $\bar{v}\bar{w}^{-1}$, $\bar{w}w^{-1}$). The function ζ makes a sixth variable. Denote the variables as α_A, $A = 1$–6. The exterior derivatives of the α_A are linear combinations of the η_i, the coefficients being functions of the α_A:

$$d\alpha_A = A_A{}^K(\alpha_B)\eta_k. \tag{102}$$

An arbitrary function, $f(\alpha_A)$, has exterior derivative

$$df = (\partial f/\partial\alpha_A)d\alpha_A = (\partial f/\partial\alpha_A)A_A{}^k\eta_k = B^k(f)\eta_k, \tag{103}$$

where the $B^k = A_A{}^k\partial/\partial\alpha_A$ are linear operators.

To search for BT's, we assume a linear transformation like Eq. (45), as before—except that now, instead of 14 functions of two variables, we have 42 functions of six variables!

$$\eta_k' = C_k{}^l(\alpha_A)\eta_l \tag{104}$$

where the families of self-dual and anti-self-dual η_i trans-

form among themselves. The problem is made tractable only by the shorthand use of the B_k, which operate on the $C_k{}^l$. Even then, progress can be made only by making an ansatz, typically that several of the $C_k{}^l$ vanish. For example, one may try $\eta_1' = F\eta_1$, or $\eta_1' = F\eta_1 + G\eta_5$. It appears likely that only two or three coefficients, in the expression for any η_k', are nonzero. However, even this approach does not work until we make the additional ansatz, made in IV, that

$$\psi\tilde{\psi} + \sigma\tilde{\sigma} - \chi\tilde{\chi} = 0. \tag{105}$$

This reduces the number of independent variables from six to five. Fortunately, the equations—expressed in terms of the operators B^k—keep the same form. Only the B^k change.

The ansatz which gives results is: In η_1', include either η_1 or η_2, but not both; either η_7 or η_8, but not both; and η_5. It is clear that we have four possible choices (η_1 and η_7, η_1 and η_8, etc). Setting certain of the coefficients to zero then induces other coefficients to be zero. One continues in a similar manner. The problem becomes possible to solve, though still very complicated.

Solutions found to date after pursuit of this approach all reduce to the single BT reported by Kramer and Neugebauer in IV.

The possible relation of this BT approach to electrovac space-times to that of Cosgrove[18] is yet to be explored.

9. FINAL REMARKS

It is interesting that the Ernst equation admits three BT's found by the method in this paper, while the Ernst–Maxwell equations admit only one (at least, known to date)! It may be that the extra equations force enough additional structure to preclude more than one BT.

The MWE approach to finding BT's appears to have promise. It suggests that it would be desirable to find a canonical set of 1-forms for any equation, or set of equations, such that a CC ideal of forms can be constructed. It has been demonstrated, for example, that CC ideals exist for the sine–Gordon and Korteweg–de Vries equations and that they can be used to derive the associated BT's.[10,19]

It is possible to formulate more complicated problems in elegant ways by using differential forms. As an example, the vacuum Einstein equations, with only one nonnull Killing vector, may be cast into an ideal of 1-, 2-, and 3-forms with most coefficients constant.[8,20] Investigation into an inverse scattering or BT formulation for this problem, using techniques due to Morris,[21] has been done—without definite results to date.[20] The formulation of the full vacuum equations in differential forms, due to Israel,[22] might be suitable for application of these methods.

The MWE approach also serves to suggest new generalizations. If the equations being studied admit similarity or invariant variables μ_α, their exterior derivatives can be written in the form

$$d\mu_\alpha = F_\alpha{}^k(\mu_\beta)\eta_k. \tag{106}$$

ζ is such a variable for the Ernst or Ernst–Maxwell equations, as noted above, and it was convenient (even necessary!) to include it as a variable in the coefficients in Eqs. (28) and (45). But others may exist. Such a generalization is currently being tried by the author and is already yielding new BT's to the (vacuum) Ernst equation,[8] although it is not yet known whether these BT's provide new solutions to the equation. These results will be reported later as they are completed.

Note added in proof: It now appears that these new BT's are combinations of Ehlers transformations and known BT's.

[1] F. J. Ernst, Phys. Rev. **167**, 1175–8 (1968).
[2] See, for example, P. Forgacs, Z. Horvath, and L. Palla, Phys. Rev. Lett. **45**, 505–8 (1980), and Phys. Lett. B **99**, 232–6; N. Sanchez, Phys. Rev. D, (to be published); A. Chakrabarti, Phys. Rev. D **25**, 3282–8 (1982).
[3] B. K. Harrison, Phys. Rev. Lett. **41**, 1197–2000 (1978) (cited as paper I in text).
[4] G. Neugebauer, J. Phys. A: Math. Gen. **12**, L67–70 (1979) (cited as paper II in text).
[5] C. M. Cosgrove, J. Math. Phys. **21**, 2417–47 (1980) (cited as paper III in text).
[6] B. K. Harrison, Proc. Utah Acad. Sci., Arts, Lett. **53**, part 1, 67–74 (1976).
[7] B. K. Harrison, in *Proceedings of the Second Marcel Grossmann Meeting on Recent Developments of General Relativity*, edited by R. Ruffini (North-Holland, Amsterdam, 1982), pp. 341–352.
[8] B. K. Harrison, in *Proceedings of the Third Marcel Grossmann Meeting on Recent Developments of General Relativity* (to be published).
[9] H. D. Wahlquist and F. B. Estabrook, Phys. Rev. Lett. **31**, 1386–90 (1973); J. Math. Phys. **16**, 1–7 (1975).
[10] F. B. Estabrook, J. Math. Phys. **23**, 2071–6 (1982).
[11] F. B. Estabrook, letter to Peter Gragert, June 1982.
[12] G. Neugebauer and D. Kramer, Ann. Phys. (Leipzig) **24**, 62 (1969).
[13] C. J. Bradley and A. P. Cracknell, *The Mathematical Theory of Symmetry in Solids* (Oxford U. P., Oxford, 1972), p. 234.
[14] C. M. Cosgrove, private communication.
[15] F. J. Ernst, Phys. Rev. **168**, 1415–7 (1968).
[16] I. Hauser and F. J. Ernst, Phys. Rev. D **20**, 1783–90 (1979).
[17] D. Kramer and G. Neugebauer, J. Phys. A: Math. Gen. **14**, L333–L338 (1981) (cited as paper IV in text).
[18] C. M. Cosgrove, J. Math. Phys. **22**, 2624–2639 (1981).
[19] B. K. Harrison, unpublished; H. D. Wahlquist, private communication.
[20] B. K. Harrison, paper given at the Ninth International Conference on General Relativity and Gravitation, Jena, DDR, 1980.
[21] H. C. Morris, J. Math. Phys. **17**, 1870–2 (1976).
[22] W. Israel, *Differential Forms in General Relativity* (Institute for Advanced Studies, Dublin, 1979), 2nd ed.

On apparent horizons and the Schwarzschild surface for a uniform fluid sphere in general relativity

Henning Knutsen[a)]
Institute of Theoretical Physics, University of Warsaw, Warsaw, Poland

(Received 20 January 1983; accepted for publication 17 March 1983)

The time history of the marginally trapped surfaces, i.e., the apparent horizons for a spherically symmetric nonstatic fluid of uniform density are studied. Generally it is found that apparent horizons may or may not exist dependent upon the choice of arbitrary functions of integration. However, it is shown in this paper that if the metric is conformally flat or if the circumference of the sphere is an increasing function of a radial coordinate, apparent horizons exist if and only if the surface is inside the Schwarzschild surface. Then there exist in fact at least two horizons: The absolute Schwarzschild surface and an apparent horizon in the interior of the fluid matter.

PACS numbers: 04.20.Jb

I. INTRODUCTION

Several authors have studied different aspects of nonstatic spheres of uniform density in general relativity. Gupta[1] has given the general solution for a shear-free, nonsingular model. Bonnor and Faulkes[2] find that pulsating spheres may exist. Bondi[3] finds conditions for an inward motion to reverse for various assumed relations between density and central pressure. Thompson and Whitrow,[4] Nariai,[5] and Banerjee[6] concentrate on the global motion of the sphere, and they find the possibility that a bounce may exist. Krishna Rao[7] shows that when a singularity at the center is removed, the space time is conformally flat. Banerjee[8] has studied the spectrum of radiation emitted from the surface of a collapsing or expanding sphere. Cook[9] has obtained the surprising result that there exist classes of solutions for which the sign of the spatial curvature may vary with time. In a recent paper, Glass[10] claims that uniform spheres always contain trapped surfaces. This last result is in fact not correct, and it is the purpose of this paper to investigate in some detail the conditions for such apparent horizons to exist.

II. TRAPPED SURFACES AND APPARENT HORIZONS

Writing the metric in the form

$$ds^2 = A^2\,dt^2 - B^2 dr^2 - R^2 d\Omega^2, \tag{1}$$

where $d\Omega^2 = d\theta^2 + \sin^2\theta\, d\phi^2$

and A and B are functions of time t and the comoving radial coordinate r, the total mass–energy contained within the sphere r at time t is given by

$$m = 4\pi \int_0^r w R^2 \frac{\partial R}{\partial \bar{r}}\, d\bar{r}, \tag{2}$$

where w is the density.

The function m was first introduced by Misner and Sharp,[11] and m may also be written (cf. Hernandez and Misner[12])

$$\frac{2m}{R} = 1 + \frac{1}{A^2}\left(\frac{\partial R}{\partial t}\right)^2 - \frac{1}{B^2}\left(\frac{\partial R}{\partial r}\right)^2. \tag{3}$$

Following Glass,[10] we define

$$\rho = -l^\alpha \nabla_\alpha \ln R, \mu = n^\alpha \nabla_\alpha \ln R, \tag{4}$$

where $l^\mu = (1/\sqrt{2})((1/A)\delta^\mu_t + (1/B)\delta^\mu_r)$,

$n^\mu = (1/\sqrt{2})((1/A)\delta^\mu_t - (1/B)\delta^\mu_r)$.

Hernandez and Misner's equation (3) may then be written

$$2m/R = 1 - 2\mu\rho R^2. \tag{5}$$

Glass then claims that a trapped surface exists when both μ and ρ are positive. From (5) we then obtain

$$2m/R < 1. \tag{6}$$

But Hernandez and Misner[12] have in fact proved that even light rays which are moving out relative to the matter are falling in in regions of a collapsing fluid where

$$2m/R > 1, \tag{7}$$

and all events satisfying the inequality (7) lie inside a Schwarzschild surface and signals sent out from these events cannot reach external observers. The condition $2m/R > 1$ is sufficient, but not necessary, for outgoing light rays to be trapped. We may have $2m/R < 1$ close to the the center, but a light ray starting from this point cannot overtake one starting farther out and will be trapped if the outer one is. We therefore think that Glass's condition for a trapped surface to exist is wrong and should be replaced by the inequality (7). This does not in fact influence the contents of Glass's paper, since Glass investigates the condition $\rho = 0$. The time history of the marginally trapped surface, i.e., the apparent horizon, is given by

$$2m/R = 1. \tag{8}$$

Equation (8) may alternatively be written

$$F + \sqrt{x}\, H - 2x\frac{\partial F}{\partial x} = 0, \tag{9}$$

where $F = 1/B, x = r^2$, and the condition $\partial B/\partial t = ABH$ for zero mass–energy flux and the condition

$$\frac{\partial B}{\partial t}\frac{1}{B} = \frac{\partial R}{\partial t}\frac{1}{R}$$

for shear-free motion have been used.

$H = H(t)$ is here an arbitrary function of integration.

III. UNIFORM DENSITY SOLUTIONS

Demanding pressure isotropy yields the follwing equation first obtained by Wyman[13]:

[a)] Present address: Institute of Theoretical Astrophysics, University of Oslo, Oslo, Norway.

 0022-2488/83/082188-03$02.50

$$\frac{\partial^2}{\partial x^2}(F) + J(x)F^2 = 0, \tag{10}$$

where J is an arbitrary function of x. All uniform density solutions are obtained by putting (Glass[10])

$$J = kx^{-5/2}, \tag{11}$$

where k is an arbitrary constant. The substitution $F = \sqrt{x}\Gamma$, $x = e^y$ allows a special solution of (10):

$$\frac{1}{2}\left(\frac{\partial}{\partial y}\Gamma\right)^2 + \frac{1}{3}k\Gamma^3 - \frac{1}{8}\Gamma^2 + C_1(t) = 0. \tag{12}$$

Glass solves Eq. (12) fo the case $C_1 = 0$. But in his discussion he claims that *all* homogeneous density solutions have the characteristic which is special for the solution of (12) with $C_1 = 0$, namely, to contain a trapped surface. This generalization we cannot accept without proof. Moreover, his solution of (12) is erroneous [Eq. (47) in Glass's paper]. The correct solution of (12) with $C_1 = 0$ is

$$F = (3/2k)xc_2(1 + \sqrt{x}\, c_2)^{-2}, \tag{13}$$

where $c_2(t)$ is an arbitrary function of time. From Eq. (13) and the relation $R = \sqrt{x}/F$, we conclude that R is not finite at the center. Using Eq. (2), the following relation is obtained:

$$2m/R = (\tfrac{8}{3}\pi w)R^2. \tag{14}$$

Hence the singularity at the center is always hidden by a trapped surface.

Using (13) in Eq. (9) for the apparent horizon, we obtain the following equation:

$$q^3 + (\alpha + 3)q^2 + (3 - \alpha)q + 1 = 0, \tag{15}$$

where $\alpha = 3/2kH$, and $q = c_2 x_{ah}^{1/2}$. The suffix ah denotes apparent horizon values. For shear-free motion we may write

$$1/F = B = (1/r)R.$$

We therefore restrict F to be positive. For positive c_2, Eq. (13) and the definition of q then yield that only the positive roots of Eq. (14) are acceptable.

The discriminant of the cubic equation (15) is given by $\alpha^2(1 - \alpha^2/108)$. When $\alpha^2 < 108$, Eq. (14) has therefore only one real root which is negative. Hence no apparent horizon exists when $\alpha^2 < 108$, and from (14), we conclude that each layer of this sphere is inside a trapped surface.

On the other hand, when $\alpha = -\frac{27}{2}$, one of the roots of (14) is $q = 2$. Hence there is an apparent horizon.

It is also easy to see that this apparent horizon may be located inside the surface of the matter distribution. At this boundary of the sphere, the pressure drops to zero, and the interior solution must be matched to the exterior vacuum Schwarzschild solution. These two conditions yield the following equation:

$$H^2 = 2Mx_b^{-3/2}F_b^3 - 4F_b\left(\frac{\partial F}{\partial x}\right)_b + 4x_b\left(\frac{\partial F}{\partial x}\right)_b^2, \tag{16}$$

where the suffix b denotes boudary values, and the constant M is total mass–energy inside the boundry.

Writing $N = 3M/2k$ and defining

$$P_6(z) = z^6 + 6z^5 + 15z^4 + 2(-N\alpha^2 + 2\alpha^2 + 10)z^3 + 15z^2 + 6z + 1, \tag{17}$$

Eq. (16) may be written

$$P_6(z) = 0, \tag{18}$$

where $z = c_2\sqrt{x_b}$. Choosing the constant $N > \frac{43}{29}(\frac{7}{6})^2$, it is seen that (18) has a root $z > 2$. Hence the apparent horizon is located inside the boundary of the matter distribution.

Using (2) and (13), it is found that for this model we further have

$$\frac{\partial}{\partial r}\left(\frac{2m}{R}\right) = \frac{32\pi w}{9}kc_2R\,\frac{\bar{y}^2 - 1}{\bar{y}^2} \tag{19}$$

where $\bar{y} = c_2\sqrt{x}$. Hence we conclude that the boundary is inside the Schwarzschild surface.

From (17), it is also seen that it is possible to have a model where $P_6(q) > 0$ and $P_6(1) < 0$. Hence we have $0 < z < 1$. Using (19), we conclude that, for this model, the boundary is outside the Schwarzschild surface.

IV. CONFORMALLY FLAT SOLUTIONS

A conformally flat space is defined as one with a metric of the form

$$ds^2 = P\eta_{\mu\nu},$$

where P is an arbitrary positive function and $\eta_{\mu\nu}$ is the Lorentz–Minkowski metric. A well known theorem then says that the Weyl tensor $C_{\mu\nu\rho\sigma}$ is zero. All the conformally flat solutions are then found upon setting $J = 0$ [see for example Eq. (14) of Glass's paper]. From (10), the solution for F is then obtained. When this solution is used in (9), $x_{ah}^{1/2}$ is found. Equation (16), combined with the relation $R = \sqrt{x}F$, now yields the following relation which is valid when the apparent horizon is located inside the matter distribution:

$$(R_b/R_{ah})^2 = 2M/R_b. \tag{20}$$

If the surface of the sphere is outside the Schwarzschild surface, i.e., $2M/R_b < 1$, Eq. (20) shows immediately that an apparent horizon *inside* the boundary cannot exist. Outside the matter configuration, there is vacuum. We thus have $m = M$ outside the sphere. Existence of an apparent horizon *outside* the matter distribution would then give

$$1 = 2m/R_{ah} = 2M/R_{ah} < 2M/R_b \tag{21}$$

in contradiction to $2M/R_b < 1$, i.e., the surface of the sphere to be outside the Schwarzschild surface. Hence in sharp contrast with Glass, who claims that all conformally flat solutions have an apparent horizon, we conclude that when the surface of the sphere is outside the Schwarzschild surface, no apparent horizon exists.

When the surface of the sphere is inside the Schwarzschild surface, i.e., $2M/R_b > 1$, there is of course an apparent horizon outside the matter, namely, the absolute Schwarzschild horizon.

From Eq. (2) we obtain, since the density w is uniform,

$$2m/R = \tfrac{8}{3}\pi wR^2. \tag{22}$$

Hence $(2m/R)_{\text{center}} = 0.$ (23)

The continuity of $2m/R$ thus yields that the following relation is valid somewhere in the interior of the sphere:

$$2m/R = 1. \tag{24}$$

Hence when the surface of the sphere is inside the Schwarzschild surface, there exists at least one interior apparent horizon.

Glass also investigates the exact solution

$$F = (x + \alpha)\, h_1(t) + \beta, \tag{25}$$

where h_1 is an arbitrary function of t and α and β are constants. For this solution we have

$$R_b = (4\pi w/3M)^{-1/3} \tag{26}$$

$$R_{ah} = (8\pi w/3)^{-1/2}. \tag{27}$$

From (26) and (27) we obtain

$$R_b = R_{ah} \Leftrightarrow 2M/R_b = 1 \tag{28}$$

Glass concludes that for a collapsing model, there is an interior apparent horizon and that after some finite time the boundary will fall through the apparent horizon and an absolute Schwarzschild horizon will remain. This interpretation is not correct, and in fact what happens is the following: First the boundary of the collapsing sphere is outside the Schwarzschild surface and no apparent horizon exists; then the boundary passes the Schwarzschild surface; thereafter there is an apparent horizon in the interior and an absolute Schwarzschild horizon outside the matter configuration. Glass also claims that for an exploding model of this kind, an apparent horizon will develop. From our previous result, we conclude that for an exploding model there exists no horizon, and Glass's interpretation is wrong once more.

V. SOLUTIONS WITH $\partial R/\partial r > 0$

One usually expects the circumference, as measured by an observer riding in a shell of matter, to be an increasing function of radial distance, i.e., $\partial R/\partial r > 0$. Papapetrou[14] also uses this as a condition for a solution to be acceptable when he discusses the collapse of a pressureless dust cloud. But situations may exist where $\partial R/\partial r > 0$ is not fulfilled (cf. Misner[15]). From (2) we now obtain, with w = constant,

$$\frac{\partial}{\partial r}\left(\frac{2m}{R}\right) = \frac{16}{3}\pi w R \frac{\partial R}{\partial r}. \tag{29}$$

We also demand

$$\frac{\partial R}{\partial r} > 0. \tag{30}$$

If the surface of the matter configuration is outside the Schwarzschild surface, from (21), (22), (28), and (29) it follows that $2m/R < 1$ in the interior of the sphere. Hence no interior apparent horizon exists. The argument showing the nonexistence of an exterior apparent horizon is the same as in the case of the conformally flat solutions, and we do not repeat it. We conclude that when the surface of the sphere is outside the Schwarzschild surface, no apparent horizon exists.

When the boundary of the sphere is inside the Schwarzschild surface, an argument similar to the one used for the conformally flat solution gives the result that there exists an interior apparent horizon. But in this case, $2m/R$ is a strictly increasing function, and thus there will exist only one apparent horizon in the interior of the sphere. Hence we conclude that when the surface of the sphere is inside the Schwarzschild surface, there exist two apparent horizons: an outer absolute Schwarzschild horizon and an interior apparent horizon.

[1] P. S. Gupta, Ann. Physik **2**, 421 (1959).
[2] W. B. Bonnor and M. C. Faulkes, Mon. Not. R. Astron. Soc. **137**, 239 (1967).
[3] H. Bondi, Mon. Not. R. Astron. Soc. **142**, 333 (1969).
[4] I. H. Thompson and G. J. Whitrow, Mon. Not. R. Astron. Soc. **139**, 499 (1968).
[5] H. Nariai, Prog. Theor. Phys. **40**, 3, 679 (1968).
[6] A. Banerjee, J. Phys. A **5**, 1305 (1972).
[7] J. Krishna Rao, Gen. Relativ. Gravit. **4**, 5 (1973).
[8] A. Banerjee, Phys. Rev. **150** (4), 1086 (1966).
[9] M. W. Cook, Aust. J. Phys. **28**, 413 (1975).
[10] E. N. Glass, J. Math. Phys. **20**, 1508 (1979).
[11] C. W. Misner and D. H. Sharp, Phys. Rev. **136** (2B), B571 (1964).
[12] W. C. Hernandez and C. W. Misner, Astrophys. J. **143** (2), 452 (1966).
[13] M. Wyman, Phys. Rev. **70**, 396 (1946).
[14] A. Papapetrou, Lectures on General Relativity (Reidel, Boston, MA, 1974).
[15] C. W. Misner, *Astrophysics and General Relativity, Proceedings of the Brandeis University Summer Institute in Theoretical Physics*, edited by M. Chretien, S. Deser, and J. Goldstein (Gordon and Breach, New York, 1968), Vol. 1, p. 113.

Symmetries of the nontwisting type-N solutions with cosmological constant

Humberto Salazar I., Alberto García D.,[a] and Jerzy F. Plebański[b]
Centro de Investigación y de Estudios Avanzados del I.P.N., Departamento de Física, Apdo. Postal 14-740, 07000 México D.F., México

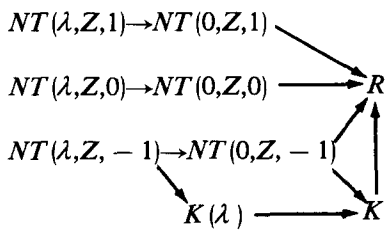
(Received 13 October 1982; accepted for publication 17 December 1982)

The conformal, homothetic, and isometric symmetries of all nontwisting type-N solutions are established. All solutions with λ allow at most the existence of two Killing vectors. All vacuum solutions, except the Robinson metric, permit maximally the existence of two (isometric and homothetic) symmetries. The Robinson solutions are the only ones which allow conformal symmetries.

PACS numbers: 04.20.Jb, 04.90. + e

1. INTRODUCTION

The aim of this paper is to describe systematically all possible symmetries of the nontwisting type-N solutions of the Einstein equations with λ (also allowing λ to be zero). This class of solutions, as was demonstrated in Ref. 1, subdivides into nine essentially different branches according to the scheme of contractions given below.



All these solutions have the property of being determined by an arbitrary complex function f which, in a chart of coordinates $\{X^\mu\} = \{\xi, \bar{\xi}, r, t\}$, depends on ξ and t only, i.e., $f = f(\xi, t)$.

The determination of the symmetries or motions of a given Riemannian space can be achieved by solving the Killing equations

$$\mathcal{L}_K g_{\mu\nu} = \kappa g_{\mu\nu} \tag{1.1}$$

for some function $\kappa = \kappa(X^\alpha)$, where $\mathcal{L}_K$ denotes the Lie derivative with respect to the vector field K. The $g_{\mu\nu}$ are the components of the metric g which, in the null tetrad formalism with signature + 2, is given as

$$g = 2e^1 \otimes e^2 + 2e^3 \otimes e^4, \quad e^2 = (\bar{e}^1), \quad e^3 = \bar{e}^3, \quad e^4 = \bar{e}^4, \tag{1.2}$$

where $e^a \in \Lambda^1$ are the null tetrad vectors. The motion is conformal if the conformal factor κ is a function of the coordinates. The motion is homothetic or isometric according to whether κ is a nonzero constant or zero. Correspondingly, one says that a Riemannian space admits a conformal Killing vector (CKV), a homothetic Killing vector (HKV), or a Killing vector (KV).

For the class of metrics being studied, the problem of searching for symmetries reduces to solving a single constraint equation depending on the vector field K and the structural function $f(\xi, t)$. For arbitrary general function $f(\xi, t)$ all these solutions, except the R metric, have no symmetries. The existence of symmetries is related just to the constraint equation; every function $f(\xi, t)$ satisfying the constraint equation allows at least one Killing direction.

In the next sections, the symmetries for each metric of the scheme of nontwisting solutions are established. The pattern we shall follow consists in giving the cotangent tetrad e^a, the form-invariance metric transformations, the components of the Killing vectors, the conformal factor κ, the constraint equation, the general solution to the constraint equation, and a table of results.

Complex functions will be designated by means of Greek symbols, while real functions will be labeled by Latin symbols; exception is made with respect to the structural function $f(\xi, t)$—complex, $\psi(\xi, \bar{\xi}, t)$—real, the real conformal factor κ, the real cosmological constant λ, and the real function $\omega(t)$ appearing in the transformation of the variable t. To designate constants we shall use the corresponding symbols with a suffix 0.

In the tables the symbols G, κ, $\mathscr{C}$, and $\mathscr{H}$ stand for the maximal group order, the conformal factor, the conformal Killing vector, and the homothetic KV, respectively; the Killing vectors are given simply by their expressions. For the sake of simplicity, the suffix 0 of the constant is dropped; a, b, k, l, m, n are used to designate real constants, while ρ, μ, ν for complex ones. The structural functions quoted in the tables have been reduced to their simplest form by using the transformations of invariance of the metric; they are essentially different from one to another.

2. THE $NT(\lambda, Z, \epsilon)$ AND $NT(0, Z, \epsilon)$ SOLUTIONS

The most general family of nontwisting N solutions, the $NT(\lambda, Z, \epsilon)$ solutions, is given in a chart of coordinates $\{\xi, \bar{\xi}, r, t\}$ by the null tetrad

$$\begin{aligned} e^1 &= \bar{e}^2 = r\,d\xi + (\psi_{\bar{\xi}} - rf)dt, \\ e^3 &= \psi(\xi, \bar{\xi}, t)dt, \\ e^4 &= dr + \left[-\psi_{\xi\bar{\xi}} + \tfrac{1}{2}r(f_\xi + \bar{f}_{\bar{\xi}}) + \tfrac{1}{6}\lambda r^2\psi \right] dt, \end{aligned} \tag{2.1}$$

where $f = f(\xi, t)$ is an arbitrary complex function depending

[a] Also at Sección de Graduados, Escuela Superior de Ingeniería Mecánica y Eléctrica del I.P.N. México, D.F., México.
[b] On leave of absence of the University of Warsaw, Warsaw, Poland.

 0022-2488/83/082191-06$02.50

on ξ and t only. Because the curvature is proportional to $f_{\xi\xi\xi}$, the space is a curved space if $f_{\xi\xi\xi}$ does not vanish. The function ψ satisfying the Liouville equation, without any loss of generality, can be given in the form

$$\psi = 1 + \epsilon\xi\bar{\xi}, \tag{2.2}$$

where the discrete parameter ϵ takes the values 1, 0, -1, depending upon whether the source lines of the gravitational waves are, respectively, timelike, null, or spacelike. The quantities λ and Z appearing in the symbol $NT(\lambda,Z,\epsilon)$ stand for the cosmological constant and the divergence of the congruence e^3.

The nonvanishing component of the conformal curvature, by choosing a coordinate gauge such that ψ is of the form (2.2), amounts to

$$C^{(1)} = -(r\psi)^{-1} f_{\xi\xi\xi}.$$

The corresponding $NT(0,Z,\epsilon)$ vacuum metric is obtained from the $NT(\lambda,Z,\epsilon)$ solutions by equating λ to zero.

The coordinate transformations, which maintain invariant the form of the metric (1.2) with the null tetrad (2.1), are given by

$$t = t(t'), \quad \xi = \xi(\xi',t'), \quad \bar{\xi} = \bar{\xi}(\bar{\xi}',t'), \quad r = (\xi'_{\xi}\bar{\xi}'_{\bar{\xi}})^{1/2} r',$$
$$\psi = t'_t(\xi'_{\xi}\bar{\xi}'_{\bar{\xi}})^{-1/2}\psi', \quad f = \xi_{\xi'} t'_t (f' - t_{t'}\partial_t \xi'). \tag{2.3}$$

For the metric with ψ of the form (2.2), these transformations reduce to

$$t = t' + t_0, \quad \xi = \frac{\overline{\alpha}(t)\xi' - \beta(t)}{\alpha(t) + \epsilon\overline{\beta}(t)\xi'}, \quad \xi' = \frac{\alpha\bar{\xi}' - \overline{\beta}}{\overline{\alpha} + \epsilon\beta\bar{\xi}},$$
$$r = |\alpha + \epsilon\overline{\beta}\xi|^2 r', \quad \alpha\overline{\alpha} + \epsilon\beta\overline{\beta} = 1,$$
$$\psi = |\alpha + \epsilon\overline{\beta}\xi'|^{-2}\psi', \tag{2.4}$$
$$f = (\alpha + \epsilon\overline{\beta}\xi')^{-2}(f' - f_0),$$
$$f_0 := \alpha\dot{\beta} - \dot{\alpha}\beta + [\dot{\alpha}\overline{\alpha} - \ddot{\alpha}\alpha + \epsilon(\dot{\beta}\overline{\beta} - \ddot{\overline{\beta}}\beta)]\xi' + \epsilon(\ddot{\beta}\overline{\alpha} - \ddot{\overline{\alpha}}\overline{\beta})\xi'^2.$$

Searching for motions of the $NT(\lambda,Z,\epsilon)$ field, from the Killing equations one obtains the components of the K vector

$$K^t = a(t) := \tfrac{1}{2}\kappa t + a_0,$$
$$K^{\xi} = \beta(\xi,t) := \epsilon\overline{\alpha}(t)\xi^2 + ib(t)\xi + \alpha(t), \tag{2.5}$$
$$K^{\bar{\xi}} = (\overline{K^{\xi}}),$$
$$K^r = -(r/2)(\beta_{\xi} + \overline{\beta}_{\bar{\xi}} - \kappa).$$

The conformal factor κ for $NT(\lambda,Z,\epsilon)$ solutions must be zero. Thus this class of solutions permits isometries only. For the $NT(0,Z,\epsilon)$ metric, κ is equal to a constant, therefore these solutions permit at most the existence of homothetic Killing vectors.

The components of the K vector and the function $f(\xi,t)$ are restricted by the equation

$$af_t + \beta f_{\xi} + (a_t - \beta_{\xi})f = \beta_t. \tag{2.6}$$

The general solution $f(\xi,t)$, which allows maximally two symmetries, is given by

$$f(\xi,t) = (1/a\Omega)(\xi - \zeta(t))^2[\Phi(\chi) + \psi(\xi,t)], \tag{2.7}$$

where $\zeta(t)$ is any particular solution of the Ricatti equation $\dot{\zeta} = a^{-1}(\alpha(t)\zeta^2 + ib(t)\zeta + \alpha(t))$ (the dot denotes the derivative with respect to t). In order to give the expressions of the other functions appearing in (2.7), it is convenient to define the auxiliary functions

$$\dot{\tilde{\beta}} := \epsilon\dot{\overline{\alpha}}(t)\zeta^2(t) + i\dot{b}(t)\zeta(t) + \dot{\alpha}(t),$$
$$\tilde{\beta}' := 2\epsilon\overline{\alpha}(t)\zeta(t) + ib(t), \tag{2.8}$$
$$\dot{\tilde{\beta}}' := 2\epsilon\dot{\overline{\alpha}}(t)\zeta(t) + i\dot{b}(t);$$

the dot over α and b denotes the derivative with respect to t, while the dot over $\overline{\beta}$ is merely a symbol. In terms of them, the structural functions present in (2.7) are

$$\Omega(t) = \exp\int a^{-1}\tilde{\beta}'\,dt,$$
$$\chi(\xi,t) = (\xi - \zeta(t))^{-1}\Omega + \Pi,$$
$$\Pi(t) = \int \epsilon a^{-1}\overline{\alpha}\Omega\,dt,$$
$$\Psi(\xi,t) = \int [\epsilon\dot{\overline{\alpha}}\Omega - \Pi\dot{\tilde{\beta}}' + \Pi^2\dot{\tilde{\beta}}]\,dt - \chi\int[2\Omega^{-1}\dot{\tilde{\beta}} - \dot{\tilde{\beta}}']\,dt + \chi^2\int\Omega^{-1}\dot{\tilde{\beta}}\,dt. \tag{2.9}$$

[For a equal to zero, the function $f(\xi,t)$ happens to be $f = \beta\int\beta^{-2}\beta_t\,d\xi$.]

For functions $f(\xi,t)$ of the form (2.7), the metric has symmetries (at most two K vectors). For $f(\xi,t)$ outside of the mentioned class, the metric has no symmetries.

Not entering into details, by using the coordinate freedom, one obtains a list of particularly interesting structural functions, which are shown in Table I together with the corresponding symmetries. As was stated before the $NT(\lambda,Z,\epsilon)$ solutions permit only the existence of Killing vectors for $f(\xi,t)$ of the form given by (2.7) with the function a being a constant. Therefore, the isometries of $NT(\lambda,Z,\epsilon)$ are the same as those of the $NT(0,Z,\epsilon)$ solutions.

3. THE $K(\lambda)$ SOLUTIONS

The tetrad of these solutions, which generalize the Kundt metric[2] and reduce to it when λ tends to zero, can be given as

$$e^1 = (\overline{e^2}) = \frac{d\xi}{\cosh x} - r\,dt,$$
$$e^3 = \mu^{-1}\tanh x\,dt, \tag{3.1}$$
$$e^4 = \frac{dr}{\cosh x} + [(\partial_{\xi} + \partial_{\bar{\xi}} - 2\mu\tanh x)(f + \bar{f}) - \mu r^2\tanh x]\,dt,$$

where $x := \mu(\xi + \bar{\xi})$, $\mu := (\frac{1}{6}\lambda)^{1/2}$.

The nonvanishing component of the conformal curvature is

$$C^{(1)} = -2\mu\frac{\cosh^3 x}{\sinh x}(\partial_{\xi}\partial_{\bar{\xi}} - 4\mu^2)f_{\xi}. \tag{3.2}$$

The transformations which maintain the metric invariant are

TABLE I. Symmetries of the $NT(\lambda, Z, \epsilon)$ and $NT(0, Z, \epsilon)$ solutions.

$f(\xi, t)$	$f_{\xi\xi\xi} \neq 0$	G	Homothetic and isometric killing vectors	$\kappa = \text{const}$
$\phi(\xi)$		1	T	0
$t^{-1}\phi(\xi)$		1	$\mathscr{H} = lL$	$2l$
$e^{i\nu t}\xi^{1+\nu}$		1	$T - R$	0
$t^{-1+2i\nu}\xi^{1+\nu}$		1	$\mathscr{H} = l(L - 2R)$	$2l$
ξ^{1+in}		2	$T, \mathscr{H} = l(L + n^{-1}R)$	$2l$
$A^{-1}\dot{A}\xi \ln \xi$		1	$A(t)R$	0
$\xi \ln \xi^n$		2	$T, e^{nt}R$	0
$t^{n-1}\xi \ln \xi^n$		2	$t^n R, \mathscr{H} = lL$	$2l$
$e^{-\xi}, \epsilon = 0$		2	$T, \mathscr{H} = l(L + D)$	$2l$
$t^{-1+\nu}e^{-\nu\xi}$,	$\epsilon = 0$	1	$\mathscr{H} = l(L + D)$	$2l$
$n(1+\xi^2)\arctan\xi$,	$\epsilon \neq 0$	2	$T, e^{nt}M(+)$	0
$nt^{-1}(1+\xi^2)\arctan\xi$,	$\epsilon \neq 0$	2	$t^n M(+), \mathscr{H} = lL$	$2l$
$n(1-\xi^2)\ln\dfrac{1+\xi}{1-\xi}$,	$\epsilon \neq 0$	2	$T, e^{2nt}M(-)$	0
$nt^{-1}(1-\xi^2)\ln\dfrac{1+\xi}{1-\xi}$,	$\epsilon \neq 0$	2	$t^{2n}M(-), \mathscr{H} = lL$	$2l$

$T := \partial_t$, $L := t\partial_t + r\partial_r$, $R = i(\xi\partial_\xi - \bar{\xi}\partial_{\bar{\xi}})$, $D = \partial_\xi + \partial_{\bar{\xi}}$,
$M(\pm) := \frac{1}{2}[(1+i) \pm \epsilon(1-i)]\{(1 \pm \xi^2)\partial_\xi \pm \epsilon(1 \pm \bar{\xi}^2)\partial_{\bar{\xi}} \mp r(\xi \pm \epsilon\bar{\xi})\partial_r\}$

$$t = \int e^{-\omega(t')}dt', \quad \xi = \xi' + iZ_0, \quad \bar{\xi} = \bar{\xi}' - iZ_0,$$
$$r = e^{\omega(t')}(r' - (1/2\mu)\omega_{t'}\sinh x'),$$
$$f = e^{2\omega}\{[f' - (1/8\mu^2)(\omega_{t't'} + \tfrac{1}{2}(\omega_{t'})^2)] + f_0\}, \tag{3.3}$$

where

$$f_0(\xi', t') := \alpha(t')e^{2\mu\xi'} + \overline{\alpha}(t')e^{-2\mu\xi'} + iT(t');$$

this function is such that $C^{(1)}(f_0) = 0$.

By integrating the Killing equations one arrives at the components of the K vector in the form

$$K^t = a(t),$$
$$K^\xi = (\overline{K^\xi}) = ib_0, \tag{3.4}$$
$$K^r = -r\dot{a} + (\ddot{a}/2\mu)\sinh x.$$

The conformal factor κ must be zero. Therefore, this family of solutions permits at most isometries.

The components of K and the function $f(\xi,t)$ are restricted by the constraint equation

$$(a^2 f)_{,t} + ib_0 a^{-1}(a^2 f)_{,\xi} = \ddot{a}a/8\mu^2 + (\dot{\alpha} + 2i\mu b_0 a^{-1})e^{2\mu\xi} + (\dot{\overline{\alpha}} - 2i\mu b_0 a^{-1}\overline{\alpha})e^{-2\mu\xi} + i\dot{T}. \tag{3.5}$$

The general solution of the equation above is given by

$$f(\xi,t) = a^{-2}\left\{\phi\left(\xi - ib_0\int\frac{dt}{a}\right) + \frac{1}{8\mu^2}(\ddot{a}a - \tfrac{1}{2}\dot{a}^2) + f_0(\xi,t)\right\}, \tag{3.6}$$

where

$$f_0(\xi,t) = \alpha(t)e^{2\mu\xi} + \overline{\alpha}(t)e^{-2\mu\xi} + iT(t).$$

Only the metric structures with functions $f(\xi,t)$ of the form (3.6) allow motions.

By using the coordinate freedom, one can establish that the maximal order of the group of isometries is 2. Concrete results are presented in Table II.

TABLE II. Symmetries of the $K(\lambda)$ solutions.

$f(\xi, t)$	$f_{\xi\xi\xi} - 4\mu^2 f_\xi \neq 0$	G	Killing vectors	$\kappa = 0$
$\phi(\xi - ibt)$		1	$T + bR$	
$A(t)\xi$		1	R	
ξ		2	T, R	
$e^{2in\xi}$		2	$T, R + nL$	

$T := \partial_t$, $R := i(\partial_\xi - \partial_{\bar{\xi}})$, $L := t\partial_t - r\partial_r$

4. THE KUNDT SOLUTIONS

The tetrad 1-forms, which describe the Kundt solutions, can be given as

$$e^1 = (\overline{e^2}) = d\xi - r\,dt,$$
$$e^3 = (\xi + \bar{\xi})\,dt, \tag{4.1}$$
$$e^4 = dr + (f + \bar{f} - r)\,dt.$$

The conformal curvature, referred to the tetrad (4.1), is given by

$$C^{(1)} = -(2/\xi + \bar{\xi})f_{\xi\xi}. \tag{4.2}$$

The metric invariant transformations are

$$t = \int e^{-\omega(t')}dt', \quad \xi = \xi' + iz_0, \quad \bar{\xi} = \bar{\xi}' - iz_0,$$
$$r = e^{\omega}[r' - \tfrac{1}{2}(\omega_{t'} - e^{-\omega} + 1)(\xi' + \bar{\xi}')],$$
$$f = e^{2\omega}\{f' + (\xi'/4)(2\omega_{t't'} + (\omega_{t'})^2 + e^{-2\omega} - 1) + iT(t')\}. \tag{4.3}$$

The components of the K vector amount to

$$K^t = a(t),$$
$$K^{\xi} = \overline{(K^{\xi})} = \beta(\xi) := (\kappa/2)\xi + ib_0, \tag{4.4}$$
$$K^r = (\kappa/2 - \dot{a})r + \tfrac{1}{2}(\xi + \bar{\xi})(\ddot{a} + \dot{a}).$$

The conformal factor κ is a constant. Hence, the Kundt solutions permit at most the existence of homothetic Killing vectors.

The components of the K vectors and the structural function $f(\xi,t)$ ought to satisfy the equation

$$af_t + \beta f_{\xi} + (2\dot{a} - \kappa/2)f + \tfrac{1}{2}\xi(\ddot{a} - \dot{a}) = ia^{-1}[\dot{T} - (a^{-1}/2)(\kappa T + 2b_0 S(t))], \tag{4.5}$$

where $S(t) := \frac{1}{4}(2\ddot{a}a - \dot{a}^2 - a^2)$. The general solution of this equation is given by

$$a^2 f = e^{(\kappa/2)\int a^{-1}dt}\phi\left(\ln\beta - \frac{\kappa}{2}\int a^{-1}\,dt\right) - \xi S(t) + iT(t). \tag{4.6}$$

For the existence of isometries, i.e., when κ is equal to zero, the function $f(\xi,t)$ happens to be

$$a^2 f(\xi,t) = \phi\left(\xi - ib_0\int a^{-1}\,dt\right) - \xi S(t) + iT(t).$$

After using the transformations (4.3), one arrives at the canonical forms of the function $f(\xi,t)$ and the corresponding symmetries shown in Table III.

5. THE ROBINSON SOLUTIONS

This section is included in the present paper for the sake of completeness. Most of the results concerned with isometries of the R solutions[3] can also be found in Ref. 4.

The tetrad of the R solutions, the most "degenerate" nontwisting N's, can be given as

$$e^1 = (\overline{e^2}) = d\xi, \quad e^3 = dt, \quad e^4 = dr + (f + \bar{f})\,dt, \tag{5.1}$$

with conformal curvature

$$C^{(1)} = f_{\xi\xi}.$$

The transformations preserving the form of the metric are

$$t = a_0 t' + t_0,$$
$$\xi = e^{-ib_0}[\xi' + \beta(t')], \quad \bar{\xi} = e^{ib_0}[\bar{\xi}' + \bar{\beta}(t')],$$
$$r = a_0^{-1}[r' + R(t') - \xi'\dot{\bar{\beta}} - \bar{\xi}'\dot{\beta}], \tag{5.2}$$
$$f = a_0^{-2}[f' + \xi'\ddot{\bar{\beta}} - \tfrac{1}{2}(\dot{R} + \dot{\beta}\dot{\bar{\beta}}) + iT(t)]$$

(dots denote the derivative with respect to t').

The components of the K vector are

$$K^t = a(t) := l_0 t^2 + m_0 t + n_0,$$
$$K^{\xi} = \overline{(K^{\xi})} = \alpha(t)\xi + \tau(t), \quad \alpha(t) := l_0 t + \alpha_0, \tag{5.3}$$
$$K^r = (\alpha + \bar{\alpha} - \dot{a})r - [\ddot{a}\xi\bar{\xi} + \dot{\tau}\bar{\xi} + \dot{\bar{\tau}}\xi + b(t)].$$

The conformal factor amounts to

$$\kappa = \alpha + \bar{\alpha} = 2(l_0 t + a_0), \quad a_0 = \mathrm{Re}\,\alpha_0. \tag{5.4}$$

The components of K and the function $f(\xi,t)$ ought to fulfill the equation

$$af_t + (\alpha\xi + \tau)f_{\xi} + [2\dot{a} - (\alpha + \bar{\alpha})]f = \ddot{\bar{\tau}}\xi + \tfrac{1}{2}\dot{b} + iP(t). \tag{5.5}$$

It should be noticed, whichever the function $f(\xi,t)$, the metric determined by (5.1), being independent of r, always has the symmetry ∂_r. This fact, as it should be, is also confirmed by the (5.5) and (5.3) equations.

Determining the general solution of (5.5) it is convenient to distinguish the cases of a being equal to zero or different from it.

For $a \neq 0$,

TABLE III. Symmetries of the Kundt solutions.

$f(\xi,t)$	$f_{\xi\xi} \neq 0$	G	Homothetic and isometric killing vectors	$\kappa = \text{const}$
$\phi(\xi - ibt)$		1	$T + bR$	0
$e^{lt}\phi(\xi e^{-lt})$		1	$\mathscr{H} = T + lD$	$2l$
$e^{2in\xi} + \frac{1}{4}\xi$		2	$T, nL + R$	0
$\xi^{n+1} + \frac{1}{4}\xi$		2	$T, \mathscr{H} = l(D - (n/2)L)$	$2l$

$T := \partial_t$, $R := i(\partial_{\xi} - \partial_{\bar{\xi}})$, $D := \xi\partial_{\xi} + \bar{\xi}\partial_{\bar{\xi}} + r\partial_r$, $L := t\partial_t + \frac{1}{2}(\xi + \bar{\xi} - 2r)\partial_r$

$$f=(a^2\sigma\bar{\sigma})^{-1}\left\{\phi\left(\zeta\sigma-\int\tau\sigma a^{-1}\,dt\right)+\xi\sigma\int\ddot{\bar{\tau}}\,\bar{\sigma}a\,dt\right.$$
$$+\frac{1}{2}\int a\dot{b}\sigma\bar{\sigma}\,dt-\int\tau\sigma a^{-1}\left(\int\ddot{\bar{\tau}}\,\bar{\sigma}a\,dt\right)dt$$
$$\left.+i\int\sigma\bar{\sigma}Pa\,dt\right\},\qquad(5.6)$$

where $\sigma(t)=\exp(-\int\alpha a^{-1}\,dt)$.

For $a=0$,

$$f=\phi(t)(\alpha_0\xi+\tau)^{1+\bar{\alpha}_0\alpha_0^{-1}}-(\ddot{\bar{\tau}}/\alpha_0)\xi-(\alpha_0+\bar{\alpha}_0)^{-1}$$
$$\times[\tfrac{1}{2}\dot{b}+(\ddot{\bar{\tau}}/\alpha_0)\tau+iP].\qquad(5.7)$$

An exhaustive list of structural functions $f(\xi,t)$ and the symmetries determined by them are shown in Table IV. The symmetry ∂_r, the existence of which does not depend on the choice of $f(\xi,t)$, will be denoted simply by K. Some Killing directions depending on τ, which in its turn obeys an ordinary linear differential equation of order s, will be designated by $K(s)$, where s is now equal to the maximal number of independent constants.

6. CONCLUSIONS

For completely arbitrary structural functions the nontwisting type-N solutions, except the R metric which always possesses one translation, have no symmetries.

The nontwisting type-N solutions with nonvanishing cosmological constant allow the existence of isometries (at most two Killing vectors) if the structural function $f(\xi,t)$ has the form given by formula (2.7) with κ equal to zero for the $NT(\lambda,Z,\epsilon)$ metrics, and by (3.6) for the $K(\lambda)$ solutions; by the use of coordinate gauges one brings the structural function $f(\xi,t)$ to the form shown in Table I ($\kappa=0$) and II, respectively. The vacuum $NT(0,Z,\epsilon)$ and Kundt solutions with $f(\xi,t)$

TABLE IV. Symmetries of the Robinson solutions.

$f(\xi,t)$	$f_{\xi\xi}\neq0$	G	Conformal, homothetic, and isometric Killing vectors	$\kappa=2(lt+a)$
$\phi(\xi,t)$		1	K	0
$\phi(\xi e^{int})$		2	$K,\ T-nR$	0
$t^{-2}\phi(\xi t^{-1})$		2	$K,\ \mathscr{C}=C$	$2lt$
$t^{-2(1+n)}\phi(\xi t^{n+ib})$		2	$K,\ \mathscr{H}=a(H-(1/n)L+(b/n)R);\ n=0$: $K,\ L-bR$	$2a$
$\frac{\xi^{2\mu}}{(t^2-n)^{1+\mu}}\left(\frac{S(+)}{S(-)}\right)^{\rho}$,	$\mu\neq1$	2	$K,\ \mathscr{C}=C+aH-nT+bR$ $\nu\rho=a(\mu-1)+ib\mu,\ \nu=\sqrt{n},\ S(\pm)=t\pm\sqrt{n}$	$2l(t+a)$
$\xi^{2m/(1+ib)}$		3	$K,\ T,\ \mathscr{H}=a(H+(1-m)L+bR)$	$2a$
$t^{-2}\xi^{-2}$		2	$K,\ L$	0
ξ^{-2}		4	$K,\ T,\ \mathscr{C}=lC+a(2L+H)$	$2(lt+a)$
$t^{\mu}\exp\nu\xi$		2	$K,\ M(\mu)=L-(1/\nu\bar{\nu})[(2+\mu)\bar{\nu}\partial_{\xi}+(2+\bar{\mu})\nu\partial_{\bar{\xi}}]$	0
$\exp\nu\xi$		3	$K,\ M(\mu=0),\ T$	0
$iA(t)\ln\xi;\ A(t)\ln\xi$		2	$K,\ R+2K\int A\,dt;\ K,\ R$	0
$(t^2-n)^{-1}\ln\xi$		3	$K,\ R,\ \mathscr{C}=C-nT-\ln(t^2-n)K$	$2lt$
$t^{-2+n}\ln\xi$,	$n\neq0$	3	$K,\ R,\mathscr{H}(n)=a((2/n)L-(2/n-1)t^{-1+n}K+H)$	$2a$
$\ln\xi$		4	$K,\ R,\ T,\ \mathscr{H}(n=2)$	$2a$
$t^{-2}\ln\xi$		3	$K,\ R,\ L$	0
$\phi(t)\xi^2$		6	$K,\ K(4),\ \mathscr{H}=H$ $K(4):=\tau\partial_{\xi}+\dot{\bar{\tau}}\partial_{\bar{\xi}}-(\dot{\tau}\bar{\xi}+\dot{\bar{\tau}}\partial_{\xi})\partial_r,\ \ddot{\bar{\tau}}=2\phi\tau$	$2a$
$\frac{\xi^2}{(t^2-n)^2}\left(\frac{S(+)}{S(-)}\right)^{ib/\nu}$,	$\nu=\sqrt{n}$	7	$K,\ K(4),\ \mathscr{C}=l(C-nT+bR)+aH$	$2(lt+a)$
$\xi^2\exp2ikt$		7	$K,\ K(4),\ T-kR,\ \mathscr{H}=aH$	$2a$
$t^{-2(1-ik)}\xi^2$		7	$K,\ K(4),\ L-kR,\ \mathscr{H}=aH$	$2a$
$\xi^2t^{-4}\exp2ikt^{-1}$		7	$K,\ K(4),\ \mathscr{C}=l(C-kR)+aH$	$2(lt+a)$

$K:=\partial_r,\ T:=\partial_t,\ D:=\xi\partial_{\xi}+\bar{\xi}\partial_{\bar{\xi}},\ L:=t\partial_t-r\partial_r,\ H:=D+2rK,\ C:=t^2T+tD-2\xi\bar{\xi}K.$

given correspondingly by formulas (2.7) and (4.6) permit maximally the existence of two Killing vectors or one homothetic and one Killing vector; see also Tables I and III, respectively. As is well known the type-N Robinson solutions permit the existence of six Killing vectors for some specific structural functions $f(\xi,t)$. The maximal group order for the conformal Killing vectors is seven, which is in agreement with the results by Collinson and French.[5] The structural functions, which allow symmetries of the R metric, are given by formulas (5.6) and (5.7); a list of representative structural functions with their symmetries are shown in Table IV.

ACKNOWLEDGMENT

One of us (H.S.I.) acknowledges the Consejo Nacional de Ciencia y Technolgía for financial assistance through a doctoral fellowship.

[1]A. García Díaz and J. F. Plebański, J. Math. Phys. **22**, 2655 (1981).
[2]W. Kundt, Z. Phys. **163**, 77 (1961).
[3]I. Robinson, Lecture at King's College, London, 1956; report to the Rayaumont Conference, 1959.
[4]J. Ehlers and W. Kundt, in *Gravitation: An Introduction to Current Research* (Wiley, New York, 1962).
[5]C. D. Collison and D. C. French, J. Math. Phys. **8**, 701 (1967); see also W. D. Halford and R. P. Kerr, *ibid.* **21**, 120 (1980).

On the group structure, GKS and FKG inequalities for Ising models

Bernhard Baumgartner[a)]
Department of Mathematics, University of Virginia, Charlottesville, Virginia 22903

(Received 20 July 1982; accepted for publication 22 October 1982)

The FKG inequality is applied to a lattice of states of the Ising model which are related to subgroups of the group of spin variables. The resulting correlation inequalities are exactly the GKS inequalities.

PACS numbers: 05.50. + q, 02.30.Nq

Both Griffiths' second inequality (also named the second GKS inequality)[1-3] and the FKG inequality[4] have the following feature in common: Given a certain classical measure space X, they ascertain that for an explicitly determined set $\mathscr{M}$ of normalized measures the functions of a certain cone $\mathscr{F}$ are positively correlated:

$$\forall \mu \in \mathscr{M}, \quad \forall f,g \in \mathscr{F}: \ \mu(fg) \geqslant \mu(f)\mu(g). \tag{1}$$

Both inequalities apply to the classical Ising model, where $X = \{-1, +1\}^S$, S a finite point set. The FKG inequality exploits an order structure, whereas the GKS inequalities are most naturally expressed and proved in terms of the group structure of the Ising model.[5-7]

The origin of the group structure is to regard $\{-1, +1\} = \mathbb{Z}_2$ as a multiplicative group and each configuration $s: S \to \mathbb{Z}_2$ as an element of the group

$$G = \mathbb{Z}_2^S$$

with the product

$$(ss')(i) = s(i)s'(i). \tag{2}$$

The coordinate functions σ_i,

$$\sigma_i(s) = s(i), \tag{3}$$

together with their products

$$\sigma_A = \prod_{i \in A} \sigma_i \tag{4}$$

constitute the dual group $\hat{G}$ of characters. This character group is by the association $\sigma_A \leftrightarrow A$ naturally isomorphic to the group of subsets of S, with the symmetric difference

$$A \square B = A \cup B \setminus A \cap B \tag{5}$$

as the group product.

We recall that each function f on G has a unique expansion in characters and that f is positive definite if and only if the coefficients of this expansion are nonnegative.

Let μ_0 be the counting measure on G. The second GKS inequality identifies the sets used in (1) as

$$\mathscr{M} = \{\mu_v : \mu_v(f) = \mu_0(e^V f)/\mu_0(e^V), \ V \text{ is positive definite}\}, \tag{6a}$$

$$\mathscr{F} = \text{set of positive definite functions.} \tag{6b}$$

The FKG inequality deals with functions on a distributive lattice L. The original proof was quite involved with lattice-theoretic notations, but we will be content with the well-known theorem[8] that every finite distributive lattice is isomorphic to a sublattice of a Boolean lattice $P(S)$. This is the set of subsets of S, the lattice functions $\wedge$ and $\vee$ are identified as "intersection" and "union," and the relation $\leqslant$ as "contained in." (A sublattice is a subset which contains with A and B also $A \wedge B$ and $A \vee B$.)

For the FKG inequality, $\mathscr{M}$ is identified as the set of normalized measures on L satisfying the *multiplicative FKG condition*

$$\mathscr{M} = \{\mu : \mu(A)\mu(B) \leqslant \mu(A \wedge B)\mu(A \vee B), \ \mu(L) = 1\}, \tag{7a}$$

and $\mathscr{F}$ is the set of nondecreasing functions,

$$\mathscr{F} = \{f : A \leqslant B \Rightarrow f(A) \leqslant f(B)\}. \tag{7b}$$

The FKG inequality has not only been put to good use in statistical mechanics[9] and quantum field theory,[10] it has also received quite a lot of attention from the side of mathematics. It has been generalized and extended in various directions, and new proofs have been invented.[11-15] The simplest proof so far can be found in a book by Glimm and Jaffe[16] for the special case of pair interactions, and in Ref. 17 for the general case.

In the original application of the FKG inequality to the Ising model, L is identified with the lattice of configurations, ordered by the product order

$$s \leqslant s' \Leftrightarrow \forall\, i : s(i) \leqslant s'(i). \tag{8}$$

$\mathscr{M}$ contains then also the important equilibrium states for ferromagnetic pair interactions, but the only functions which are common to $\mathscr{F}_{\text{FKG}}$ and $\mathscr{F}_{\text{GKS}}$ are the σ_i. In this application there is thus only a small overlap between the two types of inequalities.

We will now identify another lattice of states, so that the FKG inequality, applied to this lattice, yields exactly the GKS inequalities.

Let V be positive definite

$$V = \sum_{A \subset S} J_A \sigma_A, \quad J_A \geqslant 0, \tag{9}$$

then

$$e^{J_A \sigma_A} = \cosh J_A (1 + \sigma_A \tanh J_A).$$

With $\lambda_A = \tanh J_A$, we have

$$e^V = \left(\prod_A \cosh J_A\right) \prod_{A \in P(S)} [(1 - \lambda_A) + \lambda_A(1 + \sigma_A)]. \tag{10}$$

Lemma:

$$e^V/\mu_0(e^V) = \sum_{H \subset G} \lambda_H \sum_{\sigma \in H} \sigma, \tag{11}$$

[a)] Work supported by a Max Kade Foundation Fellowship Grant. On leave of absence from the Institute für Theoretische Physik der Universität Wien, Boltzmanng. 5, A-1090 Wien, Austria.

 0022-2488/83/082197-03$02.50

where H denotes subgroups of G, $0 \leqslant \lambda_H \leqslant 1$, $\Sigma \lambda_H = 1$.

Proof: The λ_A in (10) are positive but smaller than 1. Expanding the product in (10) yields a sum of terms $\prod_{A \in \mathscr{A} \subset P(S)} (1 + \sigma_A)$ with positive coefficients. By induction on the number of factors one sees that this yields a sum $\Sigma_H \sigma$: $\{1, \sigma_A\}$ is itself a subgroup of G, and

$$(1+\sigma_A)\sum_H \sigma = \begin{cases} 2\sum_H \sigma & \text{if } \sigma_A \in H \\ \sum_{H \cup \sigma_A H} \sigma & \text{if } \sigma_A \notin H. \end{cases}$$

$H \cup \sigma_A H$ is again a group since $\sigma_A^2 = 1$. One can even calculate the coefficients:

$$\prod_{A \in \mathscr{A} \subset P(S)} (1+\sigma_A) = 2^{|\mathscr{A}|}/|H_{\mathscr{A}}| \sum_{H_{\mathscr{A}}} \sigma, \tag{12}$$

with $H_{\mathscr{A}}$ = the subgroup of G generated by $\{\sigma_A, A \in \mathscr{A}\}$. Since $\mu_0(\Sigma_H \sigma) = 1$, the normalization of e^V makes $\Sigma \lambda_H = 1$. □

Definition: μ_H and $\mu_{\mathscr{A}}$ are defined by

$$\mu_H(f) = \mu_0\left(\sum_{\sigma \in H} \sigma f\right) \quad \text{and} \quad \mu_{\mathscr{A}} = \mu_{H_{\mathscr{A}}}. \tag{13}$$

($\mu_{\mathscr{A}}$ can be regarded as the ground state for $V = \Sigma_{A \in \mathscr{A}} \sigma_A$.) Observe that the μ_H are positive-definite measures. In fact, we have the following:

Conjecture: The μ_H are presumably the extreme points of the convex set of normalized positive and positive-definite measures.

We remark that positive definiteness plays an important role in the derivation of Griffiths' inequalities.[6] A better knowledge of the set of positive and positive-definite functions (or measures) on groups might thus lead to additional insight into the structure of ferromagnetic states. In the Ising model the group is just a product of $\mathbb{Z}_2$ factors. For other groups, the indicator functions of subgroups are also extreme points of the set in question, but certainly not the only ones.

The subgroups of G form a lattice under the order given by inclusion:

$$\begin{aligned} &H \leqslant K \Leftrightarrow H \subseteq K, \\ &H \wedge K = H \cap K, \\ &H \vee K = \text{the group generated by } H \text{ and } K. \end{aligned} \tag{14}$$

This lattice is not distributive, so we cannot immediately apply the FKG inequality. We therefore imbed this lattice in $P(P(S))$ or, rather, we associate each H with all the $\mathscr{A}$ such that $H = H_{\mathscr{A}}$. According to (10) and (12), we have

$$e^V = \left(\prod_A \cosh J_A\right) \sum_{\mathscr{A} \in P(P(S))} \prod_{B \notin A} (1-\lambda_B) \times \prod_{A \in \mathscr{A}} (2\lambda_A)/|H_{\mathscr{A}}| \mu_{\mathscr{A}}. \tag{15}$$

Lemma: The coefficients

$$\lambda_{\mathscr{A}} = \prod_{B \notin \mathscr{A}} (1-\lambda_B) \prod_{A \in \mathscr{A}} (2\lambda_A)/|H_{\mathscr{A}}| \tag{16}$$

satisfy the multiplicative FKG condition.

Proof: The factors $(1-\lambda_B)$ and $2\lambda_A$ give the same contributions to $\lambda_{\mathscr{A} \wedge \mathscr{B}} \lambda_{\mathscr{A} \vee \mathscr{B}}$ as to $\lambda_{\mathscr{A}} \lambda_{\mathscr{B}}$. It remains to show that

$$|H_{\mathscr{A} \vee \mathscr{B}}|\,|H_{\mathscr{A} \wedge \mathscr{B}}| \leqslant |H_{\mathscr{A}}|\,|H_{\mathscr{B}}|. \tag{17}$$

Let $\mathscr{C} \subset \mathscr{A} \wedge \mathscr{B}$ be a minimal generating set for $H_{\mathscr{A} \wedge \mathscr{B}}$, that is, $H_{\mathscr{C}} = H_{\mathscr{A} \wedge \mathscr{B}}$, but no proper subset of $\mathscr{C}$ generates $H_{\mathscr{A} \wedge \mathscr{B}}$. Then

$$|H_{\mathscr{A} \wedge \mathscr{B}}| = 2^{|\mathscr{C}|}. \tag{18}$$

$\mathscr{C}$ can be enlarged to minimal generating sets $\mathscr{C} \cup \mathscr{D}$ for $H_{\mathscr{A}}$ and $\mathscr{C} \cup \mathscr{E}$ for $H_{\mathscr{B}}$. Note that $\mathscr{D} \cap \mathscr{E} = \emptyset$,

$$|H_{\mathscr{A}}| = 2^{|\mathscr{C}|+|\mathscr{D}|}, \quad |H_{\mathscr{B}}| = 2^{|\mathscr{C}|+|\mathscr{E}|}. \tag{19}$$

Now $\mathscr{C} \cup \mathscr{D} \cup \mathscr{E}$ is a generating set for $H_{\mathscr{A} \vee \mathscr{B}}$, but not necessarily minimal. Thus

$$|H_{\mathscr{A} \vee \mathscr{B}}| \leqslant 2^{|\mathscr{C}|+|\mathscr{D}|+|\mathscr{E}|}. \tag{20}$$

Formulas (18)–(20) yield (17), and this completes the proof. □

Since normalization factors play no role in the FKG condition, we have now a measure $\overline{\mu}_V$ on the lattice $P(P(S))$, defined by

$$\overline{\mu}_V(\mathscr{A}) = \lambda_{\mathscr{A}} \Big/ \sum_{P(P(S))} \lambda_{\mathscr{A}}, \tag{21}$$

satisfying the multiplicative FKG condition.

Next we have to associate with each observable in the Ising model a function on $P(P(S))$. Since

$$\mu_{\mathscr{A}}(\sigma_{\mathscr{A}}) = \begin{cases} 1, & \sigma_A \in H_{\mathscr{A}}, \\ 0, & \sigma_A \notin H_{\mathscr{A}}, \end{cases}$$

we map σ_A to $\bar{\sigma}_A$ with

$$\begin{aligned} \bar{\sigma}_A(\mathscr{A}) &= 1 \quad \text{if and only if } \sigma_A \in H_{\mathscr{A}} \\ &= 0 \quad \text{otherwise.} \end{aligned} \tag{22}$$

Then we hve

$$\mu_V(\sigma_A) = \overline{\mu}_V(\bar{\sigma}_A).$$

The mapping of the σ_A can be extended to a linear transformation of the set of functions, but it is not an algebraic homomorphism. A difference arises when $\sigma_{\mathscr{A} \square \mathscr{B}} = \sigma_A \sigma_B \in H_{\mathscr{A}}$, but $\sigma_A, \sigma_B \notin H_{\mathscr{A}}$. The opposite case is impossible, since $H_{\mathscr{A}}$ is a group. Therefore,

$$\bar{\sigma}_{\mathscr{A} \square \mathscr{B}} \geqslant \bar{\sigma}_A \bar{\sigma}_B. \tag{23}$$

This inequality alone is in fact the second GKS inequality for the state μ_A. Together with the FKG inequality for $P(P(S))$ it yields the second GKS inequality for all μ_V with positive-definite V:

$$\begin{aligned} \mu_V(\sigma_A \sigma_B) &= \overline{\mu}_V(\bar{\sigma}_{\mathscr{A} \square \mathscr{B}}) \geqslant \overline{\mu}_V(\bar{\sigma}_A \bar{\sigma}_B) \geqslant \overline{\mu}_V(\bar{\sigma}_A)\overline{\mu}_V(\bar{\sigma}_B) \\ &= \mu_V(\sigma_A)\mu_V(\sigma_B). \end{aligned} \tag{24}$$

Since the $\bar{\sigma}_A$ are increasing functions,

$$\begin{aligned} \sigma_A(\mathscr{A}) = 1 &\Leftrightarrow \sigma_A \in H_{\mathscr{A}} \Rightarrow \forall\ \mathscr{B} \supset \mathscr{A} : \sigma_A \in H_{\mathscr{B}} \\ &\Rightarrow \forall\ \mathscr{B} \supset \mathscr{A} : \bar{\sigma}_A(\mathscr{B}) = 1. \end{aligned}$$

The set of increasing functions on $P(P(S))$ which are also images under Φ, the extension of $\sigma_A \to \bar{\sigma}_A$, is, of course, precisely the image of positive-definite functions. Let

$$f = \sum c_A \sigma_A;$$

then

$$\Phi f = \sum c_A \bar{\sigma}_A .$$

This is increasing if all the c_A are $\geqslant 0$. But assume $c_B < 0$; then, for $\mathscr{B} = \{1, B\}$,

$$(\Phi f)(\mathscr{B}) = c_\varnothing - c_B ,$$

but this is smaller than $(\Phi f)(\{1\}) = c_\varnothing$, so Φf is not increasing.

[1]R. Griffiths, "Correlations in Ferromagnets. I," J. Math. Phys. **8**, 478–83 (1967).

[2]R. Griffiths, "Correlations in Feromagnets. II," J. Math. Phys. **8**, 484–9 (1967).

[3]D. Kelley and S. Sherman, "General Griffiths Inequalities on Correlations in Ising Ferromagnets," J. Math. Phys. **9**, 466–84 (1968).

[4]C. Fortuin, P. W. Kasteleyn, and J. Ginibre, "Correlation Inequalities on Some Partially Ordered Sets," Comm. Math. Phys. **22**, 89–103 (1971).

[5]S. Sherman, "Cosets and Ferromagnetic Correlation Inequalitites," Comm. Math. Phys. **14**, 1–4 (1969).

[6]J. Ginibre, "General Formulation of Griffiths' Inequalities," Comm. Math. Phys. **16**, 310–28 (1970).

[7]D. Cartier, "Inegalits de Correlation en Mécanique Statistique," in *Lecture Notes in Mathematics, No. 383* (Springer-Verlag, Berlin-Heidelberg-New York, 1973), pp. 242-64.

[8]G. Birkhoff, *Lattice Theory* (American Mathematical Society, Providence, RI, 1967).

[9]J. L. Lebowitz, "Bounds on the Correlations and Analyticity Properties of Ferromagnetic Ising Spin Systems," Comm. Math. Phys. **28**, 313-21 (1972).

[10]B. Simon and R. B. Griffiths, "The $(\phi^4)_2$ Theory as a Classical Ising Model," Comm. Math. Phys. **33**, 145–64 (1973).

[11]R. Holley "Remarks on the FKG Inequalities," Comm. Math. Phys. **36**, 227–31 (1974).

[12]C. J. Preston, "A Generalization of the FKG Inequality," Comm. Math. Phys. **36**, 233–41 (1979).

[13]T. E. Harris, "A Correlation Inequality for Markov Processes in Partially Ordered State Spaces," Ann. Probability **5**, 451–59 (1977).

[14]J. H. B. Kemperman, "On the FKG Inequality for Measures on a Partially Ordered Space," Proc. K. Ned. Aka. Wet Ser. A **80**, 313–31 (1977).

[15]C. J. K. Batty and H. W. Bollman, "Generalized Holley-Preston Inequalities on Measure Spaces and Their Products," Z. Wahr. **53**, 157–73 (1980)

[16]J. Glimm and A. Jaffe, *Quantum Physics, A Functional Integral Point of View* (Princeton U.P., Princeton, NJ, 1981).

[17]G. A. Battle and L. Rosen, "The FKG Inequality for the Yukawa$_2$ Quantum Field Theory," J. Stat. Phys. **22**, 123–92 (1980).

A generalization of the proof of the triviality of scalar field theories

Tetsuya Hattori
Department of Physics, University of Tokyo, Tokyo 113, Japan

(Received 1 February 1983; accepted for publication 15 April 1983)

The proof of the triviality of ϕ^4 theory in $d>4$ dimensions is generalized. We show that the triviality of scalar field theories holds for a wide class of potentials.

PACS numbers: 11.10.Ef, 05.50. + q

I. INTRODUCTION

Recently, Aizenman[1] and Fröhlich[2] rigorously proved the triviality of scalar ϕ^4 theory in $d>4$ dimensions. Their proofs are essentially based on the observation that the propagation of scalar particles with full quantum corrections ("dressed" particles) can be represented by a weighted sum of random walks and that the interaction of two particles (the connected four-point functions) are bounded by the intersection probability of two random walks, which approaches zero in the continuum limit in $d>4$ dimensions.

It is, of course, of most interest to see how these results can be extended to $d=4$ dimensions, and the above cited authors are now making progress along this line. Here, we wish to investigate a different extension: the generalization of the form of the potential. Our main interest here is to see how far the original proof of triviality works when we try to use potentials other than ϕ^4 potentials. We expect that these trials may reveal some features of the mechanisms of triviality of scalar field theories.

In the following, for definiteness, we will follow the methods of Fröhlich. We will derive the same form of inequality which Fröhlich found for ϕ^4 theory in a wider class of potentials.

II. FORMULATION

In the following, we always work on the $d>4$ dimensional simple hypercubic lattice. On each site j, we put a field variable $\phi_j \in R$. The standard Euclidean lattice quantum scalar field theory is defined by the probability measure

$$d\mu = Z^{-1} \exp(\beta \sum \phi_i \phi_j) \prod_i d\rho(\phi_i), \tag{1}$$

where $\beta\,(>0)$ is a constant (related to the field strength renormalization constant), the sum is over pairs of nearest neighbor sites i and j, and $d\rho(\phi_i)$ is a single-site measure of the form

$$d\rho(\phi) = \exp(-V(\phi^2))d\phi, \tag{2}$$

where $V(\phi^2)$ is a real analytic even function of ϕ (the potential term). Further constraints on the form of V will be imposed later. Z is a constant which normalizes the measure

$$\int d\mu = 1.$$

The correlation functions are given by

$$\langle \phi_i \phi_j \cdots \phi_l \rangle = \int \phi_i \phi_j \cdots \phi_l \, d\mu. \tag{3}$$

The continuum limit is approached in the following way:

(1) Write the coordinates that enter Eq. (3) in physical units. For example, replace i by x_i/a.

(2) Fix the coordinates written in physical units (e.g., x_i).

(3) Let $a \to 0$.

The renormalization procedure is obtained by letting the parameters which enter Eq. (1) change as we change a, so that certain correlation functions satisfy some input conditions (renormalization conditions). Triviality is a statement that, by imposing physically acceptable renormalization conditions on two-point functions (so that it can describe a propagation of a nonnegative mass squared particle when analytically continued into Minkowski space), all other connected n-point functions vanish in the continuum limit.

Now we can summarize the renormalization conditions which Fröhlich imposed (we write the dependence on lattice spacing a explicitly in the following):

$$0 \leqslant \zeta = \zeta(a) < 1, \tag{4}$$

where $\zeta(a) = \beta(a) a^{2-d}$.

$$0 < \langle \phi_0 \phi_x \rangle_a < c/\zeta(a)(|x| + a)^{2-d} \quad \text{(for all } x\text{)}, \tag{5}$$

where x is the position of lattice site written in physical units, and c is a constant that depends only on space-time dimension d.

$$\langle \phi_0 \phi_x \rangle_a < N \quad \text{(for all } |x| > 1\text{)}, \tag{6}$$

where N is a finite constant independent of x and a.

The meaning of Eqs. (4) and (5) is the following. There exists a theorem of infrared bounds[3] in the form[4]

$$0 < \langle \phi_0 \phi_j \rangle \leqslant M^2 + c/\beta(|j| + 1)^{2-d},$$

where M is a constant that depends on β and on the form of V. Thus the renormalization condition that we imposed through Eq. (4) is that $\zeta(a)$ should not blow up as $a \to 0$ so the two-point function should not disappear, and Eq. (5) means that $M^2 = 0$, which means there should be no spontaneous magnetization: we need not redefine the measure to obtain a unique vacuum.

The last condition, Eq. (6), means that the two-point function should not blow up with increasing $|x|$. If we adopt the usual renormalization condition which we use in perturbation theory, we may impose a stronger condition that the two-point function should decrease with increasing $|x|$.

The essential step of the proof of the triviality of ϕ^4 theory due to Fröhlich is the proof of the following inequality:

$$0 > u^4(x_1, x_2, x_3, x_4) > -\beta(a)^2 \sum_p \sum_{z:z',z''} \langle \phi_{x_{p(1)}} \phi_z \rangle_a \times \langle \phi_{z'} \phi_{x_{p(2)}} \rangle_a \langle \phi_{x_{p(3)}} \phi_z \rangle_a \langle \phi_{z''} \phi_{x_{p(4)}} \rangle_a + E(a), \tag{7}$$

 0022-2488/83/082200-04$02.50

where Σ_p is a sum over possible ways of pairing x_1,x_2,x_3,x_4, $\Sigma_{z:z',z''}$ is a sum where z runs over the whole range of lattice sites, and z',z'' are nearest neighbors of z. The connected four-point function u^4 is defined by

$$u^4(x_1,x_2,x_3,x_4) = \langle \phi_{x_1}\phi_{x_2}\phi_{x_3}\phi_{x_4}\rangle_a - \sum_p \langle \phi_{x_{p(1)}}\phi_{x_{p(2)}}\rangle_a \langle \phi_{x_{p(3)}}\phi_{x_{p(4)}}\rangle_a .$$

The term $E(a)$ in Eq. (7) takes care of corrections which emerge when we consider the case where z coincides with x_i. It approaches zero as a^{d-2} under the renormalization conditions.

It is now a direct consequence of Eqs. (4)–(7) that

$$\lim u^4(x_1,x_2,x_3,x_4) = 0 \tag{8}$$

for fixed x_i's, separated from each other. Equation (8) [and similar results for arbitrary $n(\geqslant 3)$-point functions] implies triviality of scalar field theories.

The theorem of infrared bounds is known to hold for a wide range of potentials $V(x)$. It holds if $V(x)$ is continuous for $[0,\infty)$ and if $V(x)$ increases sufficiently fast as x increases [$V(x)/x \to \infty$ $(x \to \infty)$ is enough]. The first inequality (Lebowitz inequality[5]) in Eq. (7) is also known to hold for some range of potential forms[6,7]:

$$\frac{d^3V(\phi^2)}{d\phi^3} \geqslant 0 \quad \text{for } \phi \geqslant 0, \tag{9}$$

or

$$V'(x) \geqslant 0, \quad V''(x) \geqslant 0 \quad \text{for } x \geqslant 0. \tag{10}$$

On the other hand, the second inequality of Eq. (7) (Fröhlich inequality) has, so far, been proved only for the case of ϕ^4 theories:

$$V(x) = \mu x + \lambda x^2, \quad \lambda > 0. \tag{11}$$

We will, in the following, extend these previous results and see that Eq. (7) holds for the case

$$V'(x) \geqslant -m, \quad V''(x) \geqslant 0 \quad \text{for } x \geqslant 0, \tag{12}$$

where m is an arbitrary finite constant. Therefore the triviality of scalar field theory holds whenever the potential satisfies Eq. (12). The main tool we use is correlation inequalities (particularly the GKS inequality[8]) and the random walk representations.[7]

III. PROOF OF THE INEQUALITY

The Lebowitz inequality for the potentials satisfying Eq. (12) is a trivial extension of Ref. 7. There, the random walk representation for correlation functions is employed. The random walk representation expresses correlation functions through weighted sums of partition functions for locally biased potentials. Its explicit form is given by

$$\langle \phi_i F\rangle = \sum_{w:i-j} \beta^{|w|} \int d\nu_w(t) \left\langle \frac{\partial F}{\partial \phi_j}\right\rangle_t Z_t/Z, \tag{13}$$

where F is an arbitrary function of field variables, and $w:i-j$ is a random walk from site i to site j, which is an ordered set of sites:

$$w:i-j = (i = i_1,i_2,i_3,\cdots,i_n = j),$$

where i_x and i_{x+1} are nearest neighbor sites. $|w|$ is defined by

$$|w| = \sum_i n(i,w),$$

where $n(i,w)$ is the number of times that a site i appears in the left-hand side of Eq. (14). t is a short-hand notation for a set of real variables $\{t_i : i \text{ lattice sites}\}$, and $d\nu_w(t)$ is defined by

$$d\nu_w(t) = \prod_i d\nu_{n(i,w)}(t_i),$$

$$d\nu_n(x) = \begin{cases} \delta(x)dx & \text{if } n = 0 \\ x^{n-1}/(n-1)!\,\theta(x)dx & \text{if } n = 1,2,3,\ldots \end{cases}.$$

Z_t is a partition function for biased potential

$$Z_t = \int \exp(\beta \sum \phi_i\phi_j) \prod_i (\exp(-V(\phi_i^2 + 2t_i))d\phi_i).$$

$\langle\cdot\rangle$ is an expectation value with respect to the biased measure

$$\langle\cdot\rangle_t = \int \cdot \exp\left(\beta \sum \phi_i\phi_j\right) \times \prod_i \left(\exp \frac{(-V(\phi_i^2 + 2t_i))d\phi_i}{Z_t}\right). \tag{14}$$

We will use the random walk representation for the two-point function and the connected four-point function. If we put $F = \phi_j$ in Eq. (13) we obtain

$$\langle \phi_i\phi_j\rangle = \sum_{w:i-j} \beta^{|w|} \int d\nu_w(t) \frac{Z_t}{Z}. \tag{15}$$

If we put $F = \phi_j\phi_k\phi_l$, we obtain the random walk representation for the four-point function. Combining it with Eq. (15), we obtain

$$u^4(i,j,k,l) = \sum_{w:i-j} \beta^{|w|-1} \int d\nu_w(t) \times (\langle \phi_k\phi_l\rangle_t - \langle \phi_k\phi_l\rangle) Z_t/Z + (\text{terms with } j \text{ and } k \text{ interchanged}) + (j-l \text{ interchanged}). \tag{16}$$

From these definitions,

$$\partial\langle \phi_k\phi_l\rangle_t/\partial t_i = -2(\langle V'(\phi_i^2 + 2t_i)\phi_k\phi_l\rangle_t - \langle V'(\phi_i^2 + 2t_i)\rangle_t \langle \phi_k\phi_l\rangle_t)$$

since $\langle m\rangle_t = m$, where m is the constant which appeared in Eq. (12). The GKS inequality can be applied here in the form

$$\langle \phi_k\phi_l f(\phi_i)\rangle_t \geqslant \langle \phi_k\phi_l\rangle_t \langle f(\phi_i)\rangle_t$$

for $f(x)$ with $f(-x) = f(x) \geqslant 0$ and $df/dx \geqslant 0$ for $x \geqslant 0$. Under the condition (12), $f(x) = V'(x^2 + 2t_i) + m$ has the desired properties, and therefore

$$\partial\langle \phi_k\phi_l\rangle_t/\partial t_i \leqslant 0,$$

and

$$\langle \phi_k\phi_l\rangle_t \leqslant \langle \phi_k\phi_l\rangle. \tag{17}$$

From Eq. (16), $u^4 \leqslant 0$. Thus the Lebowitz inequality is proved.

Next we consider the proof of the Fröhlich inequality. First, we prove that for $t_1 = \{t_{1i} \geqslant 0\}$ and $t_2 = \{t_{2i} \geqslant 0\}$,

$$ZZ_{t_1+t_2} - Z_{t_1}Z_{t_2} \geqslant 0 \quad \text{if} \quad t_{1i}t_{2i} = 0 \quad \text{for all } i. \tag{18}$$

We prove this by noting that

$$\ln\left(\frac{ZZ_{t_1+t_2}}{Z_{t_1}Z_{t_2}}\right)=\int_0^1 ds_1\int_0^1 ds_2\,\frac{d}{ds_1}\frac{d}{ds_2}\times(\ln Z(t_{1i}s_1+t_{2i}s_2)).$$

Writing $t_i=t_{1i}s_1+t_{2i}s_2$, and performing the differentiation with respect to s_1 and s_2,

$$\ln\left(\frac{ZZ_{t_1+t_2}}{Z_{t_1}Z_{t_2}}\right)=\int_0^1 ds_1\int_0^1 ds_2\sum_i\sum_j(4t_{1i}t_{2j})\times(\langle V'(\phi_i^2+2t_i)V'(\phi_j^2+2t_j)\rangle_t-\langle V'(\phi_i^2+2t_i)\rangle_t\langle V'(\phi_j^2+2t_j)\rangle_t),$$

where the nonappearance of a factor V'' comes from $t_{1i}t_{2i}=0$. We use the GKS inequality, this time in the form

$$\langle f(\phi_i)f(\phi_j)\rangle_t\geqslant\langle f(\phi_i)\rangle_t\langle f(\phi_j)\rangle_t$$

for $f(x)$ with $f(-x)=f(x)\geqslant 0$ and $df/dx\geqslant 0$ for $x\geqslant 0$. Under the condition (12), $f(x)=V'(x^2+2t_i)+m$ has the desired properties and

$$\langle V'(\phi_i^2+2t_i)V'(\phi_j^2+2t_j)\rangle_t\geqslant\langle V'(\phi_i^2+2t_i)\rangle_t\langle V'(\phi_j^2+2t_j)\rangle_t.$$

Therefore $\ln(ZZ_{t_1+t_2}/Z_{t_1}Z_{t_2})\geqslant 0$, and $ZZ_{t_1+t_2}\geqslant Z_{t_1}Z_{t_2}$ if $t_{1i}t_{2i}=0$. The condition $t_{1i}t_{2i}=0$ means that the two random walks $w1$ and $w2$ do not intersect. This is essential for the proof of Eq. (18) because when we write Eq. (17) in the random walk representation,

$$\langle\phi_i\phi_j\rangle_{t_1}-\langle\phi_i\phi_j\rangle=\sum_{w:i\to j}\beta^{|w|}\int d\nu_w(t_2)\left(\frac{Z_{t_1+t_2}}{Z_{t_1}}-\frac{Z_{t_2}}{Z}\right)=\sum_{w:i\to j}\beta^{|w|}\int d\nu_w(t_2)\times(ZZ_{t_1+t_2}-Z_{t_1}Z_{t_2})\leqslant 0,$$

the direction of the inequality is reversed when we include the terms where the two random walks intersect. This means that the contribution from the terms where $w1$ and $w2$ do not meet (we may call it "vacuum bubble effects") tends to increase the value of u^4, so that we get a lower bound estimate by dropping these terms and taking into account only those terms where $w1$ and $w2$ intersect ("real particle effects"). This is a crucial step in applying the notion of the intersection probability of two random walks.

The remainder of the proof goes in a form similar to that in Ref. 2. With the expressions of Eqs. (15) and (16),

$$u^4=\sum_{w1:i\to j}\sum_{w2:k\to l}\beta^{|w1|+|w2|-2}\int d\nu_{w1}(t_1)d\nu_{w2}(t_2)\times\left(\frac{Z_{t_1+t_2}}{Z_{t_1}}-\frac{Z_{t_2}}{Z}\right)\frac{Z_{t_1}}{Z}+(j-k)+(j-l)$$

$$\geqslant\sum_{w1\wedge w2\neq\varnothing}\sum\beta^{|w1|+|w2|-2}\int d\nu_{w1}(t_1)d\nu_{w2}(t_2)\times\left(\frac{Z_{t_1+t_2}}{Z}-\frac{Z_{t_1}Z_{t_2}}{Z^2}\right)+(j-k)+(j-l)$$

$$\geqslant-\sum_{w1\wedge w2\neq\varnothing}\sum\beta^{|w1|+|w2|-2}\int d\nu_{w1}(t_1)d\nu_{w2}(t_2)\frac{Z_{t_1}Z_{t_2}}{Z^2}+(j-k)+(j-l)$$

$$\geqslant-\sum_z\sum_{w1\wedge w2\ni z}\sum\beta^{|w1|+|w2|-2}\int d\nu_{w1}(t_1)d\nu_{w2}(t_2)\times\frac{Z_{t_1}Z_{t_2}}{Z^2}+(j-k)+(j-l),$$

where we used Eq. (18) and the positivity of partition functions. We make use of the decomposition of random walks described in Ref. 2. The result is

$$u^4\geqslant-\beta^2\sum_{z:z',z''}\sum_{w1':i\to z}\sum_{w1'':z'\to j}\sum_{w2':k\to z}\sum_{w2'':z''\to l}\beta^{|w1'|-1}\int d\nu_{w1'}(t_1')\,\beta^{|w1''|-1}\int d\nu_{w1''}(t_1'')\frac{Z_{t_1'+t_1''}}{Z}\times\beta^{|w2'|-1}\int d\nu_{w2'}(t_2')\,\beta^{|w2''|-1}\int d\nu_{w2''}(t_2'')\frac{Z_{t_2'+t_2''}}{Z}+(j-k)+(j-l)+E(a).$$

The random walk representation for the two-point function gives

$$u^4=-\beta^2\sum_{z:z',z''}\sum_{w1':i\to z}\beta^{|w1'|-1}\int d\nu_{w1'}(t_1')\,\langle\phi_{z'}\phi_j\rangle t_1'\frac{Z_{t_1'}}{Z}\times\sum_{w2':k\to z}\beta^{|w2'|-1}\int d\nu_{w2'}(t_2')\langle\phi_{z''}\phi_l\rangle\frac{Z_{t_2'}}{Z}+(j-k)+(j-l)+E(a).$$

Using Eqs. (17) and (15), we obtain

$$u^4\geqslant-\beta^2\sum_{z:z',z''}\langle\phi_{z'}\phi_j\rangle\langle\phi_i\phi_z\rangle\langle\phi_{z''}\phi_l\rangle\langle\phi_k\phi_z\rangle+(j-k)+(j-l)+E(a)$$

which proves the Fröhlich inequality (7).

IV. CONCLUSIONS

We have seen that within the framework of Fröhlich, triviality can be proved for the scalar field theory with potential V satisfying Eq. (12). The first inequality of Eq. (12),

$$V'(x)\geqslant -m,$$

permits essentially any form of potential, since m can be as large as we wish. Typically, this leads to the consequence that for the potential of polynomial type

$$V(x)=bx+cx^2+\cdots+dx^n,\qquad (V(\phi^2)=b\phi^2+c\phi^4+\cdots+d\phi^{2n}),\tag{19}$$

b can be either positive or negative (if we had adopted the constraint Eq. (10), $V'(x)\simeq b\geqslant 0$ for $x\geqslant 0$ and negative b could not have been allowed). On the other hand, the second inequality of Eq. (12)

$$V''(x)\geqslant 0,$$

is a stronger constraint. In the case of the potentials of the form Eq. (19), this implies, especially,

$$V''(x) = c + O(x) \geqslant 0 \quad \text{for } x \text{ near zero,}$$

and the potentials with negative c are excluded.

At present, the author does not know whether this constraint is of any physical meaning. It is, however, interesting to see that in the approach of Aizenman,[1] there arise restrictions of similar kind. In this approach, Aizenman approximates the scalar field theory by Ising models, the method which is known[9] to break down, for example, in the case of ϕ^6 theories with negative ϕ^4 coefficients.

ACKNOWLEDGMENTS

The author would like to thank K. Higashijima and K. Hattori for discussions, Professor J. Fröhlich and Professor B. Simon for instructive correspondence, and Professor T. Eguchi and K. Hikasa for careful reading of the manuscript.

[1] M. Aizenman, Phys. Rev. Lett. **47**, 1 (1981); Commun. Math. Phys. **86**, 1 (1982).
[2] J. Fröhlich, Nucl. Phys. B **200 [FS4]**, 297 (1982).
[3] J. Fröhlich, B. Simon, and T. Spencer, Commun. Math. Phys. **50**, 79 (1976).
[4] A. Sokal, Ann. Inst. Henri Poincaré (to be published).
[5] J. L. Lebowitz, Commun. Math. Phys. **28**, 313 (1972).
[6] R. S. Ellis, J. L. Monroe, and C. M. Newman, Commun. Math. Phys. **46**, 167 (1976).
[7] D. Brydges, J. Fröhlich, and T. Spencer, Commun. Math. Phys. **83**, 123 (1982).
[8] R. B. Griffiths, J. Math. Phys. **8**, 478, 484 (1967); D. G. Kelly and S. Sherman, J. Math. Phys. **9**, 466 (1968); J. Ginibre, Commun. Math. Phys. **16**, 310 (1970).
[9] B. Simon and R. B. Griffiths, Commun. Math. Phys. **33**, 45 (1973).

A two-parameter matrix Riccati equation pair for a class of nonlinear sigma models based on a symmetric space

M. A. Snyder[a)]
Department of Physics, Wellesley College, Wellesley, Massachusetts 02181

(Received 1 February 1983; accepted for publication 22 April 1983)

It is shown that the infinitesimal action of a group on one of its associated symmetric spaces is, in the case of a classical off-diagonal symmetric space, in the form of a matrix Riccati equation. A two-parameter Riccati equation is then found for the nonlinear sigma model based on a symmetric space G/H.

PACS numbers: 11.10.Lm, 02.40. + m, 11.15.Kc, 02.20. + b

I. INTRODUCTION

Two-dimensional nonlinear sigma models[1] have attracted a great deal of attention in recent years, primarily because of the strong analogies[2] they possess with more realistic four-dimensional Yang–Mills gauge theories. A prominent role in the investigation of nonlinear sigma models has been played by the inverse scattering method,[3,4] despite the fact that the asymptotic conditions of the inverse scattering method for the nonlinear sigma models do not permit one[5] (at least naively) to use the inverse scattering method in the same way that it is used[4] in the case of other two-dimensional theories that are described by nonlinear evolution equations (the KdV equation,[6] sine-Gordon equation,[7] and so on). The inverse scattering method is useful, however, in deriving the fact that nonlinear sigma models have, like the other two-dimensional theories mentioned above,[8] an infinite number of conservation laws.[5,9,10] As for some (all?[11]) of the two-dimensional nonlinear evolution equations that possess an inverse scattering formalism, these conservation laws come in two varieties—local and nonlocal.[9,10] The infinite set of nonlocal conservation laws is most conveniently derived[5,12] from the linear isospectral equation of the inverse scattering method, whereas the infinite set of local conservation laws is most conveniently derived[13–15] from a pair of nonlinear equations (a so-called "Riccati" equation pair) closely associated with the linear isospectral equation. This pair of Riccati equations is constructed so that their integrability condition is equivalent to the equations of motion of the nonlinear sigma model.

In the literature, the Riccati equations for the nonlinear sigma models have usually appeared[13–15] as a sort of *deus ex machina*, with little motivation. One of the purposes of this paper is to give a geometrical interpretation for the Riccati equations, namely, that the Riccati equations represent the infinitesimal nonlinear action of a group G on one of its associated symmetric spaces[16] G/H. In so doing we significantly generalize the results of Scheler,[14] who found such Riccati equations for a certain class of symmetric spaces, to the case of an arbitrary classical off-diagonal symmetric space (this will be defined in Sec. III). This class of symmetric spaces contains the class studied by Scheler.

Having found this Riccati pair, we proceed to note that it possesses a two-parameter form which, as long as the parameters are functionally independent, has as its integrability condition the equations of motion of the nonlinear sigma model. By choosing a particular relation between the two parameters, the Riccati pair reduces to that used by Scheler.[14] Since Scheler finds his parameter-dependent Riccati pair by making a Pohlmeyer[9] transformation on the nonlinear sigma model fields, our two-parameter Riccati pair can be seen as a generalization (at least at the level of the Riccati equation) of the Pohlmeyer transformation. One may then, having in hand a parameter-dependent Riccati pair, derive *à la* Scheler[14] an infinity of local conservation laws for nonlinear sigma models based on this class of symmetric spaces.

The paper is organized as follows. In Sec. II we briefly review the "Cartan" formulation of the nonlinear sigma model. In Sec. III we define the notion of a classical off-diagonal symmetric space, and show that the infinitesimal action of a group G on such a symmetric space G/H is represented by a Riccati pair. In Sec. IV we show that a two-parameter Riccati pair can be derived, and give its relation to the previously known[14] Riccati pair. In Sec. V we summarize some of the consequences of the discovery of this Riccati pair, the detailed study of which will be pursued in a separate paper.[17]

II. CARTAN FORMULATION OF THE NONLINEAR SIGMA MODEL

The nonlinear sigma model based on a symmetric space G/H is defined through the Lagrangian density

$$\mathscr{L} = \tfrac{1}{2}\,\mathrm{tr}\, D_\mu g D^\mu g\,, \quad g \in G\,, \tag{1}$$

where

$$\begin{aligned} &D_\mu g = \partial_\mu g - C_\mu g\,,\\ &D_\mu g^{-1} = \partial_\mu g^{-1} + g^{-1} C_\mu\,,\\ &C_\mu \in \mathscr{H} = \text{Lie algebra of } H \subset G\,, \end{aligned} \tag{2}$$

with equation of motion

$$D_\mu D^\mu g - D_\mu g\, g^{-1} D^\mu g\,. \tag{3}$$

This Lagrangian density possesses manifest global invariance under the group G, and local invariance under the group H. Following Scheler's work[14] (to which the reader is referred for amplification and notation), we decompose

[a)] Submitted in partial fulfillment of the requirements for the Ph.D. at Yale University.

 0022-2488/83/082204-04$02.50

$\partial_\mu g\, g^{-1}$ as follows:

$$\partial_\mu g\, g^{-1} = C_\mu + V_\mu\,, \tag{4}$$

where

$$V_\mu \equiv D_\mu g\, g^{-1} \in \mathscr{G} \bmod \mathscr{H}\,, \quad \mathscr{G} = \text{Lie Algebra of } G\,,$$
$$C_\mu \equiv \partial_\mu g\, g^{-1} - V_\mu \in \mathscr{H}\,. \tag{5}$$

The decomposition (4) thus mirrors the canonical Cartan decomposition of a symmetric space[16]

$$\mathscr{G} = \mathscr{H} \oplus \mathscr{G} \bmod \mathscr{H}\,, \tag{6}$$

where

$$\mathscr{G} = \text{Lie Algebra of } G\,,$$
$$\mathscr{H} = \text{Lie Algebra of } H\,. \tag{7}$$

Using the fact that the coset space is a symmetric space, and changing to the light-cone coordinates (x,t),

$$x^0 = x + t\,,$$
$$x^1 = x - t\,, \tag{8}$$

we find that the equations of motion (3) can be expressed as[14]

$$\partial_x C_t - \partial_t C_x = [C_x, C_t] + [V_x, V_t]\,,$$
$$\partial_x V_t = [C_x, V_t]\,,$$
$$\partial_t V_x = [C_t, V_x]\,. \tag{9}$$

We will call this form of the equations of motion the *Cartan formulation* of the equations of motion.

III. GEOMETRY AND THE MATRIX RICCATI EQUATION PAIR

The linear isospectral equation of the inverse scattering method has the form[11]

$$\underline{d}P = \underline{\Omega} P\,, \tag{10}$$

where we are using the notation of differential forms (see Ref. 18), a differential 1-form A being denoted by an underline: $\underline{A}$. Here $\underline{d}$ is the exterior differential, $\underline{\Omega}$ is a 1-form element of the Lie algebra of some group G:

$$\underline{\Omega}(\lambda) = \Omega_x(\lambda)\underline{dx} + \Omega_t(\lambda)\underline{dt}\,, \quad \Omega_{x,t}(\lambda) \in \mathscr{G}\,. \tag{11}$$

P is either a column vector (in the case of Ablowitz, Kaup, Newell, and Segur[4] (AKNS)-type systems) or a matrix (in the case[5] of the nonlinear sigma model), and λ is a real parameter (the isospectral parameter). We shall restrict our considerations to the latter case. The whole point of the inverse scattering method is to find an equation like (10), such that the integrability condition which follows from the Poincaré lemma $\underline{d}^2 = 0$,

$$\underline{d}\underline{\Omega}(\lambda) - \underline{\Omega}(\lambda) \wedge \underline{\Omega}(\lambda) = 0\,, \tag{12}$$

is equivalent to the nonlinear equation to be solved.

We shall give (10) the geometric interpretation that it expresses the infinitesimal linear action of a group G on itself. It is well known, however, that a group can act on itself in many different ways. In particular, the group G can act on one of its coset spaces G/H, where H is a closed subgroup of G. In general, this action will be nonlinear. We shall be interested only in the case where G/H has the further property of being a symmetric space. We shall show that this infinitesimal nonlinear action is represented by a Riccati equation.

In order to show this, we must look at the ways symmetric spaces are parametrized. It is shown in the text of Gilmore[19] that except for the series of symmetric spaces

$$\mathrm{SL}(n;\mathbb{R})/\mathrm{SO}(n)\,, \quad \mathrm{SU}^*(2n)/\mathrm{USp}(2n)\,, \tag{13}$$

all of the classical symmetric spaces G/H (i.e., those in which both H and G are classical—as opposed to exceptional—groups) can be obtained as exponentials of an off-diagonal matrix

$$M_{m,n} = \begin{matrix} \\ m \\ n \end{matrix} \begin{matrix} \begin{matrix} m & n \end{matrix} \\ \begin{bmatrix} 0 & B \\ \pm B^\dagger & 0 \end{bmatrix} \end{matrix}, \tag{14}$$

where the upper (lower) sign is taken for spaces of noncompact (compact) type, respectively. We shall call all such symmetric spaces *classical off-diagonal symmetric spaces*. All of what follows is valid only for such symmetric spaces.

Upon taking the exponential of (14), we find that

$$\exp M = \begin{bmatrix} W_\pm & X_\pm \\ \pm X^\dagger_\pm & Y_\pm \end{bmatrix} \in G/H\,, \tag{15}$$

where

$$X_\pm = B(B^\dagger B)^{-1/2} \begin{cases} \sinh(B^\dagger B)^{1/2} & \text{upper}\,, \\ \sin(B^\dagger B)^{1/2} & \text{lower}\,, \end{cases}$$
$$W_\pm = [\mathbb{1}_m \pm X_\pm X^\dagger_\pm]^{1/2}\,,$$
$$Y_\pm = [\mathbb{1}_n \pm X^\dagger_\pm X_\pm]^{1/2}\,. \tag{16}$$

We thus may take the $m \times n$ matrix $X_\pm$ to parametrize such a classical off-diagonal symmetric space (we shall henceforth suppress the $\pm$ subscripts except in cases where an ambiguity may arise). The group G acts on the symmetric space G/H by left translation

$$g_0 \in G\,, \quad g_0 = \begin{bmatrix} A & B \\ C & D \end{bmatrix} : X \to X' \tag{17}$$

via

$$\underset{\in G}{\begin{bmatrix} A & B \\ C & D \end{bmatrix}} \underset{\in G/H}{\begin{bmatrix} W & X \\ \pm X^\dagger & Y \end{bmatrix}} \equiv \underset{\in G/H}{\begin{bmatrix} W' & X' \\ \pm X'^\dagger & Y' \end{bmatrix}} \underset{\in H}{\begin{bmatrix} H_1 & 0 \\ 0 & H_2 \end{bmatrix}}. \tag{18}$$

The points in the symmetric space transform in a complicated nonlinear way if the space is parametrized by X. However, if the space is parametrized instead by the projective variable Z

$$Z = XY^{-1} \tag{19}$$

(note that Z is related in a 1-1 way to X), then it can be easily seen that under the left translation (17) Z transforms via a fractional linear (Möbius, homographic) transformation:

$$g_0 : Z \to Z' = (AZ + B)(CZ + D)^{-1}\,. \tag{20}$$

Thus the action of the group G on the classical off-diagonal symmetric space G/H is nonlinear.

Let us now take the transformation (17) to be infinitesimal, that is, the group element g_0 is infinitesimally close to the identity

$$g_0 = \mathbb{1} + \begin{pmatrix} \alpha & \beta \\ \gamma & \epsilon \end{pmatrix}. \tag{21}$$

It then follows that the transformation (20) with (21) has the form

$$Z \to Z + \delta Z\,, \quad \delta Z = \beta + \alpha Z - Z\epsilon - Z\gamma Z\,. \tag{22}$$

Consequently the nonlinear counterpart to (10) is the Riccati pair

$$d\underline{Z} = \underline{\Omega}_{12} + \underline{\Omega}_{11}Z - Z\underline{\Omega}_{22} - Z\underline{\Omega}_{21}Z\,, \tag{23}$$

where

$$\underline{\Omega} = \begin{bmatrix} \underline{\Omega}_{11} & \underline{\Omega}_{12} \\ \underline{\Omega}_{21} & \underline{\Omega}_{22} \end{bmatrix}, \quad \Omega_{x,t} \in \mathscr{G}\,. \tag{24}$$

This nonlinear counterpart to the linear isospectral equation of the inverse scattering method is a matrix Riccati pair [this result could also have been found directly from (10) by writing $P = RS$, where $R \in G/H$ and $S \in H$, and then re-expressing everything in terms of Z]. The Riccati pair can thus be interpreted as expressing the infinitesimal action of the group G on the classical off-diagonal symmetric space G/H.

We can now make contact with the nonlinear sigma model based on such a symmetric space by defining a matrix N to be

$$N = \begin{bmatrix} 0 & Z \\ \pm Z^{\dagger} & 0 \end{bmatrix} \in \mathscr{G} \bmod \mathscr{H}\,. \tag{25}$$

Furthermore, if we let in (10),

$$\Omega_{x,t} = C_{x,t} + V_{x,t}\,, \tag{26}$$

where

$$C_{x,t} = \begin{bmatrix} \underline{\Omega}_{11} & 0 \\ 0 & \underline{\Omega}_{22} \end{bmatrix}_{x,t} \in \mathscr{H}\,, \tag{27}$$

$$V_{x,t} = \begin{bmatrix} 0 & \underline{\Omega}_{12} \\ \pm \underline{\Omega}_{12}^{\dagger} & 0 \end{bmatrix}_{x,t} \in \mathscr{G} \bmod \mathscr{H}\,, \tag{28}$$

and $C_{x,t}$ and $V_{x,t}$ are the nonlinear sigma model fields (5), then the integrability condition (12) of (10) is equivalent to the equation of motion (9) for the nonlinear sigma model. Since a classical symmetric space always has compact Π, i.e., $\underline{\Omega}_{11}^{\dagger} = -\underline{\Omega}_{11}$, $\underline{\Omega}_{22}^{\dagger} = -\underline{\Omega}_{22}$, it follows that the Riccati pair (23) becomes

$$\partial_x N = V_x + [C_x, N] - NV_xN\,, \tag{29}$$

$$\partial_t N = V_t + [C_t, N] - NV_tN\,. \tag{30}$$

The reader may easily check, using the Poincaré lemma and the definition of a symmetric space (see Ref. 14 or Ref. 18), that the integrability condition for the Riccati pair (29), (30) is satisfied if the equations of motion (9) for the nonlinear sigma model are satisfied. Thus, just as in AKNS-type systems, the Riccati pair has as its integrability equation the nonlinear equation to be solved. The only difference between AKNS-type systems and the nonlinear sigma model is that in the latter the Riccati pair are matrix equations, whereas in the former they are not.

As in AKNS-type systems, the Riccati pair (29), (30) implies[8] that a conserved current can be constructed. If we consider the quantity

$$\underline{J} \equiv \mathrm{tr}[N\underline{\Omega}]\,, \tag{31}$$

then one may show with a little algebra that

$$d\underline{J} = \mathrm{tr}[N(d\underline{\Omega} - \underline{\Omega} \wedge \underline{\Omega})]\,, \tag{32}$$

which vanishes if the equation of motion (12) is satisfied. Thus, the current $\underline{J}$ is a closed 1-form—a conserved current—for the nonlinear sigma model if the equation of motion (12) is satisfied.

IV. A TWO-PARAMETER RICCATI EQUATION PAIR FOR THE NONLINEAR SIGMA MODEL

We saw in the last section that the Riccati pair (29), (30) had as its integrability condition that the equations of motion for the nonlinear sigma model be satisfied, and that in this case a conserved current (31) could be constructed. However, no parameters were involved in this Riccati pair. In order to construct an infinity of conserved currents, we must have a parameter-dependent Riccati pair. Consider the two-parameter Riccati pair

$$\partial_x N = \theta V_x + [C_x, N] - \epsilon N V_x N\,, \tag{33}$$

$$\partial_t N = \epsilon^{-1} V_t + [C_t, N] - \theta^{-1} N V_t N\,, \tag{34}$$

where θ and ϵ are independent real parameters. One may show after a little algebra that the integrability condition for the Riccati pair (33), (34) is just the equations of motion (9). We now make contact with the work of Scheler by noting that if we choose

$$\theta = -\epsilon = \gamma^{-1}\,, \tag{35}$$

we arrive at the Riccati pair of Scheler [Eqs. (43) and (44) of Ref. 14]. Since Scheler arrived at his parameter-dependent Riccati pair by making a Pohlmeyer[9] transformation, we see that if the Pohlmeyer transformation is considered as a transformation of a parameter-*independent* Riccati pair to a parameter-*dependent* Riccati pair, then the transformation from (29), (30) to (33), (34) is a more general transformation, of which the Pohlmeyer transformation is simply the "diagonal" part (35) of (33), (34) (note that this interpretation can be made only at the "Riccati pair" level, and not at the "Lie algebra" level).

One may then show that the following conservation law follows from (33), (34):

$$-\epsilon^{-1}\partial_x[\mathrm{tr}(NV_t)] + \theta\partial_t[\mathrm{tr}(NV_x)] = 0\,. \tag{36}$$

This is identical to Scheler's result [Ref. 14, Eq. (45)] in the "diagonal" case (35). One can then, by an analysis identical to Scheler's, derive an infinity of local conservation laws.[17]

V. CONCLUSIONS AND FURTHER REMARKS

In this work we have given a clear geometric interpretation to the Riccati equations that in previous work have arisen *ex nihilo*. We have, in addition, found a two-parameter Riccati pair, which does not seem to have been noticed previously. Further analysis of this pair of equations may lead to a better understanding of the local and nonlocal conservation laws in nonlinear sigma models. In the continuation of this work (Ref. 17), we will give a more detailed exposition of the results in this paper, and show how the ideas advanced here may be used to find the nonlocal conservation laws as well as the local ones, via a procedure identical to that used by Sasaki and Bullough[11] for the sine-Gordon equation. We will also consider the natural extension of these methods to the case of supersymmetric nonlinear sigma models.

ACKNOWLEDGMENTS

I would like to thank my advisor, H. C. Tze, for his interest and encouragement, and F. Gürsey for helpful discussions. I would also like to thank the W. G. Lehnert foundation for financial support during the writing of this work.

[1]J. Schwinger, Ann. Phys. (N.Y.) **2**, 407 (1957); F. Gürsey, Nuovo Cimento **16**, 230 (1960); Ann. Phys. (N.Y.) **12**, 91 (1961); M. Gell-Mann and M. Levy, Nuovo Cimento **16**, 705 (1960).
[2]For a review, see F. Gürsey and H. C. Tze, Ann. Phys. (N.Y.) **128**, 29 (1980).
[3]C. S. Gardner, J. M. Greene, M. C. Kruskal, and R. M. Miura, Phys. Rev. Lett. **19**, 1095 (1967); see also G. L. Lamb, Jr., *Elements of Soliton Theory* (Wiley, New York, 1980).
[4]M. J. Ablowitz, D. J. Kaup, A. S. Newell, and H. Segur, Phys. Rev. Lett. **30**, 1262 (1973); **31**, 125 (1973).
[5]H. Eichenherr, Freiburg preprint (1980).
[6]D. J. Korteweg and G. de Vries, Phil. Mag. **39**, 422 (1895).
[7]J. Frenkel and T. Kontorova, J. Phys. (USSR) **1**, 137 (1939); T. H. R. Skyrme, Proc. R. Soc. London, Ser. A **262**, 237 (1961); J. Finkelstein and C. V. Misner, Ann. Phys. (N.Y.) **6**, 230 (1959).
[8]R. M. Miura, C. S. Gardner, and M. D. Kruskal, J. Math. Phys. **9**, 1204 (1968); M. Wadati, H. Sanuki, and K. Konno, Prog. Theor. Phys. **53**, 419 (1975).
[9]K. Pohlmeyer, Comm. Math. Phys. **46**, 207 (1976).
[10]M. Lüscher and K. Pohlmeyer, Nucl. Phys. B **137**, 46 (1978).
[11]R. Bullough and R. Sasaki, Proc. R. Soc. London, Ser. A **376**, 401 (1981).
[12]H. Eichenherr and M. Forger, Nucl. Phys. B **155**, 381 (1979).
[13]K. Scheler, Phys. Lett. B **93**, 331 (1980).
[14]K. Scheler, Zeit. Phys. C **6**, 365 (1980).
[15]H. Eichenherr, Nucl. Phys. B **135**, 1 (1978).
[16]S. Helgason, *Differential geometry, Lie Groups, and Symmetric Spaces* (Academic, New York, 1978).
[17]M. A. Snyder, "Local and non-local conservation laws in nonlinear sigma models based on classical off-diagonal symmetric spaces," Wellesley College preprint (1983).
[18]H. Flanders, *Differential Forms* (Academic, New York, 1963).
[19]R. Gilmore, *Lie Groups, Lie Algebras, and their Applications* (Wiley, New York, 1974).

Radiation gauge covariance. II. Quantization of the free electromagnetic field

Jorge Krause
Instituto de Física, Universidad Católica de Chile, Casilla 114-D, Santiago, Chile

(Received 20 July 1982; accepted for publication 10 December 1982)

The "substratum" of the electromagnetic field is defined as a two-dimensional duality isovector, satisfying the homogeneous wave equation. The components of the substratum are two Lorentz scalar fields that carry the two degrees of freedom of the field. It is then shown that the free Maxwell field becomes effectively quantized by the canonical quantization of the substratum. The equal-time canonical commutator thus obtained is precisely the same commutator one gets while quantizing the electromagnetic field within the Coulomb gauge. The whole approach is manifestly relativistic.

PACS numbers: 11.15. − q, 03.70. + k, 11.30.Cp

1. INTRODUCTION

There has been some work performed in recent years on the subject of electromagnetism in terms of its two degrees of freedom.[1] Through that work, it becomes clear that the problem set by the isolation and quantization of the true dynamical variables of the Maxwell field, in a completely gauge-independent way, raises a nontrivial and important question.

In a previous paper[2] we have been able to identify the true degrees of freedom of the radiation field in a gauge-independent manner while presenting a formalism which handles the Coulomb gauge in a manifestly Lorentz covariant fashion. In the present paper we wish to examine this matter further: we study the canonical quantization of pure radiation within the covariant formalism of the Coulomb gauge presented in Paper I.

The electromagnetic field has just two degrees of freedom. Radiation fields, however, are usually described by introducing, either, the six-component antisymmetric intensity tensor $F_{\mu\nu}(x)$, or else, the four-component potential vector $A_\mu(x)$. The use of redundant variables (i.e., the introduction of more dynamical variables than there are actually independent degrees of freedom) leads to well-known difficulties in the canonical quantization of the field. The essential difficulty already appears at the classical level of the theory. Its main features occur in classical electrodynamics, as well as in general relativity[3] and also in Yang–Mills theories.[4] In all these field-theoretic formalisms the lack of a covariant guage-independent identification of the true degrees of freedom of the system at hand brings into the picture the very same kind of troubles for having a canonical description of the field.

In quantum electrodynamics, the major trouble one has to face, in this respect, stems from the quite familiar fact that the gauge constraints one has to impose, as well as the field equations which must be satisfied, are not always consistent with the canonical commutation relations one postulates at the very beginning. Let us briefly recall this issue. For instance, if one starts the formalism by introducing a relativistic Lagrangian density, one gets the following expression for the equal-time canonical commutator[5]:

$$[A_j(t,\mathbf{x}), E_k(t,\mathbf{x}')] = i\delta_{jk}\delta(\mathbf{x}-\mathbf{x}')\,. \tag{1.1}$$

Plainly, these commutation relations are inconsistent with the first of Maxwell's equations for the field without charges, which reads

$$\nabla_k E_k(t,\mathbf{x}) = 0\,. \tag{1.2}$$

On the other hand, one may give up the relativistic appearance of the theory from the beginning, and start with the Hamiltonian formalism. Therefore, one uses the Fourier amplitudes of the field as canonical variables, and thus one quantizes directly the field's harmonic oscillators[6] (i.e., one may use the normal modes as canonical variables because of Jean's theorem[7]). In this manner one gets, instead of Eq. (1.1), the well-known commutator

$$[A_j(t,\mathbf{x}), E_k(t,\mathbf{x}')] = i\delta^T_{jk}(\mathbf{x}-\mathbf{x}')\,, \tag{1.3}$$

where now $\delta^T_{jk}(\mathbf{x})$ is the transverse delta function,[8] which satisfies

$$\nabla_j \delta^T_{jk}(\mathbf{x}) = 0\,. \tag{1.4}$$

The commutator (1.3) is thus consistent with Eq. (1.2), and, furthermore, it is also consistent with the Coulomb gauge condition; namely,

$$\nabla_j A_j(t,\mathbf{x}) = 0\,. \tag{1.5}$$

Hence, the appearance of the transverse delta function removes the inconsistency, while we impose the radiation gauge. This formalism, however, is not manifestly Lorentz covariant, although the theory itself remains relativistic. (Of course, because of the lack of manifest covariance, in order to stay within this formalism each time we perform a Lorentz transformation, we have to suitably regauge the potentials.)

When sources are present, the second method of quantization sketched above is based on the fact that the longitudinal and scalar fields can be replaced by the instantaneous interaction between all charges, while using the Coulomb gauge. Clearly, for the free field this means that the radiation gauge brings the longitudinal and scalar fields out of the picture, without further ado. Thus, from the point of view of quantum theory, this method is by far the simplest. In the usual approach, however, it has the already mentioned disadvantage of spoiling the space-time symmetric appearance of the theory.

 0022-2488/83/082208-04$02.50

In the case of the free electromagnetic field (which is indeed the only case we shall consider in this paper) this lack in a manifestly relativistic appearance is, perhaps, a mere defect in elegance. Nevertheless, quantum electrodynamics is handicapped by certain ambiguities, and it has been well known, for a long time indeed, that a powerful guidance as to how these ambiguities should be removed is obtained from the demand of relativistic covariance of the formalism.[9] For this purpose, then, it is desirable to try a covariant approach *ab initio* and to carry it as far as possible. This is done, for instance, in the Gupta-Bleuler formalism,[10] although at the cost of a cogent physical interpretation. Furthermore, we should also recall that for the computation of some complicated radiative corrections a symmetric handling of all four space-time components of the potential $A^{\mu}(x)$ affords the simplest mathematical treatment.

On these grounds, a closer investigation would be desirable as to the physical possibilities of the covariant radiation gauge formalism for handling quantum electrodynamics.[11] At this stage, we shall tentatively concentrate our attention on the free field canonical quantization within the relativistic approach to the Coulomb gauge presented in our previous work.

2. CANONICAL COVARIANT QUANTIZATION

For the sake of quantizing the true degrees of freedom of the electromagnetic field, we shall follow the same approach used by Carmeli,[12] *mutatis mutandi*. Namely, we shall apply the method of gauge-free quantization to the radiation field, adapting it to the peculiar symmetry of our own approach [i.e., we are using the Lorentz group, instead of SU(2) as used by Carmeli].

The local polarization basis we have introduced in Sec. 7 of Paper 1, has its support on the (**k**-space) light cone.[13] In what follows, we need an extension of the complete tetrad $\{v^{\mu}, k^{\mu}, \epsilon^{\mu}_{A}(k; v); A = 1,2\}$ all over momentum space-time. This can be easily achieved. Indeed, we assume Eqs. (I.7.1), (I.7.2), and (I.7.3), to hold everywhere, with $v_{\mu}v^{\mu} = 1$, where now $k^2 = k_{\mu}k^{\mu}$ and $k_{\mu}v^{\mu}$ can be anything. The completeness relation for this tetrad then reads

$$\frac{k^2 v^{\mu}v^{\nu} + k^{\mu}k^{\nu} - k_{\lambda}v^{\lambda}(v^{\mu}k^{\nu} + k^{\mu}v^{\nu})}{k^2 - (k_{\rho}v^{\rho})^2} - \delta_{AB}\epsilon^{\mu}_{A}(k; v)\epsilon^{\nu}_{B}(k; v) = \eta^{\mu\nu} \tag{2.1}$$

[which, in effect, becomes Eq. (I.7.4) on the light cone]. It is immediate that when $k^2 - (k_{\mu}v^{\mu})^2 = 0$, one has $k^{\mu} = \lambda v^{\mu}$; and therefore, while writing

$$k^{\mu} = \lambda v^{\mu} + \epsilon a^{\mu}, \tag{2.2}$$

where $a_{\mu}v^{\mu} = 0$, $a_{\mu}a^{\mu} = -1$, $a_{\mu}\epsilon^{\mu}_{A}(k; v) = 0$, $A = 1,2$, and $\epsilon^2 < \lambda^2$, one easily shows, from Eq. (2.1), that

$$\lim_{k^{\mu}\to\lambda v^{\mu}} \delta_{AB}\epsilon^{\mu}_{A}(k; v)\epsilon^{\nu}_{B}(k; v) = -\eta^{\mu\nu} - v^{\mu}v^{\nu} + a^{\mu}a^{\nu}. \tag{2.3}$$

So the aforementioned complete tetrad is all right everywhere in momentum space-time.

Let us also briefly examine the behavior of this tetrad under the inversion $k^{\mu} \to -k^{\mu}$. Clearly, $v_{\mu}\epsilon^{\mu}_{A}(-k; v) = 0$ and $-k_{\mu}v^{\mu}(-k; v) = 0$. Hence

$$\epsilon^{\mu}_{A}(-k; v) = \sigma_{AB}\epsilon^{\mu}_{B}(k; v), \tag{2.4}$$

i.e.,

$$\eta_{\mu\nu}\epsilon^{\mu}_{A}(-k; v)\epsilon^{\nu}_{B}(k; v) = -\sigma_{AB}, \tag{2.5}$$

while

$$\epsilon^{\mu}_{A}(k; v) = \sigma^{-1}_{AB}\epsilon^{\mu}_{B}(-k; v). \tag{2.6}$$

But then we get

$$\sigma^{-1}_{AB} = \sigma_{BA}. \tag{2.7}$$

Now, if we consider $v^{\mu} = (1,\mathbf{0})$, the completeness relation, Eq. (2.1), gives us

$$\frac{k^i k^j}{\mathbf{k}^2} + \sigma_{AB}\hat{\epsilon}^{i}_{A}\hat{\epsilon}^{j}_{B} = \sigma^{ij}, \tag{2.8}$$

where

$$\epsilon^{0}_{A}(k;1,\mathbf{0}) = 0, \quad \epsilon^{j}_{A}(k;1,\mathbf{0}) = \hat{\epsilon}^{j}_{A}(\mathbf{k}). \tag{2.9}$$

Thus, in the v-frame, we may consider a right-handed triad, i.e., $\boldsymbol{\epsilon}_1 \times \hat{\boldsymbol{\epsilon}}_2 = \hat{\mathbf{k}}$, with $k^j = |\mathbf{k}|\hat{k}^j$, everywhere in **k**-space. Therefore, without loss of generality, we may adopt the convention

$$\epsilon^{\mu}_{1}(-k; v) = \epsilon^{\mu}_{1}(k; v), \quad \epsilon^{\mu}_{2}(-k; v) = -\epsilon^{\mu}_{2}(k; v). \tag{2.10}$$

with respect to any Galilean working frame. That is, we set

$$\sigma_{AB} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \tag{2.11}$$

quite generally.

After these preliminaries, let us define the *substratum* (say) of the free Maxwell field as the Fourier transform of its true degrees of freedom, i.e.,

$$\phi_A(x) = \int \frac{d^4k}{(2\pi)^2}\delta(k^2)\rho_A(k)e^{-ikx}, \quad A = 1,2. \tag{2.12}$$

where $\rho_A(k)$ denotes the true dynamical variables (cf. Paper I). Thus, we have two (independent) scalar solutions of the homogeneous d'Alembert equation,

$$\Box\phi_A(x) = 0, \quad A = 1,2, \tag{2.13}$$

which carry the two degrees of freedom of the electromagnetic field. Next, we also introduce the auxiliary field

$$K^{\mu}_{A}(k; v) = \int \frac{d^4k}{(2\pi)^2}\epsilon^{\mu}_{A}(k; v)e^{-ikx} \tag{2.14}$$

such that

$$v_{\mu}k^{\mu}_{A}(x; v) = 0, \tag{2.15}$$

$$\nabla_{\mu}K^{\mu}_{A}(x; v) = 0, \tag{2.16}$$

$$K^{*\mu}_{A}(x - y; v) = K^{\mu}_{A}(y - x; v) = \sigma_{AB}K^{\mu}_{B}(x - y; v), \tag{2.17}$$

where $K^{*\mu}_{A}$ is the complex conjugate of K^{μ}_{A}, and

$$\eta_{\mu\nu}\int d^4z\, K^{\mu}_{A}(x - z; v)K^{\nu}_{B}(z - y; v) = -\delta_{AB}\delta^{(4)}(x - y). \tag{2.18}$$

Then we get a one-to-one correspondence:

$$A^{\mu}_{T}(x;v) = \delta_{AB}\int d^4y\, K^{\mu}_{A}(x-y;v)\phi_B(y)\,, \tag{2.19}$$

$$\phi_A(x) = -\eta_{\mu\nu}\int d^4y\, K^{\mu}_{A}(x-y;v)A^{\nu}_{1}(y;v)\,, \tag{2.20}$$

as it should be.

Because of our convention, Eqs. (2.4) and (2.11), the reality condition for the electromagnetic field requires

$$\rho^{*}_{A}(k) = \sigma_{AB}\,\rho_B(-k)\,. \tag{2.21}$$

Hence, for the substratum of the field we have the following complex conjugation rule:

$$\phi^{*}_{A}(x) = \sigma_{AB}\phi_B(x)\,; \tag{2.22}$$

namely, $\phi_1(x)$ is real, while $\phi_2(x)$ is pure imaginary. (Of course, one could equally well introduce a real substratum; for that matter, one changes from the "linear" polarization complex vectors $\epsilon^{\mu}_{A}(k;v)$, which are real, to a set of "circular" polarization complex vectors. However, we left this matter as it is.)

For the Lagrangian density of the substratum we take, as usual,

$$\mathscr{L} = \tfrac{1}{2}\eta^{\mu\nu}\delta_{AB}(\nabla_\mu\phi^{*}_{A})(\nabla_\nu\phi_B)\,, \tag{2.23}$$

i.e.,

$$\mathscr{L} = \tfrac{1}{2}\eta^{\mu\nu}\sigma_{AB}(\nabla_\mu\phi_A)(\nabla_\nu\phi_B)\,. \tag{2.24}$$

Clearly, the current is $J^{\mu}(x) = 0$, as it must be, while

$$\pi_A(x) = \sigma_{AB}\dot{\phi}_A(x) \tag{2.25}$$

is the canonical momentum. The Hamiltonian density of the substratum is given by

$$\mathscr{H} = \tfrac{1}{2}\delta^{\mu\nu}\sigma_{AB}(\nabla_\mu\phi_A)(\nabla_\nu\phi_B)\,. \tag{2.26}$$

Hence, for the Lagrangian and the Hamiltonian of the substratum, we have

$$L = \tfrac{1}{2}\eta^{\mu\nu}\sigma_{AB}\int d^3x(\nabla_\mu\phi_A)(\nabla_\nu\phi_B)\,, \tag{2.27}$$

and

$$H = \tfrac{1}{2}\delta^{\mu\nu}\sigma_{AB}\int d^3x(\nabla_\mu\phi_A)(\nabla_\nu\phi_B)\,, \tag{2.28}$$

respectively.

We shall prove that, if we substitute from Eq. (2.20) into Eq. (2.27), while adopting the v-frame as our working reference system, we obtain that L corresponds plainly to the Lagrangian of the Maxwell field. By the same token, the Hamiltonian H of the substratum corresponds precisely to the Hamiltonian of the radiation field in the v-frame. The physical meaning of these identifications is immediate.

To this end, let us consider

$$\begin{aligned}\delta_{AB}\int d^3x\, K^{\mu}_{A}(x-y;v)K^{\nu}_{B}(x-z;v) &= \int_{-\infty}^{\infty}\frac{dk_0}{2\pi}e^{ik_0(t_y-t_x)}\int_{-\infty}^{\infty}\frac{dk'_0}{2\pi}e^{-ik'_0(t_z-t_x)}\\ &\times\int\frac{d^3k}{(2\pi)}\delta_{AB}\epsilon^{\mu}_{A}(k_0,\mathbf{k};v)\epsilon^{\nu}_{B}(k'_0,\mathbf{k};v)e^{-i\mathbf{k}\cdot(\mathbf{y}-\mathbf{z})}\,.\end{aligned} \tag{2.29}$$

Therefore, if we take $v^{\mu} = (1,\mathbf{0})$, according to Eqs. (2.8) and (2.9), we get

$$\begin{aligned}&\delta_{AB}\int d^3x\, K^{\mu}_{A}(x-y;\mathbf{1},\mathbf{0})k^{\nu}_{B}(x-z;\mathbf{1},\mathbf{0})\\ &\quad=\delta(t_x-t_y)\delta(t_x-t_z)\delta^{jk}_{T}(\mathbf{y}-\mathbf{z})\delta^{\mu}_{j}\delta^{\nu}_{k}\,,\end{aligned} \tag{2.30}$$

where

$$\delta^{jk}_{T}(\mathbf{x}) = \{\delta^{jk}-\nabla^{-2}\nabla^{j}\nabla^{k}\}\delta(\mathbf{x}) \tag{2.31}$$

is the transverse delta function [cf. Eqs. (1.3) and (1.4)]. On the other hand, if we impose the very weak and quite natural boundary condition

$$\oint_{(\infty)} d\sigma_\mu(y)A^{\nu}_{T}(y;v)e^{iky} = 0\,, \tag{2.32}$$

where the four-vector $d\sigma_\mu(y)$ denotes the three-dimensional element of hypersurface at space-time infinity, then, clearly,

$$\nabla_\mu\phi_A(x) = -\eta_{\nu\lambda}\int d^4y\, K^{\nu}_{A}(x-y,v)A^{\lambda}_{T,\mu}(y;v)\,. \tag{2.33}$$

Hence we have, with respect to the v-frame,

$$\begin{aligned}\sigma_{AB}\int d^3x(\nabla_\mu\phi_A)(\nabla_\nu\phi_B) &= \iint d^4y\, d^4z\, A^{T}_{\lambda,\mu}(y;\mathbf{1},\mathbf{0})A^{T}_{\rho,\nu}(z;\mathbf{1},\mathbf{0})\sigma_{AB}\\ &\times\int d^3x\, K^{\lambda}_{A}(x-y;\mathbf{1},\mathbf{0})K^{\rho}_{B}(x-z;\mathbf{1},\mathbf{0})\\ &=\iint d^3y\, d^3z\, A^{T}_{j,\mu}(t_x,\mathbf{y})A^{T}_{k,\nu}(t_x,\mathbf{z})\delta^{jk}_{T}(\mathbf{y}-\mathbf{z})\,,\end{aligned} \tag{2.34}$$

where we have written $A^{T}_{j}(t,\mathbf{x})$ for $A^{T}_{j}(t,\mathbf{x};\mathbf{1},\mathbf{0})$. Now, let $g^{\mu\nu} = \mathrm{diag}(+,\pm,\pm,\pm)$. From Eqs. (2.31) and (2.34), we get (omitting t_x)

$$\begin{aligned}g^{\mu\nu}\sigma_{AB}\int d^3x(\nabla_\mu\phi_A)(\nabla_\nu\phi) &= \iint d^3y\delta^{jk}\{\dot{A}^{T}_{j}(\mathbf{y})\dot{A}^{T}_{k}(\mathbf{y})\pm\delta^{lm}A^{T}_{j,l}(\mathbf{y})A^{T}_{k,m}(\mathbf{y})\}\\ &\pm\iint d^3y\, d^3z\,\delta^{lm}A^{T}_{j,l}(\mathbf{y})A^{T}_{k,m}(\mathbf{z})\nabla^{-2}\nabla^{j}\nabla^{k}\delta(y-z)\,,\end{aligned} \tag{2.35}$$

since

$$\iint d^3y\, d^3z\,\dot{A}_j(\mathbf{y})\dot{A}_k(\mathbf{z})\nabla^{-2}\nabla^{j}\nabla^{k}\delta(\mathbf{y}-\mathbf{z}) = 0 \tag{2.36}$$

because of the Coulomb gauge constraint. By the same token,

$$\begin{aligned}&\iint d^3y\, d^3z\,\delta^{lm}A^{T}_{j,l}(\mathbf{y})A^{T}_{k,m}(\mathbf{z})\nabla^{-2}\nabla^{j}\nabla^{k}\delta(\mathbf{y}-\mathbf{z})\\ &\quad=\int d^3y\,\delta^{jk}\delta^{lm}A^{T}_{l,k}(\mathbf{y})A^{T}_{j,m}(\mathbf{y})\,.\end{aligned} \tag{2.37}$$

Finally, recalling the definitions

$$E_j = -\dot{A}^{T}_{j}\,,\quad B_j = \epsilon_{jkl}A^{T}_{l,k}\,, \tag{2.38}$$

we get the desired identifications; namely,

$$\frac{1}{2}g^{\mu\nu}\delta_{AB}\int d^3x(\nabla_\mu\phi_A)(\nabla_\nu\phi_B) = \frac{1}{2}\int d^3x(\mathbf{E}^2\pm\mathbf{B}^2)\,, \tag{2.39}$$

for the Hamiltonian (+) or the Lagrangian (−) of the Maxwell field, according as $g^{\mu\nu}$ denotes $\delta^{\mu\nu}$ or $\eta^{\mu\nu}$, respectively.

These results show the consistency of the quantization procedure we adopt for the wave equations (2.13). Without further ado, we introduce the canonical equal-time commutator:

$$[\phi_A(t,\mathbf{x}), \pi_B(t,\mathbf{y})] = i\delta_{AB}\delta(\mathbf{x}-\mathbf{y}), \tag{2.40}$$

i.e.,

$$[\phi_A(t,\mathbf{x}), \dot{\phi}_B(t,\mathbf{y})] = i\sigma_{AB}\delta(\mathbf{x}-\mathbf{y}). \tag{2.41}$$

In this manner, the quantization of the Maxwellian tensor variables is ensured automatically without assuming for them, separately, equal-time commutation relations. For instance, if we use Eq. (2.30), then Eq. (2.41) becomes precisely the well-known commutator, belonging in the Coulomb gauge [cf. Eq. (1.3)], afforded by the usual noncovariant approach.

3. CONCLUDING REMARKS

We have shown that the problem of solving Maxwell's field equations, without sources, can always be reduced to the problem of solving the homogeneous d'Alembert equation for two scalar fields. An elementary device for generating a Maxwellian field from any scalar wave function has been well known, for a long time indeed.[14] However, while using that device, it is not suggested that any arbitrary radiation field can be thus obtained. Rather, this well-known method offers a way of generating a certain special set of Maxwellian solutions. The method presented in our work [i.e., Eqs. (2.19) and (2.20)], on the contrary, is completely general.

A natural way of isolating the dynamical variables of the free electromagnetic field has been established, in a completely gauge-invariant manner, while handling the formalism in a manifestly relativistic fashion. The canonical procedure, as applied to the isolated dynamical variables, has shown that all the physical contents of a free electromagnetic field are already included in the two-component substratum $(\phi_1(x), \phi_2(x))$. This object behaves as a two-dimensional isovector under duality rotations, and as a set of two scalar fields under Lorentz transformation.

Consequently, the substratum represents the true degrees of freedom of the electromagnetic radiation field, and, therefore, the Maxwell field becomes effectively quantized by the canonical procedure of quantum field theory, as applied to the Lagrangian formalism of the substratum. No specific commutation relations have to be assumed for the different tensor variables characterizing a Maxwell field, since they come into the fore by applying the radiation gauge geometric tools to the fundamental equal-time commutator of the substratum. In this manner, a basis for the description of free photons is obtained, which seems much more consistent with all other elementary particles, since we drop the unphysical variables while retaining the relativistic invariance at every step of the formalism.

Finally, we wish to remark that, no matter how simple this formalism may look when applied to the free Maxwell field, one is the more tempted to hope a similar (i.e., projective) procedure could be applicable for isolating and quantizing the true degrees of freedom of a nonabelian Yang–Mills field. A closer investigation would be also desirable on this intriguing question.

ACKNOWLEDGMENT

This work was supported by Dirección de Investigación de la Universidad Católica de Chile (DIUC); Grant No. 50/81.

[1] S. Hojman, Ann. Phys. (N.Y.) **103**, 74 (1977); R. Gambini and S. Hojman, Ann. Phys. (N.Y.) **105**, 407 (1977).
[2] J. Krause, J. Math. Phys. **22**, 1759 (1981); hereafter referred to as Paper I.
[3] P. G. Bergman and A. Komar, Int. J. Theor. Phys. **5**, 15 (1972); also, M. Halpern and S. Malin, J. Math. Phys. **12**, 213 (1971).
[4] E. S. Abers and B. W. Lee, Phys. Rep. **9**(1), 1–41 (1973). See also A. Actor, Rev. Mod. Phys. **51**, 461 (1979) for SU(2) classical gauge theories. Also, R. Jackiw, Rev. Mod. Phys. **52**, 661 (1980).
[5] Cf., for instance, J. D. Bjorken and S. D. Drell, *Relativistic Quantum Fields* (McGraw-Hill, New York, 1965), Chap. 14.
[6] E. Fermi, Rev. Mod. Phys. **4**, 87 (1932).
[7] J. H. Jeans, Philos. Mag. **10**, 91 (1905).
[8] Cf., J. D. Bjorken and S. D. Drell, *Relativistic Quantum Fields* (McGraw-Hill, New York, 1965).
[9] Cf., W. Heitler, *The Quantum Theory of Radiation*, 3rd ed. (Oxford University, New York, 1966), Chap. VI.
[10] S. N. Gupta, Proc. Phys. Soc. (London) A **63**, 681 (1950); K. Bleuler, Helv. Phys. Acta **23**, 567 (1950).
[11] In this sense, one should recall the work of S. Osaki, Progr. Theor. Phys. **14**, 511 (1955), on quantum electrodynamics without subsidiary conditions.
[12] M. Carmeli, J. Math. Phys. **10**, 1699 (1969); M. Carmeli, Nuovo Cimento B **67**, 103 (1970).
[13] For full details, the reader is referred to our previous paper, Paper I. Thus, we refer to Eq. (I.7.1), for instance, to denote Eq. (7.1) of that paper.
[14] J. L. Synge, *Relativity: The Special Theory*, 2nd ed. (North-Holland, Amsterdam, 1965), p. 362.

Fermion currents in 1 + 1 dimensions

A. L. Carey
Department of Pure Mathematics, University of Adelaide, G.P.O. Box 498, Adelaide, South Australia 5001, Australia

C. A. Hurst and D. M. O'Brien
Department of Mathematical Physics, University of Adelaide, G.P.O. Box 498, Adelaide, South Australia 5001, Australia

(Received 28 May 1982; accepted for publication 10 December 1982)

Let $\mathcal{F}$ denote the Fock space for a free massless fermion field in two space-time dimensions. If Q and Q_5 denote the charge and chiral charge operators, then $\mathcal{F} = \oplus_{n_1,n_2} \mathcal{F}_{n_1,n_2}$, where $\mathcal{F}_{n_1,n_2}$ is the joint eigenspace of Q and Q_5 with corresponding eigenvalues n_1 and n_2. The smeared time-zero free fermion currents formally given by $J_\mu(f) = \int :\bar{\psi}(x)\gamma_\mu\psi(x): f(x)\,dx$ [$\mu = 0,1; f \in \mathcal{S}(\mathbb{R})$] are in fact self-adjoint densely defined operators on $\mathcal{F}$ and their exponentials generate a C^* algebra which leaves each subspace $\mathcal{F}_{n_1,n_2}$ invariant. If a periodic box cutoff for the fermions is introduced, Uhlenbrock has shown that this C^* algebra acts irreducibly in each "sector" $\mathcal{F}_{n_1,n_2}$. We prove this same result without any cutoffs and determine the properties of the representations in each sector. The importance of this result for the problem of constructing fermion fields from "observables" in two-dimensional models, such as those of Thirring and Schwinger, is discussed.

PACS numbers: 11.40.Dw, 11.10.Mn, 11.10.Qr

1. INTRODUCTION

The main reason that soluble fermion models exist in 1 + 1 dimensions is that the fermion currents can be defined as "fields" acting on the representation space for the fermions and their algebraic properties may be exploited to yield the solution. Similarly, the success of the "bosons into fermions" programme[1–3] lies in the fact that the bosons in question are just the currents and the fermions are essentially determined algebraically by their commutation relations with them. The most complete account of the latter argument appears in Ref. 2. There is a gap, however, in these discussions which occurs first at the level of the free fermion field.

In 1 + 1 dimensions, the free fermion currents are definable as the Wick-ordered products

$$J_\mu(f) = \int_{-\infty}^{\infty} :\bar{\psi}(x)\gamma_\mu\psi(x): f(x)dx, \quad f \in \mathcal{S}(\mathbb{R}),\ \mu = 0,1 \tag{1.1}$$

of the free fermion fields. Under the action of these operators $J_\mu(f)$, the fermion Fock space $\mathcal{F}$ should decompose into a direct sum of invariant subspaces. When the fermion mass is zero it is a folk theorem that this decomposition has the form $\mathcal{F} = \oplus_{n,n'} \mathcal{F}_{n,n'}$, the labeling by integers n, n' (charge and chiral charge quantum numbers). Moreover, on $\mathcal{F}_{n,n'}$ the action of the currents is believed to be irreducible.

Dell-Antonio, Frischman, and Zwanziger[2] show how to reconstruct fermion fields from a representation of "currents" having the form hypothesized above. These fermion fields have different dynamics depending on the precise form of imposed commutation relations with the currents. For example one can obtain free fermions or ones satisfying the equation of motion of the massless Thirring model. This construction was also applied to the Schwinger model.[3] However, the decomposition of the free fermion Fock space $\mathcal{F}$ into these invariant subspaces $\mathcal{F}_{n,n'}$, irreducible under the action of the currents, has not been shown to exist. The only positive result in this direction is that of Uhlenbrock[4] which shows that this decomposition occurs when a box cutoff is imposed on the fermions, except that the representations of the currents in the $\mathcal{F}_{n,n'}$ are all equivalent. This latter fact indicates that removing the cutoff will be difficult since we expect inequivalent representations in its absence. Nevertheless, Lundberg's analysis[5] of the "vacuum sector" (i.e., the cyclic subspace generated from the fermion vacuum by the currents) suggested to us that a variant of Uhlenbrock's proof may work.

Our aim then is to prove that, for free massless fermions, the above decomposition into subspaces irreducible under the action of the currents is valid and, moreover, to show that the representations which occur are "displaced Fock representations" (cf. Ref. 6) being all inequivalent.

The difficult part of the proof is the irreducibility and this occupies Secs. 5–7 of the paper. After setting up the problem in Secs. 2 and 3, the rest of the argument is dealt with quite easily in Sec. 4. As we have indicated here and will further discuss in Sec. 8, this then supports the analysis of Refs. 2 and 3 for the massless Thirring and Schwinger models.

One can clearly ask similar questions for the case of massive fermions. We will indicate which results carry over easily to the massive case, although there, only the ordinary charge operator is defined and so at best we could only expect a decomposition of $\mathcal{F}$ into charge eigenspaces. The question of irreducibility again appears to be the most difficult.

Notation: We write γ^0, γ^1 for the Dirac matrices in two dimensions with $\gamma^5 = \gamma^0\gamma^1$. The choice

$$\gamma^0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \gamma^5 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

 0022-2488/83/082212-10$02.50

will be made when necessary. Note that

$$\gamma_\mu = g_{\mu\nu}\gamma^\nu, \quad g^{\mu\nu} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

2. CAR ALGEBRA

Our methods depend on some previous work[7] and we need to discuss the results of the paper (henceforth referred to as CHOB). So let $\mathscr{H}$ be a complex Hilbert space and $\mathscr{A}(\mathscr{H})$ be the CAR algebra over $\mathscr{H}$ generated by

$$\{a(g), a^*(g) | g \in \mathscr{H}\},$$

where

$$a(h)a^*(g) + a^*(g)a(h) = \langle h, g\rangle I,$$

$$a(h)a(g) + a(g)a(h) = 0.$$

We can represent this algebra in the usual way by its action on the antisymmetric Fock space $\mathscr{F}_{\mathscr{H}}$ over $\mathscr{H}$, and we will use the same symbols $a(g)$, $a^*(g)$ to denote the CAR generators in this representation. If P is an orthogonal projection on $\mathscr{H}$, then a new representation, π_P of $\mathscr{A}(\mathscr{H})$ acting on $\mathscr{F}_{\mathscr{H}}$, is defined by

$$\pi_P(a(h)) = a^*(\overline{Ph}) + a((1-P)h),$$

where the bar denotes a complex conjugation in $\mathscr{H}$.

If U is a unitary operator on $\mathscr{H}$, then U defines an automorphism of $\mathscr{A}(\mathscr{H})$ by $a(h) \to a(Uh)$. We say U is implementable in π_P if there is an operator $\Gamma(U)$ on $\mathscr{F}_{\mathscr{H}}$ such that

$$\Gamma(U)\pi_P(a(h))\Gamma(U)^{-1} = \pi_P(a(Uh)).$$

The case of interest to us is when $\mathscr{H} = L^2(\mathbb{R},\mathbb{C}^2)$ and $P = P_-$ is the spectral projection of the free Dirac Hamiltonian corresponding to the interval $(-\infty, -m]$ with m = fermion mass. In this case P_- is given by its action on the Fourier transform space by multiplication by the function

$$P_-(k) = \frac{1}{2}\left(1 + \frac{k}{\omega(k)}\gamma^5 - \frac{m}{\omega(k)}\gamma^0\right),$$

where γ^0, γ^1 are the two-dimensional Dirac matrices, $\gamma^5 = \gamma^0\gamma^1$, and $\omega_k = (k^2 + m^2)^{1/2}$. For $m = 0$ this reduces to

$$P_-(k) = \tfrac{1}{2}(1 - \epsilon(k)\gamma_5), \quad \epsilon(k) = \begin{cases} 1 & k \geqslant 0, \\ -1 & k \leqslant 0. \end{cases}$$

We let $1 - P_- = P_+$ and write $\psi(g) = \pi_{P_-}(a(g))$, interpreting $\psi(g)$ as the fermion field operator. The space $\mathscr{F}_{\mathscr{H}}$ in this case is then identified with the space $\mathscr{F}$ of the Introduction. The automorphisms of interest in this case are defined by multiplication operators on $L^2(\mathbb{R},\mathbb{C}^2)$ as follows. For each $f_\mu \in \mathscr{S}(\mathbb{R})$, $\mu = 0,1$ (i.e., real-valued Schwartz functions), we define

$$(U_{tf}g)(x) = \exp\{it[f_0(x) + \gamma^5 f_1(x)]\}g(x), \quad t \in \mathbb{R}.$$

Then $t \to U_{tf}$ is a uniformly continuous one-parameter group on $L^2(\mathbb{R},\mathbb{C}^2)$. The corresponding automorphisms of the CAR algebra over $L^2(\mathbb{R},\mathbb{C}^2)$ are implementable (see Ref. 5 or CHOB). We may fix the phase of the implementing unitary $\Gamma(U_{tf})$ by requiring

$$\langle \Omega, \Gamma(U_{tf})\Omega\rangle \geqslant 0, \tag{2.1}$$

where Ω is the vacuum vector ($\|\Omega\| = 1$) in $\mathscr{F}$. On the other hand, from Ref. 8 we know that there exists a possibly different choice such that $t \to \Gamma(U_{tf})$ is a strongly continuous one parameter group on $\mathscr{F}$. Using this second choice, the relation

$$\Gamma(U_{tf})\psi(g)\Gamma(U_{tf})^{-1} = \psi(U_{tf}g) \tag{2.2}$$

is the "integrated" form of the commutation relations

$$[J^0(f_0), \psi(g)] = \psi(f_0 g), \tag{2.3}$$

$$[J^1(f_1), \psi(g)] = \psi(\gamma^5 f_1 g), \tag{2.4}$$

where $J^0(f_0)$ and $J_1(f_1)$ are the generators of the one-parameter groups $t \to \Gamma(U_{t\tilde{f}})$ with $\tilde{f} = (f_0, 0)$ and $\tilde{f} = (0, f_1)$, respectively. We will see later that $J(f) = J^0(f_0) + J^1(f_1)$ is identifiable with the current (1.1) defined as a Wick-ordered product. We now develop the results of CHOB relevant to this discussion.

Associated with these currents J^0 and J^1 are two "charge operators" which we define as follows. Let $(U_0(\theta)g)(x) = e^{i\theta}g(x)$ and $(U_5(\theta)g)(x) = e^{i\gamma^5\theta}g(x)$ for $g \in L^2(\mathbb{R},\mathbb{C}^2)$, $\theta \in \mathbb{R}$. Then $\theta \to U_0(\theta)$ and $\theta \to U_5(\theta)$ are implementable groups of unitaries in π_{P_-} (Ref. 5, CHOB), when $m = 0$ and we fix the phase of $\Gamma(U_0(\theta))$ and $\Gamma(U_5(\theta))$ as in (2.1). Then we identify the generators of $\theta \to \Gamma(U_0(\theta))$ and $\theta \to \Gamma(U_5(\theta))$, say Q and Q_5, as the charge and chiral charge operators. When $m \neq 0$, only $U_0(\theta)$ is implemented and hence only Q is defined.

We intend to show that on the joint eigenspaces of Q and Q_5 the set of operators $\{\Gamma(U_f) | f_\mu \in \mathscr{S}(\mathbb{R}), \mu = 0,1\}$ acts irreducibly. To do this, we need another way of constructing these eigenspaces. We introduce another pair of operators on $L^2(\mathbb{R},\mathbb{C}^2)$ following Ref. 9 and CHOB. Let

$$(U^{\pm}g)(x) = \exp(i\gamma_{\pm}\eta(x))g(x),$$

where $\gamma_{\pm} = \frac{1}{2}(1 \pm \gamma^5)$ and $\eta(x) = (2\tan^{-1}x + \pi)$. The operators $U^{\pm}$ are implemented in π_{P_-} (see CHOB) independently of the value of the fermion mass m and in the case $m = 0$, the vectors

$$\Omega_{n_1,n_2} = \Gamma(U^+)^{n_1}\Gamma(U^-)^{n_2}\Omega \tag{2.5}$$

are eigenvectors of Q and Q_5 with eigenvalues $n_1 - n_2$ and $n_1 + n_2$, respectively. It follows from CHOB that Q and Q_5 commute with $\Gamma(U_f)$ for all $f_\mu \in \mathscr{S}(\mathbb{R})$ so that the eigenspaces of Q and Q_5 reduce $\{\Gamma(U_f) | f_\mu \in \mathscr{S}(\mathbb{R}), \mu = 0,1\}$.

The physical interpretation of the integers n_1, n_2 can be obtained from a knowledge of the explicit form of Q and Q_5. Some discussion of this may be found in Ref. 9 (although our notation is different). We remark only that the number of fermion or antifermion states occupied with probability one in Ω_{n_1,n_2} is given by

$n_1 > 0$, n_1 "left going" antifermion states,

$n_1 < 0$, $-n_1$ "left going" fermion states,

$n_2 < 0$, $-n_2$ "right going" antifermion states,

$n_2 > 0$, n_2 "right going" fermion states.

Here "left" or "right" refers to the support of the appropriate functions being on the negative or positive half of the momentum axis (see Sec. 5 for details).

3. THE C^* ALGEBRA FOR THE CURRENTS

We have seen that for each $f=(f_0,f_1)$, with $f_\mu\in\mathcal{S}(\mathbb{R})$ there is a unitary operator U_f on $L^2(\mathbb{R},\mathbb{C}^2)$. Now $\{U_f|\ f_\mu\in\mathcal{S}(\mathbb{R})\}$ form an abelian group and $U_f\to\Gamma(U_f)$ is a multiplier representation of this group. Lundberg has determined the multiplier explicitly[8]:

$$\Gamma(U_f)\Gamma(U_{f'})\Gamma(U_f)^{-1}=\exp[iB(f,f')]\Gamma(U_{f'}), \tag{3.1}$$

where

$$B(f,f')=-2\ \mathrm{Im}\ \mathrm{tr}(P_-j(f)P_+j(f')P_-),$$

with $j(f)$ denoting the operator

$$(j(f)g)(x)=(f_0(x)+\gamma^5 f_1(x))g(x),\quad g\in L^2(\mathbb{R},\mathbb{C}^2) \tag{3.2}$$

and tr being the trace of the nuclear operator $P_-j(f)P_+j(f')P_-$. A straightforward calculation yields

$$\mathrm{tr}(P_-j(f)P_+j(f')P_-)$$
$$=\frac{1}{4\pi}\int dk\,dl\sum_{\mu,\nu}\hat f^*_\mu(l-k)c_{\mu\nu}(k,l)\hat f'_\nu(l-k), \tag{3.3}$$

where

$$f(x)=\frac{1}{\sqrt{2\pi}}\int dk\,\hat f(k)e^{ikx}$$

and

$$C(k,l)=(c_{\mu\nu}(k,l))$$
$$=\begin{pmatrix}1-(kl+m^2)/\omega_k\omega_l & k/\omega_k-l/\omega_l\\ k/\omega_k-l/\omega_l & 1-(kl-m^2)/\omega_k\omega_l\end{pmatrix}.$$

When $m=0$, all of this simplifies considerably:

$$\mathrm{tr}(P_-j(f)P_+j(f')P_-)=\frac{1}{2\pi}\int dq\,\Sigma_{\mu\nu}\hat f^*_\mu(q)c_{\mu\nu}(q)\hat f'_\nu(q), \tag{3.4}$$

where

$$C(q)=(C_{\mu\nu}(q))=\begin{pmatrix}|q| & q\\ q & |q|\end{pmatrix},$$

and

$$B(f,f')=\frac{i}{\pi}\int dq\,q[\hat f'_0(q)\hat f_1(q)^*+\hat f_0(q)^*\hat f'_1(q)] \tag{3.5}$$

[using $\hat f_\mu(q)^*=\hat f_\mu(-q)$].

We may associate with $\{U_f|f_\mu\in\mathcal{S}(\mathbb{R})\}$ a C^* algebra defined as follows. Firstly regard B as a nondegenerate symplectic form on

$$M=\{f=(f_0,f_1)|\ f_\mu\in\mathcal{S}(\mathbb{R}),\ \mu=0,1\}.$$

Then we can form $C^*(M,B)$, the C^* algebra of the CCR over (M,B).[10] When $m=0$, B is the imaginary part of the inner product on M given by

$$f,f'\to\mathrm{tr}(P_-j(f)P_+j(f')P_-)$$

because we can define a complex structure L on M by

$$Lf_\mu(k)=-i\epsilon(k)f_{1-\mu}(k), \tag{3.6}$$

so that

$$B(f,Lf)=2\ \mathrm{tr}(P_-j(f)P_+j(f)P_-)=\|f\|^2_M. \tag{3.7}$$

Write δ_f for the element of $C^*(M,B)$ given by

$$\delta_f(f')=\begin{cases}0 & f\neq f',\\ 1 & f=f',\end{cases}$$

and these elements satisfy

$$\delta_f\delta_{f'}=\exp(i/2)B(f,f')\delta_{f+f'}. \tag{3.8}$$

Notice that in the massless case by virtue of (3.7), the Fock representation of $C^*(M,B)$ has generating functional

$$\exp[-\tfrac14 B(f,Lf)]$$
$$=\exp\left[-\frac{1}{4\pi}\int dq\,|q|(|f_0(q)|^2+|f_1(q)|^2)\right]. \tag{3.9}$$

Finally we write $\bar M$ for the completion of M in the norm defined by (3.7). Combining the above with the previous section, we obtain the following Proposition.

Proposition 3.1: There is a choice of phase for the $\Gamma(U_f)$ such that the map $\delta_f\to\Gamma(U_f)$ is a representation of $C^*(M,B)$ with $\lambda\to\Gamma(U_{\lambda f})$ strongly continuous in $\lambda\in\mathbb{R}$.

4. REPRESENTATIONS OF $C^*(M,B)$

We assume throughout this section that $m=0$. We saw in Proposition 3.1 that the C^* algebra $C^*(M,B)$ is represented in the fermion Fock space $\mathcal{F}$ over $L^2(\mathbb{R},\mathbb{C}^2)$. Our aim in this section is to decompose this representation. Recall the definition (2.5) of the vectors $\Omega_{n_1,n_2}\in\mathcal{F}$.

Lemma 4.1: $\langle\Omega_{n_1,n_2},\Gamma(U_f)\Omega\rangle=0$ for all n_1, $n_2\in\mathbb{Z}$ (not both zero) and all $f\in M$.

Proof: Since Q and Q_5 are self-adjoint and Ω_{n_1,n_2} and $\Gamma(U_f)\Omega$ are eigenvectors of Q and Q_5 with distinct eigenvalues when n_1 and n_2 are not both zero, the result follows.

Now let $\mathcal{F}_{n_1,n_2}$ denote the subspace of $\mathcal{F}$ generated by the action of the $\Gamma(U_f)$ on Ω_{n_1,n_2} as f ranges over M. From Lemma 4.1 we deduce that the $\mathcal{F}_{n_1,n_2}$ are all mutually orthogonal subspaces of $\mathcal{F}$. On $\mathcal{F}_{n_1,n_2}$ we have a cyclic representation of $C^*(M,B)$ with generating functional

$$\begin{aligned}\phi_{n_1,n_2}(f)&=\langle\Omega_{n_1,n_2},\Gamma(U_f)\Omega_{n_1,n_2}\rangle\\ &=\langle\Omega,\Gamma(U^-)^{-n_2}\Gamma(U^+)^{-n_1}\\ &\quad\times\Gamma(U_f)\Gamma(U^+)^{n_1}\Gamma(U^-)^{n_2}\Omega\rangle\\ &=\sigma_{n_1,n_2}(f)\langle\Omega,\Gamma(U_f)\Omega\rangle,\end{aligned} \tag{4.1}$$

where $\sigma_{n_1,n_2}(f)$ is a complex number of modulus one. [This follows from the fact that $\Gamma(U_f)$ and

$$\Gamma(U^-)^{-n_2}\Gamma(U^+)^{-n_1}\Gamma(U_f)\Gamma(U^+)^{n_1}\Gamma(U^-)^{n_2}$$

implement the same automorphism of the CAR algebra.] Letting $\phi_{0,0}(f)=\langle\Omega,\Gamma(U_f)\Omega\rangle$, we can determine this generating functional $\phi_{0,0}$ in a number of ways, one of which we will outline in the next section. We record the answer here for the ensuing discussion:

$$\phi_{0,0}(f)=\exp\left(-\frac{1}{4\pi}\int dq\,|q|(|f_0(q)|^2+|f_1(q)|^2\right). \tag{4.2}$$

We immediately recognize this as the generating functional for the Fock representation of $C^*(M,B)$. [The fact that, on $\mathcal{F}_{0,0}$ the representation $f\to\Gamma(U_f)$ is Fock, was established by Lundberg.[5]] We can write, for each $f\in M$, $\sigma_{n_1,n_2}(f)=\exp i(\delta(f))$ with $\delta(f)\in\mathbb{R}$.

From the linearity of $f \to J(f)$ and the relation

$$\langle \Omega_{n_1,n_2}, J(f)\Omega_{n_1,n_2}\rangle = \frac{d}{d\lambda}\,\delta(\lambda f)\Big|_{\lambda=0},$$

it follows that δ is a linear functional on M. From (4.1) and (4.2) we conclude that the generating functionals ϕ_{n_1,n_2} have the form expected for "displaced Fock" representations.

In order to show that $\mathcal{F} = \oplus_{n_1,n_2} \mathcal{F}_{n_1,n_2}$ it is sufficient to show that the joint eigenspaces of Q and Q_5 coincide with the space $\mathcal{F}_{n_1,n_2}$.

Lemma 4.2: The joint eigenspace of Q and Q_5 corresponding to the eigenvalues $n_1 - n_2$ and $n_1 + n_2$, respectively, coincides with $\mathcal{F}_{n_1,n_2}$ provided $\mathcal{F}_{0,0}$ coincides with the subspace of $\mathcal{F}$ annihilated by Q and Q_5.

Proof: Notice that

$$\Gamma(U_f)\Omega_{n_1,n_2} = \sigma_{n_1,n_2}(f)^{-1}\Gamma(U^+)^{n_1}\Gamma(U^-)^{n_2}\Gamma(U_f)\Omega.$$

Thus $\Gamma(U^+)^{n_1}\Gamma(U^-)^{n_2}$ maps the 0,0 eigenspace of Q, Q_5 (assumed to be $\mathcal{F}_{0,0}$) into the $n_1 - n_2$, $n_1 + n_2$ eigenspace (cf. CHOB). Conversely, the inverse of this operator maps the $n_1 - n_2$, $n_1 + n_2$ eigenspace of Q, Q_5 into $\mathcal{F}_{0,0}$ (CHOB), proving the result.

Thus in order to establish the decomposition $\mathcal{F} = \oplus_{n_1,n_2} \mathcal{F}_{n_1,n_2}$, we need only prove the following proposition.

Proposition 4.3: The subspace of $\mathcal{F}$ consisting of vectors v with $Qv = Q_5 v = 0$, coincides with $\mathcal{F}_{0,0}$.

The next three sections are devoted to the proof of this result. There is one final proposition which we have to establish before stating our main theorem.

Proposition 4.4: The representation of $C^*(M,B)$ occurring in the above decomposition are all mutually inequivalent.

Proof: It is sufficient to show that the representation in $\mathcal{F}_{n_1,n_2}$ is inequivalent to that in $\mathcal{F}_{0,0}$. Moreover we will deal, for simplicity, with the case n_1, n_2 both positive. Then a typical state in the $n_1 - n_2$, $n_1 + n_2$ eigenspace of Q and Q_5, say (2.5), has the form

$$\Xi = \psi(g_1)^*\cdots\psi(g_{n_1})^*\psi(h_1)^*\cdots\psi(h_{n_2})^*\Omega,$$

where we have g_i, $h_j \in L^2(\mathbb{R},\mathbb{C}^2)$, pairwise orthogonal and of norm one with $P_- g_i = g_i$, $P_+ h_j = h_j$ and the support of $\hat{g}_i$ and $\hat{h}_j$ (the Fourier transform) being in the left and right half-axis, respectively. Then

$$\langle \Xi, J(f)\Xi\rangle = \Sigma_i - \langle g_i, j(f)g_i\rangle + \Sigma_r \langle h_r, j(f)h_r\rangle.$$

With our choice of Dirac matrices, we can write

$$\hat{g}_i(k) = \frac{1}{\sqrt{2}}\,\tilde{g}_i(k)\begin{pmatrix}1\\ \epsilon(k)\end{pmatrix}, \quad \hat{h}_r(k) = \frac{1}{\sqrt{2}}\,\tilde{h}_r(k)\begin{pmatrix}1\\ -\epsilon(k)\end{pmatrix},$$

where $\tilde{g}_i(k)$, $\tilde{h}_r(k) \in \mathbb{C}$ so that

$$\langle g_i, j(f)g_i\rangle = 2\int \tilde{g}_i(k)^*(\hat{f}_0(k-l) + \epsilon(k-l)\hat{f}_1(k-l))\tilde{g}_i(l)\,dl\,dk,$$

$$\langle h_r, j(f)h_r\rangle = 2\int \tilde{h}_r(k)^*(\hat{f}_0(k-l) - \epsilon(k-l)\hat{f}_1(k-l))\tilde{h}_r(l)\,dl\,dk.$$

Rewriting these expressions, we get

$$\langle g_i, j(f)g_i\rangle = 2\int dk\,(\hat{f}_0(k) + \epsilon(k)\hat{f}_1(k))\int dl\,\tilde{g}_i(k+l)^*\tilde{g}_i(l),$$

$$\langle h_r, j(f)h_r\rangle = 2\int dk\,(\hat{f}_0(k) - \epsilon(k)\hat{f}_1(k))\int dl\,\tilde{h}_r(k+l)^*\tilde{h}_r(l).$$

Since we are regarding $f \to \langle \Xi, J(f)\Xi\rangle$ as a linear functional on M, we know that the representation in $\mathcal{F}_{n_1,n_2}$ of $C^*(M,B)$ will be equivalent to that in $\mathcal{F}_{0,0}$ if and only if the functions

$$u{:}k \to \frac{1}{|k|}\left(\Sigma_i - \int \tilde{g}_i(k+l)^*\tilde{g}_i(l)\,dl + \Sigma_r\int \tilde{h}_r(k+l)^*\tilde{h}_r(l)\,dl\right),$$

$$v{:}k \to \frac{\epsilon(k)}{|k|}\left(\Sigma_i - \int \tilde{g}_i(k+l)^*\tilde{g}_i(l)\,dl - \Sigma_r\int \tilde{h}_r(k+l)^*\tilde{h}_r(l)\,dl\right)$$

lie in $\bar{M}$. Now near $k = 0$ u and v are approximately $(n_1 + n_2)/|k|$ and $(n_1 - n_2)/|k|$, respectively, and consequently cannot have finite $\bar{M}$-norm. It follows then that all the representations of $C^*(M,B)$ occurring in the decomposition of $\mathcal{F}$ are inequivalent.

Combining the previous propositions, we obtain the following theorem.

Theorem 4.5: The Fermion Fock space $\mathcal{F}$ decomposes as a direct sum $\oplus_{n_1,n_2} \mathcal{F}_{n_1,n_2}$ of subspaces invariant under the representation $\delta_f \to \Gamma(U_f)$ of $C^*(M,B)$. $\mathcal{F}_{n_1,n_2}$ is the subspace of $\mathcal{F}$ on which Q and Q_5 take the values $n_1 - n_2$ and $n_1 + n_2$, respectively. Moreover the representation of $C^*(M,B)$ on $\mathcal{F}_{n_1,n_2}$ is of displaced Fock type, hence irreducible, and is not equivalent to any of the other representations occurring in the decomposition.

Remark 4.5: In Ref. 2 Dell-Antonio, Frischman, and Zwanziger described the reconstruction of fermion fields from representations of the current operators. They based their reconstruction on the assumption that one has a Hilbert space carrying a representation of the currents precisely of the form described in Theorem 4.5. Given such a representation one can find "fermion fields" acting on the same space and having the vacuum expectation values of free fermion fields (and satisfying the Dirac equation). Thus Theorem 4.5 shows that in the case of free massless fermions, the currents do indeed have the representation on the fermion Fock space required by their analysis.

Remark 4.6: As we noted in the Introduction, the difficult part of the proof of Theorem 4.5 is Proposition 4.3, which is the only result remaining to establish. The next three sections are devoted to the proof.

5. THE CURRENTS AS WICK-ORDERED PRODUCTS

We introduce the usual notation for the fermion fields as operator-valued distributions:

$$\psi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}dk\,e^{ikx}\left[a_+(k)\frac{1}{\sqrt{2}}\begin{pmatrix}1\\ \epsilon(k)\end{pmatrix}+a_-(-k)^*\frac{1}{\sqrt{2}}\begin{pmatrix}1\\ -\epsilon(k)\end{pmatrix}\right], \tag{5.1}$$

$$\bar{\psi}(x)=\psi(x)^*\gamma^0.$$

The Wick-ordered product (1.1) is now expressible as

$$\begin{aligned}&\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}dk\int_{-\infty}^{\infty}dl\,\{a_+(k+l)^*a_+(l)[(\theta(-k)\theta(k+l)\\ &+\theta(k)\theta(l))(\hat{f}_0(k)+\hat{f}_1(k))\\ &+(\theta(-k)\theta(-l)+\theta(k)\theta(-k-l))\\ &\times(\hat{f}_0(k)-\hat{f}_1(k))]\\ &-a_-(-l)^*a_-(-k-l)[(\theta(-k)\theta(-l)\\ &+\theta(k)\theta(-k-l))(\hat{f}_0(k)+\hat{f}_1(k))\\ &+(\theta(-k)\theta(k+l)+\theta(k)\theta(l))(\hat{f}_0(k)-\hat{f}_1(k))]\\ &+a_+(k+l)^*a_-(-l)^*[\theta(k+l)\theta(-l)(\hat{f}_0(-k)\\ &+\hat{f}_1(-k))+\theta(l)\theta(-k-l)(\hat{f}_0(k)-\hat{f}_1(k))]\\ &+a_-(-k-l)a_+(l)[\theta(l)\theta(-k-l)(\hat{f}_0(-k)\\ &+\hat{f}_1(-k))+\theta(k+l)\theta(-l)(\hat{f}_0(k)-\hat{f}_1(k))]\},\end{aligned} \tag{5.2}$$

where $\hat{f}_\mu$ ($\mu=0,1$) is of fast decrease and satisfies $\hat{f}_\mu(-k)^*=\hat{f}_\mu(k)$. We assign the following meaning to (5.2). We define it by its formal action on the domain $\mathscr{D}$ consisting of vectors in the fermion Fock space with a finite number of particles present and fast decrease in momentum space. Direct calculation verifies that $\mathscr{D}$ is invariant and that (5.2) satisfies the same commutation relations with the fermion fields as does the operator $J(f)=J_0(f_0)+J_1(f_1)$ defined in Sec. 2 provided $\hat{f}_\mu$ is the Fourier transform of f_μ. By our choice of phase factor in the definition of $\Gamma(U_f)$, it follows that on the vacuum state, (5.2) coincides with $J(f)\Omega$ [cf. Eq. (3.25) of Ref. 8]. Thus $J(f)$ coincides in $\mathscr{D}$ with the operator $\tilde{J}(f)$ defined by (5.2). One may then verify the bound

$$\|\tilde{J}(f)G_r\|\leqslant(r\|j(f)\|+\|f\|_M)\|G_r\|,$$

where G_r is an r-particle state in $\mathscr{D}$. Induction on n gives

$$\|\tilde{J}(f)^nG_r\|\leqslant C_n\|G_r\|,$$

where

$$\frac{C_n}{n!}=\frac{2^{n-1}}{n}\prod_{k=1}^{n-1}[1+(r\|j(f)\|+\|f\|_M)/2k]$$

so that for small t, $\Sigma_1^\infty t^n\|\tilde{J}(f)^nG_r\|/n!$ converges, verifying that the states in $\mathscr{D}$ are analytic vectors for $\tilde{J}(f)$. It follows that $\tilde{J}(f)$ is essentially self-adjoint on $\mathscr{D}$ and that $J(f)$ is its unique self-adjoint extension. Now introduce annihilation and creation operators for the Bose field $J(f)$ as $b(f)=\frac{1}{2}(J(f)-iJ(Lf))$ and $b(f)^*$, with L given by (3.6).

It is important to note that the smearing function in the expression for $b(f)$ contains a step function which restricts the support to the negative half-axis (in momentum space) while $b(f)^*$ is restricted to the positive half-axis. There is a simplification which we can make at this point by observing that (5.2) splits into two terms depending, respectively, on $f_0\pm f_1$. Thus we define subspaces M_+ and M_- on which $f_0=f_1$ and $f_0=-f_1$, respectively. It follows that the closure $\bar{M}$ is the direct sum of $\bar{M}_+$ and $\bar{M}_-$

By setting $\frac{1}{2}f_+=f_0=f_1$, we see that

$$J(f_+)=b(f_+)+b(f_+)^*,\quad f_+\in M_+ \tag{5.3}$$

and $f_+\to J(f_+)$ is a boson field with test functions from M_+ and (5.3) is its decomposition into creation and annihilation parts. Similar remarks apply to M_-, and if $f_-\in M_-$ we have $[J(f_+),J(f_-)]=0$. Notice also that

$$[b(f_+),b(g_+)^*]=\frac{1}{\pi}\int_{-\infty}^{\infty}dk\,k\theta(k)f_+^*(k)g_+(k),\quad f_+,g_+\in M_+ \tag{5.4}$$

and

$$b(f_+)\Omega=0. \tag{5.5}$$

In fact from (5.3), (5.4), and (5.5) one can calculate the generating functional $\langle\Omega,\exp J(f_+)\Omega\rangle$ explicitly and obtain (4.2) (for $f_+\in M_+$). By restricting B to M_+ and M_- separately, we obtain nondegenerate symplectic forms on each of these spaces and hence representations in $\mathscr{F}$ of $C^*(M_+,B)$ and $C^*(M_-,B)$. We will see in the next section that in order to prove Proposition 4.3 it is sufficient to consider separately the representations of $C^*(M_+,B)$ and $C^*(M_-,B)$ in $\mathscr{F}_{0,0}$. We record from the above discussion the following essential result.

Proposition 5.1: The cyclic representation of $C^*(M_+,B)$ generated from the fermion vacuum state Ω by $\{\Gamma(U(f_+))|f_+\in M_+\}$ is equivalent to the Fock representation with vacuum state Ω, and generating functional

$$\rho_{0,0}(f_+)=\exp\left[-\frac{1}{2\pi}\int_{-\infty}^{\infty}dk\,k\theta(k)|f_+(k)|^2\right].$$

A similar result holds for $C^*(M_-,B)$.

6. THE FERMION, ANTIFERMION PAIR SPACE

Acting on Ω we know that $C^*(M,B)$ generates a subspace of the joint eigenspace of Q and Q_5 corresponding to the eigenvalue zero. In order to show that it generates all of this eigenspace we need a convenient representation of those states v in $\mathscr{F}$ with $Qv=Q_5v=0$. It is clear from the expressions in the previous section for ψ and $\bar{\psi}$ that we can take $L^2(\mathbb{R},\mathbb{C})$ for the one-particle fermion or antifermion space as the spinors have only one independent component. Moreover, as we are interested in the zero eigenspace of both Q and Q_5, any state annihilated by these two operators must contain equal numbers of particles and antiparticles either all moving to the right or all moving to the left. Because of our conventions for the creation and annihilation operators in Sec. 5, we may identify a state with n particles and n antiparticles both moving to the right as an element of $R_+^{(n)}\otimes R_-^{(n)}$, where

$$R_\pm^{(n)}=L^2(\mathbb{R}_\pm)\overset{(n)}{@}\cdots@L^2(\mathbb{R}_\pm),$$

where @ denotes antisymmetrization and $\mathbb{R}_\pm=\{q\in\mathbb{R}|q\gtrless 0\}$. Here the first n variables are the momentum variables for particle states and the second n are those for

antiparticle states. The space of all such states is

$$R_1 = \sum_{n=0}^{\infty} R_+^{(n)} \otimes R_-^{(n)}, \quad R_\pm^{(0)} = \mathbb{C}.$$

There is a similar space R_2 for particles and antiparticles moving to the left and the zero eigenspace of Q, Q_S is the sum $R_1 + R_2$ with $R_1 \cap R_2 = R_\pm^{(0)} = \mathbb{C}$, the vacuum subspace. If $F \in R_+^{(n)} \otimes R_-^{(n)}$ and $\mathbf{p} = (p_n,...,p_1)$, $\mathbf{q} = (q_n,...,q_1)$ then F is totally antisymmetric in the p_i's and q_i's separately with norm

$$\|F\|^2 = (n!)^2 \int_{p_i>0} d\mathbf{p} \int_{q_i<0} d\mathbf{q} |F(\mathbf{p},\mathbf{q})|^2. \tag{6.1}$$

The C^* algebra for the currents, $C^*(M_+,B)$, generates a subspace of R_1 by acting on the vacuum state while $C^*(M_-,B)$ generates a subspace of R_2. It is sufficient therefore to prove that $C^*(M_+,B)$ generates a dense subspace of R_1 in order to prove Proposition 4.3, the argument for $C^*(M_-,B)$ and R_2 being essentially the same.

We record for convenience the action of the boson creation operators $b(f)^*, f \in M_+$, on R. If $F = (F_n)$ with $F_n \in R_+^{(n)} \otimes R_-^{(n)}$, then

$$\begin{aligned}(b(f)^*F)_n(\mathbf{p},\mathbf{q}) &= \int_{-\infty}^{\infty} dk \hat{f}_+(k) \sum_{i=1}^{n} [\theta(p_i - k)F_n(p_n,...,p_i - k,...,p_1,\mathbf{q}) \\ &\quad - \theta(-q_i - k)F_n(\mathbf{p},q_n,...,q_i + k,...,q_1)] \\ &\quad - \frac{1}{(n!)^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (-1)^{i+j} \theta(p_i)\theta(-q_j)\hat{f}_+(p_i - q_j) \\ &\quad \times F_{n-1}(p_n,...,\hat{p}_i,...,p_1,q_n,...,\hat{q}_j,...,q_1),\end{aligned} \tag{6.2}$$

where $\hat{f}_+$ denotes the restriction of $\hat{f}$ to the positive half-axis. [This formula is established directly from the formal action of (5.2).] We note an important fact. Since we are dealing with the Fock representation of $C^*(M_+,B)$ we can extend our test functions for the $b(f)^*$ from those f in M_+ to those in $\bar{M}_+$. The action of $b(f)^*$ for $f \in \bar{M}_+$ is still given by (6.2). By Fourier transforming

$$F_n^\vee(\mathbf{x},\mathbf{y}) = \frac{1}{(2\pi)^{n/2}} \int F_n(\mathbf{p},\mathbf{q}) e^{i\mathbf{p}\cdot\mathbf{x} + i\mathbf{q}\cdot\mathbf{y}} \, d\mathbf{p} \, d\mathbf{q},$$

we obtain the action of $b(f)^*$ in this representation as

$$\begin{aligned}(b(f)^*F)_n^\vee(\mathbf{x},\mathbf{y}) &= \sqrt{2\pi} \sum_{i=1}^{n} [(\hat{f}_+)^\vee(x_i)F_n^{\vee(i+)}(\mathbf{x},\mathbf{y}) - (\hat{f}_+)^\vee(-y_i)F_n^{\vee(i-)}(\mathbf{x},\mathbf{y})] \\ &\quad + \frac{i}{\sqrt{2\pi}(n!)^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (-1)^{i+j} \frac{(\hat{f}_+)^\vee(x_i) - (\hat{f}_+)^\vee(-y_j)}{x_i + y_j} F_{n-1}^\vee(\hat{\mathbf{x}}_i,\hat{\mathbf{y}}_j),\end{aligned} \tag{6.3}$$

where the notation $F_n^{\vee(i+)}$ means that the Fourier transform has support in $\mathbb{R}_\pm$ in the ith fermion and antifermion momentum variable, $\hat{\mathbf{x}}_i$, $\hat{\mathbf{y}}_j$ means that the coordinates x_i, y_j are missing and

$$(\hat{f}_+)^\vee(x) = \frac{1}{\sqrt{2\pi}} \int dk \hat{f}_+(k) e^{ikx}.$$

For convenience we will write $\tilde{f}(x) = (\hat{f}_+)^\vee(x)$.

Now from (6.3) we can deduce an expression for the states in R_1 which is obtained from the vacuum by the action of polynomials in the boson creation operators.

Lemma 6.1: Let F_{mn} denote the component in $R_+^{(n)} \otimes R_-^{(n)}$ of a vector obtained by the application of a product of m boson operators to the vacuum: $b(f_1)^*,...,b(f_m)^*, f_j \in M_+$. Then

$$\begin{aligned}F_{mn}^\vee(\mathbf{x},\mathbf{y}) &= \frac{i^n}{(n!)^2} (2\pi)^{(n-2m)/2} \sum_{\text{partitions}} \sum_{p,q} (-1)^{p+q} \prod_{s=1}^{n} (1/x_{s_p} + y_{s_q}) \\ &\quad \times \prod_{t=1}^{r_s} (\tilde{f}_{i_t^{(s)}}(x_{s_p}) - \tilde{f}_{i_t^{(s)}}(-y_{s_q})), \quad n \leqslant m,\end{aligned} \tag{6.4}$$

where

(a) $\Sigma_{\text{partitions}}$ means partitions of m into n blocks, so $m = r_1 + r_2 + \cdots + r_n$ with $r_i > 0$;

(b) $\Sigma_{p,q}(-1)^{p+q}$ means sum over all permutations $p(q)$ of the n arguments $x_i(y_j)$ with sign -1 for odd permutations;

(c) x_{s_p}, y_{s_q} denotes the coordinates assigned to the block s, $i_t^{(s)}$ is the label of the r_s function in the block s.

Proof: This is a tedious induction argument. For $m = 0$, take

$$F_{oo}^\vee = 1, \quad F_{on}^\vee = 0, \quad n > 0.$$

From (6.3) we obtain

$$F_{11}^\vee(x,y) = i(\tilde{f}(x) - \tilde{f}(-y))/\sqrt{2\pi}(x + y),$$

which is (6.4) in this case. We sketch the induction step omitting details. We represent partitions by Young diagrams where each of the m boxes in a diagram corresponds to one of the functions f_i (all i distinct, $i = 1,...,m$) and each row corresponds to a pair of coordinates (x_k, y_l) with all x_k, y_l distinct and k not necessarily equal to l. Equation (6.3) describes the effect of the inclusion of one more function f_m. The first term in that equation contains those contributions for which a partition with n rows and $m - 1$ boxes is changed to one with n rows and m boxes with f_m being assigned to one of the existing rows so that there is no change in the number of pairs of coordinates. The row containing f_m must therefore have at least two boxes. The second term contains contributions in which a partition with $(n - 1)$ rows [and hence $(n - 1)$ coordinate pairs] is changed to one with n rows (and so one more pair) with f_m assigned to a new row with only a single box. If therefore (6.4) is correct for $(m - 1)$, (6.3) correctly generates all the terms required to make (6.4) correct for (m).

In order to compare the fermion-antifermion pair states with the boson states, we exploit the fact that we are dealing always with functions having support in one half of the momentum space axis. This allows us to use a Cayley transform

to realize the test functions as elements of Hardy spaces. As before, we detail the argument for R_1 only.

Definition 6.2: We write $H^2_{\pm}$ for the Hilbert space of functions h: $\mathbb{T}\to C$, which are the boundary values (on the unit circle $\mathbb{T}$) of functions holomorphic inside (outside) the unit disc D with the norm

$$\|h\|^2_{\pm} = \int_0^{2\pi} dt\, |h(t)|^2.$$

In terms of the Fourier transform, $h(t) = \Sigma_{n=-\infty}^{\infty} a_n e^{int}$ and $h \in H_+(H_-)$ if $a_n = 0$, $n < 0$ ($a_n = 0$, $n \geqslant 0$) and $\Sigma |a_n|^2 < \infty$.

Lemma 6.3: If $\hat{f}_{\pm} \in L^2(\mathbb{R}_{\pm})$ with

$$f_{\pm}(x) = \frac{1}{\sqrt{2\pi}} \int e^{ikx} \hat{f}_{\pm}(k)\, dk,$$

then the maps $S_{\pm}$ defined by

$$(S_{\pm}\hat{f}_{\pm})(t) = \sqrt{2} f_{\pm}\left(i\frac{1-e^{it}}{1+e^{it}}\right)(1+e^{\pm it})^{-1}$$

are isometries of $L^2(\mathbb{R}_{\pm})$ onto $H^2_{\pm}$. In particular, if $\hat{f}_{\pm} \in L^2(\mathbb{R}_{\pm})$, then the continuation of $f_{\pm}$ to the upper (lower) half-plane is holomorphic, and hence $S_{\pm} f_{\pm}$ are holomorphic in the interior (exterior) of the unit disc.

Proof: This is fairly standard; see, for example, Devinatz.[11]

We now regard $S_{\pm}$ as acting on $R^{(1)}_{\pm}$ and "lift" them in the obvious way to R_1. So we have an isometry S_1 of R_1 onto a new Hilbert space, which for convenience we denote $S_1 R_1$, where

$$S_1 R_1 = \bigoplus_{n=0}^{\infty} S^{(n)} R^{(n)} = \bigoplus_{n=0}^{\infty} S^{(n)}(R^{(n)}_{+} \otimes R^{(n)}_{-})$$

with

$$S^{(n)}(R^{(n)}_{+} \otimes R^{(n)}_{-}) = \left(\textcircled{@}_1^n S_+ R^{(1)}_{+}\right) \otimes \left(\textcircled{@}_1^n S_- R^{(1)}_{-}\right).$$

The elements of $S^{(n)}R^{(n)}$ are functions $\tilde{F}_n$ of $\mathbf{t} = (t_n,\ldots,t_1)$, $\mathbf{s} = (s_n,\ldots,s_1)$, which are totally antisymmetric in $\mathbf{t}$ and $\mathbf{s}$ separately and are the boundary values of functions regular in the interior of $\Pi_1^n D$ and $\Pi_1^n(\mathbb{C}\setminus D)$, respectively. The norm is given by

$$\|\tilde{F}_n\|^2 = (n!)^2 \int \mathbf{dt}\, \mathbf{ds} |\tilde{F}_n(\mathbf{t},\mathbf{s})|^2. \tag{6.5}$$

For each n, an orthonormal basis is the set of trigonometric polynomials

$$\psi_{\{m\}\{m'\}}(\mathbf{t},\mathbf{s}) = \frac{1}{(n!)^2} \mathrm{Asy}\left(\prod_{j=1}^{n} \frac{e^{im_j t_j}}{\sqrt{2\pi}}\right) \mathrm{Asy}\left(\prod_{l=1}^{n} \frac{e^{-im'_l s_l}}{\sqrt{2\pi}}\right), \tag{6.6}$$

where $\{m\} = (m_n,\ldots,m_1)$, $\{m'\} = (m'_n,\ldots,m'_1)$ are unordered sequences of nonnegative integers with no two integers equal and Asy denotes antisymmetry in the arguments. If we write $e^{it_j} = \zeta_j$, $e^{is_l} = \eta_l$, then the basis vectors (6.6) can be written as the product of two determinants:

$$\frac{(2\pi)^{-n/2}}{n!} \begin{vmatrix} \zeta_1^{m_1} & \cdots & \zeta_n^{m_1} \\ \vdots & & \\ \zeta_1^{m_n} & \cdots & \zeta_n^{m_n} \end{vmatrix} \cdot \frac{(2\pi)^{-n/2}}{n!} \begin{vmatrix} \eta_1^{-m'_1} & \cdots & \eta_n^{-m'_1} \\ \vdots & & \\ \eta_1^{-m'_n} & \cdots & \eta_n^{-m'_n} \end{vmatrix}, \tag{6.7}$$

and we note the ordering conventions $0 \leqslant m_1 < m_2 < \cdots < m_n$, $0 \leqslant m'_1 < m'_2 < \cdots < m'_n$.

Lemma 6.4: Let $N(n,|m|,|m'|)$ be the number of basis states (6.6), which are trigonometric polynomials of order $|m| = \Sigma_{i=j}^{n} m_i$, $|m'| = \Sigma_{i=1}^{n} m'_i$ in the fermion and antifermion variables, respectively, with $0 \leqslant m_1 < m_2 < \cdots < m_n$, $0 \leqslant m'_1 < \cdots < m'_n$, and let

$$H_n(t) = \sum_{|m| = (1/2)n(n-1)}^{\infty} \sum_{|m'| = (1/2)n(n-1)}^{\infty} N(n,|m|,|m'|) t^{|m|+|m'|}. \tag{6.8}$$

Then

$$H_n(t) = t^{n(n-1)} \sum_{r=1}^{n} \frac{1}{(1-t^r)^2}. \tag{6.9}$$

Proof: Each determinant in (6.7) can be labeled by a partition (or Young diagram) with n rows of boxes and $0 \leqslant m_1 < m_2 < \cdots < m_n$ or $0 \leqslant m'_1 < m'_2 < \cdots < m'_n$ boxes in each row. The lowest-degree polynomial corresponds to the diagram with $[0,1,2,\ldots,(n-1)]$ boxes in the successive rows and will be of degree $\frac{1}{2}n(n-1)$. All other diagrams are obtained by adding an equal number of boxes to the last r rows with $1 \leqslant r \leqslant n$. The generating function for adding a box to each of the last r rows is $1 + t^n + t^{2n} + \cdots = 1/(1-t^n)$. Hence the generating function for all possible determinants will be

$$\begin{aligned} H'_n(t) &= t^{n(n-1)/2}(1+t+t^2+\cdots)(1+t^2+t^4+\cdots) \\ &\quad \cdots(1+t^n+t^{2n}+\cdots) \\ &= t^{n(n-1)/2} \prod_{r=1}^{n} \frac{1}{1-t^r}. \end{aligned}$$

Because (6.7) is a product of two determinants, we have

$$H_n(t) = (H'_n(t))^2 = t^{n(n-1)} \prod_{r=1}^{n} \frac{1}{(1-t^r)^2},$$

which is (6.9).

7. BOSONS AND HARDY SPACES

There is one point in the discussion in Sec. 5 which we did not make explicit there and that is that $\bar{M}_+$ equipped with the complex structure (3.6) is isomorphic, as a complex Hilbert space, to $L^2(\mathbb{R}_+, d\nu)$, where $d\nu(k) = k\theta(k)dk$. Since we know that, on the cyclic subspace of R_1 generated from the vacuum by $\{b(f_+)^* | f_+ \in \bar{M}_+\}$ we are dealing with the Fock representation of $C^*(\bar{M}_+, B)$, we can identify this subspace of R_1 with the usual symmetric Fock space over $\mathscr{H}_B = L^2(\mathbb{R}_+, d\nu)$:

$$\mathscr{F}_B = \mathbb{C} + \mathscr{H}_B \oplus \mathscr{H}_B \textcircled{S} \mathscr{H}_B \oplus \mathscr{H}_B \textcircled{S} \mathscr{H}_B \textcircled{S} \mathscr{H}_B \oplus \cdots.$$

The annihilation and creation operators act in $\mathscr{F}_B$ in the standard fashion. For $F = (F_n) \in F_B$, $F_n \in \overset{(n)}{\mathscr{H}_B \textcircled{S} \cdots \textcircled{S} \mathscr{H}_B}$, we have

$$(b(f)F)_n(k_1,\ldots,k_n)$$

$$= \sqrt{n+1}\int dk\,\hat{f}(k)^* k\theta(k)F_{n+1}(k,k_1,\ldots,k_n),$$

$$(b(f)^*F)_n(k_1,\ldots,k_n) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\hat{f}(k_j)F_{n-1}(k,\ldots,{}_1\hat{k}_j,\ldots,k_n)$$

with the commutation relations

$$[b(f),b(g)^*] = \int dk\,\theta(k)k\hat{f}(k)^*\hat{g}(k),$$

[cf. (5.4)], $f, g\in\overline{M}_+$. Once again there is a Cayley transform which will map $\mathscr{H}_B$ over into a Hardy space.

Definition 7.1: The Hardy space $H^2(\mathbb{T},\mu)$ is the Hilbert space of functions defined on T via the Fourier transform

$$h(t) = \sum_{n=1}^{\infty} a_n e^{\mathrm{int}},$$

from sequences (a_n) with $\|(a_n)\|^2 = 2\pi\,\Sigma_{n=1}^{\infty} n|a_n|^2 < \infty$. The functions satisfy

$$\|h\|^2 = i\int_0^{2\pi} dt\,\frac{dh^*}{dt}(t)h(t) < \infty,$$

and are the boundary values of functions analytic in the interior of the unit disc.

Lemma 7.1: The map $S_B: L^2(\mathbb{R}_+,\nu)\to H^2(\mathbb{T},\mu)$ given by

$$(S_B\hat{f}_+)(t) = f_+\left(i\frac{1-e^{it}}{1+e^{it}}\right),\quad \hat{f}_+\in L^2(\mathbb{R}_+,\nu),$$

where

$$f_+(x) = \frac{1}{\sqrt{2\pi}}\int_0^{\infty} dk\,e^{ikx}\hat{f}_+(k)$$

is an isometry.

Proof: This is a straightforward variation of Lemma 6.3. We note in particular that using the fact that f_+ is the Fourier transform of a function with support on the positive half-axis, we know it has an analytic extension to the upper half-plane so that $S_B\hat{f}_+$ is analytic on the interior of the unit disc. Using S_B, we can now identify $\mathscr{F}_B$ with the Fock space over $H^2(\mathbb{T},\mu)$.

We now return to Lemma 6.1 where we have an expression for the states in R_1 constructed by applying the boson creation operators to the fermion vacuum. Using S, we can represent these states in the Hardy space picture. Suppose that $\hat{f}_+\in L^2(\mathbb{R}_+,d\nu)$ so that

$$(S_B\hat{f}_+)(t) = \sum_{1}^{\infty} a_n e^{\mathrm{int}}.$$

Consider the one-particle boson state in r, given by

$$F_{11}(x,y) = \frac{\tilde{f}(x)-\tilde{f}(-y)}{x+y},$$

where

$$\tilde{f}(x) = \frac{1}{\sqrt{2\pi}}\int_0^{\infty} e^{ikx}\hat{f}_+(k)\,dk.$$

Then

$$SF_{11}(t,s) = \frac{S_B\hat{f}_+(t) - S_B\hat{f}_+(-s)}{2i(1-e^{i(t+s)})}\cdot\frac{(1+e^{it})(1+e^{is})}{(1+e^{it})(1+e^{-is})}$$

$$= -\frac{\Sigma_1^{\infty}a_n(e^{\mathrm{int}}-e^{-\mathrm{ins}})}{2i(e^{it}-e^{-is})}.$$

Observe that if $S_B\hat{f}_+$ is a trigonometric polynomial, then so, too, is SF_{11}. Now n-particle boson states are built up from additional factors of the form $x_1,y_1\to\tilde{f}'(x_1)-\tilde{f}'(-y_1)$ and these map under S into factors of the form $\Sigma_{n=1}^{\infty} b_n(e^{\mathrm{int}} - e^{-\mathrm{ins}})$, where

$$S_B\hat{f}'_+(t) = \sum_{n=1}^{\infty} b_n e^{\mathrm{int}}.$$

We recall the notation $(\hat{f}_j)_+$ for the restriction of the Fourier transform of $f_j\in\overline{M}_+$ to the positive half-axis and record the preceding discussion as the following lemma.

Lemma 7.2: Boson states in R_1 of the form given by (6.4), in which the test functions $f_1,\ldots,f_m$ in $\overline{M}_+$ have the property that $S_B(\hat{f}_j)_+$ is also a trigonometric polynomial, map under S into elements of SR_1, which are also trigonometric polynomials.

In the Fock space over $H^2(\mathbb{T},\mu)$, the n-particle space is given by functions $\tilde{F}_n$, symmetric in their arguments, for which

$$\|\tilde{F}_n\|^2 = i^n n!\int_0^{2\pi} dt_1\cdots dt_n\prod_{j=1}^{n}\frac{\partial}{\partial t_j}\tilde{F}_n(t_1,\ldots,t_n)^*\tilde{F}_n(t_1,\ldots,t_n).$$

$$< \infty.$$

An orthonormal basis for this space is provided by the trigonometric polynomials $\phi_{m_1,\ldots,m_n}$, where

$$\phi_{m_1,\ldots,m_n}(t_1,\ldots,t_n) = \frac{1}{n!}\,\mathrm{Sy}\,\frac{1}{N_{\{m\}}}\prod_{j=1}^{n} e^{im_jt_j}\Big/(2\pi m_j)^{1/2}, \tag{7.1}$$

where $\{m\} = \{m_1,\ldots,m_n\}$ is an unordered sequence of non-zero positive integers and $N_{\{m\}}$ is the normalization constant, and Sy denotes symmetrization.

Lemma 7.3: If $N_B(p)$ is the number of independent trigonometric polynomials of order p of the form (7.1), where $p = \Sigma_{i=1}^{n} m_i$ and $G(t) = \Sigma_{|m|=1}^{\infty} N_B(p)t^{|m|}$, then

$$G(t) = \prod_{r=1}^{\infty}\left(\frac{1}{1-t^r}\right) - 1. \tag{7.2}$$

Proof: $N_B(p)$ is the number of partitions of p excluding $p = 0$ and (7.2) is the well-known formula for the generating function.

Let us return now to Lemmas 7.2 and 6.1 Let $\xi = (e^{it_1},\ldots,e^{it_n})$ and $\eta^{-1} = (e^{-is_1},\ldots,e^{-is_n})$. Consider the state in R_1, $F^{\vee}_{m_n}$, obtained by applying m boson operators to the vacuum and considering the component containing $n\leqslant m$ fermion, antifermion pairs.

When the test functions f_j are such that $S_B(\hat{f}_j)_+$ is a trigonometric polynomial, $S^{(n)}F^{\vee}_{mn}$ is a polynomial in ξ_i, η_i^{-1} $(i = 1,\ldots,n)$. Moreover, in the case where the product of the numerator factors $\tilde{f}_i$ is a homogeneous polynomial of degree p, because there are n first-degree factors in ξ_i and η_i^{-1} in the denominator, $S^{(n)}F^{\vee}_{mn}$ will be a homogeneous polynomial of degree $p-n$. An m-boson state is always a

direct sum of states containing n pairs with $n \leqslant m$, say,

$$\psi_m = \sum_{n=1}^{m} \oplus (S^{(n)} F^{\vee}_{mn})(\xi, \eta^{-1}).$$

By taking linear combinations of ψ_m for different choices of $\tilde{f}_i$ we can construct a linearly independent set of polynomial states ϕ_{mn} with a maximum number of pairs equal to n with $1 \leqslant n \leqslant m$. Noting that states with different values of m must also be linearly independent, we can let $N(m,n,p)$ be the number of independent polynomial states so obtained with m bosons, a maximum of n pairs, and polynomial degree of the n-pair state p.

Now let $N_B(m,p)$ denote the number of m boson states of degree p. By definition $N_B(p) = \Sigma_{m=1}^{p} N_B(m,p)$, where $N_B(p)$ is defined in Lemma 7.3. Note that the summation over m must terminate at p because each single boson state is of degree $\geqslant 1$. From their definition, we have

$$N_B(m,p) = \sum_{n=1}^{m} N(m,n,p-n), \tag{7.3}$$

where the summand in (7.3) is zero if $n \geqslant p$. Note that the entry $p-n$ appears because an n-pair state of degree p' arises from a m-boson state of degree $p'+n$, as explained above.

We now have the inequality

$$\sum_{m=n}^{p-n(n-1)} N(m,n,p) \leqslant N_p(n,p), \tag{7.4}$$

where $N_p(n,p)$ is the coefficient of t^p in (6.9) (i.e., the numbers of independent n-pair states of total polynomial degree p). The inequality follows because $N(m,n,p)$ counts the number of independent states of degree p with n pairs and m bosons, and the total number of such states, when summed over m, is certainly less than $N_p(n,p)$. Then for $0 < t < 1$ we have

$$\begin{aligned} G(t) &= \sum_{p=1}^{\infty} N_B(p) t^p = \sum_{p=1}^{\infty} \sum_{m=1}^{\infty} N_B(m,p) t^p \\ &= \sum_{n=1}^{\infty} \sum_{p=n^2}^{\infty} \sum_{m=n}^{p-n(n-1)} N(m,n,p-n) t^p \\ &\leqslant \sum_{n=1}^{\infty} \sum_{p=n^2}^{\infty} N_p(n,p-n) t^p \\ &= \sum_{n=1}^{\infty} \sum_{p=n(n-1)}^{\infty} N_p(n,p) t^{p+n} \\ &= \sum_{n=1}^{\infty} t^{n^2} \prod_{r=1}^{n} \frac{1}{(1-t^r)^2} \quad \text{from (6.9)} \\ &= G(t) \end{aligned}$$

from Euler's identity in the theory of partitions. Hence

$$\sum_{m=n}^{p-n(n-1)} N(m,n,p-n) = N_p(n,p-n),$$

and this means that there are sufficient polynomial pair states produced by the boson operators to approximate any pair state. This completes the proof of Proposition 4.3.

8. APPLICATION TO SOLVABLE MODELS

There are a number of alternative approaches to the "bosons into fermions" programme which we have not touched on. A discussion of these is contained in Ringwood.[12] He remarks that the various other methods of reconstructing fermions from bosons (for example, Ref. 13) do not in fact yield fermion fields satisfying the CAR's. These approaches all depend on expressing the fermion field essentially as an exponential of a boson field. The difficulty with this has been pointed out by Ruijsenaars[14] and our calculations confirm his finding that this only makes sense when interpreted in terms of quadratic forms and not as operators. Naturally, statements about vacuum expectation values of quadratic forms do not immediately imply anything about algebraic properties of associated operators (which need not even exist).

On the other hand, the approach of Ref. 2 apparently yields fermion fields with the correct anticommutation relations. In view of Refs. 12 and 14, it is certainly worth checking, at least for the free massless fermion field, how the reconstructed fermions relate to the original ones (so that one has an operator equivalence and not just an equivalence of vacuum expectation values). We intend to investigate this point, as well as the corresponding argument for the massless Thirring model elsewhere, using the results of this paper. In our analysis of the Schwinger model,[3] we employed some of the results of Ref. 2 for the free massless fermion field. This paper justifies in part that application of Ref. 2. However, Ref. 3 remains incomplete since it would be much more satisfactory to have a method of defining the fermion solutions of the Schwinger model equations of motion directly rather than in an *ad hoc* way.

A second factor which motivated this analysis was our interest in $(QED)_2$ (i.e., massive fermions). The only results we are aware of here involve the charge zero vector[15] or deal with various cutoff versions of the model.[16,17] Ideally, one would like a working version of $(QED)_2$ in which a representation of the local U(1) gauge group would be defined in the Hilbert space of the model. Presumably this would give some qualitative understanding of the effect of local gauge invariance in quantum field theory. Such an analysis would require an understanding of the charge sectors for $(QED)_2$, as well as a treatment of the electromagnetic potential in general gauges and therefore goes well beyond.[15-17]

As to how the results of this paper apply, we believe that an understanding of the massive free fermion currents will be a first step in analyzing these aspects of $(QED)_2$. The main result of the previous analysis for massless fermions is that in the zero charge sector ($Q = Q_5 = 0$), the currents are free massless Bose fields, and that the other sectors would be obtained from this one by the action of an automorphism of the C^* algebra of the currents. In the massive case, Q_5 does not exist and one is left with the charge operator Q alone. It is a straightforward matter to compute the two-point function for the representation of the massive time-zero currents in the $Q = 0$ sector. However, the time evolution of the currents (being determined by the evolution of the fermions) is not an automorphism of the C^* algebra generated by the time-zero currents. Consequently, one cannot expect a precise analog, in the massive case, of the result presented here. However, it does appear that the massive currents for all times act cyclicly in the charge-zero sector, although possibly not irreducibly. Whether the analysis of this "current

algebra" will assist in makng sense of the interaction term $j_\mu(x)A^\mu(x)$ of $(\mathrm{QED})_2$ remains to be seen. We intend to report on this subsequently.

[1]R. F. Streater, in *Physical Reality and Mathematical Description*, edited by C. P. Enz and J. Mehra (Reidel, Boston, 1974) and references therein.
[2]G. F. Dell-Antonio, Y. Frischmann, and D. Zwanziger, Phys. Rev. D **6**, 988–1007 (1972).
[3]A. L. Carey and C. A. Hurst, Commun. Math. Phys. **80**, 1–21 (1981).
[4]D. A. Uhlenbrock, Commun. Math. Phys. **4**, 64–76 (1967).
[5]L.-E. Lundberg, "Observable Algebra Approach to the Thirring–Schwinger Model," Copenhagen preprint, 1979.
[6]P. Basarab-Horwath, R. F. Streater, and J. Wright, Commun. Math. Phys. **68**, 195–207 (1979).
[7]A. L. Carey, C. A. Hurst, and D. M. O'Brien, J. Funct. Anal. **48**, 360–393 (1982).
[8]L.-E. Lundberg, Commun. Math. Phys. **50**, 103–112 (1976).
[9]A. K. Raina and G. Wanders, Ann. Phys. **132**, 404–26 (1981).
[10]J. Manuceau, Ann. Inst. Henri Poincaré **8**, 139–161 (1968).
[11]A. Devinatz, in *Functional Analysis*, edited by B. Gelbaum (Thompson, Washington, 1967).
[12]G. A. Ringwood, Nuovo Cimento A **54**, 483–494 (1979).
[13]R. Seiler and D. A. Uhlenbrock, Ann. Phys. **115**, 81–110 (1977).
[14]S. N. M. Ruijsenaars, Ann. Phys. **132**, 328–382 (1981).
[15]J. Fröhlich and E. Seiler, Helv. Phys. Acta **49**, 889–924 (1976).
[16]D. Weingarten and J. Challifour, Ann. Phys. **123**, 61–101 (1979).
[17]K. R. Ito, "Construction of Euclidean $(\mathrm{QED})_2$ via Lattice Gauge Theory," Bedford College preprint, 1981.

Current responses of first and second order in a collisionless plasma. III. Three-wave interaction in periodic plasmas

Jonas Larsson
Department of Plasma Physics, Umeå University, S-901 87 Umeå, Sweden

(Received 6 April 1982; accepted for publication 14 January 1983)

Three-wave interaction in a possibly strongly inhomogeneous plasma was considered in Paper II. These results are now generalized in order to treat also time-dependent periodic systems. The coupled mode equations as given in this paper are applicable to a very large class of unperturbed systems including those in Paper II and plasmas with self-consistent or in external oscillating field and also plasmas with some crystal symmetry in space-time. In the particular case when the relativistic Vlasov equation is used to describe the plasma we give symmetric expressions for the coupling coefficients such that the Manley–Rowe relations follow if wave–particle interaction is neglected.

PACS numbers: 52.40.Db, 52.25. – b, 03.65. – w

1. INTRODUCTION

In Paper II we considered resonant three-wave interaction in a possibly strongly inhomogeneous but stationary plasma. It is of considerable interest also to treat oscillating backgrounds, like plasmas with a self-consistent large amplitude electromagnetic wave present or plasmas in some external oscillating field. The importance of such systems is evident, for example, in connection with laser fusion or radio frequency heating of magnetically confined plasmas.

In II the background was assumed to have one or several directions of homogeneity in space-time with at least one such direction timelike. A mathematically natural way to generalize this so that also oscillating systems are included is to replace the condition of "homogeneity in some direction" by "translational invariance with respect to some 4-vector." In this Paper III we thus consider plasmas that unperturbed are unchanged by the translation in space-time by some timelike 4-vector. There may also be other 4-vectors of translational invariance. This defines a very large class of unperturbed systems including those in II as well as oscillating systems and also plasmas with the symmetry of a perfect crystal in some linear subspace of space-time.

The purpose of this paper is to give a generalization of the coupled mode equations in II valid for the large class of unperturbed systems discussed above. The coefficients in the coupled mode equations are expressed in terms of the 4-current response operators $\delta J^{(1)}[\phi]$ and $\delta J^{(2)}[\phi,\phi]$, giving the 4-current due to the perturbation ϕ in the 4-potential, and in terms of the operator $\delta J^{(1-)}$ defined by (3.9) (I). So far we have not specified a model for the plasma particles, only that the electromagnetic field and the 4-current may be expressed in terms of a 4-potential field. The coupled mode equations, in this form, are given in Sec. 2.

An essential motivation for deriving the result in Sec. 2 is the existing general expressions[1–4] for the 4-current response operators in the particular case when the plasma is described by the Vlasov equation. These expressions exhibit the mathematical structure behind the Manley–Rowe relations, which for a Vlasov plasma are valid when wave–particle interactions may be neglected. In the case with a stationary background this mathematical structure was considered in some detail in I. When the particle orbits had certain almost periodic properties, it was then explicitly seen that the symmetries followed if certain pole contributions were neglected in close analogy to the well-known homogeneous case. It is sufficiently evident that these considerations in Sec. 4(I) can be generalized to include the class of backgrounds considered in this part III, and it seems hardly worthwhile doing it explicitly.

The results of this paper have been formulated with as few new notations as possible. The possibility of background plasmas with "crystal symmetry," however, makes concepts (well-known from solid-state physics) like "reciprocal lattice" and "unit cell" unavoidable. In the particular case of one-dimensional crystal symmetry, these things become easier than in higher dimensions, and, since the one-dimensional case is of considerable physical interest including plasmas with periodically oscillating fields, the one-dimensional case has been given some separate attention in Sec. 2. The mathematical details for the higher-dimensional case are given in Sec. 4 B. In Sec. 3 we consider a plasma described by the relativistic Vlasov equation. The derivations are given in Sec. 4.

The application of the results in this paper as an instrument for treating three-wave interactions in a great variety of situations as well as an unifying tool is further discussed and exemplified in Paper IV.

2. THE COUPLED MODE EQUATIONS

The notations in I, II, and Ref. 4 and the set-theoretic notations in Sec. 4 A are used. Thus space-time is a four-dimensional real affine space E with the space V of 4-vectors as the associated vector space of translations. As in I and II we choose an arbitrary event $O\in E$ as origin. Then to each vector $x\in V$ there corresponds an event $P\in E$ such that $P = O + x$. Often in notation we identify x and P and write $x\in E$, etc. The Lorentz metric (called a scalar product in Ref. 4; cf. Sec 4 A) is given as a symmetric and nondegenerate bilinear form $\cdot$ on V: $a\cdot b\in R$ = real numbers for $a,b\in V$. The

 0022-2488/83/082222-09$02.50

Lorentz metric has negative index of inertia 1 and positive index 3.

A. The unperturbed plasma

Definition 1: (a) T is the set of all 4-vectors of translational invariance for the unperturbed system.

(b) V_t is the vector space spanned by T; V_h is the vector space of all directions of homogenity for the unperturbed system:

$$V_i = V_t^{\perp}, \quad V_p = V_h^{\perp} \cap V_t.$$

(c) $T_p = T \cap V_p$.

Assumptions: The unperturbed systems we consider in this paper satisfy the following two properties:

(a) V_t contains a timelike vector;

(b) V_h is a nondegenerate subspace of V.

Lemma 1: $V = V_h + V_p + V_i$, $V = V_t + V_i$, $V_t = V_h + V_p$,

$$V_h \cap V_p = V_h \cap V_i = V_p \cap V_i = V_t \cap V_i = \{0\}.$$

V_t, V_h, V_p, and V_i are nondegenerate subspaces of V.

Remark 1: Assumption (a) gives a timelike structure needed in the definition of a useful wave concept. It is also easy to see that assumption (a) implies that V_t is nondegenerate subspace of V. This fact and assumption (b) are sufficient for Lemma 1 to be valid. Note that degeneracy of V_t or V_h would imply that $V_t \cap V_i$ or $V_h \cap V_p$ contains nonzero vector (let, for example, V_h be degenerate: this means by definition the existence of a nonzero $a \in V_h$ such that $a \cdot V_h = \{0\}$; then $a \in V_h^{\perp} \cap V_h = V_h \cap V_p$), and Definition 2 below would then not make sense.

It is easy to show that assumption (b) means that V_h either contains a timelike vector or only spacelike vectors. Thus V_h containing lightlike, but no timelike vectors are excluded.

Remark 2: The vector space V is thus divided in three parts V_h, V_p, and V_i, which reflects the translational invariance structure of the background. In I and II the particular case dim $V_p = 0$ was considered. In this paper we deal with the general case with arbitrary dimension on V_p. As was noted in the Introduction, the cases dim $V_p \geqslant 2$ involves some terminology from the theory of crystals which is introduced in Sec. 4 B.

Definition 2: (a) Subscripts t, h, p, or i on a vector $a \in V$ are defined by $a_t + a_i = a$, $a_h + a_p + a_i = a$ and $a_t \in V_t$, $a_i \in V_i$, $a_h \in V_h$, and $a_p \in V_p$ (see Lemma 1).

(b) $\nabla_E = \nabla_t + \nabla_i = \nabla_h + \nabla_p + \nabla_i$ defines ∇_t, ∇_i, ∇_h, and ∇_p in analogy to (a) above.

Lemma 2: The set T_p spans V_p and have the translational invariant structure of a perfect crystal; i.e., there exists a basis $\{v_1,\ldots,v_n\}$ of V_p such that

$$T_p = \{s_1 v_1 + \cdots + s_n v_n \,|\, s_j \text{ integers for } j = 1,2,\ldots,n\}.$$

Definition 3: (a) For dim $V_p = 1$ we define the 4-vector of periodicity p so that $T_p = \{sp | s = 0, \pm 1, \cdots\}$ (see Lemma 2). The vector p is uniquely determined except for sign. If p is timelike, we take p to be future-oriented. Also define $\hat{p} = 2\pi (p \cdot p)^{-1} p$ so that $p \cdot \hat{p} = 2\pi$ and $\hat{T}_p = \{s\hat{p} | s = 0, \pm 1, \cdots\}$. Define $U = [0, p]$ and $\hat{U} = [0, \hat{p}]$.

(b) För arbitrary dim $V_p \geqslant 1$ we define the set of reciprocal vectors to T_p as

$$\hat{T}_p = \{v \in V_p \,|\, v \cdot T_p \subset \{2\pi s | s = 0, \pm 1, \cdots\}\}.$$

A unit cell for T_p is denoted U and a unit cell for $\hat{T}_p$ is denoted $\hat{U}$. The details are given in Sec. 4 B.

B. The current response operators

We consider the current response operators $\delta J^{(1)}$ and $\delta J^{(2)}$ giving the 4-current response on a perturbation ϕ of the 4-potential so that the change in 4-current due to ϕ is to second order in ϕ equal to $\delta J^{(1)}[\phi] + \delta J^{(2)}[\phi,\phi]$. The operator $\delta J^{(1)}$ is linear and $\delta J^{(2)}$ bilinear and symmetric. The perturbation ϕ is a 4-vector field on space-time, i.e., $\phi E \to V$, and, in order to get causal operators, ϕ is assumed to vanish towards the past [$\phi \in L_0(E,V)$ in the notation of Ref. 4]. From the equations governing the motion of the plasma particles, we do not specify these equations in this section, the current responses due to ϕ may be calculated, and causality follows then from the condition that also $\delta J^{(1)}[\phi](x)$ and $\delta J^{(2)}[\phi,\phi](x)$ vanish when the point x in space-time is moved towards the past. It will, however, be very convenient to have the response operators defined also on some perturbations ϕ that do not vanish towards the past.

Definition 4: (a) Let α be a function, $\alpha: V_p + V_i \to C$ (C = complex numbers) such that $\alpha(v + a) = \alpha(v)$ for $v \in V_p + V_i$ and $a \in T_p$. Define

$$\langle \alpha \rangle = \int_{V_i} \alpha(x_i)\, dx_i \quad \text{for dim } V_p = 0,$$

$$\langle \alpha \rangle = \int_U dx_p \int_{V_i} \alpha(x_p + x_i)\, dx_i \Big/ \int_U dx_p \quad \text{for dim } V_p \geqslant 1.$$

In particular, for dim $V_p = 1$, we obtain

$$\langle \alpha \rangle = |p \cdot p|^{-1/2} \int_{[0,p]} dx_p \int_{V_i} \alpha(x_p + x_i)\, dx_i = \int_0^1 ds \int_{V_i} \alpha(sp + x_i)\, dx_i. \tag{2.1}$$

(b) $P = P(V_p + V_i, V^+)$ is a set of functions $\psi: V_p + V_i \to V^+$ such that $\psi(v + a) = \psi(v)$ for $v \in V_p + V_i$ and $a \in T_p$ and such that $\langle |b \cdot \psi| \rangle < \infty$ for all 4-vectors b.

(c) Let $L = (e_0, e_1, e_2, e_3)$ be a Lorentz frame.[5] Define $P_L = \{\psi \in P \,|\, e_0 \cdot \psi \equiv 0\}$.

(d) $P(\kappa) = P \cdot e^{i\kappa \cdot x} = \{\phi \,|\, \phi: E \to V^+, \phi(x) = \psi(x_p + x_i) e^{i\kappa \cdot x}\}$, where $\kappa \in V_t$:

$$P_L(\kappa) = \{\phi \in P(\kappa) | e_0 \cdot \phi = 0\}.$$

(e) Define the bilinear from $\langle\ ,\ \rangle$ on $P(\kappa)$ by $\langle \phi_1, \phi_2 \rangle = \langle \phi_1^* \cdot \phi_2 \rangle$.

Remark 3: It is natural to consider perturbations and normal modes of the form $\phi + \phi^*$, where $\phi \in P(\kappa)$. In this way we take the translational invariance structure of the unperturbed system into account. In the particular case of homogeneous plasma we observe that $P(\kappa)$ is just the plane waves. We now want to have the operators $\delta J^{(1)}$, $\delta J^{(1-)}$, and $\delta J^{(2)}$ defined on $P(\kappa)$ and $P(\kappa) \times P(\kappa)$ respectively. Clearly, we have

to extend the domains of definition for these operators. First we extend the operator $\delta J^{(1)}$ to domain $L_0(E,V^+) = L_0(E,V) + iL_0(E,V)$ so that its linear property is preserved; $\phi \in L_0(E,V^+)$ is expressed as $\phi = \phi_1 + i\phi_2$ with $\phi_1,\phi_2 \in L_0(E,V)$ and $\delta J^{(1)}[\phi] = \delta J^{(1)}[\phi_1] + i\delta J^{(1)}[\phi_2]$.[6] We define $\delta J^{(1-)}$ on $L^0(E,V^+)$ and $\delta J^{(2)}$ on $L_0(E,V^+)\times L_0(E,V^+)$ in an analogous way. However, since the functions in $P(\kappa)$ do not vanish towards the past or the future, we have $P(\kappa)\cap L_0(E,V^+) = P(\kappa)\cap L^0(E,V^+) = \{0\}$. Let e be a future-oriented timelike 4-vector. For $\phi \in P(\kappa)$ we define $\phi_s(x) = \phi(x)e^{-se\cdot x}$. Then $\phi_s \in L_0(E,V^+)$ for $s>0$ and $\phi_s \in L^0(E,V^+)$ for $s<0$ and $\phi_s \to \phi$ when $s \to 0$.

Definition 5: Take $\phi \in P(\kappa)$, $\phi_j \in P(\kappa_j)$, $j = 1,2$ and $\kappa,\kappa_j \in V_t$. Take ϕ_s as in the remark above. Define

$$\delta J^{(1)}[\phi] = \lim_{s\to 0+} \delta J^{(1)}[\phi_s],$$

$$\delta J^{(1-)}[\phi] = \lim_{s\to 0-} \delta J^{(1-)}[\phi_s],$$

$$\delta J^{(2)}[\phi_1,\phi_2] = \lim_{s\to 0+} \delta J^{(2)}[\phi_{1s},\phi_{2s}].$$

Lemma 3: $\delta J^{(1)}: P(\kappa) \to P(\kappa)$; $\delta J^{(1-)}: P(\kappa) \to P(\kappa)$ for $\kappa \in V_t$; $\delta J^{(2)}: P(\kappa_1)\times P(\kappa_2) \to P(\kappa_1+\kappa_2)$ for $\kappa_1,\kappa_2 \in V_t$.

Lemma 4: Let $L = \{e_0,e_1,e_2,e_3\}$ be a Lorentz frame[5] and Π_L the projection on V taking the spatial part; $\Pi_L v = v + e_0\cdot v e_0$. Then $\Pi_L \circ (\delta J^{(1)} + \delta J^{(1-)})$ is an Hermitian operator on $P_L(\kappa)$.

Remark 4: The bilinear form $\langle\ ,\ \rangle$ is positive definite on $P_L(\kappa)$ but not on $P(\kappa)$. We consider $P_L(\kappa)$ to be a Hilbert space with this inner product.

C. The coupled mode equations

The electromagnetic wave equation for the perturbation ϕ of the 4-potential is to second order in ϕ

$$\nabla_E\cdot(\nabla_E \wedge \phi) = -(\mu_0/\epsilon_0)^{1/2}(\delta J^{(1)}[\phi] + \delta J^{(2)}[\phi,\phi]). \tag{2.2}$$

Decomposing the linear part in (2.2) as the sum of one essentially Hermitian and one anti-Hermitian part (see Lemma 5 below) yields

$$H[\phi] + ih[\phi] = -(\mu_0/\epsilon_0)^{1/2}\delta J^{(2)}[\phi,\phi], \tag{2.3}$$

where

$$H[\phi] = \nabla_E\cdot(\nabla_E \wedge \phi) + 2^{-1}(\mu_0/\epsilon_0)^{1/2}(\delta J^{(1)}[\phi] + \delta J^{(1-)}[\phi]), \tag{2.4}$$

$$h[\phi] = -i2^{-1}(\mu_0/\epsilon_0)^{1/2}(\delta J^{(1)}[\phi] - \delta J^{(1-)}[\phi]). \tag{2.5}$$

Remark 5: In a Lorentz frame $L = (e_0,e_1,e_2,e_3)$ denoting $x \in E$ with $x = (x^i) = x^i e_i$ ($= \Sigma_0^3 x^i e_i$; we use the summation convention) and raising and lowering indices means $e_0 = -e^0, e_i = e^i, x^0 = -x_0$, and $x^i = x_i$ for $i = 1,2,3$. We then have

$$\nabla_E = e^i\frac{\partial}{\partial x^i} = e_i\frac{\partial}{\partial x_i}.$$

The wave operator $\Box$ is defined:

$$\Box = \nabla_E\cdot\nabla_E = \frac{\partial^2}{\partial x^i\,\partial x_i}. \tag{2.6}$$

The operator on the left-hand side in (2.2) is equal to $\Box$ if we use the Lorentz gauge ($\nabla_E\cdot\phi = 0$) since

$$\nabla_E\cdot(\nabla_E \wedge \phi) = (\nabla_E\cdot\nabla_E)\phi - \nabla_E(\nabla_E\cdot\phi). \tag{2.7}$$

Lemma 5: Let L be a Lorentz frame.[5] Then $\Pi_L \circ H$ and $\Pi_L \circ h$ are Hermitian operators on $P_L(\kappa)$.

Lemma 6: Let $L = (e_0,e_1,e_2,e_3)$ be a Lorentz frame[5] and take $\kappa \in V_t$ such that $e_0\cdot(\kappa + a) \neq 0$ for all $a \in T_p$. Then for each $\phi \in P(\kappa)$ there exist a unique $\chi \in P_L(\kappa)$ such that $\nabla_E \wedge \phi = \nabla_E \wedge \chi$.

Remark 6: Thus from Lemma 6 it is seen that we may choose to consider perturbations only in the subspace $P_L(\kappa)$ of $P(\kappa)$. This just means that we have choosen a particular electromagnetic gauge, namely, a zero scalar potential in frame L. It is after we have choosen gauge that Hilbert space and Hermitian operators appear in Lemmas 4 and 5.

Result 1: Let $\phi_j \in P(\kappa_j)$ be given such that $\kappa_1 + \kappa_2 + \kappa_3 = 0$, $\kappa_j \in V_1$, and $H(\phi_j) = 0$. Also assume that h may, to first order, be neglected in comparison with H in (2.3) so that $\phi_j + \phi_j^*$ may be regarded as normal modes.

If the wave equation (2.3) with second order terms in ϕ and damping (h is not neglected) have a solution of the form

$$\sum_{j=1}^{3} A_j(x_t)\phi_j(x) + \text{c.c.}, \tag{2.8}$$

where $A_j(x_t)$ is varying slowly in x_t (in comparison with ϕ_j) due to resonant wave interaction and damping, then A_j satisfy the coupled mode equations.

$$\left(\frac{\partial}{\partial\kappa}\langle\chi_{3\kappa}^*\cdot H[\chi_{3\kappa}]\rangle\right)_{\kappa=\kappa_3}\cdot\nabla_t A_3^* - \langle\phi_3^*\cdot h[\phi_3]\rangle A_3^* = 2i(\mu_0/\epsilon_0)^{1/2}\langle\phi_3\cdot\delta J^{(2)}[\phi_1,\phi_2]\rangle A_1 A_2, \tag{2.9}$$

and the two equations are obtained by permuting subscripts 1, 2, and 3. Here $\partial/\partial\kappa = \nabla_t$ acting on $\kappa \in V_t$ [in coordinates $L = (e_0,e_1,e_2,e_3)$, $\kappa = (\omega/c)e_0 + \mathbf{k}$, $\nabla_t = -e_0 c\partial/\partial\omega + \partial/\partial\mathbf{k}$ while ∇_t acting on $x_t = cte_0 + \mathbf{r}$ is $\nabla_t = -e_0 c^{-1}\partial/\partial t + \partial/\partial\mathbf{r}$). We define $\chi_{3\kappa}(x) = \phi_3(x)\exp[i(\kappa-\kappa_3)\cdot x]$. Equation (2.9) is gauge invariant, i.e., the coefficients are independent of the gauge choosen for ϕ_1, ϕ_2, and ϕ_3, the equation is also written in a covariant (coordinate-free) form. The coefficients on the left-hand side of (2.9) are real. The 4-vector $((\partial/\partial\kappa)\langle\chi_{3\kappa}^*\cdot H[\chi_{3\kappa}]\rangle)_{\kappa=\kappa_3}$ is proportional to the group 4-velocity $u_{g3} \in V_t \cap S$.

Remark 7: The relation corresponding to (2.18) (II) does not in the general case with time-dependent unperturbed state involve the wave energy but rather the wave action $N_j|A_j(x_t)|^2$, where

$$N_j = \epsilon_0\left(\frac{\partial}{\partial\omega}\langle\chi_{j\kappa}^*\cdot H[\chi_{j\kappa}]\rangle\right)_{\kappa=\kappa_j}. \tag{2.10}$$

In the particular case with a time-dependent background we have $N_j = W_j/\omega_j$. Also note that a factor c^{-1} will be inserted on the left-hand side in (2.15) (II); there we have $c^{-1}e_0\cdot\nabla_h = \partial/\partial\omega$ and the so-corrected (2.15) (II) is consistent with (2.18) (II).

Remark 8: The ansatz (2.8) is not always sufficiently general. The discussion in the introduction in II is, with mi-

nor modifications, relevant also for the more general situation considered in the present paper. We may, for example, have to consider functions A_j depending not only on x_t but also on x_i. If so, the coupled mode equations (2.9) must be replaced with more complicated ones. On the left-hand side we must then have some linear operator which take the x_i dependence into account.

Corollary 1: If $h = 0$ and $\langle \phi_3 \cdot \delta J^{(2)}[\phi_1,\phi_2]\rangle = \langle \phi_2 \cdot \delta J^{(2)}[\phi_1,\phi_3]\rangle = \langle \phi_1 \cdot \delta J^{(2)}[\phi_2,\phi_3]\rangle$, then the Manley–Rowe relations are valid, i.e.,

$$\left[\frac{\partial}{\partial \kappa}\langle \chi^*_{j\kappa} \cdot H[\chi_{j\kappa}]\rangle\right]_{\kappa=\kappa_j} \cdot \nabla_t A_j(x_t) \tag{2.11}$$

is independent of $j \in \{1,2,3\}$.

Remark 9: The assumptions in Corollary 1 mean for a Vlasov plasma that wave-particle interactions are neglected.

Corollary 2: The mode coupling equation (2.9) may, in terms of the Lorentz frame mentioned in Result 1, be written as

$$\left(\frac{\partial}{\partial t} + \mathbf{v}_{g3} \cdot \frac{\partial}{\partial \mathbf{r}} + \nu_3\right) A_3^*(t,r) = \frac{i}{N_3} V(3,1,2) A_1(t,r) A_2(t,r). \tag{2.12}$$

We define

$$N_j = \epsilon_0 \left(\frac{\partial}{\partial \omega}\langle \chi^*_{j\kappa} \cdot H \chi_{j\kappa}\rangle\right)_{\kappa=\kappa_j}, \tag{2.13}$$

$$\mathbf{v}_{gj} = -\epsilon_0 \left(\frac{\partial}{\partial \mathbf{k}}\langle \chi^*_{j\kappa} \cdot H \chi_{j\kappa}\rangle\right)_{\kappa=\kappa_j} N_j^{-1}, \tag{2.14}$$

$$\nu_j = \epsilon_0 \langle \phi_j^* \cdot h \phi_j\rangle N_j^{-1}, \tag{2.15}$$

$$V(3,1,2) = -2c^{-1}\langle \phi_3 \cdot \delta J^{(2)}[\phi_1,\phi_2]\rangle. \tag{2.16}$$

3. THE COEFFICIENTS IN THE COUPLED MODE EQUATIONS IN A VLASOV PLASMA

We obtain from Sec. 2 (I) the following formulas for the coefficients in (2.9) when the plasma is described by the relativistic Vlasov and Maxwell equations. The particle species index σ is for notational convenience omitted in most places below. The unperturbed plasma is determined by the distribution function $f_0(x_p,x_i,u)$ (one for each particle species) and the 4-potential $\Phi_0(x)$. They satisfy the translational invariance property

$$f_0(x_p + a,x_i,u) = f_0(x_p,x_i,u),$$
$$\nabla_E \wedge \Phi_0(x_p + a,x_i) = \nabla_E \wedge \Phi_0(x_p,x_i) \quad \text{for } a \in T_p \tag{3.1}$$

and the Vlasov equation[4]

$$u \cdot \nabla_E f_0 + q m_0^{-1} c^{-2} (\nabla_E \wedge \Phi_0 \cdot u) \cdot \nabla_S f_0 = 0. \tag{3.2}$$

Result 2:

(a) Let $\phi_j \in P(\kappa_j)$ for $j = 0,1$, $\kappa_0 + \kappa_1 = 0$, and $\kappa_0,\kappa_1 \in V_t$. Then

$$\langle \phi_0 \cdot \delta J^{(1)}[\phi_1]\rangle = \langle \phi_1 \cdot \delta J^{(1-)}[\phi_0]\rangle = 2^{-1} q c \left\langle \int_S f_0(x_p,x_i,u)[\delta x(1) \cdot \nabla_E (u \cdot \phi_0) + \delta u(1) \cdot \phi_0 + \delta x(0) \cdot \nabla_E (u \cdot \phi_1) + \delta u(0) \cdot \phi_1]\, du \right\rangle \tag{3.3}$$

with δx and δu defined in (c) below.

(b) Let $\phi_j \in P(\kappa_j)$ for $j = 0,1,2$, $\kappa_0 + \kappa_1 + \kappa_2 = 0$, and $\kappa_0,\kappa_1,\kappa_2 \in V_t$. Then

$$\langle \phi_0 \cdot \delta J^{(2)}[\phi_1,\phi_2]\rangle = \sum_{\substack{\alpha,\beta,\gamma=0 \\ \alpha\neq\beta\neq\gamma\neq 0}}^{2} 2^{-1} q c \left\langle \int_S f_0(x_p,x_i,u) \times [6^{-1}\delta x(0) \otimes \delta x(1) \otimes \delta x(2) : \nabla_E \otimes \nabla_E \otimes \nabla_E (u \cdot \Phi_0) + 2^{-1}\delta x(\alpha) \otimes \delta x(\beta) \otimes \delta u(\gamma) : \nabla_E \otimes \nabla_E \otimes \Phi_0 + 2^{-1}\delta x(\alpha) \otimes \delta x(\beta) : \nabla_E \otimes \nabla_E (u \cdot \phi_\gamma) + \delta x(\alpha) \otimes \delta u(\beta) : \nabla_E \otimes \phi_\gamma]\, du \right\rangle. \tag{3.4}$$

(c) $$\delta x(j) = -q m_0^{-1} c^2 \{\nabla_S D_0^{-1}(u \cdot \phi_j) + D_0^{-1}(u \cdot D_0 \nabla_S D_0^{-1}(u \cdot \phi_j))u\}, \tag{3.5a}$$

$$\delta u(j) = D_0 \delta x(j). \tag{3.5b}$$

We must here specify boundary conditions in order to define D_0^{-1}. For this purpose, we replace κ_j in ϕ_j by $\kappa_j + ise$, where e is a timelike future oriented 4-vector and we take $s < 0$ for $j = 0$ and $s > 0$ for $j = 1,2$. Then the resulting functions ϕ_{js} vanish towards the future for $j = 0$ and towards the past for $j = 1,2$. Now D_0^{-1} is determined by the condition that $\delta x(0)$ vanishes towards the future and $\delta x(1)$ and $\delta x(2)$ towards the past. Calculating D_0^{-1} by integrating along unperturbed orbits, this means that for $j = 0$ we shall integrate to the infinite future and for $j = 1,2$ to the infinite past. Afterwards we take the limit $s \to 0-$ for $j = 0$ and $s \to 0+$ for $j = 1,2$ (cf. Definition 5).

4. SOME MATHEMATICS AND DERIVATIONS FOR SECTIONS 2 AND 3

A. Vector space with a metric tensor

Let X be a finite-dimensional real vector space with a *metric tensor* or *scalar product* or *inner product* or whatever one likes to call a symmetric and nondegenerate bilinear form $\cdot$ on X: $v \cdot w \in R$ for $v,w \in X$. The terminology differs between the textbooks; in Ref. 7 the above structure on X is called metric tensor in Ref. 8, inner product, and in Ref. 9, scalar product or metric tensor. In Ref. 7 it is, however, an inner product always positive or negative definite.

A subspace A of X is called nondegenerate if the bilinear form on A inherited from X is nondegenerate. In this paper we have the four-dimensional vector space V with Lorentz metric with the nondegenerate subspacex V_h, V_p, V_t and V_i.

Let A and B denote subsets of X and v, w vectors in X. We write

$$A + B = \{a + b \,|\, a \in A, b \in B\}, \tag{4.1}$$

$$v + A = A + v = \{a + v | a \in A\}, \tag{4.2}$$

$$v \cdot A = A \cdot v = \{a \cdot v | a \in A\}, \tag{4.3}$$

$$A \cdot B = \{a \cdot b \,|\, a \in A, b \in B\}, \tag{4.4}$$

$$[v,w] = \{sv + (1-s)w | 0 \leqslant s \leqslant 1\}, \tag{4.5}$$

$$A^\perp = \{v | v \in X, v \cdot A = \{0\}\}. \tag{4.6}$$

The set A is an orthonormal set in X if A is a subset of X and $|a\cdot a| = 1$ for each $a\in A$ and $a\cdot b = 0$ for $a,b\in A$, $a\neq b$. An orthonormal set is always linearly independent. A basis B for X is a linearly independent set that spans X. An orthonormal basis is a basis that is an orthonormal set. The following two lemmas are needed in the proof of Lemma 1.

Lemma 7: Each orthonormal set in X is contained in some orthonormal basis for X. In particular, there exist, an orthonormal basis for X.

Lemma 8: Let A be a subspace of X. Then A is nondegenerate if and only if $A + A^{\perp} = X$ and $A\cap A^{\perp} = \{0\}$.

The proofs of Lemmas 7 and 8 may be found in Ref. 7 or 8.

B. The periodic structure of the background

We now consider the vector space V_p with the metric tensor inherited from the Lorentz metric on V. We assume dim $V_p = n\geqslant 2$. The subset T_p of V_p (see Definition 1) has the following properties:

(a) T_p spans V_p.

(b) T_p is a subgroup of V_p if we regard V_p as a group under addition.

(c) T_p contains no accumulation point; now, regarding V_p as a topological space with the natural topology, it has as a finite-dimensional vector space (Ref. 7, p. 183).

It is evident that T_p satisfies (a) and (b). Properly (c) follows from physical considerations. It is not difficult to show that the properties (a), (b), and (c) imply that there exists a basis $B = \{v_1,...,v_n\}$ for V_p such that

$$T_p = \{s_1v_1 + \cdots + s_nv_n \,|s_1,...,s_n \text{ integers}\}. \tag{4.7}$$

We call a set satisfying (a), (b), and (c) a *lattice* and a basis $B = \{v_1,...,v_n\}$ such that (4.7) is a valid *basis* for the lattice T_p. The *unit cell* U of T_p with respect to B is

$$U = [0,v_1] + \cdots + [0,v_n]. \tag{4.8}$$

The *reciprocal lattice* $\hat{T}_p$ is defined as

$$\hat{T}_p = \{a\in V_p \,|a\cdot T_p \subset \{2\pi s|s \text{ integer}\}\}. \tag{4.9}$$

We give V_p an orientation (it is unimportant which) in order to have the *-operator defined[8] as

$$*: \quad V_p^{\wedge(n-1)} \longrightarrow V_p. \tag{4.10}$$

This is an isometry and, consequently, a bijection. The reciprocal lattice may be represented in terms of the basis $B = \{v_1,...,v_n\}$ for T_p as

$$\hat{T}_p = \{s_1\hat{v}_1 + \cdots + s_n\hat{v}_n \,|s_1,...,s_n \text{ integers}\}, \tag{4.11}$$

where

$$\hat{v}_j = 2\pi(\text{vol } U)^{-1} *(v_1\wedge\cdots\wedge v_{j-1}\wedge v_{j+1}\wedge\cdots\wedge v_n), \tag{4.12}$$

$$\text{vol } U = |\langle v_1\wedge\cdots\wedge v_n, v_1\wedge\cdots\wedge v_n\rangle|^{1/2} = |\det(v_i\cdot v_j)|^{1/2}. \tag{4.13}$$

It follows easily from basic properties of the *-operator[8] that

$$|v_i\cdot\hat{v}_j| = 2\pi\delta_{i,j} \tag{4.14}$$

and that $\hat{B} = \{\hat{v}_1,...,\hat{v}_n\}$ is a basis for V_p. The proof that $\hat{T}_p$ defined by (4.9) is a lattice and that (4.11) is true is now easy. The unit cell $\hat{U}$ for the lattice $\hat{T}_p$ with respect to the basis $\hat{B}$ is

$$\hat{U} = [0,\hat{v}_1] + \cdots + [0,\hat{v}_n]. \tag{4.15}$$

We have

$$\begin{aligned}&(\text{vol } U)\cdot(\text{vol } \hat{U})\\ &\quad = (\text{vol } U)|\langle \hat{v}_1\wedge\cdots\wedge\hat{v}_n, \hat{v}_1\wedge\cdots\wedge\hat{v}_n\rangle|^{1/2}\\ &\quad = (\text{vol } U)|\langle *\hat{v}_1, \hat{v}_2\wedge\cdots\wedge\hat{v}_n\rangle|\\ &\quad 2\pi|\langle v_2\wedge\cdots\wedge v_n, \hat{v}_2\wedge\cdots\wedge\hat{v}_n\rangle| = (2\pi)^n.\end{aligned} \tag{4.16}$$

If B and B' are two bases for T_p and U and U' corresponding unit cells, it may be shown that vol U = vol U'.

C. Transforms

Let G be a function defined on V_p and taking values in some vector space. Let G have the same periodic structure as the background plasma, i.e.,

$$G(x_p + a) = G(x_p) \quad \text{for } a\in T_p. \tag{4.17}$$

The G may be expressed as a Fourier series

$$G(x_p) = \sum_{a\in\hat{T}_p} H(a)\exp(ia\cdot x_p), \tag{4.18}$$

where

$$H(a) = (\text{vol } U)^{-1}\int_U G(x_p)\exp(-ia\cdot x_p)\,dx_p. \tag{4.19}$$

We state without proof that the Fourier coefficients $H(a)$ do not depend on the choice of unit cell U. Formula (4.19) may also be written in terms of the basis $B = \{v_1,...,v_n\}$ for T_p as

$$\begin{aligned}H(a) = \int_0^1 ds_1\cdots\int_0^1 ds_n\; &G(s_1v_1 + \cdots + s_nv_n)\\ &\times\exp[-i(s_1v_1 + \cdots + s_nv_n)\cdot a].\end{aligned} \tag{4.20}$$

We obtain (4.20) from (4.19) by means of the change of variables $g:[0,1]^{\times n}\to U$ defined $g(s_1,...,s_n) = s_1v_1 + \cdots + s_nv_n$. Then

$$\begin{aligned}dx_p &= \left|\left\langle\frac{\partial g}{\partial s_1}\wedge\cdots\wedge\frac{\partial g}{\partial s_n}, \frac{\partial g}{\partial s_1}\wedge\cdots\wedge\frac{\partial g}{\partial s_n}\right\rangle\right|^{1/2} ds_1\cdots ds_n\\ &= |\langle v_1\wedge\cdots\wedge v_n, v_1\wedge\cdots\wedge v_n\rangle|^{1/2} ds_1\cdots ds_n\\ &= (\text{vol } U)ds_1\cdots ds_n.\end{aligned} \tag{4.21}$$

We now define the Fourier transform and the $\tilde{}$ transform on functions from V_t to some vector space. A function ϕ defined on V may be regarded as a function on V_t depending on a parameter in V_i, we write $\phi(x) = \phi(x_t + x_i) = \phi(x_t;x_i)$. The Fourier transform and its inverse are defined

$$\hat{\phi}(\kappa;x_i) = \int_{V_t}\phi(x_t;x_i)e^{-i\kappa\cdot x_t}\,dx_t, \quad \kappa\in V_t, \tag{4.22}$$

$$\phi(x_t;x_i) = (2\pi)^{-m}\int_{V_t}\hat{\phi}(\kappa,x_i)e^{i\kappa\cdot x_t}\,d\kappa, \tag{4.23}$$

where $m = \dim V_t$. The $\tilde{}$ transform and its inverse are defined

$$\tilde{\phi}(\kappa,x_p;x_i) = \sum_{a\in\hat{T}_p}\hat{\phi}(\kappa + a;x_i)e^{ia\cdot x_p}, \quad \kappa\in V_t, \tag{4.24}$$

$$\phi(x_t;x_i) = (2\pi)^{-m}\int_{V_h + \hat{U}}\tilde{\phi}(\kappa,x_p;x_i)e^{i\kappa\cdot x_t}\,d\kappa, \tag{4.25}$$

By substituting (4.24) in (4.25), changing variables $\kappa + a \to \kappa$ and then using

$$\sum_{a\in\hat{T}_p} \int_{V_h + \hat{U} + a} = \int_{V_t}, \tag{4.26}$$

we see that (4.25) follows from (4.23). We will need some simple properties of the $\tilde{}$ transform:

(1) $\tilde{\phi}(\kappa,x_p + a;x_i) = \tilde{\phi}(\kappa,x_p;x_i)$ for $a\in T_p$. (4.27)

(2) $\tilde{\phi}(\kappa + a,x_p;x_i) = \tilde{\phi}(\kappa,x_p;x_i)e^{-ia\cdot x_p}$ for $a\in\hat{T}_p$. (4.28)

(3) Replacement of $V_h + \hat{U}$ by $V_h + \hat{U} + v$ with $v\in V_t$ does not change the value of the integral in (4.25).

(4) Let $\phi(x) = A(x_t)\psi(x_p;x_i)$, where A is a complex-valued function on V_t and $\psi\in P$; then

$$\tilde{\phi}(\kappa,x_p;x_i) = \tilde{A}(\kappa,x_p)\psi(x_p;x_i). \tag{4.29}$$

(5) Take A as in (4) above, $\kappa_1\in V_t$ and $B(x_t) = A(x_t)e^{i\kappa_1\cdot x_t}$. Then

$$\tilde{B}(\kappa,x_p) = \tilde{A}(\kappa - \kappa_1,x_p). \tag{4.30}$$

(6) $\tilde{\phi}(\kappa,x_p;x_i)^* = (\phi^*)^{\sim}(-\kappa^*,x_p;x_i)$. (4.31)

Here κ is an element in $V_t^+ = V_t + iV_t$.

(7) Let A, B, and C be functions A, $B: V_t \to V^+$ and C: $V_t \to V^+ \otimes V^+$, where $C(x_t) = A(x_t)\otimes B(x_t)$. Then

$$\tilde{C}(\kappa,x_p) = (2\pi)^{-m}\int_{V_h+\hat{U}} \tilde{A}(\kappa',x_p)\otimes\tilde{B}(\kappa - \kappa',x_p)\,d\kappa'. \tag{4.32}$$

(8) With A, B, and C as in (7) above,

$$\tilde{C}(\kappa,x_p) = (2\pi)^{-m}\sum_{a\in\hat{T}_p}\int_{V_h+\hat{U}} d\kappa_1 \times\int_{V_h+\hat{U}}\delta(\kappa - \kappa_1 - \kappa_2 - a)\tilde{A}(\kappa_1,x_p) \otimes\tilde{B}(\kappa_2,x_p)e^{-ia\cdot x_p}\,d\kappa_2. \tag{4.33}$$

The properties (1) and (2) follow directly from the definitions.

Proof of (3): The integrands omitted below are the same as in (4.25). It follows from (4.28) that

$$\int_A = \int_{A+a} \quad \text{for } A\subset V_t,\ a\in\hat{T}^p. \tag{4.34}$$

With the notations in Sec. 4 B is it thus sufficient to prove property (3) with $v\in\hat{U} = [0,\hat{v}_1] + \cdots + [0,\hat{v}_n]$. Write $v = s_1\hat{v}_1 + \cdots + s_n\hat{v}_n$ with $0\leqslant s_j\leqslant 1$. We have from (4.34)

$$\int_{V_h + [0,s_1\hat{v}_1] + \cdots + [0,s_n\hat{v}_n]} = \int_{V_h + [\hat{v}_1,(1+s_1)\hat{v}_1] + \cdots + [\hat{v}_n,(1+s_n)\hat{v}_n]}. \tag{4.35}$$

The equality

$$\int_{V_h+\hat{U}+v} = \int_{V_h+\hat{U}} \tag{4.36}$$

follows now from (4.35) and

$$(V_h + \hat{U} + v)\setminus(V_h + \hat{U}) = V_h + [\hat{v}_1,(1+s_1)\hat{v}_1] + \cdots + [\hat{v}_n,(1+s_n)\hat{v}_n], \tag{4.37}$$

$$(V_h + \hat{U})\setminus(V_h + \hat{U} + v) = V_h + [0,s_1\hat{v}_1] + \cdots + [0,s_n\hat{v}_n]. \tag{4.38}$$

Proof of (4): We expand ψ in a Fourier series:

$$\psi(x_p;x_i) = \sum_{a\in\hat{T}_p}\psi_a(x_i)e^{ia\cdot x_p}, \tag{4.39}$$

$$\begin{aligned}\tilde{\phi}(\kappa,x_p;x_i) &= \sum_{b\in\hat{T}_p}\hat{\phi}(\kappa + b;x_i)e^{ib\cdot x_p}\\ &= \sum_{a,b\in\hat{T}_p}\hat{A}(\kappa - a + b)\psi_a(x_i)e^{ib\cdot x_p}\\ &= \sum_{a\in\hat{T}_p}\tilde{A}(\kappa,x_p)\psi_a(x_i)e^{ia\cdot x_p}\\ &= \tilde{A}(\kappa,x_p)\psi(x_p;x_i).\end{aligned} \tag{4.40}$$

The properties (5) and (6) follows directly from the definitions.

Proof of property (7):

$$\begin{aligned}&\int_{V_h+\hat{U}}\tilde{A}(\kappa',x_p)\otimes\tilde{B}(\kappa - \kappa',x_p)\,d\kappa'\\ &= \sum_{a,b\in\hat{T}_p}\int_{V_h+\hat{U}}\hat{A}(\kappa' + a)\otimes\hat{B}(\kappa - \kappa' + b)e^{i(a+b)\cdot x_p}\,d\kappa'\\ &= \sum_{a,b}\int_{V_h+\hat{U}}\hat{A}(\kappa' + a - b)\otimes\hat{B}(\kappa - \kappa' + b)e^{ia\cdot x_p}\,d\kappa'\\ &= \sum_{a,b}\int_{V_h+\hat{U}+(a-b)}\hat{A}(\kappa')\otimes\hat{B}(\kappa - \kappa' + a)e^{ia\cdot x_p}\,d\kappa'\\ &= \sum_a\int_{V_t}\hat{A}(\kappa')\otimes\hat{B}(\kappa + a - \kappa')e^{ia\cdot x}\,d\kappa'\\ &= (2\pi)^m\sum_a\hat{C}(\kappa + a)e^{ia\cdot x_p} = (2\pi)^m\tilde{C}(\kappa,x_p).\end{aligned} \tag{4.41}$$

Proof of (8): By first performing the κ_2 integration and the summation, we obtain from (4.33)

$$\tilde{C}(\kappa,x_p) = (2\pi)^{-m}\int_{V_h+\hat{U}} d\kappa_1\,\tilde{A}(\kappa_1,x_p) \otimes\tilde{B}(\kappa - \kappa_1 - a,x_p)e^{-ia\cdot x_p}, \tag{4.42}$$

where a is a function of $\kappa - \kappa_1$ determined by the condition $\kappa - \kappa_1 - a\in V_h + \hat{U}$. From (4.42) and (4.28) we obtain (4.32).

Definition 6: For ψ, ψ_1, $\psi_2\in P$ and κ, κ_1, $\kappa_2\in V_t$ define

$$\delta J_\kappa[\psi](x_t) = \delta J^{(1)}[\phi](x)\exp(-i\kappa\cdot x), \tag{4.43}$$

$$\delta J_{\kappa_-}[\psi](x_t) = \delta J^{(1-)}[\phi]\exp(-i\kappa\cdot x), \tag{4.44}$$

$$\delta J_{\kappa_1,\kappa_2}[\psi_1,\psi_2](x_t) = \delta J^{(2)}[\phi_1,\phi_2](x)\exp[-i(\kappa_1 + \kappa_2)\cdot x], \tag{4.45}$$

where

$$\phi(x) = \psi(x_t)\exp(i\kappa\cdot x),\quad \phi_j(x) = \psi_j(x_t)\exp(i\kappa_j\cdot x). \tag{4.46}$$

Lemma 3, which is proved in subsection D below, is equivalent to

$$\delta J_\kappa: P\to P,\quad \delta J_{\kappa_-}: P\to P,\quad \delta J_{\kappa_1,\kappa_2}: P\times P\to P, \tag{4.47}$$

i.e., the right-hand sides of (4.43)–(4.45) are independent of x_i. We also need the properties

(9) $(\delta J^{(1)}[\phi])^{\sim}(\kappa,x_p;x_i) = \delta J_\kappa[\tilde{\phi}_\kappa](x_p;x_i)$, (4.48)

$(\delta J^{(1-)}[\phi])^{\sim}(\kappa,x_p;x_i) = \delta J_{\kappa-}[\tilde{\phi}_\kappa](x_p;x_i)$, (4.49)

where we have used the notation

$\tilde{\phi}_\kappa(x_p;x_i) = \tilde{\phi}(\kappa,x_p;x_i)$. (4.50)

(10) $(\delta J^{(2)}[\phi,\phi])^{\sim}(\kappa,x_p;x_i)$

$$= (2\pi)^{-m}\int_{V_h+\hat{U}} \delta J_{\kappa',\kappa-\kappa'}[\tilde{\phi}_{\kappa'},\tilde{\phi}_{\kappa-\kappa'}]\,d\kappa'. \quad (4.51)$$

(11) Let $a,b\in\hat{T}_p$, then

$$\delta J_{\kappa_1+a,\kappa_2+b}[\tilde{\phi}_{\kappa_1+a},\tilde{\phi}_{\kappa_2+b}](x_p;x_i) = \delta J_{\kappa_1,\kappa_2}[\tilde{\phi}_{\kappa_1},\tilde{\phi}_{\kappa_2}](x_p;x_i)e^{-i(a+b)\cdot x_p}. \quad (4.52)$$

(12) Let $\kappa_1,\kappa_2\in V_t^+$; then

$$\langle\psi_3^*,\delta J_{\kappa_1,\kappa_2}[\psi_1,\psi_2]\rangle^* = \langle\psi_3,\delta J_{-\kappa_1^*,-\kappa_2^*}[\psi_1^*,\psi_2^*]\rangle. \quad (4.53)$$

Proof of (9): Take the inverse $^{\sim}$ transform on (4.48) and (4.49).

Proof of (10): Denote $\tilde{\phi}_\kappa(x_p;x_i)\exp(i\kappa\cdot x) = \phi_\kappa(x)$. Then the inverse $^{\sim}$ transform on (4.51) yields

$$\begin{aligned}\delta J^{(2)}[\phi,\phi](x) &= (2\pi)^{-2m}\int_{V_h+\hat{U}} d\kappa e^{i\kappa\cdot x}\\ &\times\int_{V_h+\hat{U}}\delta J_{\kappa',\kappa-\kappa'}[\tilde{\phi}_{\kappa'},\tilde{\phi}_{\kappa-\kappa'}]\,d\kappa'\\ &= (2\pi)^{-2m}\int_{V_h+\hat{U}} d\kappa\int_{V_h+\hat{U}}\delta J^{(2)}[\phi_\kappa,\phi_{\kappa-\kappa'}]\,d\kappa',\end{aligned} \quad (4.54)$$

We have

$$\int_{V_h+\hat{U}}\phi_{\kappa-\kappa'}\,d\kappa = \int_{V_h+\hat{U}-\kappa'}\phi_\kappa\,d\kappa = \int_{V_h+\hat{U}}\phi_\kappa\,d\kappa, \quad (4.55)$$

where the last equality follows from property (3) above. Now (4.54) follows easily by means of (4.55) and (4.25).

Proof of (11): This is a simple consequence of (4.28).

Proof of (12): This follows easily from $(\delta J^{(2)}[\phi_1,\phi_2])^* = \delta J^{(2)}[\phi_1^*,\phi_2^*]$ and (4.45).

D. Proofs of the lemmas and the results in Secs. 2 and 3

Lemma 1 follows easily from Lemma 8 if V_t is nondegenerate. By assumption (a) in Sec. 2 and Lemma 7 we may choose a Lorentz frame $L = (e_0,e_1,e_2,e_3)$ such that $e_0\in V_t$. If $a\in V_t$ has the property $a\cdot V_t = \{0\}$, it follows that $a\cdot e_0 = a\cdot a = 0$ or in coordinates $a = a^i e_i$ that $a^0 = 0$ and $(a^1)^2+(a^2)^2+(a^3)^2 = 0$ and thus $a = 0$. This proves that V_t is nondegenerte. Lemma 2 is considered in Sec. 4 B. Lemma 3 is equivalent to (4.47). These are proven exactly like the corresponding relations in paper I [see remark 5(I)] with the difference that now the translation v may be a 4-vector of periodicity. Lemma 4 follows from [see (3.9) (I)]

$$\langle\phi_0,\delta J^{(1)}[\phi_1]\rangle = \langle\delta J^{(1-)}[\phi_0],\phi_1\rangle,\quad \phi_0,\phi_1\in P(\kappa). \quad (4.56)$$

Lemma 5 follows from (4.56) and

$$\langle\phi_0,\nabla_E\cdot(\nabla_E\wedge\phi_1)\rangle = \langle\nabla_E\cdot(\nabla_E\wedge\phi_0),\phi_1\rangle,\quad \phi_0,\phi_1\in P(\kappa). \quad (4.57)$$

Here (4.57) is derived by means of the equality

$$\begin{aligned}&\phi_0^*\cdot(\nabla_E\cdot(\nabla_E\wedge\phi_1)) + \nabla_i(\phi_0^*\cdot\nabla_E\wedge\phi_1)\\ &\quad = \phi_1\cdot(\nabla_E\cdot(\nabla_E\wedge\phi_0^*)) + \nabla_i\cdot(\phi_1\cdot\nabla_E\wedge\phi_0^*),\quad \phi_0,\phi_1\in P(\kappa).\end{aligned} \quad (4.58)$$

Proof of Lemma 6: We have $\phi(x) = \psi(x_p;x_i)e^{i\kappa\cdot x}$, where $\psi\in P$. Expand ψ in Fourier series [see (4.18)]

$$\psi(x_p;x_i) = \sum_{a\in\hat{T}_p}\psi_a(x_i)\exp(ia\cdot x_p). \quad (4.59)$$

Define χ by $\chi(x) = \xi(x_p;x_i)e^{i\kappa\cdot x}$, where

$$\xi(x_p;x_i) = \sum_{a\in\hat{T}_p}\xi_a(x_i)\exp(ia\cdot x_p), \quad (4.60)$$

$$\xi_a(x_i) = [e_0\cdot(\kappa+a)]^{-1}e_0\cdot(\kappa+a)\wedge\psi_a(x_i). \quad (4.61)$$

It is easy to see that (4.61) is equivalent to

$$e_0\cdot\xi_a = 0 \quad\text{and}\quad (\kappa+a)\wedge\xi_a = (\kappa+a)\wedge\psi_a \quad a\in\hat{T}_p. \quad (4.62)$$

The existence and uniqueness of χ now follows.

Proof of result 1: The gauge invariance of the coefficients in (2.9) is proved in essentially the same way as in II. Also the reality of the coefficients on the left-hand side of (2.9) follows as in II. We may thus choose $\phi_j\in P_L(\kappa_j)$ for some arbitrary Lorentz frame L. We use the symbol $O(\epsilon^n)$ to denote any quantity of order ϵ^n or smaller, where ϵ is a small parameter. We assume that

$$A_j = O(\epsilon),\quad \phi_j = O(1),\quad \Delta\kappa = O(\epsilon),$$
$$H = O(1),\quad h = O(\epsilon),\quad j = 1,2,3. \quad (4.63)$$

We define H_κ and h_κ by [cf. (4.43)–(4.45)]

$$H_\kappa[\psi](x_t) = H[\phi](x)\exp(-i\kappa\cdot x), \quad (4.64)$$

$$h_\kappa[\psi](x_t) = h[\phi](x)\exp(-i\kappa\cdot x), \quad (4.65)$$

where $\phi\in P(\kappa)$. We assume that ψ, ψ_j are related to ϕ, ϕ_j by (4.46). The $^{\sim}$ transform on (2.8) yields, with the use of (4.29) and (4.30),

$$\tilde{\phi}(x_p;x_i) = \sum_{j=1}^{3}[\tilde{A}_j(\kappa-\kappa_j,x_p)\psi_j(x_p;x_i) + (A_j^*)^{\sim}(\kappa+\kappa_j,x_p)\psi_j^*(x_p,x_i)], \quad (4.66)$$

where $\phi(x)$ denotes the expression (2.8). Now, due to the slow variation of A_j, $\tilde{A}_j(\kappa-\kappa_j)\neq 0$ only for $\kappa-\kappa_j-a = O(\epsilon)$ with $a\in\hat{T}_p$. Let us now consider κ in the interval $\kappa-\kappa_3 = O(\epsilon)$. Then

$$\langle\psi_3,(H_\kappa+ih_\kappa)\tilde{\phi}_\kappa\rangle = \langle\psi_3,[(\kappa-\kappa_3)\cdot(\partial H_\kappa/\partial\kappa)_{\kappa=\kappa_3} + ih_{\kappa_3} + O(\epsilon^2)]\tilde{\phi}_\kappa\rangle, \quad (4.67)$$

where we have used $\langle\psi_3,H_{\kappa_3}\cdots\rangle = 0$ {see (4.64), (4.56), (4.57), and $H[\phi_3] = 0$}. We obtain from (4.24) and $\hat{A}(\kappa)\neq 0$ only for $\kappa = O(\epsilon)$ that

$$\tilde{A}_3(\kappa-\kappa_3,x_p) = \hat{A}_3(\kappa-\kappa_3)\quad\text{for } \kappa-\kappa_3 = O(\epsilon). \quad (4.68)$$

Thus we have

$$\tilde{\phi}_\kappa(x_p;x_i) = \hat{A}_3(\kappa-\kappa_3)\psi_3(x_p,x_i) \quad \text{for } \kappa-\kappa_3 = O(\epsilon). \tag{4.69}$$

Substituting (4.69) in (4.67) and integrating over the κ interval $\kappa = \kappa_3 + O(\epsilon)$ in V_t after multiplying with a factor yield

$$\begin{aligned}
&(2\pi)^{-m}\int_{\kappa=\kappa_3+O(\epsilon)} \langle\psi_3,(H_\kappa + ih_\kappa)\tilde{\phi}_\kappa\rangle \exp[i(\kappa-\kappa_3)\cdot x_t]\,dx_t \\
&= (2\pi)^{-m}\int_{\kappa=\kappa_3+O(\epsilon)} \{(\partial\langle\psi_3,H_\kappa\psi_3\rangle/\partial\kappa)_{\kappa=\kappa_3}\cdot(-i)(\partial\exp[i(\kappa-\kappa_3)\cdot x_t]/\partial x_t) + i\langle\psi_3,h_{\kappa_3}\psi_3\rangle \\
&\quad\times\exp[i(\kappa-\kappa_3)\cdot x_t] + O(\epsilon^2)\exp[i(\kappa-\kappa_3)\cdot x_t]\}\hat{A}_3(\kappa-\kappa_3)\,d\kappa \\
&= -i(\partial\langle\psi_3,H_\kappa\psi_3\rangle/\partial\kappa)_{\kappa=\kappa_3}\cdot\nabla_t A_3(x_t) + i\langle\psi_3,h_{\kappa_3}\psi_3\rangle A_3(x_t) + O(\epsilon^3).
\end{aligned} \tag{4.70}$$

For obtaining the $O(\epsilon)^3$ term it is used that the $O(\epsilon)^2$ operator varies slowly as a function of κ for $\kappa\approx\kappa_3$ while $A_3(\kappa-\kappa_3)$ varies fast. From (4.48), (4.49), and (4.51) we obtain the $\tilde{}$ transform of the electromagnetic wave equation (2.3) as

$$\begin{aligned}
&(H_\kappa + ih_\kappa)\tilde{\phi}_\kappa \\
&= -\left(\frac{\mu_0}{\epsilon_0}\right)^{1/2}(2\pi)^{-m}\int_{V_h+\hat{U}} \delta J_{\kappa',\kappa-\kappa'}[\tilde{\phi}_{\kappa'},\tilde{\phi}_{\kappa-\kappa'}]\,d\kappa',
\end{aligned} \tag{4.71}$$

and from (4.71) we get

$$\begin{aligned}
&(2\pi)^{-m}\int_{\kappa=\kappa_3+O(\epsilon)} \langle\psi_3,(H_\kappa + ih_\kappa)\tilde{\phi}_\kappa\rangle \exp[i(\kappa-\kappa_3)\cdot x_t]\,d\kappa \\
&= -\left(\frac{\mu_0}{\epsilon_0}\right)^{1/2}(2\pi)^{-m}\int_{\kappa=\kappa_3+O(\epsilon)} d\kappa \\
&\quad\times\int_{V_h+\hat{U}} \langle\psi_3,\delta J^{(2)}_{\kappa',\kappa-\kappa'}[\tilde{\phi}_{\kappa'},\tilde{\phi}_{\kappa-\kappa'}]\rangle \\
&\quad\times\exp[i(\kappa-\kappa_3)\cdot x_t]\,d\kappa'.
\end{aligned} \tag{4.72}$$

Since $\tilde{\phi}_\kappa \neq 0$ only for $\kappa = \pm\kappa_j + a + O(\epsilon)$ for $j = 1,2,3$ and $a\in\hat{T}_p$, it follows that the κ' integral in (4.72) gets contributions only for κ' in the two intervals $\kappa' = -\kappa_1 + a_1 + O(\epsilon)$ and $\kappa' = -\kappa_2 + a_2 + O(\epsilon)$, where a_1 and a_2 are determined by $a_1,a_2\in\hat{T}_p$ and $-\kappa_1 + a_1$, $-\kappa_2 + a_2\in V_h + \hat{U}$. The contributions from the two κ' intervals are equal, and the right-hand side of (4.72) may be rewritten

$$-2\left(\frac{\mu_0}{\epsilon_0}\right)^{1/2}(2\pi)^{-2m}\int_{\kappa=\kappa_3+O(\epsilon)} d\kappa\int_{\kappa'=-\kappa_1+a_1+O(\epsilon)} \cdots d\kappa', \tag{4.73}$$

with the same integrand in the κ' integral as in (4.72). The change of variables $\kappa'\to a_1 + \kappa'$ in (4.73) and the use of (4.52) show that the right-hand side of (4.73) (and thus of (4.72)] may be rewritten as

$$-2\left(\frac{\mu_0}{\epsilon_0}\right)^{1/2}(2\pi)^{-2m}\int_{\kappa=\kappa_3+O(\epsilon)} d\kappa\int_{\kappa'=-\kappa_1+O(\epsilon)} \cdots d\kappa', \tag{4.74}$$

where the integrand in the κ' integral is the same as in (4.72). We have

$$\tilde{\phi}_{\kappa'} = \widehat{(A^*_1)}(\kappa'+\kappa_1)\psi^*_1, \quad \tilde{\phi}_{\kappa-\kappa'} = \widehat{(A^*_2)}(\kappa-\kappa'+\kappa_2)\psi^*_2$$
$$\text{for } \kappa = \kappa_3 + O(\epsilon) \text{ and } \kappa' = -\kappa_1 + O(\epsilon) \tag{4.75}$$

[cf. (4.69)], and, by substituting (4.75) in (4.74) and observing that only $(A^*_1)(\kappa'+\kappa_1)$ and $(A^*_2)(\kappa-\kappa'+\kappa_2)$ vary within the narrow intervals over which we integrate so that $\langle\psi_3,\delta J^{(2)}_{\kappa',\kappa-\kappa'}[\psi_1,\psi_2]\rangle$ may be taken outside the integrals, we finally obtain that (4.74) is equal to

$$-2(\mu_0/\epsilon_0)^{1/2}\langle\psi_3,\delta J_{-\kappa_1,-\kappa_2}[\psi^*_1,\psi^*_2]\rangle A^*_1(x_t)A^*_2(x_t). \tag{4.76}$$

Now combining (4.70) and (4.72) and replacing the right-hand side of (4.72) with (4.76) yields

$$\begin{aligned}
&(\partial\langle\psi_3,H_\kappa\psi_2\rangle/\partial\kappa)_{\kappa=\kappa_3}\cdot\nabla_t A^*_3 - \langle\psi_3,h_{\kappa_3}\psi_3\rangle A^*_3 \\
&= 2i(\mu_0/\epsilon_0)^{1/2}\langle\psi^*_3,\delta J_{\kappa_1,\kappa_2}[\psi_1,\psi_2]\rangle A_1 A_2.
\end{aligned} \tag{4.77}$$

Here also (4.53) has been used. It is easy to see that (4.77) is equivalent to (2.9).

The statement concerning the group 4-velocity in Result 1 is proved in much the same way as the corresponding statement in II. The proof involves relations corresponding to and formally identical to (4.16)–(4.20) (II).

Proof of Result 2: Choose $\phi\in L^0(E,V^+)$ such that $\tilde{\phi}(\kappa_0,x_p;x_i) = \phi_0(x)\exp(-i\kappa_0\cdot x)$ and so that (4.78) converge. Then we have the equality

$$\int_V \phi\cdot\delta J^{(2)}[\phi_1,\phi_2]\,dx = \langle\phi_0,\delta J^{(2)}[\phi_1,\phi_2]\rangle. \tag{4.78}$$

We derive (4.78) by expanding in a Fourier series

$$\begin{aligned}
&\delta J^{(2)}[\phi_1,\phi_2](x)\exp[-i(\kappa_1+\kappa_2)\cdot x] \\
&= \sum_{a\in\hat{T}_p} H(a;x_i)\exp(ia\cdot x_p),
\end{aligned} \tag{4.79}$$

in accordance with (4.18), observing (see Lemma 3) that the left-hand side of (4.79) has the periodic structure (4.17). By substituting (4.79) in (4.78) and applying (4.22) on the left-hand side and (4.24) on the right-hand side we obtain both sides equal to $\int_V dx_i \Sigma_{a\in\hat{T}_p}\hat{\phi}(\kappa_0 - a;x_i)\cdot H(a;x_i)$ and the equality (4.78) is thus obtained.

An expression for the left-hand side in (4.78) for a Vlasov plasma is obtained from (2.2) (I). The equality between the right-hand side of (2.2) (I) with ϕ_0 replaced by ϕ and the right-hand side of (3.4) is now proved in essentially the same way as (4.78). We then need results corresponding to Lemma 3, (4.48), and (4.49) for the operators[10] $\delta x^{(1)}$, $\delta u^{(1)}$, $\delta x^{(1-)}$, and $\delta u^{(1-)}$ in place of $\delta J^{(1)}$ and $\delta J^{(1-)}$.

[1] J. Larsson, J. Math. Phys. **23**, 176 (1982).
[2] J. Larsson, J. Math. Phys. **23**, 183 (1982).

[3]J. Larsson, J. Math. Phys. **24**, 2231 (1983).
[4]J. Larsson, J. Math. Phys. **20**, 1321 (1979).
[5]We consider only Lorentz frames with $e_0,\ldots,e_{m-1}\in V_t$ and $e_m,\ldots,e_4\in V_i$.
[6]The spaces $L^0(E,V)$ and $L_0(E,V)$ are defined in Ref. 4. The space V^+ stands for the complexification of V, formally $V^+ = V + iV$.
[7]C. T. J. Dodson and T. Poston, *Tensor Geometry* (Pitman, London, 1977).
[8]H. Flanders, *Differential Forms with Applications to the Physical Sciences* (Academic, New York, 1963).
[9]C. W. Misner, K. S. Thorn, and J. A. Wheeler, *Gravitation* (Freeman, San Fransisco, 1973).
[10]The operators $\delta x^{(1)}$ and $\delta u^{(1)}$ are defined in Definition 4 in Ref. 4, and $\delta x^{(1-)}$ and $\delta u^{(1-)}$ are defined in Remark 7 (I).

Current responses of first and second order in a collisionless plasma. IV. Applications

Jonas Larsson
Department of Plasma Physics, Umeå University, S-901 87 Umeå, Sweden

(Received 8 April 1982; accepted for publication 14 January 1983)

The purpose of this paper is to illustrate with a few concrete examples how the general formulas given in the preceding papers I–III may be used. We consider three different background plasmas: a homogeneous static magnetized plasma, a homogeneous magnetized plasma in an oscillating electric field, and a semi-infinite homogeneous static magnetized plasma. The coupled mode equations governing resonant three-wave interaction are given in each case.

PACS numbers: 52.40.Db, 52.25. – b, 03.65. – w

1. INTRODUCTION

A purpose of the companion Papers I–III[1–3] is, for the particular cases of first- and second-order responses, to exploit and to simplify the exploitation of the general formulas for the current response operators found in Ref. 4. A second-order process of basic importance is a resonant three-wave interaction, and it is considered for inhomogeneous and/or periodic plasmas in II and III. The formulas in I–III may be very useful tools in the study of first- and second-order processes in general and the resonant three-wave interaction in particular.

Considering inhomogeneous and/or time-dependent background plasmas, the linear problem of finding the normal modes becomes in general a difficult task requiring numerical methods at an early stage of the investigation. In most cases we have to manage without dispersion relations in closed forms which are so essential and convenient for the study of homogeneous-static plasmas. We wish, however, to illustrate the general formulas in I–III without the use of computers, and, therefore, we choose to study a couple of the few known particular cases of time-dependent or inhomogeneous Vlasov plasmas, where it is indeed possible to derive analytic explicit expressions for the dispersion relations. We consider, besides the homogeneous-static case (Sec. 2), a homogeneous magnetized plasma in an electric field $\mathbf{E}_0 \sin \Omega t$ (Sec. 3) and a semi-infinite homogeneous magnetized plasma (Sec. 4). We consider in Secs. 3 and 4 potential waves since then the dispersion relations have nice forms and are also much studied.[5–11] We derive in Secs. 3 and 4 the dispersion relations by means of explicit expressions for the eigenfunctions and the corresponding eigenvalues of the charge density response operators. This method makes these derivations transparent and is also very convenient when we apply Corollary 2 (III) and obtain the coupled mode equations in algebraically nice forms. Our results for the coupled mode equations are given by (2.13)–(2.17) for a homogeneous and static plasma and for the two other systems we summarize at the end of Secs. 3 and 4.

The results in Secs. 2 and 4 are merely an illustration of some formulas in I–III and a check of the agreement with previous results. In Sec. 3 we generalize Ref. 12 so that we now treat a *magnetized* plasma where $\mathbf{k}_1$, $\mathbf{k}_2$, $\mathbf{E}_0 \sin \Omega t$ and $\mathbf{B}_0$ have *arbitrary directions*.

This paper could easily be extended in some directions. The main problem, and the reason we do not include these generalizations, is that we get more complicated and less studied dispersion relations. Before considering nonlinear theory, one should study the linear theory in some detail. It must be stressed, however, that there are no qualitatively new problems involved. The assumption of potentiality in Secs. 3 and 4 may be removed. We can still use essentially the same technique, but we use the relations (B1) and (B2) or (C1) and (C2) instead of (3.8) and (3.9) or (4.4) and (4.5), respectively. Another interesting unperturbed system for which we easily find the eigenfunctions for the charge density response operators is the combined Sec. 3–Sec. 4 system obtained if we take the unperturbed semi-infinite plasma as in Sec. 4 and apply an electric field $\mathbf{E}_0 \sin \Omega t$ parallel to the surface. Denoting this unperturbed system by subscripts OS (since the systems in Secs. 3 and 4 are denoted with subscripts O and S, respectively), we have the relations

$$(\delta \rho^{(1)}_{OS}[U]; z \geqslant 0) = R^{-1} \circ \delta \rho^{(1)}_{O} \circ R[U; z \geqslant 0], \tag{1.1}$$

$$(\delta \rho^{(2)}_{OS}[U,U]; z \geqslant 0) = R^{-1} \circ \delta \rho^{(2)}_{O}[R[U; z \geqslant 0], R[U; z \geqslant 0]] \tag{1.2}$$

corresponding to (3.8)–(3.9) or (4.4)–(4.5). It is not difficult to proceed from here in a similar way as in Secs. 3 and 4.

There are many aspects on the formulas in I–III which it would be desirable to illustrate; we have, for example, not considered relativistic phenomena at all. We have, however, considered both time-dependent and space-dependent unperturbed states and also the resonant interaction both between bulk modes and between surface modes. We have used the convenient technique of eigenfunction expansion both in the linear and nonlinear theory and treated the oscillating and the semibounded plasma in a unified way. In Appendix A we have demonstrated how to take the unrelativistic limit of the formulas in I–III. In the Appendices B and C the general current response operator formulas for the Vlasov–Maxwell model of a plasma are used in the derivation of the relations (B1), (B2), (C1), and (C2).

2. THREE-WAVE INTERACTION IN A HOMOGENEOUS PLASMA

It is instructive to check that Corollary 2 (III) yields the expected formulas for the coupled mode equations when ap-

 0022-2488/83/082231-09$02.50

plied to a homogeneous stationary plasma.[13] Let us consider the electric field

$$\sum_{j=1}^{3} A_j(t, \mathbf{r})\hat{\mathbf{E}}_j \exp(-i\omega_j t + i\mathbf{k}_j\cdot\mathbf{r}) + \text{c.c.}, \tag{2.1}$$

of three resonantly interacting normal modes, i.e.,

$$\omega_1 + \omega_2 + \omega_3 = 0, \quad \mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 = 0, \tag{2.2}$$

$$\mathsf{D}'(\omega_j, \mathbf{k}_j)\cdot\hat{\mathbf{E}}_j = 0, \quad j = 1, 2, 3. \tag{2.3}$$

Here D' denotes the Hermitian part of the dispersion tensor $\mathsf{D} = \mathsf{D}' + i\mathsf{D}''$. The anti-Hermitian part $i\mathsf{D}''$ vanishes for a lossless medium, and for the modes we consider it is assumed that $i\mathsf{D}''$ may be neglected to first order. The polarization vectors $\hat{\mathbf{E}}_j$ of the electric field may be normalized so that $\hat{\mathbf{E}}_j^*\cdot\hat{\mathbf{E}}_j = 1$. In terms of the linear conductivity $\boldsymbol{\sigma}_\kappa^{(1)}$ [$\kappa = (\omega/c)e_0 + \mathbf{k}$, but here we use κ just as a shorthand notation for $(\omega, \mathbf{k})$[14] we express the dispersion tensor as

$$\mathsf{D}(\omega, \mathbf{k})\cdot\hat{\mathbf{E}} = \hat{\mathbf{E}} + c^2\omega^{-2}\mathbf{k}\times(\mathbf{k}\times\hat{\mathbf{E}}) + i(\epsilon_0\omega)^{-1}\boldsymbol{\sigma}_\kappa^{(1)}(\hat{\mathbf{E}}). \tag{2.4}$$

Take the 4-potentials ϕ_j in Corollary 2 (III) as

$$\phi_j(t, \mathbf{r}) = -ic\omega_j^{-1}\hat{\mathbf{E}}_j \exp(-i\omega_j t + i\mathbf{k}_j\cdot\mathbf{r}), \tag{2.5}$$

and thus

$$\chi_{j\kappa}(t, \mathbf{r}) = -ic\omega_j^{-1}\hat{\mathbf{E}}_j \exp(-i\omega t + i\mathbf{k}\cdot\mathbf{r}). \tag{2.6}$$

Then ($\circ$ is composition of functions or operators)

$$\Pi_L\circ\delta J^{(1)}[\chi_{j\kappa}] = (\omega/\omega_j)\boldsymbol{\sigma}_\kappa^{(1)}[\hat{\mathbf{E}}_j] \exp(-i\omega t + i\mathbf{k}\cdot\mathbf{r}), \tag{2.7}$$

$$\Pi_L\circ\delta J^{(1-)}[\chi_{j\kappa}] = (\omega/\omega_j)\boldsymbol{\sigma}_\kappa^{(1-)}[\hat{\mathbf{E}}_j] \exp(-i\omega t + i\mathbf{k}\cdot\mathbf{r}), \tag{2.8}$$

where $i\boldsymbol{\sigma}_\kappa^{(1-)}$ is the Hermitian conjugate of $i\boldsymbol{\sigma}_\kappa^{(1)}$, i.e.,

$$\mathbf{a}^*\cdot i\boldsymbol{\sigma}^{(1)}(\mathbf{b}) = (i\boldsymbol{\sigma}^{(1-)}(\mathbf{a}))^*\cdot\mathbf{b}, \tag{2.9}$$

We obtain from (2.4)–(2.5) (III), (2.4), (2.7), and (2.8) that

$$\Pi_L\circ H[\chi_{3\kappa}](t, \mathbf{r}) = -i\omega^2\omega_3^{-1}c^{-1}\mathsf{D}'(\omega, \mathbf{k})\cdot\hat{\mathbf{E}}_3 \times \exp(-i\omega t + i\mathbf{k}\cdot\mathbf{r}), \tag{2.10}$$

$$\Pi_L\circ h[\phi_3](t, \mathbf{r}) = -i\omega_3 c^{-1}\mathsf{D}''(\omega_3, \mathbf{k}_3)\cdot\hat{\mathbf{E}}_3 \times \exp(-i\omega_3 t + i\mathbf{k}_3\cdot\mathbf{r}). \tag{2.11}$$

From (2.12) (III) we now obtain

$$\left(\frac{\partial}{\partial t} + \mathbf{v}_{g3}\cdot\frac{\partial}{\partial \mathbf{r}} + \nu_3\right)A_3^* = \frac{i\omega_3}{W_3}V(3,1,2)A_1A_2, \tag{2.12}$$

which together with the two equations obtained by even permutations of (1,2,3) are the coupled mode equations. Here (cf. Ref. 13)

$$W_j = \epsilon_0\omega_j(\partial(\hat{\mathbf{E}}_j^*\cdot\mathsf{D}'(\omega, \mathbf{k}_j)\cdot\hat{\mathbf{E}}_j)/\partial\omega)_{\omega=\omega_j}, \tag{2.13}$$

$$\mathbf{v}_{gj} = -\epsilon_0\omega_j W_j^{-1}(\partial(\hat{\mathbf{E}}_j^*\cdot\mathsf{D}'(\omega_j, \mathbf{k})\cdot\hat{\mathbf{E}}_j)/\partial\mathbf{k})_{\mathbf{k}=\mathbf{k}_j}, \tag{2.14}$$

$$\nu_j = \epsilon_0\omega_j W_j^{-1}\hat{\mathbf{E}}_j^*\cdot\mathsf{D}''(\omega_j, \mathbf{k}_j)\cdot\hat{\mathbf{E}}_j, \tag{2.15}$$

$$V(3,1,2) = (2i/\omega_3)\hat{\mathbf{E}}_3^*\cdot\boldsymbol{\sigma}_{\kappa_1,\kappa_2}[\hat{\mathbf{E}}_1, \hat{\mathbf{E}}_2]. \tag{2.16}$$

For electrostatic waves we may write (2.13)–(2.15) as

$$W_j = \epsilon_0\omega_j\left(\frac{\partial\epsilon_r}{\partial\omega}\right)_j, \quad \mathbf{v}_{gj} = -\left(\frac{\partial\epsilon_r/\partial\omega}{\partial\epsilon_r/\partial\mathbf{k}}\right)_j,$$

$$\nu_j = \left(\frac{\epsilon_i}{\partial\epsilon_r/\partial\omega}\right)_j, \tag{2.17}$$

where the dielectric response function $\epsilon(\omega, \mathbf{k}) = k^{-2}\mathbf{k}\cdot\mathsf{D}(\omega, \mathbf{k})\cdot\mathbf{k}$, $\epsilon_r = \text{Re}\,\epsilon$, $\epsilon_i = \text{Im}\,\epsilon$, and $(\ )_j$ means that $\omega = \omega_j$ and $\mathbf{k} = \mathbf{k}_j$ shall be inserted.

Formulas for the dispersion tensor may, for important plasma models, be found in standard textbooks. The second order conductivities have, often implicitly, been calculated for many different particular cases in papers on three-wave interaction. General formulas for $\boldsymbol{\sigma}^{(2)}_{\kappa_1,\kappa_2}$ of a magnetized multicomponent relativistic or unrelativistic plasma are given in Ref. 15 for the Vlasov–Maxwell model.

3. THREE-WAVE INTERACTION IN THE PRESENCE OF AN OSCILLATING ELECTRIC FIELD

Consider a spatially homogeneous plasma with two components, we take the species index $\sigma = e$ or i for electrons or ions, respectively. A uniform static magnetic field $\mathbf{B}_0 = B_0 e_3$ and a uniform oscillating electric field $\mathbf{E}_0 \sin\Omega t = (E_{0x}e_1 + E_{0z}e_3)\sin\Omega t$ are present.[14] Define

$$\xi_\sigma(t) = -\frac{q}{m}\left[\left(\frac{E_{0x}}{\Omega^2 - \omega_c^2}e_1 + \frac{E_{0z}}{\Omega^2}e_3\right)\sin\Omega t + \frac{\omega_c}{\Omega}\frac{E_{0x}}{\Omega^2 - \omega_c^2}e_2\cos\Omega t\right], \tag{3.1}$$

Here, as in many places below, we have omitted the particle species index σ (on q, m, and ω_c). This unperturbed oscillating physical system will sometimes be labeled with a subscript O to distinguish it from the corresponding homogeneous and stationary system obtained in the absence of the external oscillating electric field but with the same magnetic field $\mathbf{B}_0$. This latter unperturbed state will be labeled with a subscript H. It is easy to check that $\mathbf{r}_\sigma(t) + \xi_\sigma(t)$ is an unperturbed orbit in the O system if $\mathbf{r}_\sigma(t)$ is an unperturbed orbit in the H system. We have the following relation between the unperturbed particle distribution functions:

$$f_O^\sigma(t, \mathbf{v}) = f_H^\sigma(\mathbf{v} - \xi_\sigma'(t)). \tag{3.2}$$

We are going to reconsider linear theory in a way which is suitable for the application of Corollary 2 (III). We consider perturbations of the form $a(t)\exp(-i\omega t + i\mathbf{k}\cdot\mathbf{r})$, where $a(t) = a(t + 2\pi/\Omega)$ is periodic. We introduce the usual scalar product between complex valued functions $a(t)$ and $b(t)$ with period $2\pi/\Omega$ as

$$\langle a, b\rangle = \langle a^*b\rangle = \frac{\Omega}{2\pi}\int_0^{2\pi/\Omega} a^*(t)b(t)\,dt. \tag{3.3}$$

Note that this use of the symbol $\langle\ \rangle$ is consistent with Definition 4(a) (III).[16] We define $\mu[=\mu(\mathbf{k})]$ and $\theta[=\theta(\mathbf{k})]$ so that

$$\mathbf{k}\cdot(\xi_e(t) - \xi_i(t)) = -\mu\sin(\Omega t + \theta). \tag{3.4}$$

The equality (3.4) does not uniquely determine μ and θ. In order to make the comparison with Ref. 12 as simple as possible, we take

$$\mu = -(A^2 + B^2)^{1/2}\,\text{sgn}\,A, \quad \theta = \arctan(B/A), \tag{3.5}$$

where $\text{sgn}\,A = A/|A|$ and

$$A = \frac{q_i}{m_i}\left(\frac{k_z E_{0z}}{\Omega^2} + \frac{k_x E_{0x}}{\Omega^2 - \omega_{ci}^2}\right) - \frac{q_e}{m_e}\left(\frac{k_z E_{0z}}{\Omega^2} + \frac{k_x E_{0x}}{\Omega^2 - \omega_{ce}^2}\right), \tag{3.6}$$

$$B=\frac{k_y E_{0x}}{\Omega}\left(\frac{q_i}{m_i}\frac{\omega_{ci}}{\Omega^2-\omega_{ci}^2}-\frac{q_e}{m_e}\frac{\omega_{ce}}{\Omega^2-\omega_{ce}^2}\right). \quad (3.7)$$

We prove in Appendix B that the current response operators of the H and the O systems are simply related to each other. In particular, we obtain for the charge density responses on a perturbation $U(t,\mathbf{r})$ of the scalar potential that

$$\delta\rho_{\mathrm{O}}^{(\sigma,1)}[U]=T_\sigma^{-1}\circ\delta\rho_{\mathrm{H}}^{(\sigma,1)}\circ T_\sigma[U], \quad (3.8)$$

$$\delta\rho_{\mathrm{O}}^{(\sigma,2)}[U,U]=T_\sigma^{-1}\circ\delta\rho_{\mathrm{H}}^{(\sigma,2)}[T_\sigma[U],T_\sigma[U]], \quad (3.9)$$

where the operator T_σ is defined on any function $a(t,\mathbf{r},\mathbf{v})$ by

$$T_\sigma[a](t,\mathbf{r},\mathbf{v})=a(t,\mathbf{r}+\xi_\sigma(t),\mathbf{v}+\xi_\sigma'(t)). \quad (3.10)$$

The relations (3.8) and (3.9) may alternatively be motivated by means of the oscillating reference frame technique.[7,9,17] It follows directly from (3.8) and (3.10) that

$$\delta\rho_{\mathrm{O}}^{(\sigma,1)}[\beta_n^\sigma]=-\epsilon_0 k^2\chi^\sigma(\omega+n\Omega,\mathbf{k})\beta_n^\sigma, \quad (3.11)$$

where

$$\beta_n^\sigma(t,\mathbf{r})=\alpha_n^\sigma(t)\exp(-i\omega t+i\mathbf{k}\cdot\mathbf{r}), \quad (3.12)$$

$$\alpha_n^\sigma(t)=\exp[-in(\Omega t+\theta)-i\mathbf{k}\cdot\xi_\sigma(t)], \quad (3.13)$$

n is an integer, and $\chi^\sigma(\omega,\mathbf{k})$ is the susceptibility of the H system, i.e.,[18]

$$\delta\rho_{\mathrm{H}}^{(\sigma,1)}[\exp(-i\omega t+i\mathbf{k}\cdot\mathbf{r})](t,\mathbf{r})=-\epsilon_0 k^2\chi^\sigma(\omega,\mathbf{k})\exp(-i\omega t+i\mathbf{k}\cdot\mathbf{r}). \quad (3.14)$$

The functions α_n^σ satisfy

$$\langle\alpha_m^\sigma,\alpha_n^\sigma\rangle=\delta_{m,n}, \quad (3.15)$$

$$\langle\alpha_m^e,\alpha_n^i\rangle=\langle\alpha_n^i,\alpha_m^e\rangle=J_{m-n}(\mu). \quad (3.16)$$

The factor $\exp(-in\theta)$ in the definition (3.13) is motivated by the simple result (3.16).

It is now easy to derive the dispersion relation. Write the potential

$$U(t,\mathbf{r})=u(t)\exp(-i\omega t+i\mathbf{k}\cdot\mathbf{r})+\text{c.c.}, \quad (3.17)$$

where $u(t)$ has period $2\pi/\Omega$. We have

$$u(t)=\sum_n u_n^\sigma\alpha_n^\sigma(t),\quad u_n^\sigma=\langle\alpha_n^\sigma,u\rangle. \quad (3.18)$$

Here and below $\Sigma_n=\Sigma_{n=-\infty}^{\infty}$. The linearized Poisson's equation

$$\Delta U=-(1/\epsilon_0)(\delta\rho_{\mathrm{O}}^{(e,1)}[U]+\delta\rho_{\mathrm{O}}^{(i,1)}[U]). \quad (3.19)$$

Substitution of (3.17) and (3.18) in (3.19) yields after the use of (3.11) that

$$\sum_n[(1+\chi_n^e)u_n^e\alpha_n^e(t)+\chi_n^i u_n^i\alpha_n^i(t)]=0, \quad (3.20)$$

where $\chi_n^\sigma=\chi^\sigma(\omega+n\Omega,\mathbf{k})$. Scalar multiplication from the left with α_l^e on (3.20) and then multiplication with $(1+\chi_l^e)^{-1}\alpha_l^e$ and summation over l give

$$u(t)+\sum_{n,l}u_n^i\chi_n^i(1+\chi_l^e)^{-1}\langle\alpha_l^e,\alpha_n^i\rangle\alpha_l^e(t)=0. \quad (3.21)$$

Scalar multiplication from the left on (3.21) with α_m^i and the substitution of (3.16) yield

$$\sum_n M_{m,n}u_n^i=0, \quad (3.22)$$

where

$$M_{m,n}=\delta_{m,n}+\chi_n^i\sum_l(1+\chi_l^e)^{-1}J_{l-m}(\mu)J_{l-n}(\mu). \quad (3.23)$$

The condition that (3.22) has a nontrivial solution gives the dispersion relation in terms of an infinite determinant[6–9]

$$\det(M_{m,n})=0. \quad (3.24)$$

Note, however, that Ref. 9 gives the correct expression for μ only in the particular case when $E_{0x}=0$, i.e., when $\mathbf{E}_0$ is parallel to $\mathbf{B}_0$.

Let us now consider three normal modes with potentials

$$u_j(t)\exp(-i\omega_j t+i\mathbf{k}_j\cdot\mathbf{r})+\text{c.c.},\quad j=1,2,3, \quad (3.25)$$

where ω_j and $\mathbf{k}_j$ satisfy the dispersion relation

$$\det(M_{m,n}(\omega_j,\mathbf{k}_j))=0 \quad (3.26)$$

and the resonance conditions

$$\omega_1+\omega_2+\omega_3=0,\quad \mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3=0. \quad (3.27)$$

For $(\omega,\mathbf{k})=(\omega_j,\mathbf{k}_j)$ we write $\alpha_{j,n}^\sigma,\beta_{j,n}^\sigma,u_{j,n}^\sigma,\theta_j$, and μ_j instead of $\alpha_n^\sigma,\beta_n^\sigma,u_n^\sigma,\theta$, and μ. From (3.9) and (3.13) we obtain

$$\delta\rho_{\mathrm{O}}^{(\sigma,2)}[\beta_{1,n_1}^\sigma,\beta_{2,n_2}^\sigma](t,\mathbf{r})=(\beta_{3,n_3}^\sigma(t))^*\exp(-in_1\theta_1-in_2\theta_2-in_3\theta_3)\times C^{(\sigma,2)}(\omega_1+n_1\Omega,\mathbf{k}_1,\omega_2+n_2\Omega,\mathbf{k}_2),\ n_3=-n_1-n_2, \quad (3.28)$$

where $C^{(\sigma,2)}$ is the second-order capacitance for the H system, i.e.,[18,19]

$$\delta\rho_{\mathrm{H}}^{(\sigma,2)}[\exp(-i\omega_1 t+i\mathbf{k}_1\cdot\mathbf{r}),\exp(-i\omega_2 t+i\mathbf{k}_2\cdot\mathbf{r})](t,\mathbf{r})=C^{(\sigma,2)}(\omega_1,\mathbf{k}_1,\omega_2,\mathbf{k}_2)\times\exp(-i(\omega_1+\omega_2)t+i(\mathbf{k}_1+\mathbf{k}_2)\cdot\mathbf{r}). \quad (3.29)$$

By means of the continuity equation for the second-order current and charge density [i.e., (3.2) (II)] we obtain $C^{(\sigma,2)}$ in terms of the second-order conductivity $\sigma^{(\sigma,2)}$ for the H system as

$$C^{(\sigma,2)}(\kappa_1,\kappa_2)=-\omega_3^{-1}\mathbf{k}_3\cdot\sigma_{\kappa_1,\kappa_2}^{(\sigma,2)}[\mathbf{k}_1,\mathbf{k}_2]. \quad (3.30)$$

Useful expressions for the conductivities may be found in Ref. 15.

It is now straightforward to obtain the coupled mode equation by means of Corollary 2 (III). Assume that the potential in the plasma may be written

$$\sum_{j=1}^{3}A_j(t)u_j(t)\exp(-i\omega_j t+i\mathbf{k}_j\cdot\mathbf{r})+\text{c.c.}, \quad (3.31)$$

where $A_j(t)$ is slowly varying due to resonant wave–wave interaction. We neglect for simplicity wave–particle interactions. The functions $\chi_{3,\kappa}$ and ϕ_3 in Corollary 2 (III) are

$$\chi_{3,\kappa}(t,\mathbf{r})=e_0 u_3(t)\exp(-i\omega t+i\mathbf{k}\cdot\mathbf{r}), \quad (3.32)$$

$$\phi_3(t,\mathbf{r})=e_0 u_3(t)\exp(-i\omega_3 t+i\mathbf{k}_3\cdot\mathbf{r}). \quad (3.33)$$

Neglecting wave–particle interactions means $\delta J^{(1)}=\delta J^{(1-)}$ and thus $h=0$ and the operator $H=\nabla_E\cdot(\nabla_E\wedge\cdots)+(\mu_0/\epsilon_0)^{1/2}\delta J_{\mathrm{O}}^{(1)}$. It follows from (3.11), (3.12), and (3.18) that

$$\langle\chi_{3\kappa}^*\cdot H\chi_{3\kappa}\rangle_{\mathbf{k}=\mathbf{k}_3}=k_3^2\tilde{\epsilon}(\omega,\mathbf{k}_3), \quad (3.34)$$

where

$$\tilde{\epsilon}(\omega, \mathbf{k}_3) = \sum_n \{ [1 + \chi^e(\omega + n\Omega, \mathbf{k}_3)] |u^e_{3,n}|^2 + \chi^i(\omega + n\Omega, \mathbf{k}_3) |u^i_{3,n}|^2 \}. \quad (3.35)$$

From (3.12), (3.18), (3.28), (3.34), and Corollary 2 (III) we obtain

$$V(3,1,2) = 2\sum_\sigma \sum_{n_1 + n_2 + n_3 = 0} C^{(\sigma,2)}(\omega_1 + n_1\Omega, \mathbf{k}_1, \omega_2 + n_2\Omega, \mathbf{k}_2) \times \prod_{j=1}^{3} [u^\sigma_{j,n_j} \exp(-in_j\theta_j)] \quad (3.36)$$

and the mode coupling equation

$$\frac{dA^*_3}{dt} = \frac{iV(3,1,2)}{\epsilon_0 k_3^2 (\partial\tilde{\epsilon}(\omega, \mathbf{k}_3)/\partial\omega)_{\omega=\omega_3}} A_1 A_2, \quad (3.37)$$

The Manley–Rowe relations [cf. Corollary 1 (III)] follows the symmetries

$$C^{(\sigma,2)}(\omega_j + n_j\Omega, \mathbf{k}_j, \omega_l + n_l\Omega, \mathbf{k}_l) = C^{(\sigma,2)}(\omega_1 + n_1\Omega, \mathbf{k}_1, \omega_2 + n_2\Omega, \mathbf{k}_2) \quad \text{for } j \neq 1 \text{ and } j, l \in \{1,2,3\}, \quad (3.38)$$

and the symmetries (3.38) follow from (3.30), Ref. 15, and the neglect of wave–particle interactions. The dispersion relation (3.24) has the form of an infinite determinant. A standard simplifying assumption is that $\chi^i(\omega_j + n\Omega, \mathbf{k}_j) = 0$ for $n \neq 0$.[7–9] We then consider low frequency modes $|\omega_j| \ll \Omega$. Correspondingly, we also take $C^{(i,2)}(\omega_1 + n_1\Omega, \mathbf{k}_1, \omega_2 + n_2\Omega, \mathbf{k}_2) = 0$ for $(n_1, n_2) \neq (0,0)$. With these assumptions we can solve (3.22) and obtain

$$u^i_m = -\sum_l \chi^i_0 (1 + \chi^e_l)^{-1} J_{l-m}(\mu) J_l(\mu), \quad u^e_m = -\chi^i_0 (1 + \chi^e_m)^{-1} J_m(\mu), \quad (3.39)$$

where we have taken $u^i_0 = 1$ (this is just a convenient normalization). From $u^i_0 = 1$ in (3.39) yields the dispersion relation

$$(1/\chi^i_0) + \sum_l J_l(\mu)^2/(1 + \chi^e_l) = 0. \quad (3.40)$$

We now obtain

$$\left(\frac{\partial\tilde{\epsilon}(\omega, \mathbf{k}_3)}{\partial\omega}\right)_{\omega=\omega_3} = -[\chi^i(\omega_3, \mathbf{k}_3)]^2 \left(\frac{\partial}{\partial\omega}\left[\frac{1}{\chi^i(\omega, \mathbf{k}_3)} + \sum_n \frac{J_n(\mu_3)^2}{1 + \chi^e(\omega + n\Omega, \mathbf{k}_3)}\right]\right)_{\omega=\omega_3}, \quad (3.41)$$

$$V(3,1,2) = -\sum_{n_1 + n_2 + n_3 = 0} 2C^{(e,2)}(\omega_1 + n_1\Omega, \mathbf{k}_1, \omega_2 + n_2\Omega, \mathbf{k}_2) \times \prod_{j=1}^{3} \frac{\chi^i(\omega_j, \mathbf{k}_j) J_{n_j}(\mu_j) \exp(-in_j\theta_j)}{1 + \chi^e(\omega_j + n_j\Omega, \mathbf{k}_j)} + 2C^{(i,2)}(\omega_1, \mathbf{k}_1, \omega_2, \mathbf{k}_2). \quad (3.42)$$

The resonant interaction of three waves satisfying the dispersion relation (3.40) is considered in Ref. 12 for particular case when $\mathbf{k}_j$ is parallel to $\mathbf{E}_0$ for $j = 1, 2, 3$ and the plasma is unmangetized. We recover their results if we take $\mathbf{k}_j = k_{jz} e_3 \equiv k_{jz}\hat{\mathbf{z}}$ for $j = 1, 2, 3$ and $E_{0x} = 0$. From (3.8) and (3.9) in Ref. 15 we obtain

$$\hat{\mathbf{z}}\cdot\sigma^{(\sigma,2)}_{\kappa_1,\kappa_2}(\hat{\mathbf{z}}, \hat{\mathbf{z}}) = \frac{\omega_3 q^3}{2m^2} \int \frac{\partial f_H}{\partial v_z} \times \frac{d\mathbf{v}}{(\omega_1 - k_{1z}v_z)(\omega_2 - k_{2z}v_z)(\omega_3 - k_{3z}v_z)}. \quad (3.43)$$

It is now straightforward to compare with Ref. 12. Note, however, that in (6) of Ref. 12 some notational ambiguities and a printing error have appeared, and thus we should make the following replacements in (6) of Ref. 12:

$$[\chi_i(\omega, k)]^3 \to \chi_i(\omega_1, k_1)\chi_i(\omega_2, k_2)\chi_i(\omega_3, k_3),$$
$$J_{n_j}(\mu) \to J_{n_j}(\mu_j) \quad \text{and} \quad q_e^2 \to q_e^3.$$

In the limit $\mathbf{E}_0 \to 0$ we obtain $V(3,1,2) = 2\Sigma_\sigma C^{(\sigma,2)}(\omega_1, \mathbf{k}_1, \omega_2, \mathbf{k}_2)$ and $\partial\tilde{\epsilon}/\partial\omega = \partial\epsilon/\partial\omega$ from (3.41) and (3.42) or (3.35) and (3.36). Then (3.37) is the usual mode coupling equation for electrostatic waves in a homogeneous and stationary plasma.

In conclusion, we have in this section obtained the coupled mode equations for the interaction of three potential waves (3.31) satisfying the dispersion relation (3.26). The coupled mode equations are then (3.37) plus the two equations obtained by even permutations of (1,2,3). The coefficients in these equations are obtained from (3.18), (3.35), and (3.36). In the particular case when the three interacting waves satisfy the dispersion relation (3.40), we obtain the coefficients in (3.37) from (3.41) and (3.42). The Manley–Rowe relations are satisfied in both cases.

4. THE INTERACTION OF SURFACE WAVES IN A SEMI-BOUNDED PLASMA

We consider a plasma confined to the region $z \geqslant 0$.[14] The plasma may have several particle components, but we omit the particle species index σ. We assume that the particles are

specularly reflected at the boundary.[10,11,20] A uniform magnetic field $\mathbf{B}_0 = B_0 e_3$ may be present.[14] The unperturbed distribution function f_S is

$$\begin{aligned} f_S(z, \mathbf{v}) &= f_H(\mathbf{v}) \quad \text{for } z \geqslant 0 \\ f_S(z, \mathbf{v}) &= 0 \qquad \text{for } z < 0. \end{aligned} \tag{4.1}$$

Here f_H is a solution of the Vlasov equation for an unbounded plasma, i.e., $\mathbf{v} \times \mathbf{B}_0 \cdot (\partial / \partial \mathbf{v}) f_H(\mathbf{v}) = 0$, such that $f_H(v_x, v_y, v_z) = f_H(v_x, v_y, -v_z)$. Then f_S is a solution of the Vlasov equation satisfying the right boundary condition in accordance with the specular reflection model. The subscript S is used for the semibounded unperturbed state, and the subscript H for the corresponding unbounded unperturbed state. For any function $a(t, \mathbf{r}, \mathbf{v})$ defined on $\{(t, \mathbf{r}, \mathbf{v})\}$ (essentially $= R^7$), we use the notation $(a; z \geqslant 0)$ for the restriction of a to the halfspace $\{(t, \mathbf{r}, \mathbf{v}) | z \geqslant 0\}$. The operator R is defined on functions which have this half-space as domain by

$$\begin{aligned} &R[a; z \geqslant 0](t, \mathbf{r}, \mathbf{v}) \\ &\quad = a(t, x, y, |z|, v_x, v_y, (\operatorname{sgn} z) v_z), \end{aligned} \tag{4.2}$$

where $\operatorname{sgn} z = 1$ for $z \geqslant 0$ and $\operatorname{sgn} z = -1$ for $z < 0$. Thus the operator R extends the function $(a; z \geqslant 0)$ defined on a halfspace to a function on the whole space by specular symmetry in the argument. The inverse operator R^{-1} is then defined on the specularly symmetric functions $a(t, \mathbf{r}, \mathbf{v})$, i.e., functions a such that $a(t, x, y, z, v_x, v_y, v_z) = a(t, x, y, -z, v_x, v_y, -v_z)$ for $z \neq 0$ and then $R^{-1}[a] = (a; z \geqslant 0)$. The operator R plays the same role in this section as the operators T_σ in Sec. 3. We note that $(f_S; z \geqslant 0) = R^{-1}[f_H]$. We introduce the scalar product between complex-valued functions $a(z)$ and $b(z)$ as [cf. (3.4)]

$$\langle a, b \rangle = \langle a^* b \rangle = \int_{-\infty}^{\infty} a^*(z) b(z)\, dz. \tag{4.3}$$

This is consistent with Definition 4(a) (III). Corresponding to (3.8) and (3.9), we have the relations (see Appendix C),

$$(\delta \rho_S^{(1)}[U]; z \geqslant 0) = R^{-1} \circ \delta \rho_H^{(1)} \circ R\,[U; z \geqslant 0], \tag{4.4}$$

$$\begin{aligned} &(\delta \rho_S^{(2)}[U, U]; z \geqslant 0) \\ &\quad = R^{-1} \circ \delta \rho_H^{(2)}[R\,[U; z \geqslant 0], R\,[U; z \geqslant 0]]. \end{aligned} \tag{4.5}$$

It follows that[18] [cf. (3.11)]

$$\begin{aligned} &\delta \rho_S^{(1)}[(\cos k_z z) \exp(-i\omega t + ik_x x + ik_y y)](t, \mathbf{r}) \\ &\quad = -\epsilon_0 k^2 \chi(\omega, \mathbf{k})(\cos k_z z) \exp(-i\omega t + ik_x x + ik_y y) \end{aligned} \tag{4.6}$$

$$\text{for } z \geqslant 0.$$

We are going to use the cosine transform

$$a_c(k_z) = \int_0^{\infty} a(z) \cos k_z z\, dz, \tag{4.7}$$

$$a(z) = \frac{2}{\pi} \int_0^{\infty} a_c(k_z) \cos k_z z\, dk_z, \tag{4.8}$$

in order to analyze the linearized Poisson equation

$$\Delta U + (1/\epsilon_0) \delta \rho_S^{(1)}[U] = 0. \tag{4.9}$$

We look for solutions to (4.9) of the form

$$U(t, \mathbf{r}) = u(z) \exp(-i\omega t + ik_x x + ik_y y) + \text{c.c}, \quad u(0) = 1, \tag{4.10}$$

with $u(z)$ and $u'(z)$ continuous across the boundary $z = 0$, i.e., there is no surface charge. Substitution of (4.10) in (4.9), the cosine transformation of the equation obtained, and making use of (4.6) yields

$$-u'(0) - k^2 u_c(k_z) - k^2 \chi(\omega, \mathbf{k}) u_c(k_z) = 0. \tag{4.11}$$

From (4.9) for $z \leqslant 0$, i.e., Laplace's equation, we obtain

$$\begin{aligned} &u(z) = \exp(k_\perp z) \quad \text{for } z \leqslant 0, \\ &u'(0) = k_\perp = (k_x^2 + k_y^2)^{1/2}. \end{aligned} \tag{4.12}$$

We obtain from (4.11) and (4.12)

$$\begin{aligned} u_c(k_z) &= -\frac{k_\perp}{k^2 \epsilon(\omega, \mathbf{k})}, \\ u(z) &= -\frac{2k_\perp}{\pi} \int_0^{\infty} \frac{\cos(k_z z)}{k^2 \epsilon(\omega, \mathbf{k})}\, dk_z, \quad z \geqslant 0, \end{aligned} \tag{4.13}$$

and the well-known dispersion relation is obtained from $u(0) = 1$ in (4.13)[10,11,21]

$$0 = 1 + \frac{2k_\perp}{\pi} \int_0^{\infty} \frac{1}{k^2 \epsilon(\omega, \mathbf{k})}\, dk_z \equiv \zeta(\omega, k_x, k_y). \tag{4.14}$$

Let us now consider three normal modes

$$\begin{aligned} &u_j(z) \exp(-i\omega_j t + ik_{jx} x + ik_{jy} y) + \text{c.c.}, \\ &u_j(0) = 1, \quad j = 1,2,3, \end{aligned} \tag{4.15}$$

where $\zeta(\omega_j, k_{jx}, k_{jy}) = 0$ and where

$$\begin{aligned} &\omega_1 + \omega_2 + \omega_3 = 0, \quad k_{1x} + k_{2x} + k_{3x} = 0, \\ &k_{1y} + k_{2y} + k_{3y} = 0. \end{aligned} \tag{4.16}$$

It is now straightforward to obtain the coupled mode equations from Corollary 2 (III). Assume that the potential in the plasma may be written

$$\sum_{j=1}^{3} A_j(t) u_j(z) \exp(-i\omega_j t + ik_{jx} x + ik_{jy} y) + \text{c.c.}, \tag{4.17}$$

where $A_j(t)$ is slowly varying due to resonant wave–wave interaction. The functions $\chi_{j\kappa}$ and ϕ_j in Corollary 2 (III) are[14]

$$\chi_{j\kappa}(t, \mathbf{r}) = e_0 u_j(z) \exp(-i\omega t + ik_x x + ik_y y), \tag{4.18}$$

$$\begin{aligned} \phi_j(t, \mathbf{r}) &= e_0 u_j(z) \exp(-i\omega_j t + ik_{jx} x + ik_{jy} y) \\ &\equiv e_0 \tilde{u}(t, \mathbf{r}), \end{aligned} \tag{4.19}$$

where $\kappa = (\omega/c) e_0 + k_x e_1 + k_y e_2$ is used as a shorthand notation for (ω, k_x, k_y).

We obtain by means of (4.6)–(4.8) and (4.18) that [cf. (3.34)–(3.36)]

$$\langle \chi_{3\kappa}^* \cdot H \chi_{3\kappa} \rangle = \frac{2}{\pi} \int_0^{\infty} |u_{jc}(k_z)|^2 k^2 \epsilon(\omega, \mathbf{k})\, dk_z. \tag{4.20}$$

Inserting (4.13) in (4.20) yields

$$\left[\frac{\partial}{\partial \omega} \langle \chi_{3\kappa}^* \cdot H \chi_{3\kappa} \rangle\right]_{\kappa = \kappa_3} = -k_{3\perp} \left[\frac{\partial \zeta(\omega, k_{3x}, k_{3y})}{\partial \omega}\right]_{\omega = \omega_3}. \tag{4.21}$$

We obtain from (4.19) and (4.5) that

$$V(3,1,2) = -2c^{-1}\langle \phi_3 \cdot \delta J_S^{(2)}[\phi_1, \phi_2]\rangle = 2\int_0^\infty \tilde{u}_3 \delta \rho_S^{(2)}[\tilde{u}_1, \tilde{u}_2]\, dz$$

$$= \int_{-\infty}^{\infty} R\,[\tilde{u}_3; z \geqslant 0] \delta \rho_H^{(2)}[R\,[\tilde{u}_1; z \geqslant 0], R\,[\tilde{u}_2; z \geqslant 0]]\, dz. \tag{4.22}$$

Since $f_H(v_x, v_y, v_z) = f_H(v_x, v_y, -v_z)$, we have also $\epsilon(\omega, k_x, k_y, k_z) = \epsilon(\omega, k_x, k_y, -k_z)$, and thus we obtain from (4.13) that

$$R\,[\tilde{u}_j; z \geqslant 0] = -\frac{k_{j\perp}}{\pi} \int_{-\infty}^{\infty} \frac{\exp(-i\omega_j t + i\mathbf{k}_j \cdot \mathbf{r})}{k_j^2 \epsilon(\omega_j, \mathbf{k}_j)}\, dk_{jz}. \tag{4.23}$$

We substitute (4.23) in (4.22) and obtain, by means of (3.29),

$$V(3,2,1) = -\frac{k_{1\perp} k_{2\perp} k_{3\perp}}{\pi^3} \int_{-\infty}^{\infty} dz \int_{-\infty}^{\infty} \frac{C^{(2)}(\omega_1, \mathbf{k}_1, \omega_2, \mathbf{k}_2) \exp[i(k_{1z} + k_{2z} + k_{3z})z]}{k_1^2 k_2^2 k_3^2 \epsilon(\omega_1, \mathbf{k}_1) \epsilon(\omega_2, \mathbf{k}_2) \epsilon(\omega_3, \mathbf{k}_3)} dk_{1z} dk_{2z} dk_{3z}. \tag{4.24}$$

By reversing the order of integration in (4.24) and by use of the identity $\int_{-\infty}^{\infty} \exp[i(k_{1z} + k_{2z} + k_{3z})z]\, dz = 2\pi\delta(k_{1z} + k_{2z} + k_{3z})$, we obtain

$$V(3,2,1) = -\frac{k_{1\perp} k_{2\perp} k_{3\perp} 2}{\pi^2} \int_{-\infty}^{\infty} dk_{1z} dk_{2z} \times \frac{C^{(2)}(\omega_1, \mathbf{k}_1, \omega_2, \mathbf{k}_2)}{k_1^2 k_2^2 k_3^2 \epsilon(\omega_1, \mathbf{k}_1) \epsilon(\omega_2, \mathbf{k}_2) \epsilon(\omega_3, \mathbf{k}_3)}. \tag{4.25}$$

In (4.25) we define $k_{3z} = -k_{1z} - k_{2z}$. From Corollary 2 (III) we may now write the coupled mode equations

$$\frac{dA_3^*}{dt} = \frac{-iV(3,1,2)}{\epsilon_0 k_{3\perp} (\partial \zeta(\omega, k_{3x}, k_{3y})/\partial\omega)_{\omega=\omega_3}} A_1 A_2 \tag{4.26}$$

together with the two equations obtained by even permutations of (1,2,3). The coefficients in (4.26) are given by (4.14) and (4.25). It may be of interest that these formulas may be used as they stand also for a warm fluid plasma since then we still have (4.4)–(4.5) satisfied and we have no surface charge. It is, however, not correct to use them when there is surface charge present, as, for example, in the cold fluid model. Still (4.14) happens to give the correct dispersion relation when the cold model $\epsilon(\omega, \mathbf{k})$ is inserted. Inserting the cold model quantities in (4.24) yields a divergent integral, but we may use (4.25) instead. The result so obtained is, however, not exactly correct, and some information concerning how important the neglected terms are is given in Ref. 21, where the cold-fluid coupled mode equations are obtained by calculating the corresponding warm-fluid formula before taking the cold limit.

APPENDIX A

Papers I, II, and III are written for a relativistic plasma. It is in many applications possible to choose a Lorentz frame $L = (e_0, e_1, e_2, e_3)$, in which the plasma particles have unrelativistically low speeds so that the relativistic mass correction may be neglected. Let us choose such a Lorentz frame L, and from now on we do not change reference frame. The formulas for the coupled mode equations in terms of the response operators work, of course, also when the plasma particles are treated unrelativistically; the changes enter in the formulas for the response operators, i.e., we must modify Results 1 and 2 (I) and accordingly also Result 3 (I) and Sec. 3 (III). The other parts of Paper III and all of II may be kept essentially unchanged. We must, however, keep in mind that we now work in a specific coordinate system $L = (e_0, e_1, e_2, e_3)$. The scalar product induced from L by $|e_i \cdot e_j| = \delta_{ij}$, $-e_0 \cdot e_0 = e_1 \cdot e_1 = e_2 \cdot e_2 = e_3 \cdot e_3 = 1$ is now essentially coordinate dependent, unlike the situation in relativistic space–time where the Lorentz metric is an intrinsic coordinate-independent structure.

We shall now consider the modifications of the response operator formulas in the unrelativistic limit. We take two different approaches to this problem. First, we give new unrelativistic interpretations of certain symbols so that most formulas may be left formally unchanged. In this way we get best possible contact between relativistic and unrelativistic theory. It is, however, convenient to have formulas for the unrelativistic response operators in more traditional notations, and therefore, secondly, we also give such formulas.

We concentrate on Results 1 and 2 in I since it is then easy to see how the other response operator formulas shall be modified in the unrelativistic limit. In this limit we redefine S as [see (4.6) (III)]

$$S = e_0 + \{e_0\}^\perp. \tag{A1}$$

Physically we associate with each $u \in S$ the ordinary velocity $\mathbf{v}$ by the relation

$$u = e_0 + \mathbf{v}/c. \tag{A2}$$

We note that S is now a three-dimensional affine subspace of V. The indefinite scalar product on V induces a measure on V and on $S \subset V$ as described in Ref. 4, Sec. 2C. In particular for $A \subset S$ and a function $a(u)$ defined on A

$$\int_A a(u)\, du = \int_B a(e_0 + u^1 e_1 + u^2 e_2 + u^3 e_3)\, du^1 du^2 du^3, \tag{A3}$$

where $B \subset R^3$ defined by $B = \{(u^1, u^2, u^3) | e_0 + u^1 e_1 + u^2 e_2 + u^3 e_3 \in A\}$. The unperturbed distribution function $f_0(x,u)$ in (2.1)–(2.2) (I) is normalized so that if we integrate f_0 over a subset A of phasespace, i.e., $A \subset \Pi_L(V) \times S$, we have

$$\int_A f_0(x,u)\,dA = \int_B f_0(x^0e_0 + x^1e_1 + x^2e_2 + x^3e_3,\ e_0 + u^1e_1 + u^2e_2 + u^3e_3)$$
$$\times dx^1dx^2dx^3du^1du^2du^3 = \text{number of particles in A at time } x^0/c, \tag{A4}$$

where

$$B = \{(x^1, x^2, x^3, u^1, u^2, u^3) | (x^1e_1 + x^2e_2 + x^3e_3,$$
$$e_0 + u^1e_1 + u^2e_2 + u^3e_3) \in A\}.$$

We must also redefine ∇_S in the evident way

$$\nabla_S = \sum_{i=1}^{3} e_i \frac{\partial}{\partial u^i}. \tag{A5}$$

Now (2.1) and (2.2) in I may be used as they stand for an unrelativistic plasma if we redefine $\delta x(j)$ and $\delta u(j)$ by replacing (2.7) (I) with

$$\delta x(j) = -qm_0^{-1}c^{-2}\nabla_S D_0^{-1}(u\cdot\phi_j) \tag{A6}$$

and keeping (2.3) (I) unchanged. The corresponding modification of (2.4) (I) is

$$D_0\delta u(j) - qm_0^{-1}c^{-2}\Pi_L[\nabla_E \wedge \Phi_0\cdot\delta u(j)$$
$$+ \delta x(j)\cdot\nabla_E(\nabla_E \wedge \Phi_0\cdot u)] = qm_0^{-1}c^{-2}\Pi_L(\nabla_E \wedge \phi_j\cdot u). \tag{A7}$$

We now write these response operator formulas for an unrelativistic plasma in more traditional notations. We normalize the distribution fucntion $f_0(t, \mathbf{r}, \mathbf{v})$ so that $f_0(t, \mathbf{r}, \mathbf{v})\,d\mathbf{r}d\mathbf{v}$ is the number of particles in $d\mathbf{r}d\mathbf{v}$. Take $\phi_0 \in L^0(E,V)$ and ϕ_1, $\phi_2 \in L_0(E,V)$ [see Result 1 (I)], $\phi_j = U_je_0 + c\mathbf{A}_j$ and $\Phi_0 = U^0e_0 + c\mathbf{A}^0$. Then

$$\int \phi_0\cdot\delta J^{(1)}[\phi_1]\,dtd\mathbf{r}$$
$$= \tfrac{1}{2}qc\int dtd\mathbf{r}d\mathbf{v}\, f_0(t, \mathbf{r}, \mathbf{v})$$
$$\times[\mathbf{r}_1\cdot\nabla(\mathbf{v}\cdot\mathbf{A}_0 - U_0) + \mathbf{v}_1\cdot\mathbf{A}_0$$
$$+ \mathbf{r}_0\cdot\nabla(\mathbf{v}\cdot\mathbf{A}_1 - U_1) + \mathbf{v}_0\cdot\mathbf{A}_1], \tag{A8}$$

$$\int \phi_0\cdot\delta J^{(2)}[\phi_1,\phi_2]\,dtd\mathbf{r}$$
$$= \sum qc\int dtd\mathbf{r}d\mathbf{v}\, f_0(t, \mathbf{r}, \mathbf{v})$$
$$\times[\tfrac{1}{12}\mathbf{r}_0\mathbf{r}_1\mathbf{r}_2 \vdots \nabla\nabla\nabla(\mathbf{v}\cdot\mathbf{A}^0 - U^0) + \tfrac{1}{4}\mathbf{r}_0\mathbf{r}_1\mathbf{v}_2 \vdots \nabla\nabla\mathbf{A}^0$$
$$+ \tfrac{1}{4}\mathbf{r}_0\mathbf{r}_1 : \nabla\nabla(\mathbf{v}\cdot\mathbf{A}_2 - U_2) + \tfrac{1}{2}\mathbf{r}_0\mathbf{v}_1 : \nabla\mathbf{A}_2], \tag{A9}$$

where the summation is over the six permutations of the subscripts 0,1,2 and we use **a b:c d = a·c b·d** and **a b c ⋮ d e f = a·d b·e c·f**.

The vector fields $\mathbf{r}_j(t, \mathbf{r}, \mathbf{v})$ and $\mathbf{v}_j(t, \mathbf{r}, \mathbf{v})$ are defined by

$$\mathbf{r}_j = -\frac{q}{m_0c}\frac{\partial}{\partial\mathbf{v}} D_0^{-1}(\mathbf{v}\cdot\mathbf{A}_j - U_j), \tag{A10}$$

$$cD_0 = \frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla + \frac{q}{m_0}(\mathbf{E}^0 + \mathbf{v}\times\mathbf{B}^0)\cdot\frac{\partial}{\partial\mathbf{v}}, \tag{A11}$$

$$\mathbf{v}_j = cD_0\mathbf{r}_j, \tag{A12}$$

where $\mathbf{E}^0$ and $\mathbf{B}^0$ are the unperturbed electromagnetic fields with 4-potential Φ^0. In place of (A10) we may use

$$cD_0\mathbf{v}_j + (q/m_0)\mathbf{B}^0\times\mathbf{v}_j - (q/m_0)\mathbf{r}_j\cdot\nabla(\mathbf{E}^0 + \mathbf{v}\times\mathbf{B}^0)$$
$$= (q/m_0)(\mathbf{E}_j + \mathbf{v}\times\mathbf{B}_j). \tag{A13}$$

APPENDIX B

The purpose of this appendix is to derive a relation between the current response operators in the O and the H systems defined in Sec. 3. The formulas (3.8) and (3.9) follow as a particular case. We will prove that

$$\int \phi_0\cdot\delta J_O^{(1)}[\phi_1]\,dtd\mathbf{r} = \int \tilde{\phi}_0\cdot\delta J_H^{(1)}[\tilde{\phi}_1]\,dtd\mathbf{r}, \tag{B1}$$

$$\int \phi_0\cdot\delta J_O^{(2)}[\phi_1,\phi_2]\,dtd\mathbf{r} = \int \tilde{\phi}\cdot\delta J_H^{(2)}[\tilde{\phi}_1, \tilde{\phi}_2]\,dtd\mathbf{r}, \tag{B2}$$

where ϕ_0, ϕ_1, and ϕ_2 are as in (A8) and (A9), $\phi_j = U_je_0 + c\mathbf{A}_j$, and

$$\tilde{\phi}_j = \tilde{U}_je_0 + c\tilde{\mathbf{A}}_j, \tag{B3}$$

$$\tilde{U}_j(t, \mathbf{r}) = (U_jt, \mathbf{r} + \xi(t)) - \zeta'(t)\cdot\mathbf{A}_j(t, \mathbf{r} + \zeta(t)), \tag{B4}$$

$$\tilde{\mathbf{A}}_j(t, \mathbf{r}) = \mathbf{A}_j(t, \mathbf{r} + \xi(t)). \tag{B5}$$

We have omitted the particle species index σ and $\xi(t)$ is defined by (3.1). In terms of the operator T defined by (3.10) and the linear operator G

$$G[\phi_j] = (I - c^{-1}e_0 \otimes \xi'(t))\cdot\phi_j = \phi_j - c^{-1}\xi'(t)\cdot\phi_je_0, \tag{B6}$$

we can write $\tilde{\phi}_j$ as

$$\tilde{\phi}_j = T\circ G[\phi] = G\circ T[\phi_j]. \tag{B7}$$

We now obtain from (B1) and (B2) that

$$\delta J_O^{(1)}[\phi_1] = G^{tr}\circ T^{-1}\circ\delta J_H^{(1)}\circ G\circ T[\phi_1], \tag{B8}$$

$$\delta J_O^{(2)}[\phi_1, \phi_2] = G^{tr}\circ T^{-1}\circ\delta J_H^{(2)}[G\circ T[\phi_1], G\circ T[\phi_2]], \tag{B9}$$

where G^{tr} is the operator obtained by the transposition of G, i.e.,

$$G^{tr} = (I - c^{-1}\xi'(t)\otimes e_0). \tag{B10}$$

We prove (B1) and (B2) by direct application to the H and O systems of the explicit general formulas (A8)–(A10) and (A12) for the current response operators. We need the following relations between the unperturbed Vlasov operators of the H and the O systems

$$D_O^{-1} = T^{-1}\circ D_H^{-1}\circ T, \quad D_O = T^{-1}\circ D_H\circ T \tag{B11}$$

and also the integral equality involving respective unperturbed distribution function

$$\int f_O(t, \mathbf{r}, \mathbf{v})a(t, \mathbf{r}, \mathbf{v})\,dtd\mathbf{r}d\mathbf{v}$$
$$= \int f_H(t, \mathbf{r}, \mathbf{v})T[a](t, \mathbf{r}, \mathbf{v})\,dtd\mathbf{r}d\mathbf{v}. \tag{B12}$$

It is easy to see that (B11) is equivalent to

$$[D_{\mathrm{H}}, T] = T\circ(D_{\mathrm{O}} - D_{\mathrm{H}}), \tag{B13}$$

where the commutator of two operators A and B is written $[A,B] = A\circ B - B\circ A$. We have

$$D_{\mathrm{H}} = c^{-1}\left(\frac{\partial}{\partial t} + \mathbf{v}\cdot\frac{\partial}{\partial \mathbf{r}} + \frac{q}{m}\,\mathbf{v}\times\mathbf{B}_0\cdot\frac{\partial}{\partial \mathbf{v}}\right), \tag{B14}$$

$$D_{\mathrm{O}} = D_{\mathrm{H}} + \frac{q}{mc}(\sin \Omega t)\mathbf{E}_0\cdot\frac{\partial}{\partial \mathbf{v}}. \tag{B15}$$

By straightforward calculations

$$\left[\frac{\partial}{\partial t}, T\right] = \left(\xi'\cdot\frac{\partial}{\partial \mathbf{r}} + \xi''\cdot\frac{\partial}{\partial \mathbf{v}}\right)\circ T, \tag{B16}$$

$$\left[\mathbf{v}\cdot\frac{\partial}{\partial \mathbf{r}} + \frac{q}{m}\,\mathbf{v}\times\mathbf{B}_0\cdot\frac{\partial}{\partial \mathbf{v}}, T\right] = -\left(\xi'\cdot\frac{\partial}{\partial \mathbf{r}} + \frac{q}{m}\,\xi'\times\mathbf{B}_0\cdot\frac{\partial}{\partial \mathbf{v}}\right)\circ T, \tag{B17}$$

and now (B13) easily follows from (B14)–(B17), the fact that T and D_{O}–D_{H} commute and

$$\xi''(t) = (q/m)(\mathbf{E}_0 \sin \Omega t + \xi'(t)\times\mathbf{B}_0). \tag{B18}$$

The equality (B12) is seen by the change of variables $(\mathbf{r}, \mathbf{v})\to(\mathbf{r}+\xi, \mathbf{v}+\xi')$ on the left-hand side of (B12) and by use of (3.2) and (3.10).

Let us now demonstrate how to make use of the general current response operator formulas. A typical quantity contained in the formula for $\int\phi_0\cdot\delta J_{\mathrm{O}}^{(1)}[\phi_1]\,dtd\mathbf{r}d\mathbf{v}$ obtained from (A8) and (A10) is

$$\int dtd\mathbf{r}d\mathbf{v}\, f_{\mathrm{O}}(t,\mathbf{v})\frac{\partial}{\partial \mathbf{v}} D_{\mathrm{O}}^{-1}(\mathbf{v}\cdot\mathbf{A}_1 - U_1)\cdot\nabla(\mathbf{v}\cdot\mathbf{A}_0 - U_0). \tag{B19}$$

Applying (B12), (B11), $[T, \partial/\partial \mathbf{v}] = [T, \nabla] = 0$, and $T[\mathbf{v}\cdot\mathbf{A}_j - U_j] = \mathbf{v}\cdot\tilde{A}_j - \tilde{U}_j$ yields (B19) equal to

$$\int dtd\mathbf{r}d\mathbf{v}\, f_{\mathrm{H}}(\mathbf{v})\frac{\partial}{\partial \mathbf{v}} D_{\mathrm{H}}^{-1}(\mathbf{v}\cdot\tilde{\mathbf{A}}_1 - \tilde{U}_1)\cdot\nabla(\mathbf{v}\cdot\tilde{\mathbf{A}}_0 - \tilde{U}_0). \tag{B20}$$

This is easily seen to be a part in the formula for $\int\tilde{\phi}_0\cdot\delta J_{\mathrm{H}}^{(1)}[\tilde{\phi}_1]\,dtd\mathbf{r}d\mathbf{v}$ obtained from (A8) and (A10). The relations (B1) and (B2) follow by treating each term in a similar way.

APPENDIX C

The purpose of this appendix is to derive a relation between the S and the H systems defined in Sec. 4. We will prove that

$$\int_{z>0}\phi_0\cdot\delta J_{\mathrm{S}}^{(1)}[\phi_1]\,dtd\mathbf{r} = \frac{1}{2}\int\tilde{\phi}_0\cdot J_{\mathrm{H}}^{(1)}[\tilde{\phi}_1]dt\,d\mathbf{r}, \tag{C1}$$

$$\int_{z>0}\phi_0\cdot\delta J_{\mathrm{S}}^{(2)}[\phi_1,\phi_2]\,dtd\mathbf{r} = \frac{1}{2}\int\tilde{\phi}_0\cdot\delta J_{\mathrm{H}}^{(2)}[\tilde{\phi}_1,\tilde{\phi}_2]\,dtd\mathbf{r}, \tag{C2}$$

where ϕ_0, ϕ_1, and ϕ_2 are as in (A8) and (A9), $\phi_j = U_j e_0 + c\mathbf{A}_j$, and

$$\tilde{\phi}_j(t,\mathbf{r}) = \tilde{U}_j(t,\mathbf{r}) + c\tilde{\mathbf{A}}_j(t,\mathbf{r}), \tag{C3}$$

$$\tilde{U}_j(t,\mathbf{r}) = U_j(t,x,y,|z|), \tag{C4}$$

$$\tilde{\mathbf{A}}_j(t,\mathbf{r}) = (A_{jx}(t,x,y,|z|), A_{jy}(t,x,y,|z|), (\operatorname{sgn} z)A_{jz}(t,x,y,|z|,)), \tag{C5}$$

where $\operatorname{sgn} z = 1$ for $z\geqslant 0$ and $\operatorname{sgn} z = -1$ for $z<0$. We have also used the notation $\mathbf{A} = (A_x, A_y, A_z) = A_x e_1 + A_y e_2 + A_z e_3$.[14] We note that (4.4) and (4.5) follow from (C1)–(C4). In terms of the operator R defined by (4.2) and the linear operator K

$$K[\phi_j] = (I + (\operatorname{sgn} z - 1)e_3\otimes e_3)\cdot\phi_j, \tag{C6}$$

we can write $\tilde{\phi}_j$ as

$$\tilde{\phi}_j = K\circ R\,[\phi_j, z\geqslant 0]. \tag{C7}$$

We now obtain from (C1) and (C2) that

$$(\delta J_{\mathrm{S}}^{(1)}[\phi_1], z\geqslant 0) = R^{-1}\circ K\circ\delta J_{\mathrm{H}}^{(1)}\circ K\circ R\,[\phi_1, z\geqslant 0], \tag{C8}$$

$$(\delta J_{\mathrm{S}}^{(2)}[\phi_1,\phi_2], z\geqslant 0) = R^{-1}\circ K\circ\delta J_{\mathrm{H}}^{(2)}[K\circ R\,[\phi_1, z\geqslant 0], K\circ R\,[\phi_2, z\geqslant 0]]. \tag{C9}$$

We prove (C1) and (C2) by direct application to the H and the S systems of the explicit general formulas (A8)–(A10) and (A12) for the current response operators. We need the relation

$$D_{\mathrm{S}}^{-1} = R^{-1}\circ D_{\mathrm{H}}^{-1}\circ R. \tag{C10}$$

The operator D_{S}^{-1} may be defined in terms of the unperturbed orbits in the S system, where a particle is confined to the region $z\geqslant 0$ and bounces elastically at $z = 0$. In terms of boundary conditions, this means that, for a function $a(t, \mathbf{r}, \mathbf{v})$ defined on the half-space $z\geqslant 0$, we have $b = D_{\mathrm{S}}^{-1}[a]$ determined by the conditions $D_{\mathrm{S}} b = a$ and $b(t,x,y,0,v_x,v_y,v_z) = b(t,x,y,0,v_x,v_y,-v_z)$, i.e., $R[b]$ is continuous across the boundary. It is easy to prove (C10) if we define D_{S}^{-1} in terms of unperturbed orbits so we omit these details.

Let $a(t, \mathbf{r}, \mathbf{v})$ be defined in the half-space $z\geqslant 0$. We have the obvious equality

$$\int_{z>0} f_{\mathrm{S}} a\,dtd\mathbf{r}d\mathbf{v} = \frac{1}{2}\int f_{\mathrm{H}} R[a]\,dtd\mathbf{r}d\mathbf{v}. \tag{C11}$$

The proof of (C1) and (C2) by means of (C10), (C11), and the general formulas (A8)–(A10) and (A12) is almost identical to the corresponding proof in Appendix B, and so we omit it. Just note that in one passage of the derivation we need

$$\left(R\circ\frac{\partial}{\partial v_z}\right)[a]\cdot\left(R\circ\frac{\partial}{\partial z}\right)[b] = \left(\frac{\partial}{\partial v_z}\circ R\right)[a]\cdot\left(\frac{\partial}{\partial z}\circ R\right)[b], \tag{C12}$$

where $a(t, \mathbf{r}, \mathbf{v})$ and $b(t, \mathbf{r}, \mathbf{v})$ are defined in the half-space $z\geqslant 0$. This result follows from

$$R\circ\frac{\partial}{\partial v_z} = (\operatorname{sgn} z)\frac{\partial}{\partial v_z}\circ R,$$

$$R\circ\frac{\partial}{\partial z} = (\operatorname{sgn} z)\frac{\partial}{\partial z}\circ R. \tag{C13}$$

[1] J. Larsson, J. Math. Phys. **23**, 176 (1982), companion paper.
[2] J. Larsson, J. Math. Phys. **23**, 183 (1982), companion paper.
[3] J. Larsson, J. Math. Phys. **24**, 2222 (1983).
[4] J. Larsson, J. Math. Phys. **20**, 1321 (1979).

[5]Y. M. Aliev, V. P. Silin, and C. Watson, Zh. Eksp. Teor. Fiz. **23**, 626 (1966) [Sov. Phys. JETP **50**, 943 (1966)].
[6]V. P. Silin, Zh. Eksp. Teor. Fiz. **51**, 1842 (1966) [Sov. Phys. JETP **24**, 1242 (1967)].
[7]J. R. Sanmartin, Phys. Fluids **13**, 1533 (1970).
[8]A. A. Ivanov, *Reviews of Plasma Physics*, edited by M. A. Leontovich (Consultants Bureau, New York, 1975), Vol. 6, pp. 161–235.
[9]P. K. Kaw, *Advances in Plasma Physics*, edited by A. Simon and W. B. Thompson (Wiley, New York, 1976), Vol. 6, pp. 179–206.
[10]A. Shivarova and I. Zhelyazkov, Plasma Phys. **20**, 1049 (1978).
[11]H. C. Barr and T. J. M. Boyd, J. Phys. A: Gen. Phys. **5**, 1108 (1972).
[12]J. Larsson and L. Stenflo, Phys. Lett. A **76**, 396 (1980).
[13]A. Bers, in *Plasma Physics (Les Houches)*, edited by C. DeWitt and J. Peyrand (Gordon and Breach, New York, 1975), p. 113.
[14]In this paper we always work in some chosen coordinate system $L = (e_0, e_1, e_2, e_3)$. We write $x^0 e^0 + x^1 e_1 + x^2 e_2 + x^3 e_3 = cte_0 + xe_1 + ye_2 + ze_3$. Thus, e_1, e_2, and e_3 are the unit vectors in the x, y, and z directions, respectively. The operator Π_L stands here as in I–III for the operator which takes the spatial part of a 4-vector: $\Pi_L v = v + e_0 \cdot v e_0$.
[15]J. Larsson, J. Plasma Phys. **21**, 519 (1979).
[16]For the O-system we have $V_i = \{0\}$, $T_p = \{(2\pi cn/\Omega)e_0 | n \text{ integer}\}$, $\hat{T}_p = \{(n\Omega/c)e_0 | n \text{ integer}\}$, $V_h = \{xe_1 + ye_2 + ze_3 | x, y, \text{and } z \text{ real numbers}\}$, $V_p = \{cte_0 | t \text{ real numbers}\}$, $T = V_h + T_p$; cf. Sec. 2, A (III).
[17]A. A. Galeev and R. Z. Sagdeev, Nucl. Fusion **13**, 603 (1973).
[18]More correct notation in (3.14) would be $\delta\rho_H^{(\sigma,1)}[\gamma] = -\epsilon_0 k^2 \chi_\sigma(\omega, \mathbf{k})\gamma$, where $\gamma(t, \mathbf{r}) = \exp(-i\omega t + i\mathbf{k}\cdot\mathbf{r})$. the corresponding remark can be made concerning (3.29) and (4.6).
[19]The response functions $C^{(\sigma,n)}(\kappa_1, \kappa_2)$ are defined by $\rho^\sigma(\kappa) = C^{(\sigma,1)}(\kappa)U_\kappa + (2\pi)^{-4}\int C^{(\sigma,2)}(\kappa_1, \kappa_2)\delta(\kappa - \kappa_1 - \kappa_2)U_{\kappa_1} U_{\kappa_2} d\kappa_1 d\kappa_2 + \cdots$; where $\rho^{(\sigma)}(\kappa)$ and U_κ are the total σ-charge density response and U_κ the electrostatic perturbation in κ-space, respectively.
[20]A. G. Sitenko and V. N. Pavlenko, Zh. Eksp. Teor. Fiz. **74**, 128 (1978) [Sov. Phys. JETP **47**, 65 (1978)].
[21]T. Lindgren, J. Larsson, and L. Stenflo, Plasma Phys. **26**, 407 (1981).

Comparison of the extended boundary condition method and perturbation method for scattering from a sinusoidal sea surface[a)]

R. H. Ott and D. J. Fang
Propagation Studies Department, COMSAT Laboratories, Clarksburg, Maryland 20871

(Received 23 November 1982; accepted for publication 17 March 1983)

Scattering from a lossy sinusoidal surface is examined by using two methods: extended boundary condition (EBC) and perturbation. The range of applicability of the perturbation approach is established by comparison with EBC. The range of applicability of EBC is established through numerical examples. A convergence criterion based on the sequence of partial sums of the scattered energy is suggested for the application of EBC to lossy surfaces. New analytical expressions for the perturbation solutions are given.

PACS numbers: 92.10.Hm, 41.10.Hv

I. INTRODUCTION

In L band ship-to-satellite communication , it is frequently observed that a multipath component from the sea can occur and interfere with the direct ray between the ship and the satellite.[1,2] The multipath component does not always behave as a simple reflected wave from a plane sea surface, but rather as a random wave scattering from a multitude of scatterers of small sizes. This scattering component causes random signal fades, enhancements, and, if the signal is digitally modulated, symbol errors or intersymbol interference. It is known that the sea surface is a mixture of gravity and capillary waves, frequently crowned by breaking sprays. The sea swell has a stable dominant direction, not necessarily influenced by the wind. On the other hand, the local and composite wind field modifies the sea surface by producing short gravity and capillary waves. The interaction of these waves and the swell cause localized small perturbations of the sea surface,[3] which are believed to be primarily responsible for microwave scatterings as observed at L-band frequencies.[4,5]

As a step in modeling the microwave scattering from the slightly rough surface, this paper studies the scattering of radio waves by lossy sinusoidal surfaces for which the height (peak-to-trough) varies from zero to approximately the radio wavelength. The surface is assumed to be very lossy (i.e., $\epsilon_r = 72 + i52$ at 1.6 GHz). Waterman's formulation of the extended boundary condition (EBC)[6] is used to investigate the range of applicability and computational efficiency of the method.

EBC asserts that the field radiated by the induced surface currents cancels the incident wave below the surface. In the sense that the surface integral cancels the incident wave, EBC is similar to a perturbation solution. The range of parameters in which the two methods yield comparable results is examined here.

Recently, Chuang and Kong used EBC with several choices for the spanning set that represented the surface fields.[7] Their paper also contains an extensive literature survey of this important boundary value problem.

In our study, a spanning set that is complete on the extended surface, rather than on the water surface, is used for the surface field. The convergence of this set is examined by noting that the partial sums for the scattered energy increases monotonically with the number of terms, are bounded above by unity, and must therefore converge. This convergence criterion is used because conservation of energy, which is only a necessary condition, is not applicable in the EBC method when the scatterer is lossy. The reason conservation of energy is not applicable is that the solution is only valid outside the region formed by crest and trough, and for a lossy surface, energy is dissipated between crest and trough. Also, checking the accuracy of the solution in satisfying the boundary conditions is not an option with EBC, because the EBC solution is not valid on the actual water surface.

Some new analytical results are given for the coefficients in the perturbation solution in terms of Bessel functions. The argument of the Bessel functions is the height of the surface in wavelengths projected onto the plane perpendicular to the direction of the incident wave vector. A three-significant-figure agreement is obtained between the coefficients in the EBC method and those in the perturbation method when the height of the surface is less than 0.1λ, where λ is the radio wavelength. When the surface height reaches 0.2λ, the corresponding amplitudes of the specular component calculated by the two methods differ by about 30%.

II. EBC ANALYSIS

The formulation of EBC in this section follows Waterman's presentation[6] (see Fig. 1).

From Green's theorem, for an observation point $\mathbf{r}$ below $z = \sigma(x)$, integration over the upper hemisphere yields

$$0 = \psi_i(\mathbf{r}) - \int_{\sigma(x')} dx' [G(k_0 R)\mathbf{e}_n \cdot \nabla' \psi_+(\mathbf{r}') - \psi_+(\mathbf{r}')\mathbf{e}_n \cdot \nabla' G(k_0 R)] \quad (1)$$

and for $\mathbf{r}$ above $z = \sigma(x)$, integration over the lower hemisphere yields

$$0 = \int_{\sigma(x')} dx' [G(KR)\mathbf{e}_n \cdot \nabla' \psi_-(\mathbf{r}') - \psi_-(\mathbf{r}')\mathbf{e}_n \cdot \nabla' G(KR)], \quad (2)$$

where

[a)] This paper is based upon work performed at COMSAT Laboratories under the sponsorship of Communications Satellite Corporation.

 0022-2488/83/082240-10$02.50

$$\mathbf{e}_n \cong l_0 K_0 \sin(K_0 x)\mathbf{e}_x + \mathbf{e}_z, \tag{3}$$

$K_0 = 2\pi/\Lambda$ and $R = |\mathbf{r} - \mathbf{r}'|$. The integrations in Eqs. (1) and (2) can be thought of as extended operators in the sense

$$\int_{\sigma(x)} dx' [G\mathbf{e}_n \cdot \nabla' \psi_+ - \psi_+ \mathbf{e}_n \cdot \nabla' G]$$

because the chosen spanning functions for ψ_+ do not satisfy orthogonality conditions on the surface $z = \sigma(x)$ but, instead, on the extended surfaces $z = \pm l_0$, shown in Fig. 1. The Green's function for the boundary value problem is stated as

$$G\left(\begin{Bmatrix} k_0 \\ K \end{Bmatrix} |\mathbf{r} - \mathbf{r}'|\right) = \frac{\exp[ik_0(x - x')\sin\theta]}{2i\Lambda} \times \sum_{m=-\infty}^{\infty} \frac{\exp\left[i(2\pi m/\Lambda)(x - x') \pm i \begin{Bmatrix} \gamma_m \\ \hat{\gamma}_m \end{Bmatrix} (z - z')\right]}{\begin{Bmatrix} \gamma_m \\ \hat{\gamma}_m \end{Bmatrix}}, \tag{4}$$

where the plus sign for the z dependence correponds to up-going waves and the minus sign, to down-going waves.[8]

An alternative representation to Eq. (4) in terms of Hankel functions of order zero is given in the form of the following theorem:

Spherical waves have the Fourier representation

$$\frac{\exp[ik_0(x^2 + z^2)^{1/2}]}{(x^2 + z^2)^{1/2}} e^{ik_0 x \sin\theta} \sum_{n=-\infty}^{\infty} e^{i2n\pi x/\Lambda} H_0^{(1)}(k_0 z \cos\phi_n), \tag{5}$$

where

$$\cos\phi_n = [1 - ((2n\pi/k_0\Lambda) + \sin\theta)^2]^{1/2}. \tag{6}$$

Proof: Consider the Fourier expansion of the spherical wave

$$\frac{\exp[ik_0(x^2 + z^2)^{1/2}]}{(x^2 + z^2)^{1/2}} = e^{ik_0 x \sin\theta} \sum_{n=-\infty}^{\infty} a_n e^{i2n\pi x/\Lambda}$$

with coefficients given by

$$a_n = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\exp(-i\alpha_n u)\exp(ik_0(z^2 + u^2)^{1/2})}{(z^2 + u^2)^{1/2}} du,$$

where

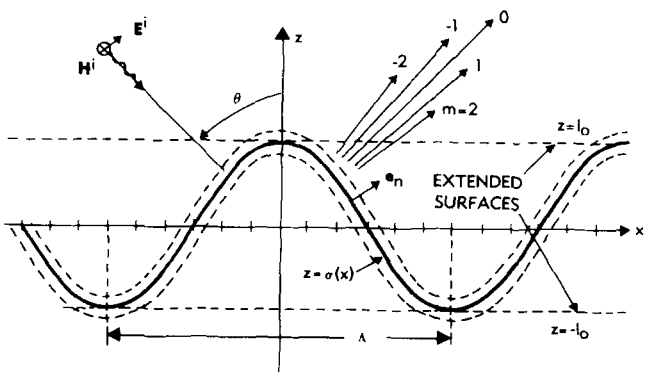


FIG. 1. Geometry for scattering from a sinusoidal surface.

$$\alpha_n = 2n\pi/\Lambda + k_0 \sin\theta.$$

Let $u = z \sinh\theta$, $du = z\cosh\theta\, d\theta$; then

$$a_n = \tfrac{1}{2} i H_0^{(1)}[z(k_0^2 - \alpha_n^2)^{1/2}] = \tfrac{1}{2} i H_0^{(1)}(k_0 z \cos\phi_n).$$

The following results are consequences of Eq. (4):

$$K^2 = \hat{\gamma}_m^2 + k_x^2, \tag{7a}$$

$$k_0^2 = \gamma_m^2 + k_x^2, \tag{7b}$$

$$\hat{\gamma}_m^2 = K^2 - (2\pi m/\Lambda + \sin\theta)^2 = K^2 \cos^2\hat{\phi}_m, \tag{7c}$$

$$\hat{\gamma}_m^2 = K\cos\hat{\phi}_m = k_0[\epsilon_r - (2\pi m/\Lambda k_0 + \sin\theta)^2]^{1/2}, \tag{7d}$$

$$\mathrm{Im}(\hat{\gamma}_m) > 0 \text{ for } \mathrm{Im}(\epsilon_r) > 0, \tag{7e}$$

$$\gamma_m^2 = k_0^2 - k_x^2 = k_0^2 - (2\pi m/\Lambda + \sin\theta)^2 = k_0^2 \cos^2\phi_m, \tag{7f}$$

$$\cos\phi_m = k_0 \begin{cases} \left[1 - \left(\frac{2\pi m}{\Lambda} + \sin\theta\right)^2\right]^{1/2}, & \frac{2\pi m}{k_0\Lambda} + \sin\theta \leqslant 1, \\ i\left[\left(\frac{2\pi m}{k_0\Lambda} + \sin\theta\right)^2 - 1\right]^{1/2}, & \frac{2\pi m}{k_0\Lambda} + \sin\theta > 1, \end{cases} \tag{7g}$$

$$k_0 \sin\phi_m = 2\pi m/\Lambda + k_0 \sin\theta, \tag{7h}$$

$$K \sin\hat{\phi}_m = 2\pi m/\Lambda + k_0 \sin\theta. \tag{7i}$$

The boundary conditions for TM polarization (ψ represents H_y in Fig. 1) are

$$\psi_+(\mathbf{r}') = \psi_-(\mathbf{r}'), \tag{8}$$

$$\mathbf{e}_n \cdot \nabla' \psi_+(\mathbf{r}') = (1/\epsilon_r)\mathbf{e}_n \cdot \nabla' \psi_-(\mathbf{r}'), \tag{9}$$

where in Eqs. (8) and (9), the + and − subscripts correspond to a representation valid just above and below the surface $z = \sigma(x)$, respectively, and the prime on the gradient operator refers to differentiation with respect to the primed coordinates.

The spanning functions

$$\psi_n(x) = \exp(ik_0 x \sin\phi_n) \tag{10}$$

are orthogonal on the extended surfaces $z = \pm l_0$, but not on the surface $z = \sigma(x)$, where the extended operator is defined. Therefore, if another vector ψ_{n+1} is added to the approximating set of vectors, the previously computed set of coefficients may change. For the surface fields,

$$\psi_+(\mathbf{r}') = 2 \sum_{n=-\infty}^{\infty} \alpha_n e^{ik_0 x' \sin\phi_n}, \tag{11}$$

$$\mathbf{e}_n \cdot \nabla' \psi_+(\mathbf{r}') = 2ik_0 \sum_{n=-\infty}^{\infty} \beta_n e^{ik_0 x' \sin\phi_n}. \tag{12}$$

The incident field in Eq. (1) is

$$\psi_i(\mathbf{r}) = \exp[ik_0(x \sin\theta - z\cos\theta)]/(\cos\theta)^{1/2}. \tag{13}$$

The following results are needed for down-going waves:

$$\nabla' G(k_0 R) = (-i\mathbf{e}_x k_0 \sin\phi_m + i\gamma_m \mathbf{e}_z) G(k_0 R) \tag{14}$$

and

$$\nabla' G(KR) = (-i\mathbf{e}_x K \sin\hat{\phi}_m - i\hat{\gamma}_m \mathbf{e}_z) G(KR) \tag{15}$$

for up-going waves.

Substituting Eqs. (4), (7), (8), (9), (11), (12), and (14) into (1) gives

$$\frac{\exp[ik_0(x\sin\theta - z\cos\theta)]}{(\cos\theta)^{1/2}}$$

$$= \frac{k_0}{\Lambda}\sum_{m=-\infty}^{\infty}\exp(ik_0 x\sin\phi_m - iz\gamma_m)\left\{\sum_{n=-\infty}^{\infty}\beta_n\int_{-\Lambda/2}^{\Lambda/2}\frac{dx'}{\gamma_m^{1/2}}\exp[i(2\pi x'/\lambda)(n-m) + ik_0 l_0\cos(K_0x')\cos\phi_m]\right.$$

$$\left. + \sum_{n=-\infty}^{\infty}\alpha_n\int_{-\Lambda/2}^{\Lambda/2}\frac{dx'}{\gamma_m^{1/2}}[-l_0K_0\sin(K_0x')\sin\phi_m + \cos\phi_m]\exp[i(2\pi x'/\Lambda)(n-m) + ik_0l_0\cos(K_0x')\cos\phi_m]\right\}$$

$$= \sum_{m=-\infty}^{\infty} a_m\frac{\exp(ik_0x\sin\phi_m + iz\gamma_m)}{(\cos\phi_m)^{1/2}}, \tag{16}$$

where the last equation defines a_m. The following T-matrix coefficients are generated from the integrations in Eq. (16):

$$\Gamma^{(1)}_{mn} = \frac{i^{|n-m|}}{(\cos\phi_m)^{1/2}}J_{|n-m|}(k_0l_0\cos\phi_m) \quad \text{and} \quad \Gamma^{(2)}_{mn} = \frac{(-1+\sin\phi_m\sin\phi_n)}{\cos\phi_m}\Gamma^{(1)}_{mn}. \tag{17}$$

Substituting Eqs. (4), (7), (8), (9), (11), (12), and (13) into Eq. (2) gives

$$0 = \frac{k_0\epsilon_r}{\Lambda}\sum_{m=-\infty}^{\infty}\frac{\exp(iKx\sin\hat\phi_m + i\hat\gamma_m z)}{\hat\gamma_m^{1/2}}$$

$$\times\left\{\sum_{n=-\infty}^{\infty}\beta_n\int_{-\Lambda/2}^{\Lambda/2}\frac{dx'}{\hat\gamma_m^{1/2}}\exp\{-i[2\pi(m-n)x'/\Lambda] - i\sqrt{\epsilon_r}k_0\cos\hat\phi_m l_0\cos(K_0x')\} - \frac{1}{\sqrt{\epsilon_r}}\sum_{n=-\infty}^{\infty}\alpha_n\int_{-\Lambda/2}^{\Lambda/2}\frac{dx'}{\hat\gamma_m^{1/2}}\right.$$

$$\left.\times[-l_0K_0\sin(K_0x')\sin\hat\phi_m - \cos\hat\phi_m]\exp\{i[2\pi(m-n)x'/\Lambda] - i\sqrt{\epsilon_r}k_0l_0\cos\hat\phi_m\cos(K_0x')\}\right\} \tag{18}$$

$$= \sum_{m=-\infty}^{\infty} b_m\frac{\exp(iKx\sin\hat\phi_m + iz\hat\gamma_m)}{(\cos\hat\phi_m)^{1/2}},$$

where the last equation defines b_m. The T-matrix coefficients generated from the integrations in Eq. (18) are

$$\Gamma^{(3)}_{mn} = \frac{(-i)^{|m-n|}}{(\cos\hat\phi_m)^{1/2}}J_{|n-m|}(\sqrt{\epsilon_r}k_0l_0\cos\hat\phi_m) \tag{19a}$$

and

$$\Gamma^{(4)}_{mn} = \frac{(-1+\sin\hat\phi_m\sin\hat\phi_n)}{\sqrt{\epsilon_r}\cos\hat\phi_m}\Gamma^{(3)}_{mn}. \tag{19b}$$

Note from Eqs. (17) and (19) the "near orthogonality" of the spanning functions in Eq. (10) on the actual surface $z = \sigma(x)$. That is, for $n = m$, the Bessel functions are "near" zero. Also, including the z dependence in the basis functions in Eq. (10) would not yield an orthogonal set on the surface $z = \sigma(x)$. An additional reason for not including the z dependence in the basis functions is the ambiguity in sign [i.e., $\exp(\pm ik_0z\cos\phi_n)$] as the boundary $z = \sigma(x)$ is crossed. There exists a jump in the z dependence. A spanning set consisting of up-going and down-going waves in the z direction of the form

$$\psi_+(\mathbf{r}') = \sum_{n=-\infty}^{\infty} e^{ik_0x'\sin\phi_n}\times\{\alpha_n e^{ik_0z\cos\phi_n} + \beta_n e^{-ik_0z\cos\phi_n}\},$$

when substituted into Eq. (1), yields zero on the right-hand side of Eq. (16) unless $n = m$; which in turn implies $\beta_n = -1, \alpha_n = 0$. This result is interesting but does not give a consistent solution. The variables z and x are not independent on the surface $z = \sigma(x)$.

The numerical implementation for the Bessel functions in Eqs. (17) and (19) is

$$J_n(z) = i^n I_n(-iz), \quad -\pi < \arg(z) \leqslant \pi/2, \tag{20}$$

where $I_n(z)$ is a modified Bessel function of the first kind. The I_n's were calculated by using recursion relations.

From Eq. (16), the only nonzero a_m is for $m = 0$; i.e.,

$$a_m = \begin{cases}1, & m = 0,\\ 0, & m \neq 0,\end{cases} \tag{21}$$

and similarly from Eq. (18), all $b_m = 0$. These two results yield the following $2M+1$ equations in $2M+1$ unknowns:

$$\underbrace{[\Gamma^{(1)}_{mn}]}_{(2M+1)\times(2M+1)}\cdot\underbrace{[\beta_n]}_{(2M+1)\times 1} + \underbrace{[\Gamma^{(2)}_{mn}]}_{(2M+1)\times(2M+1)}\cdot\underbrace{[\alpha_n]}_{(2M+1)\times 1}$$

$$= \begin{bmatrix}0\\ \vdots\\ 0\\ 1\\ 0\\ \vdots\\ 0\end{bmatrix}\begin{matrix}(-m)\\ \\ \\ (0)\\ \\ \\ (m)\end{matrix}, \tag{22}$$

$$[\Gamma^{(3)}_{mn}][\beta_n] - [\Gamma^{(4)}_{mn}][\alpha_n] = \begin{bmatrix}0\\ \vdots\\ 0\\ \vdots\\ 0\end{bmatrix}\begin{matrix}(-m)\\ \\ 0\\ \\ (m)\end{matrix}. \tag{23}$$

At this point, EBC is identical with point matching where the weighting functions are delta functions or Galekin's method where the set of weighting functions coincides with the spanning set. The inner product of both sides of Eqs. (16) and (18) is taken with respect to the functions in the spanning set.

The scattered vector is obtained from Eq. (1) by taking the observation point above the surface $z = \sigma(x)$. This requires up-going waves in the Green's function in Eq. (4). The scattered field is

$$\psi^s(\mathbf{r}) = \sum_{m=-\infty}^{\infty} \hat{a}_m \frac{\exp(ik_0 x \sin\phi_m + iz\gamma_m)}{(\cos\phi_m)^{1/2}}, \tag{24a}$$

with

$$\hat{a}_m = -\frac{k_0}{\Lambda(\cos\phi_m)^{1/2}} \sum_{n=-\infty}^{\infty} \left\{ \beta_n \int_{-\Lambda/2}^{\Lambda/2} dx' \exp\{i[2\pi x'(m-n)/\Lambda] - ik_0 l_0 \cos(K_0 x')\cos\phi_m\} + \alpha_n \int_{-\Lambda/2}^{\Lambda/2} dx' [-l_0 K_0 \sin(K_0 x')\sin\phi_m - \cos\phi_m] \exp\{i[2\pi x'(m-n)/\Lambda] - ik_0 l_0 \cos(K_0 x')\cos\phi_m\} \right\}. \tag{24b}$$

The T-matrix coefficients generated from the integrations in Eq. (24) are

$$\hat{\Gamma}^{(1)}_{mn} = \frac{(i)^{|n-m|}}{(\cos\phi_m)^{1/2}} J_{|n-m|}(k_0 l_0 \cos\phi_m), \quad \hat{\Gamma}^{(2)}_{mn} = \frac{(1 - \sin\phi_m \sin\phi_n)}{\cos\phi_m} \hat{\Gamma}^{(1)}_{mn}. \tag{25}$$

Similarly, the transmitted vector is obtained from Eq. (2) by taking the observation point below the surface $z = \sigma(x)$. This requires down-going waves in the Green's function in Eq. (4). The transmitted field is

$$\psi^T(\mathbf{r}) = \sum_{m=-\infty}^{\infty} \hat{b}_m \frac{\exp(ikx \sin\hat{\phi}_m - iz\hat{\gamma}_m)}{(\cos\hat{\phi}_m)^{1/2}}, \tag{26a}$$

with

$$\hat{b}_m = \frac{K_0}{\Lambda(\cos\hat{\phi}_m)^{1/2}} \sum_{n=-\infty}^{\infty} \left\{ \beta_n \int_{-\Lambda/2}^{\Lambda/2} dx' \{\exp[i(2\pi x'/\Lambda)(n-m) + i\sqrt{\epsilon_r} k_0 l_0 \cos(K_0 x')\cos\hat{\phi}_m]\} - \alpha_n \int_{-\Lambda/2}^{\Lambda/2} [-l_0 k_0 \sin(K_0 x')\sin\hat{\phi}_m + \cos\hat{\phi}_m] \exp[i(2\pi x'/\Lambda)(n-m) + i\sqrt{\epsilon_r} k_0 l_0 \cos(K_0 x')\cos\phi_m] \right\}. \tag{26b}$$

The following T-matrix coefficients are generated from the integrations in Eq. (26):

$$\hat{\Gamma}^{(3)}_{mn} = \frac{i^{|n-m|}}{(\cos\hat{\phi}_m)^{1/2}} J_{|n-m|}(\sqrt{\epsilon_r} k_0 l_0 \cos\hat{\phi}_m), \tag{27a}$$

$$\hat{\Gamma}^{(4)}_{mn} = \frac{(1 - \sin\hat{\phi}_m \sin\hat{\phi}_n)}{\sqrt{\epsilon_r}\cos\hat{\phi}_m} \hat{\Gamma}^{(3)}_{mn}. \tag{27b}$$

The scattered and transmitted vectors can be writeen in matrix notation as

$$\begin{aligned} [\hat{a}_m] &= -([\hat{\Gamma}^{(1)}_{mn}][\beta_n] + [\hat{\Gamma}^{(2)}_{mn}][\alpha_n]), \\ [\hat{b}_m] &= [\hat{\Gamma}^{(3)}_{mn}][\beta_n] - [\hat{\Gamma}^{(4)}_{mn}][\alpha_n]. \end{aligned} \tag{28}$$

It is desired to approximate the scattered and transmitted vectors in terms of the linear independent sets in Eqs. (24) and (26). The best approximation is considered to be the linear combination

$$\sum_{m=-M}^{M} a_m \phi^s_m(x,z),$$

which is closest to $\psi^s(\mathbf{r})$; that is, for which the quantity

$$\left|\left| \psi^s(\mathbf{r}) - \sum_{m=-M}^{M} a_m \frac{\exp(ik_0 x \sin\phi_m + iz\gamma_m)}{(\cos\phi_m)^{1/2}} \right|\right|^2$$

is the smallest. The problem is that ψ^s is unknown. However, the functions

$$\exp(ik_0 x \sin\phi_m + iz\gamma_m)/(\cos\phi_m)^{1/2}$$

are orthogonal on the extended surfaces $z = \pm l_0$. For any finite M, integrating on the surface $z = l_0$ yields

$$\left|\left| \psi^s - \sum_{m=-M}^{M} a_m \frac{\exp(ik_0 x \sin\phi_m + iz\gamma_m)}{(\cos\phi_m)^{1/2}} \right|\right|^2 = \|\psi^s\|^2 - \sum_{m=-M}^{M} |\hat{a}_m|^2 \geqslant 0,$$

which implies

$$\sum_{m=-M}^{M} |a_m|^2 \leqslant \|\psi^s\|^2 \leqslant 1. \tag{29}$$

The infinite series

$$\sum_{-\infty}^{\infty} |a_m|^2$$

is the limit of a sequence of partial sums that increase monotonically with M and are bounded above by the unknown number $\|\psi^s\|^2$. The infinite series must, therefore, converge. Now $\|\psi^s\|^2 \leqslant 1$; so the partial sum

$$\sum_{-M}^{M} |a_m|^2$$

is taken until the partial sums in the sequence change by less than a prescribed amount in a given significant figure. Experience with several numerical examples has shown that the test in Eq. (29) gives an accurate convergence criterion. Because the series converges, it follows that $\lim_{m\to\infty} a_m = 0$ (the Riemann–Lebesque lemma), and this fact is consistent with numerical results shown later.

III. PERTURBATION ANALYSIS

The incident, scattered, and transmitted fields for TM polarization are

$$\mathbf{E}^i = (\mathbf{e}_x k_0 \cos\phi + \mathbf{e}_z k_0 \sin\theta) \times \exp[ik_0(x\sin\phi - z\cos\theta)] \tag{30a}$$

$$\mathbf{H}^i = -(k_0/Z_0)\mathbf{e}_y \exp[ik_0(x\sin\theta - z\cos\theta)] \tag{30b}$$

$$\mathbf{E}^s = k_0 \sum_{m=-\infty}^{\infty} a_m(-\mathbf{e}_x\cos\phi_m + \mathbf{e}_z\sin\phi_m) \times \exp[ik_0(x\sin\phi_m + z\cos\phi_m)] \tag{30c}$$

$$\mathbf{H}^s = -\frac{k_0}{Z_0}\mathbf{e}_y \times \sum_{m=-\infty}^{\infty} a_m \frac{\exp[ik_0(x\sin\phi_m + z\cos\phi_m)]}{(\cos\phi_m)^{1/2}} \tag{30d}$$

$$\mathbf{E}^T = k_0\sqrt{\epsilon_r} \sum_{m=-\infty}^{\infty} b_m(\mathbf{e}_x\cos\hat{\phi}_m + \mathbf{e}_z\sin\hat{\phi}_m) \times \frac{\exp[ik_0\sqrt{\epsilon_r}(x\sin\hat{\phi}_m - z\cos\hat{\phi}_m)]}{(\cos\hat{\phi}_m)^{1/2}} \tag{30e}$$

$$\mathbf{H}^T = -\frac{k_0\epsilon_r}{Z_0}\mathbf{e}_y \times \sum_{m=-\infty}^{\infty} b_m \frac{\exp[ik_0\sqrt{\epsilon_r}(x\sin\hat{\phi}_m - z\cos\hat{\phi}_m)]}{(\cos\hat{\phi}_m)^{1/2}}. \tag{30f}$$

The similarity to Eqs. (24) and (26) is evident, and only the method for solving for the Fourier coefficients a_m and b_m differs from the EBC method. The boundary condition [Eq. (8)] and the Rayleigh hypothesis gives

$$\frac{\exp[ik_0(x\sin\theta - z\cos\theta)]}{(\cos\theta)^{1/2}} + \sum a_m \frac{\exp[ik_0(x\sin\phi_m + z\cos\phi_m)]}{(\cos\phi_m)^{1/2}} = \epsilon_r \sum b_m \frac{\exp[ik_0\sqrt{\epsilon_r}(x\sin\hat{\phi}_m - z\cos\hat{\phi}_m)]}{(\cos\hat{\phi}_m)^{1/2}} \tag{31}$$

and after some algebra, the boundary condition (9) gives

$$[-l_0k_0\sin(K_0x)\sin\theta + \cos\theta]\frac{\exp[-ik_0l_0\cos\theta\cos(K_0x)]}{(\cos\theta)^{1/2}} - \sum a_m[l_0k_0\sin(K_0x)\sin\phi_m + \cos\phi_m]\frac{\exp[i(2m\pi x/\Lambda) + ik_0l_0\cos\phi_m\cos(K_0x)]}{(\cos\theta_m)^{1/2}} = \sqrt{\epsilon_r}\sum b_m[-l_0k_0\sin(K_0x)\sin\hat{\phi}_m + \cos\hat{\phi}_m]\frac{\exp[i(2m\pi x/\Lambda) - ik_0l_0\sqrt{\epsilon_r}\cos\hat{\phi}_m\cos(K_0x)]}{(\cos\hat{\phi}_m)^{1/2}}. \tag{32}$$

At this point, this study departs somewhat from classical perturbation analysis[5] and uses the orthogonality of the spanning functions

$$e^{-i2n\pi x/\Lambda}$$

over the interval $|x| < \Lambda/2$. That is, both sides of Eqs. (31) and (32) are multiplied by the spanning functions in Eq. (10), and the following table of integrals is used:

$$\int_{-\Lambda/2}^{\Lambda/2} \exp[i(2\pi x/\Lambda)(m-n) + ik_0l_0\cos\phi_m\cos(K_0x)]dx = \Lambda i^{|m-n|}J_{|m-n|}(k_0l_0\cos\phi_m),$$

$$\int_{-\Lambda/2}^{\Lambda/2} \exp[i(2\pi x/\Lambda)(m-n) - ik_0l_0\sqrt{\epsilon_r}\cos\phi_m\cos(K_0x)]dx = \Lambda(-i)^{|m-n|}J_{|m-n|}(k_0l_0\sqrt{\epsilon_r}\cos\hat{\phi}_m),$$

$$\int_{-\Lambda/2}^{\Lambda/2} l_0k_0\sin(K_0x)\exp[i(2\pi x/\Lambda)(m-n) + ik_0l_0\cos(K_0x)]dx = \Lambda\frac{(\sin\phi_m - \sin\phi_n)}{\cos\phi_m}i^{|m-n|}J_{|m-n|}(k_0l_0\cos\phi_m),$$

$$\int_{-\Lambda/2}^{\Lambda/2} l_0k_0\sin(K_0x)\exp[i(2\pi x/\Lambda)(m-n) - ik_0l_0\sqrt{\epsilon_r}\cos\hat{\phi}_m\cos(K_0x)]dx = \Lambda\frac{(\sin\hat{\phi}_n - \sin\hat{\phi}_m)}{\cos\hat{\phi}_m}(-i)^{|m-n|}J_{|m-n|}(k_0l_0\sqrt{\epsilon_r}\cos\hat{\phi}_m). \tag{33}$$

Next, the ratio of surface height to slope is assumed to be small; i.e.,

$$\epsilon = \lambda/\Lambda < 0.1. \tag{34}$$

Substituting Eq. (33) into Eqs. (31) and (32) and combining the two equations yield a factor on the right-hand side that is approximated as (for $|\epsilon_r| \gg 1$)

$$\frac{1 - \sin\hat{\phi}_m\sin\hat{\phi}_n}{\cos\hat{\phi}_m} \cong \frac{(\epsilon_r - \sin^2\theta)^{1/2}}{\sqrt{\epsilon_r}}. \tag{35}$$

The approximation in Eq. (35) is tight because $|\epsilon_r|$ is on the order of 100 in the case of sea water at 1.6 GHz.

Substituting Eqs. (33) and (35) into Eqs. (31) and (32) yields

$$\sum_{m=-M}^{M} a_m \left((\epsilon_r - \sin^2\theta)^{1/2} + \frac{\epsilon_r(1 - \sin\phi_m \sin\phi_n)}{\cos\phi_m} \right) \frac{i^{|m-n|} J_{|m-n|}(k_0 l_0 \cos\phi_m)}{(\cos\phi_m)^{1/2}}$$

$$= [\epsilon_r(\cos\theta - n\epsilon \tan\theta) - (\epsilon_r - \sin^2\theta)^{1/2}](i)^{|n|} \frac{J_{|n|}(k_0 l_0 \cos\theta)}{(\cos\theta)^{1/2}}, \tag{36}$$

$$n = 0, \pm 1, \pm 2, \ldots, \pm M$$

These equations can now be solved in a perturbation manner. If $n = 0$, there is one equation in one unknown, and the zeroth order result for a_0 is

$$a_0^{(0)} = \frac{\epsilon_r \cos\theta - (\epsilon_r - \sin^2\theta)^{1/2}}{\epsilon_r \cos\theta + (\epsilon_r - \sin^2\theta)^{1/2}}, \tag{37}$$

which is the Fresnel reflection coefficient.

For $n = \pm 1$, there are three equations in three unknowns, which give the first- and second-order correction to the coefficients a_m. These equations are

$$a_{-1}^{(1)} [(\epsilon_r - \sin^2\theta)^{1/2} + \epsilon_r \cos\phi_{-1}] \frac{J_0(k_0 l_0 \cos\phi_{-1})}{(\cos\phi_{-1})^{1/2}} + a_0^{(1)} \left[(\epsilon_r - \sin^2\theta)^{1/2} + \frac{\epsilon_r(1 - \sin\theta \sin\phi_{-1})}{\cos\phi_{-1}} \right] \frac{iJ_1(k_0 l_0 \cos\theta)}{(\cos\theta)^{1/2}}$$

$$+ O = [\epsilon_r(\cos\theta + \epsilon \tan\theta) - (\epsilon_r - \sin^2\theta)^{1/2}] \frac{(-i)J_1(k_0 l_0 \cos\theta)}{(\cos\theta)^{1/2}}$$

$$a_{-1}^{(1)} \left[(\epsilon_r - \sin^2\theta)^{1/2} + \frac{\epsilon_r(1 - \sin\phi_1 \sin\theta)}{\cos\theta} \right] \frac{iJ_1(k_0 l_0 \cos\phi_{-1})}{(\cos\phi_{-1})^{1/2}} + a_0^{(1)} \left[(\epsilon_r - \sin^2\theta)^{1/2} + \epsilon_r \cos\theta \frac{J_0(k_0 l_0 \cos\theta)}{(\cos\theta)^{1/2}} \right]$$

$$+ a_1^{(1)} \left[(\epsilon_r - \sin^2\theta)^{1/2} + \frac{\epsilon_r(1 - \sin\theta \sin\phi_1)}{\cos\theta} \right] \frac{iJ_1(k_0 l_0 \cos\phi_1)}{(\cos\phi_1)^{1/2}} = [\epsilon_r \cos\theta - (\epsilon_r - \sin^2\theta)^{1/2}] \frac{J_0(k_0 l_0 \cos\theta)}{(\cos\theta)^{1/2}}$$

$$O + a_0^{(1)} \left[(\epsilon_r - \sin^2\theta)^{1/2} + \frac{\epsilon_r(1 - \sin\theta \sin\phi_1)}{\cos\phi_1} \right] \frac{iJ_1(k_0 l_0 \cos\theta)}{(\cos\theta)^{1/2}} + a_1^{(1)} [(\epsilon_r - \sin^2\theta) + \epsilon_r \cos\phi_1] \frac{J_0(k_0 l_0 \cos\phi_1)}{(\cos\phi_1)^{1/2}}$$

$$= [\epsilon_r(\cos\theta - \epsilon \tan\theta) - (\epsilon_r - \sin^2\theta)^{1/2}] \frac{(-i)J_1(k_0 l_0 \cos\theta)}{(\cos)^{1/2}} \tag{38}$$

The determinant of the matrix involved in Eq. (38) is

$$\Delta \cong [(\epsilon_r - \sin^2\theta)^{1/2} + \epsilon_r \cos\phi_{-1}][(\epsilon_r - \sin^2\theta)^{1/2} + \epsilon_r \cos\theta][(\epsilon_r - \sin^2\theta)^{1/2} + \epsilon_r \cos\phi_1]$$
$$\times \frac{J_0(k_0 l_0 \cos\phi_{-1})}{(\cos\phi_{-1})^{1/2}} \frac{J_0(k_0 l_0 \cos\theta)}{(\cos\theta)^{1/2}} \frac{J_0(k_0 l_0 \cos\phi_1)}{(\cos\phi_1)^{1/2}}. \tag{39}$$

From Eq. (38), new results are obtained (which in spirit agree with those of Rice,[4] Wait,[9] and Fang.[10] To first order,

$$a_{-1}^{(1)} = (-i) \frac{\{2[\epsilon_r \cos\theta - (\epsilon_r - \sin^2\theta)^{1/2}] + \epsilon_r \epsilon \tan\theta\}}{[(\epsilon_r - \sin^2\theta)^{1/2} + \epsilon_r \cos\phi_{-1}]}$$
$$\times \left(\frac{\cos\phi_{-1}}{\cos\theta} \right)^{1/2} \frac{J_1(k_0 l_0 \cos\theta)}{J_0(k_0 l_0 \cos\phi_{-1})} \tag{40}$$

and

$$a_1^{(1)} = (-i) \frac{\{2[\epsilon_r \cos\theta - (\epsilon_r - \sin^2\theta)^{1/2}] - \epsilon_r \epsilon \tan\theta\}}{[\epsilon_r \cos\theta + (\epsilon_r - \sin^2\theta)^{1/2}]}$$
$$\times \left(\frac{\cos\phi_1}{\cos\theta} \right)^{1/2} \frac{J_1(k_0 l_0 \cos\theta)}{J_0(k_0 l_0 \cos\phi_1)}. \tag{41}$$

To second order, a_0 is

$$a_0 \cong a_0^{(0)} + a_0^{(2)} = a_0^{(0)}$$
$$\times \frac{\left[1 - \frac{J_1(k_0 l_0 \cos\theta)}{J_0(k_0 l_0 \cos\theta)} \frac{J_1(k_0 l_0 \cos\phi_{-1})}{J_0(k_0 l_0 \cos\phi_{-1})} \right]}{\left[1 + \frac{J_1(k_0 l_0 \cos\theta)}{J_0(k_0 l_0 \cos\theta)} \frac{J_1(k_0 l_0 \cos\phi_{-1})}{J_0(k_0 l_0 \cos\phi_{-1})} \right]} \tag{42}$$

which is a new result. To second order, the determinant for three equations in three unknowns is

$$\Delta = [(\epsilon_r - \sin^2\theta)^{1/2} + \epsilon_r \cos\phi_{-1}]$$
$$\times [(\epsilon_r - \sin^2\theta)^{1/2} + \epsilon_r \cos\theta]$$
$$\times [(\epsilon_r - \sin^2\theta)^{1/2} + \epsilon_r \cos\phi_1] \frac{J_0(k_0 l_0 \cos\phi_{-1})}{\cos\phi_{-1}}$$
$$\times \left\{ 1 + \frac{J_1(k_0 l_0 \cos\theta)}{J_0(k_0 l_0 \cos\theta)} \frac{J_1(k_0 l_0 \cos\phi_{-1})}{J_0(k_0 l_0 \cos\phi_{-1})} \right\}. \tag{43}$$

This process could be continued to yield the higher order nonzeros coefficients $a_{\pm 2}^{(2)}, a_{\pm 1}^{(3)}, a_0^{(4)}, a_{\pm 2}^{(4)}, a_{\pm 4}^{(4)}, \ldots$, [namely, all the $a_m^{(p)}$ are zero except those in the set $a_m^{(p)} | p = 0,1,2,3,\ldots; m = \pm p, \pm(p-2), \pm(p-4), \ldots, \pm \mathrm{mod}(p,2)\}$.[11]

IV. NUMERICAL RESULTS

Examples will be shown in which the number of modes is increased for a given set of surface parameters, the surface height in wavelengths is varied (all other parameters constant), and the surface period is varied (all other parameters constant). Finally, the EBC results will be compared with the perturbation results where applicable.

Before numerical results are presented, some information will now be given about the computer used to generate the numerical results. The assembler architecture is IBM 370 (third generation) with a 3350 storage device (i.e., 572 070 bytes per cylinder, 1024 bytes per record). The 8-byte word length option was used, providing about 14 significant figures. (Early calculations showed that the single word length of 4 bytes, or about 6.2 significant figures, was not adequate for the large matrices in this problem.)

Figure 2 shows an interesting phenomenon that occurs when the chosen truncation interval of the Fourier representation for the surface fields is not equal to a multiple of the period of the sinusoid. In Fig. 2, the magnitude of the scattered vectors, a_m, is plotted as a function of mode order m for eight different choices of the total number of basis functions, $2M + 1$. The parameters for the example are $l_0 \cong \lambda/3$, $\Lambda \cong 67\lambda$, $K_0 \cong 0.5\ \mathrm{m}^{-1}$, $\theta = 10°$, and $\epsilon_r \cong 72 + i72 + i52$ at 1.6 GHz. For these parameters, propagating modes occur for $-78 \leqslant m \leqslant 55$. The sidelobe occurring for mode orders in the range $-20 \leqslant m \leqslant -8$ is a result of not choosing the truncation, or total number of basis functions, to be nearly a multiple of the period of the sinusoid (i.e., $4\pi\ \mathrm{m}^{-1}$ in the example). Truncation of the infinite matrix results in the $[\sin(2\pi m/M)]/M$ function, which in a convolution with Green's function (that is, used in the extended operator), introduces additional lobing. Fortunately, however, the scattered energy, $\Sigma|a_m|^2$, is not semiconvergent, as can be seen in Fig. 2. That is, two significant figure accuracy in the scattered energy can be realized when 55 modes are included. Note that the nearest multiples of the wave period, 4π, are 37 and 51. The results for 37 and 53 modes agree reasonably well. For this set of parameters, the method of physical optics is applicable. However, a method applicable to cases in which the angle of incidence is near grazing is of interest.

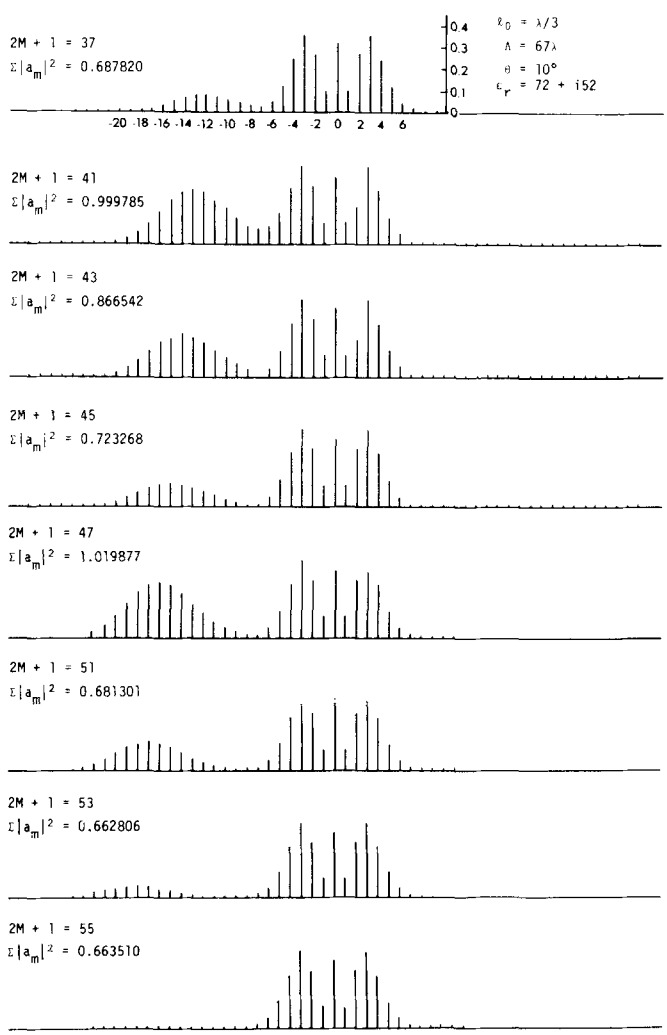


FIG. 2. Scattered vector a_m versus mode order m for eight different numbers of basis functions (occurrence of the sidelobe for m where $-20 \leqslant m \leqslant 8$ results from truncating the infinite matrix).

For the case of a lossless surface having the same parameters as those in Fig. 2 (i.e., $\epsilon_r = 72 + i0$), the sidelobe resulting from truncation of the infinite matrix does not occur. This suggests that in the EBC method the requirement for near cancellation of the fields at the surface of a good conductor results in numerical instability. The scattered and transmitted vectors for the case of a lossless surface are given in Table I.

For the case of a lossless surface, conservation of energy

$$\sum |a_m|^2 + \sqrt{\epsilon_r} \sum |b_m|^2 = 1 \tag{44}$$

provides a good numerical check on the accuracy of the results. For the case of Fig. 2, the check yielded 1.0001373. The total number of basis functions for this case was on the order of Λ/λ.

In Fig. 3, the sinusoid period is reduced by a factor of 3 from its value in Fig. 2. The total number of basis functions in Fig. 3 is approximately 15 times the period of the sinusoid. The $(\sin m)/m$ function is reduced by at least a factor of 3 from its value in Fig. 2, and the presence of any sidelobes is thereby reduced. Also, as seen in Fig. 3, three significant figure accuracy is obtained for the scattered energy, with about 7 parts error in the fourth significant figure.

Figure 4 shows the behavior of the scattered vectors as the height of the sinusoid is varied. The angle of incidence is

TABLE I. Scattered and transmitted vectors for $l_0 = \lambda/3$, $\Lambda = 67\lambda$, $\theta = 10°$, $\epsilon_r = 72 + i0$, $2M + 1 = 55$.

m	a_m	b_m
−7	−i0.015 294 78	−i0.285 746
−6	0.046 936 28	−0.342 853
−5	i0.119 571 74	−i0.024 687
−4	−0.239 881 84	−0.358 538
−3	−i0.341 877 59	i0.157 469
−2	0.255 037 87	−0.298 550
−1	−i0.094 549 13	i0.233 576
0	0.301 923 2	−0.268 661
1	−i0.091 389 19	i0.233 927
2	0.259 465 5	−0.298 106
3	−i0.340 792 7	i0.158 812
4	−0.234 215 85	−0.358 364
5	i0.113 834 24	−i0.022 455
6	0.043 342 754	−0.344 260
7	−i0.013 625 494	−i0.284 013

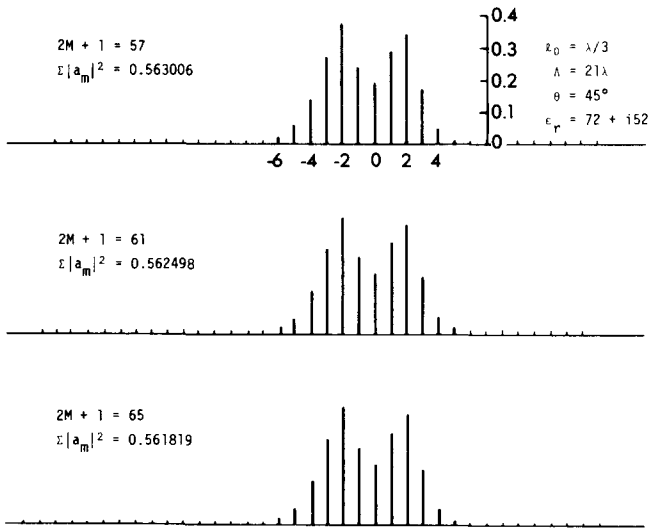


FIG. 3. Scattered vectors a_m versus mode order m for three different choices of total number of basis functions (in contrast to Fig. 2, no sidelobe occurs).

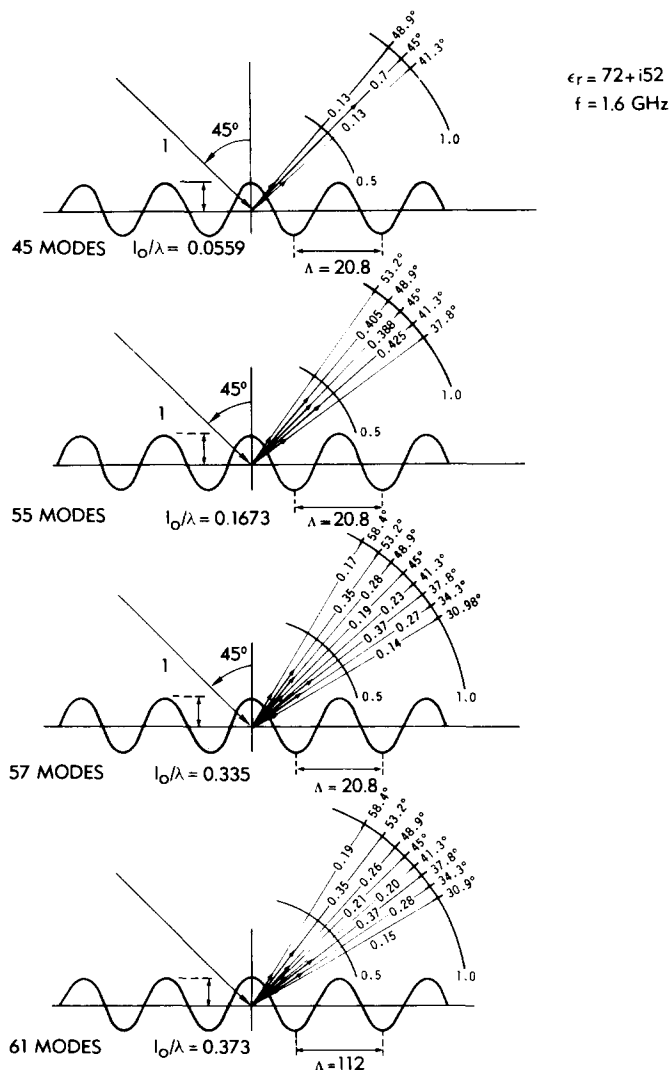


FIG. 4. Scattered vectors a_n versus scattering angle ϕ_m and height of sinusoid.

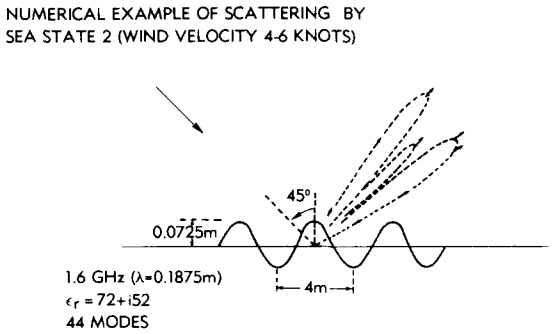


FIG. 5. Scattering pattern for sea state 2 modeled as a sinusoid. (The distance from the origin to the tip of each vector represents the magnitude of the scattered wave in that direction. The dashed curve between each vector represents a conceptual lobing phenomenon.)

45°, and the complex dielectric constant of the surface is $72 + i52$. The width of the sidelobes in this figure is a straight line and is only symbolic. As may be seen from these figures, diffuse scattering takes place for $l_0/\lambda = 0.1673$, and the maximum lobe does not lie in the specular direction. The magnitude and phase of the specular component for the smallest roughness (i.e., $l_0/\Lambda = 0.0559$) is $0.7e^{i5.4^\circ}$, which is in good agreement with the Fresnel reflection coefficient for vertical polarization (TM), which yields $0.74e^{i5^\circ}$.

Figure 5 shows a plot of the magnitude of the scattered vectors, a_m, for a sea state 2 (wind velocity approximately 4 to 6 knots). The electrical height of the sinusoid for this example is 0.387λ.

Figure 6 plots the magnitude of the scattered vectors,

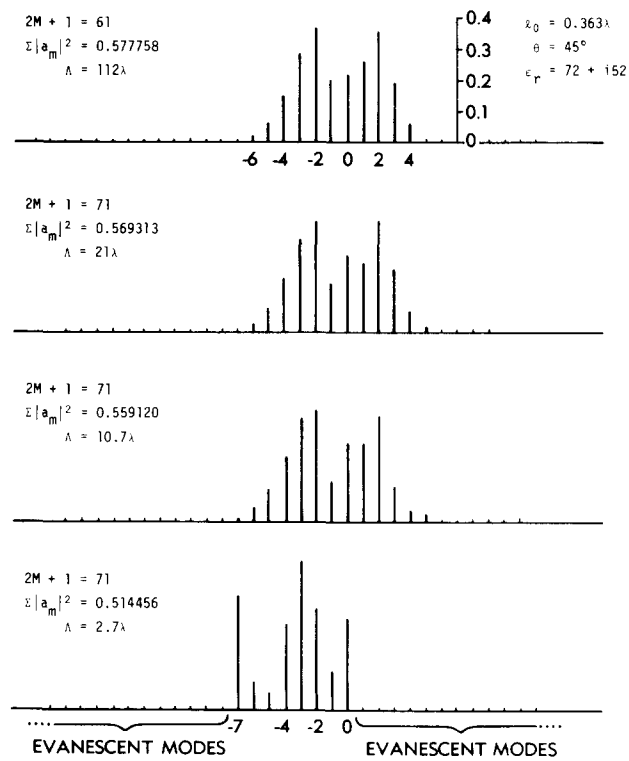


FIG. 6. Scattered vectors a_m versus mode order m and period of the sinusoid.

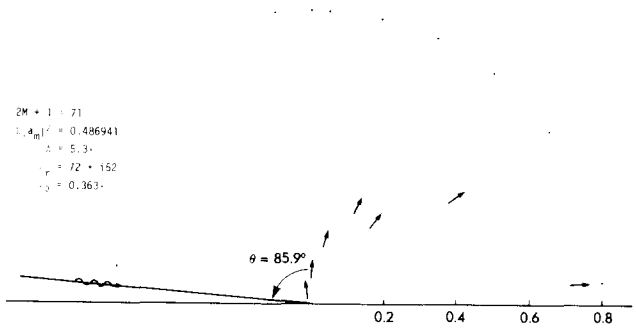


FIG. 7. Scattering pattern for near grazing incidence (The distance from the origin to the tip of each vector represents the magnitude of the scattered wave in that direction.)

a_m, as the period of the sinusoid is varied. For the angle of incidence of 45°, the asymmetry in a_m for $\pm m$ increases as the period decreases, as expected.

Figure 7 plots the magnitude of the scattered vectors as a function of the scattering angles ϕ_m for a case of near grazing incidence. This example is outside the range of applicability of the physical optics method.

Table II compares the results obtained from EBC with the perturbation method for a sinusoidal height of 0,056λ.

From Table II, four significant figure accuracy with 3.5 parts error in the fourth significant figure is obtained for the magnitude of a_{-1}. Two significant figure accuracy with 1.6 parts error in the second significant figure is obtained for the magnitude of a_1. Two significant figure accuracy with 4 parts error in the second signficant figure is obtained for the magnitude of a_0.

Table III compares the results obtained from EBC with the perturbation method for a sinusoidal height of 0.167λ.

From Table III, one significant figure accuracy with 9.2 parts error in the second significant figure is obtained for the magnitude of a_{-1}. One significant figure is obtained for the magnitude of a_1. About 30% error occurs in the magnitude of a_0.

When the surface height increases to 0.335λ, the perturbation method gives 0.75 for the magnitude of a_{-1}, while the EBC predicts essentially a null, i.e., 0.043. The total height of the surface is 0.67λ, well beyond the expected range of validity for the perturbation method.

From these comparisons, the perturbation analysis appears reasonably accurate for surface heights less than 0.2λ (peak to trough heights of 0.4λ). The authors are unaware of any numerical checks on the validity of the perturbation analysis in the literature. It is hoped that these results will provide useful criteria for the use of the perturbation method.

TABLE II. Comparison of EBC and perturbation for $l_0 = 0.056\lambda$, $\theta = \pi/4$, $\Lambda = 21.3\lambda$, $\epsilon_r = 72 + i52$.

	a_{-1}	a_0	a_1
EBC	$0.18822e^{i84.780°}$	$0.70e^{i5.4°}$	$0.18903e^{i84.669°}$
Perturbation	$0.18787e^{i84.659°}$	$0.74e^{i5.0°}$	$0.17331e^{i84.345°}$

TABLE III. Comparison of EBC and perturbation for $l_0 = 0.167\lambda$, $\theta = \pi/4$, $\Lambda = 21.3\lambda$, $\epsilon_r = 72 + i52$.

	a_{-1}	a_0	a_1
EBC	$0.4265e^{i84.782°}$	$0.388e^{i5.4°}$	$0.40731e^{i84.348°}$
Perturbation	$0.51892e^{i84.659°}$	$0.505e^{i5.4°}$	$0.5221e^{i84.669°}$

V. CONCLUSIONS

New numerical results have been presented for which the EBC method was used for scattering from very lossy sinusoidal surfaces. For a sinusoidal surface with $l_0 = 0.427\lambda$, the EBC method required 99 complex modes to reach a convergent result for the scattered energy, and to eliminate the (sin m)/m sidelobe arising from truncating the infinite matrix. The computer CPU time was 3 min 10 s on the IBM 370. This example appears to be near the limit of computational efficiency in terms of time and storage (99 complex modes require 198 storage locations).

The EBC method does fill a void for applicable methods when the surface period becomes as large as 100λ and the angle of incidence approaches grazing. There does not appear to be any results in the literature for scattering from very lossy sinusoidal surfaces for which the point matching technique was used. It would be interesting to compare the computational efficiency of the point-matching technique with the EBC method for scattering from lossy sinusoidal surfaces. One advantage of EBC versus point matching is its applicability to periodic surfaces having singularities similar to a rectangular tooth or a thin comb.

New analytical expressions have also been presented for the coefficients of the scattered vector obtained by the perturbation method. Results obtained by the perturbation method have been compared with those obtained by EBC for surfaces where the height is less than 0.2λ. These numerical checks should be useful in understanding the range of applicability when applying the perturbation method to scattering problems.

[1]D. J. Fang, F. Tseng, and T. O. Calvit, "A Low Elevation Angle Propagation Measurement of 1.5 GHz Satellite Signals in the Gulf of Mexico," IEEE Trans. Antennas and Propag. **AP-30**, 10–20 (1982).
[2]D. J. Fang, F. T. Tseng, and T. O. Calvit, "A Measurement of the MARISAT L-Band Signals at Low Elevation Angles On-Board Mobil Aero," IEEE Trans. Commun. **COM-30**, 359–365 (1982).
[3]B. Kinsman, *Wind Waves, Their Generation and Propagation on the Ocean Surface* (Prentice-Hall, Englewood Cliffs, NJ, 1965).
[4]S. O. Rice, "Reflection of Electromagnetic Waves from Slightly Rough Surfaces," in *The Theory of Electromagnetic Waves*, edited by M. Kline (Dover, New York, 1951) [paper was originally in Commun. Pure App. Math. **3** (4), 355–499 (1940)].
[5]P. Beckmann and A. Spizzichino, *The Scattering of Electromagnetic Waves from Rough Surfaces* (Pergamon, Oxford, 1963).
[6]P. C. Waterman, "Scattering by Periodic Surfaces," J. Acoust. Soc. Am. **57**, 791–802 (1975).
[7]S. L. Chuang and J. A. Kong, "Scattering of Waves from Periodic Surfaces," Proc. IEEE **69**, 1132–1144 (1981).

[8]R. I. Mosel and R. P. Merrill, "Quantum Scattering from a Sinusoidal Hard Wall: Atomic Diffraction from Solid Surfaces," Phys. Rev. B **12**, 5545–5551 (1975).
[9]J. R. Wait, "Perturbation Analysis for Reflection from Two-Dimensional Periodic Sea Waves," Radio Sci. **6**, 387–391 (1971).
[10]D. J. Fang, "Scattering from a Perfectly Conducting Sinusoidal Surface," IEEE Trans. Antenna Prop. **AP-20** (3), 388–390 (1972).
[11]R. K. Rosich and J. R. Wait, "A General Periodic Solution for Reflection from Two-Dimensional Periodic Surfaces," Radio Sci. **12**, 719–729 (1977).

Erratum: Irreducible representations of SU(*m*/*n*) [J. Math. Phys. 24, 157 (1983)]

J.-P. Hurni and B. Morel
Université de Genève, Département de Physique Théorique, 1211 Genève 4, Switzerland

(Received 30 March 1983; accepted for publication 8 April 1983)

PACS numbers: 02.20.Qs, 11.30.Pb, 99.10. + g

In the tableau of the bottom of page 161 showing the representations of SU(1/8), instead of a_2 one should read a_1 and for the value $a_1 = 8$ the correct content is

$$\bar{8} + 63 + 216 + 420 + 504 + 378 + 168 + 36.$$

In the line above, the three last members of the "typical" representation are $(36 + 28) + \bar{8}$ instead of $(26 + 28) + \bar{8}$.

As a consequence, in the special case $a_1 = 8$, one recovers a representation satisfying the trace condition of Ref. 8, i.e., J. Ellis, M. K. Gaillard, and B. Zumino, Phys. Lett. B **94**, 345 (1980).

We thank J. Thierry-Mieg for pointing out to us this mistake.

On page 160, right-hand column, ninth line from the top, read:

$$a_2 = 2 \quad 2 + 3 \text{ of SU(2)}.$$

Erratum: New sum rule for products of Bessel functions with application to plasma physics [J. Math. Phys. 23, 1278 (1982)]

Barry S. Newberger
Los Alamos National Laboratory, Los Alamos, New Mexico 87545

(Received 19 May 1983; accepted for publication 26 May 1983)

PACS numbers: 02.30.Gp, 99.10. + g

The extension of the results beyond their original domain of restriction by analytic continuation in the parameters α and β in Sec. III is false. The original series can be shown to diverge outside of this domain. The statement with regard to S_2 that the condition $\mathrm{Re}(\alpha + \beta) \geqslant p$ is sufficient but not necessary remains true, as the case $\alpha = \beta = 0$, $\gamma = 1$ shows; otherwise, the restrictions imposed in the proofs must be observed.

As a consequence, the results for Turkins functions, Eq. (4.6), must be restricted to $m \geqslant 0$. For $m \leqslant 0$, the restrictions are satisfied for the representation of Turkin's function

$$T(z, \alpha) = -\sum_{n=-\infty}^{\infty} (-1)^n \frac{J_n J_{-(m+n)}}{n + \alpha},$$

yielding the result

$$T(z, \alpha) = -(\pi/\sin \pi\alpha) J_{-\alpha} J_{\alpha - m}, \quad m \leqslant 0.$$

Correspondence with M. Bakker and N. M. Temme in this regard is greatly appreciated. Attention is called to their report which discusses this issue as well as provides interesting new results [Mathematisch Centrum Report TN103/83, revised (to appear)].

Erratum: Symmetry of the complete second-order nonlinear conductivity tensor for an unmagnetized relativistic turbulent plasma [J. Math. Phys. 24, 1332 (1983)]

Howard E. Brandt
U. S. Army Electronics Research and Development Command, Harry Diamond Laboratories, Adelphi, Maryland 20783

(Received 16 May 1983; accepted for publication 26 May 1983)

PACS numbers: 52.60. + h, 52.35.Ra, 52.25.Fi, 52.35.Mw, 99.10. + g

In Eq. (42) there is a typographical error in the last sign which should be plus instead of minus. Also in Eq. (35) u should be μ.